

Мультимодальная лингвистика

А. А. КИБРИК

Любое использование языка
неизбежно является мультимодальным.

(Scollon, 2006, p. 386)

1 Вводные замечания

В течение XX в. лингвистика пыталась осмыслить себя как самостоятельная наука, обособленная от других наук о человеке, в том числе психологии, антропологии, социологии и т. д. Этот процесс, протекавший достаточно болезненно, включал в себя в том числе и формирование автономного объекта исследований. Наиболее известные лингвисты, внесшие решающий вклад в автономизацию и в конечном счете инкапсуляцию лингвистики – Ф. де Соссюр и Н. Хомский. Понимаемый автономно язык необходимо было отделить от прочих проявлений человеческой деятельности на разных уровнях – и на уровне собственно единиц анализа (например, слова в протиположность жестам), и даже на уровне мозгового субстрата («языковой орган»). И в настоящее время большинство лингвистов продолжает полагать, что лингвистика должна изучать лишь фонемы, морфемы, слова, синтаксические конструкции и другие *вербальные сегментные единицы*, а выход за эти пределы неизбежно приводит нас в запретную или по крайней мере нереспектабельную область «паралингвистики».¹

Между тем, становится все более очевидно, что попытки принципиального отделения языка от коммуникации, мышления и поведения малопродуктивны, искусственны, обусловлены лишь логикой развитием науки, но никак не самой природой вещей. Вопрос о том, что входит в язык и в лингвистику, а что не входит и потому не заслуживает просвещенного интереса лингвистов, оказывается все менее интересным.

¹ Исследование осуществлено при поддержке гранта РГНФ №08-04-00165а. Данная статья частично основана на материалах двух докладов, прочитанных на 3-й Международной конференции по когнитивной науке (Kibrik, Podlesskaya, 2008; Kibrik, Èl'bert, 2008).

В данной работе я хотел бы привлечь внимание к некоторым исследованиям – моим собственным и принадлежащим моим коллегам, – которые показывают, что учет несегментных и невербальных явлений дает более объемную картину и позволяет по-новому взглянуть в том числе и на традиционные лингвистические вопросы.

Термин «мультимодальный», вынесенный в заголовок данной статьи, опирается на понимание *модальности*, принятое в психологии, нейрофизиологии и информатике: модальность – это тип внешнего стимула, воспринимаемого одним из чувств человека, в первую очередь зрением и слухом.

2 Информационные каналы

Как хорошо известно, в звуковом коде, помимо сегментного, вербального, компонента, имеется также несегментный, *просодический*, компонент. Метафорически можно сказать, что вербальный и просодический каналы соотносятся как цифровой (дискретный) и аналоговый (недискретный) коды. Просодия очень разнообразна – существуют десятки просодических явлений и большинство из них семантически нагружены. К числу наиболее очевидных относятся паузация, акценты, тоны в акцентах, темповые и долготные характеристики, тональные регистры, степени редукции, гортанные признаки, громкость и т. д. (см., напр.: Кодзасов 1996, 2001). Чтобы убедиться в информативности просодического канала, достаточно представить себе устную речь без просодии и интонации. Или наоборот, вспомнить о том, что разговор за стеной, в котором слышен лишь просодический контур, но не слышен вербальный компонент, бывает отчасти понятен. Хотя в существовании просодии никто не сомневается, тем не менее эта проблематика остается на периферии внимания лингвистов (о причинах такого положения дел см.: Николаева Ю. В., 2008).

В последние десятилетия исследователи коммуникации изучают не только звуковые информационные каналы, но и канал *визуальный*, включающий жесты, направление взгляда, мимику, позы, проксемику, культурный символизм и вообще «язык тела» (см., напр.: McNeill, 1992; Крейдлин, 2002; Бутовская, 2004; Николаева Ю. В., 2004). Визуальные средства играют очень большую роль в коммуникации и могут полностью менять содержание вербального компонента – например, если человек делает комплимент с недовольной гримасой. Отмечу, что связь между зрением и языком изучается и в совершенно иной связи – так называемая психолингвистическая парадигма «визуальный мир», использующая приборы слежения за движениями глаз говорящего или слушающего, опирается на важное наблюдение: семантическая обработка языковых элементов тесно связана со зрительным вниманием к соответствующему объекту, находящемуся в поле зрения (см. обзор по «визуальному миру» в: Федорова, 2008).

В данном исследовании выделяется три информационных канала: сегментный (вербальный), просодический и визуальный. В параграфе 3 я покажу, что просодия может играть принципиальную роль для базовых лингвистических вопросов – таких, например, как статус понятия «предложение». В параграфе 4 обсуждается важность одного из визуальных компонентов –

жестового – для такого центрального языкового явления, как референция. В параграфе 5 будет рассмотрен вопрос об относительной значимости трех информационных каналов. Параграф 6 посвящен жестовым языкам глухих, в которых визуальный канал выходит на первое место в качестве основного канала передачи информации. Параграф 7 носит обзорный характер: в нем кратко рассматриваются прикладные лингвистические исследования, ориентированные на мультимодальную коммуникацию. Некоторые итоги будут подведены в параграфе 8.

3 Просодия

Просодия является исключительно важным компонентом устной речи (дискурса) – основной и исходной формы использования языка. В первую очередь, просодия отвечает за *сегментацию* устной речи. Как известно, дискурс продвигается вперед не плавным потоком, а шагами, квантами – *элементарными дискурсивными единицами* (Chafe, 1994; Кибрик, Подлесская, 2006). Элементарные дискурсивные единицы (ЭДЕ) отождествляются при помощи комплекса просодических критериев, в том числе паттернов темпа, громкости, частоты основного тона, а также наличия просодического центра и пауз. Наиболее типичный семантико-грамматический коррелят ЭДЕ – клауза, или предикация: более половины ЭДЕ реализуются в устном дискурсе именно как предикатно-аргументные структуры.

Другая важнейшая характеристика устного дискурса, также выражаемая просодически – *фаза* (Кодзасов, 2002). Семантическая категория фазы имеет два основных значения: конец и не-конец. Конечная, или финальная, фаза указывает на то, что данная ЭДЕ является завершающей в некотором дискурсивном фрагменте, а нефинальная – что, напротив, продолжение следует. Имеются различные иерархические группы, для которых релевантно различие по фазе, однако здесь я хотел бы остановиться на одной из них – иллокутивном акте. На этом уровне фазовое различие состоит в том, является ли данная ЭДЕ финальной в рамках иллокутивного акта. Финальная ЭДЕ просодически оформляется при помощи направления тона в главном акценте. Так, для иллокутивного акта сообщения, как правило, используется нисходящий тон. Нефинальные ЭДЕ зеркально адаптируются к последующей финальной ЭДЕ, и в сообщениях чаще всего оформляются восходящим тоном. Ср. следующий пример из работы: Рассказы о сновидениях..., 2009, с. 98:

(1) 051z¹

0.5	2.	...(1.0) мм (0.2) ..(0.4) /сплю я,
2.6	3.	'и /вижу,
3.3	4.	...(0.2) как ... (0.8) мм (0.1) ... (0.9) две-е /женщины какие-то ..(0.1) в /складе в каком-то ложатся \спать.

1 Примеры в этом разделе заимствованы из корпуса устной речи «Рассказы о сновидениях» (2009). В заголовке примера указывается номер рассказа в корпусе. Строки в примерах соответствуют членению на ЭДЕ. О принципах транскрибирования устной речи, использованных в корпусе, см. указанную монографию. Перед текстом каждой ЭДЕ указывается ее номер в рамках данного рассказа, а в самой левой колонке – время в секундах от начала рассказа.

Несущие акценты, обозначаемые в транскрипции подчеркиванием ударных гласных, падают в этом примере на слова сплю, вижу и спать. Значимые тоны, реализованные на этих словах, обозначены знаками / и \ перед акцентированными словами. Легко видеть, что в финальной ЭДЕ сообщения (строка 4) используется нисходящий несущий акцент, в нефинальных (строки 2, 3) – восходящий.

Подобные примеры показывают, что в устной речи можно с определенной степенью надежности выделять такую единицу, как *предложение*, подобно тому как это делается в письменном тексте (ср. пунктуационные знаки, аналогичные письменной пунктуации, хотя и расставляемые на основе просодических характеристик, в конце каждой строки примера (1)). В данном параграфе я остановлюсь именно на вопросе о выделимости предложения в устной речи и на роли просодии в решении этого вопроса.

Как известно, понятие «предложение» рассматривается как базовое во многих направлениях лингвистики. Однако при ближайшем рассмотрении выясняется, что идентификация предложений в устной речи сталкивается с существенными проблемами. В исследованном нами корпусе устных рассказов о сновидениях (Кибрик, Подлесская, 2003; Рассказы о сновидениях..., 2009), помимо двух названных типов просодической огласовки ЭДЕ (т. е. финального падения и нефинального подъема), часто встречается и третий – так называемое *нефинальное падение*, например:

(2) 009z

9.8	4.	...(0.6) А /тогда уже д= ..(0.1) закрывались \двери,
12.8	5.	...(0.1) и /'Аня не –успела \сесть.

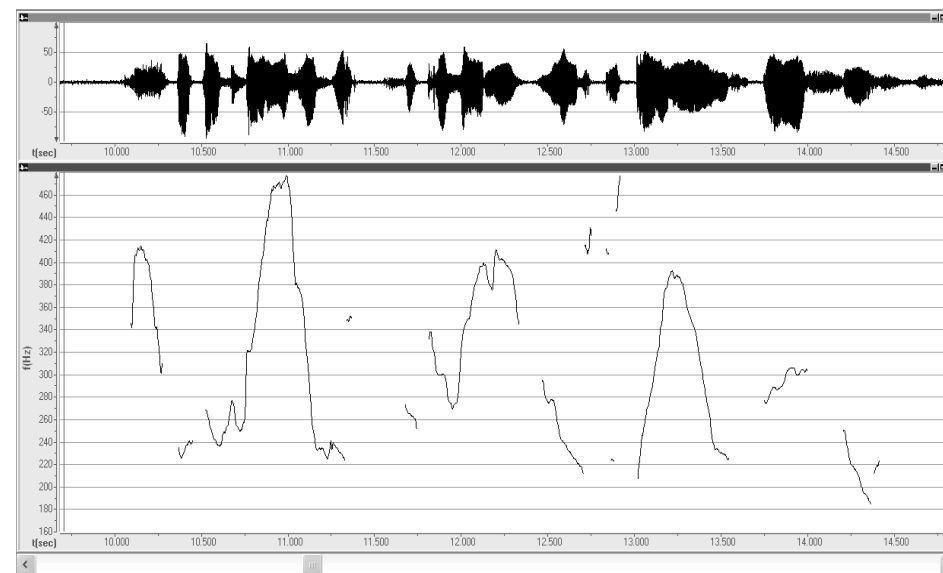


Рис. 1. Целевые уровни падения в примере (2): нефинального (обозначено пунктирной стрелкой) и финального (обозначено сплошной стрелкой)

Нефинальное падение содержится в строке 4 примера (2), в отличие от финального в строке 5 (и в отличие от нефинальных подъемов в примере (1)). Основная разница между двумя типами падения состоит в том, что нефинальное падение происходит в значительно более высокий целевой уровень тона. В данном случае – в уровень 210 Гц, в отличие от 180 Гц при финальном падении. Это различие можно наглядно видеть на рис. 1.

Такого рода явления широко распространены в устном дискурсе (см. ниже), и это ставит применимость понятие «предложение» к устной речи под вопрос: если конец предложения нельзя отождествить на основе направления тонального акцента, следовательно, границы предложений неопределенны.

Однако в действительности оказывается, что в речи большинства говорящих финальное и нефинальное падение представляют собой две семиотически противопоставленные просодические модели. При транскрибировании устной речи релевантны несколько просодических критериев, в совокупности позволяющих различить два типа падения (см. таблицу 1).

Таблица 1

КРИТЕРИИ РАЗЛИЧЕНИЯ ТИПОВ ПАДЕНИЯ

№	Критерий	Финальное падение	Нефинальное падение
А.	Целевой уровень движения нисходящего тона*	В нижний уровень, ком-фортный для голоса данного говорящего	В уровень, семиотически противопоставленный нижнему и приподнятый на несколько полутонов
Б.	Движение тона на послеакцентных слогах	Обязательно продолжается падение	Необязательно продолжается падение и, напротив, часто возникает загиб тональной кривой вверх
В.	Наличие паузы между данной ЭДЕ и следующей	Как правило, есть пауза	Паузы часто нет
Г.	Подхват vs ресет	В следующей ЭДЕ – ресет (начало тональной кривой) с типичного для данного говорящего тонального уровня	В следующей ЭДЕ подхват на том тональном уровне, на котором закончилась данная ЭДЕ

Примечание: * Каждый говорящий имеет не точечный уровень целевого падения тона, а определенный диапазон, например от 175 до 200 Гц.

Имеются и дополнительные критерии различения типов падений, в том числе «крутизна» падения, интервал падения и характер тонального движения (простой vs смешанный). Более подробно явление нефинального падения, в том числе вызывающие его функциональные контексты, рассматривается в другой нашей работе (Кибрик, 2008а). Здесь я хотел бы лишь отметить высокую частотность этого типа просодического оформления ЭДЕ. Как можно видеть на рисунке 2, по своей частотности нефинальное падение соизмеримо с каноническими фазовыми огласовками.

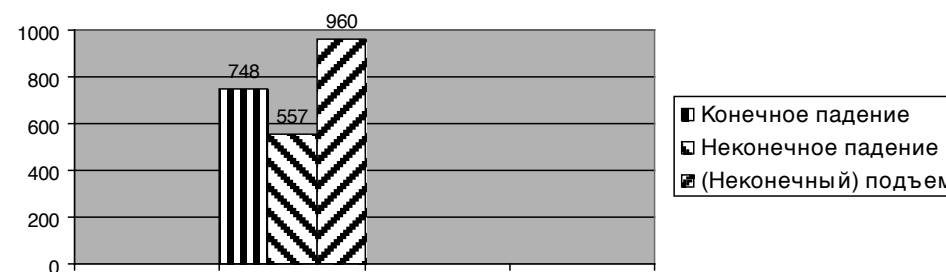


Рис. 2. Частотность фазовых огласовок в дискурсе

Итак, в речи большинства рассказчиков паттерны запятой (с подъемом или с падением) и точки достаточно четко противопоставлены и их удается различить при помощи серии просодических критериев. Это позволяет утверждать, что в устном нарративном дискурсе понятие предложения сохраняет свою валидность: говорящие, так сказать, знают, когда они заканчивают предложение, а когда нет. Очевидно, предложение устной речи является прототипом письменного предложения.

Тем не менее, выделение предложений возможно лишь на основе сложной аналитической процедуры, включающей построение просодического портрета конкретного говорящего. Если от говорящего зафиксировано малое количество данных, то надежное определение границ предложений оказывается невозможным. Кроме того, процедура, выработанная в данном случае, может потребовать значительной настройки при приложении к другим типам дискурса, другим типам иллокуций и, тем более, другим языкам. Все эти обстоятельства не позволяют утверждать, что предложение относится к базовым языковым единицам. Это довольно трудноуловимая, промежуточная языковая единица, находящаяся между базовыми уровнями: локальной структурой дискурса (отдельные ЭДЕ) и глобальной структурой (эпизоды, абзацы).

Итак, предложение является реальной, но небазовой единицей в основной форме использования языка – в устной речи. Этот вывод, имеющий большое значение для лингвистической теории, был получен на основе анализа просодических фактов. Обсуждение столь центральных вопросов грамматики оказывается невозможным без учета несегментных аспектов звука, т. е. просодического информационного канала. Разумеется, есть и десятки других семантически и грамматически важных явлений, которые могут быть поняты лишь с учетом просодического компонента. В настоящее время совершенно очевидно, что исследование языка без учета широкого спектра просодических явлений – это крайне редуционистский подход, и лингвистам предстоит осваивать несегментные аспекты речи, как бы это ни было непривычно.

Хотя вербальный и просодический компоненты обслуживаются, в узком смысле слова, одной и той же модальностью – звуковой, они качественно отличаются между собой. Как уже было сказано выше, противопоставление

между дискретным сегментным и недискретным просодическим компонентами можно рассматривать как различие цифрового и аналогового кодов. Представляется, что это различие достаточно велико, чтобы можно было распространить на него представление о множественности информационных каналов, т. е. понятие мультимодальности.

4 Жестикуляция

Визуальный канал имеет сложную внутреннюю организацию и, как и звук, мог бы быть расщеплен на несколько каналов. В этом разделе я остановлюсь только на *жестикуляции* как на ведущем компоненте визуального информационного канала. Жесты – это коммуникативное *кинетическое* поведение, в первую очередь при помощи рук, способное передавать смысл от говорящего к слушающему.

Среди разнообразных типов жестов (Крейдлин, 2002; Николаева Ю. В., 2004) хорошо выделяется класс *указательных* жестов. Указательные жесты играют исключительную роль в онтогенезе коммуникативных навыков и тесно связаны с таким базовым языковым явлением, как референция.

Указательные жесты развиваются у детей в конце первого года жизни – в этом возрасте дети начинают использовать директивные, экспрессивные и информирующие указательные жесты, привлекая внимание окружающих взрослых к референтам или направлениям (Tomasello et al., 2007). На этапе формирования синтаксиса (так называемая стадия двусловных конструкций) отмечаются комбинации «слово + указательный жест», где слово, например, является предикатом «открой», а указательный жест кодирует актанта «коробка» – вся конструкция в этом случае означает «открой коробку» (Butcher and Goldin-Meadow, 2000, p. 253). Таким образом, две модальности участвуют в формировании одной клаузы.

Во взрослой речи указательные жесты также активно используются при референции. Схематическая структура акта указания изображена на рисунке 3.

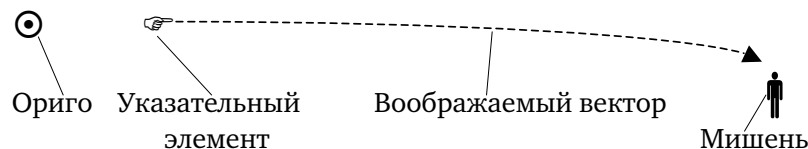


Рис. 3. Элементы канонического указательного акта

Шире всего известно использование указательных жестов при дейксисе, т. е. в ситуации привлечения внимания собеседника к некоторому референту. Указательный жест является почти необходимым элементом дейктической процедуры. Однако указательные жесты используются и при экзофоре – упоминании ранее активированных референтов, являющихся визуально доступными, например:

(3)

- а. Мы с Анатолием уже много лет очень /работаем,
<три промежуточных клаузы>
д. он много раз /завязывал,
(Николаева, 2003, с. 82)



В этом примере из телевизионного шоу говорящий указывает на своего друга, присутствующего в коммуникации речи, который весьма высоко активирован в момент осуществления указательного жеста (строка д). Вопрос о том, насколько часто указательные жесты встречаются при экзофоре, ждет детального исследования. На сегодняшний день лишь ясно, что это весьма распространенное явление (см.: O'Neill, 1996; Levy, 2000; Николаева, 2003).

Указательные жесты могут быть направлены не только на реальные, но и на воображаемые, виртуальные мишени-референты. В этом случае имеет место *виртуальное указание*. Например, при пересказе фильма рассказчик может конструировать пространство вокруг себя как аналог воображаемого физического пространства (имевшегося в фильме) и использовать это пространство при виртуальном указании. Референты занимают в этом пространстве фиксированные (хотя и меняющиеся) локусы, и указание на соответствующий локус сопровождает вербальное упоминание референта. Согласно данным Ю. В. Николаевой (личное сообщение), в ее базе данных по жестам при пересказе фильма встречается в среднем 4,2 случая виртуального указания на 100 ЭДЕ. При этом случаи референции к данному и к новому/доступному распределяются примерно в равной пропорции.

На рисунке 4 можно видеть, как рассказчица, пересказывая ранее виденный фильм, одновременно использует две руки для указания на двух активированных, но виртуальных референтов – мальчика (правая рука, вербально-субъектное местоимение в начале клаузы) и девочку (левая рука, вербально-объектное местоимение в конце клаузы).



он /засмотрелся на неё

Рис. 4. Виртуальное указание при пересказе фильма

В целом, следует заключить, что понимание указательных жестов принципиально важно для нескольких центральных разделов лингвистики, в том числе для онтогенеза речи и для исследования референции в дискурсе.

Важное направление работы – установление связей между исследованиями просодии и жестикуляции. Эти два информационных канала имеют много общего – иногда о просодии даже говорят в терминах «просодических жестов» (С. В. Кодзасов, устное сообщение). К этому направлению относятся исследования Н. В. Суховой, в частности, работа (Сухова, 2008), посвященная взаимосвязи между паузацией и «кинетическими фразами» в жестикуляции. В работе Леви (Levy, 2000) было предложено зонтичное понятие *энергозатраты* (energy expenditure), включающее как жестовые (указательные жесты), так и просодические (акцентирование) аспекты дейктической референции.

Объемлющая модель коммуникации должна учитывать не только звуковой, но и визуальный канал. Эти каналы достаточно жестко разграничены в научных традициях, но не в реальных коммуникативных процессах. Будущее лингвистики (или более широкой науки о естественной коммуникации) – одновременный учет всех одновременно используемых людьми информационных каналов.

5 Относительный вклад информационных каналов

Итак, принципиальная значимость просодического и визуального каналов и их релевантность для языка и коммуникации не подлежат сомнению. Однако мнения об относительном вкладе вербального, просодического и визуального каналов сильно разнятся. Практика основных направлений лингвистики явно имеет своей презумпцией полный приоритет вербального канала. Другие каналы, если и учитываются, то лишь как маргинальные и вспомогательные. Между тем, в некоторых других дисциплинах, также интересующихся коммуникацией, фигурируют совсем иные оценки. Так, в практической психологии часто указываются следующие цифры (исначально из работы: Mehrabian, 1971): визуальный канал передает 55% информации, просодия – 38%, а на долю вербального канала остается лишь 7%. В ряде исследований (Эльберт, 2007; Kibrik, Èl'bert, 2008) была сделана попытка эмпирически оценить относительный вклад трех информационных каналов.

Мы использовали отрывок из телесериала (3 минуты 20 секунд), которому предшествовал контекстный отрывок (8 минут). Три компонента – вербальный, просодический и визуальный – были изолированы и представлены во всех возможных комбинациях. Каждая из этих комбинаций (всего их восемь) была показана группе испытуемых (от 10 до 17 человек). Например, одна из групп смотрела исходный, полный видеоматериал (все три канала), другая группа слышала лишь просодию (как при разговоре за стеной; один канал), третья видела изображение и титры (два канала – визуальный и вербальный), и т. д. Восьмая группа вообще не видела релевантного отрывка, а лишь контекстный отрывок.

Степень понимания дискурса была оценена при помощи анкеты из 23 вопросов. В таблице 2 показана средняя степень понимания для каждой из восьми экспериментальных групп.

Таблица 2

Доля правильных ответов в восьми экспериментальных группах

Номер группы	1	2	3	4	5	6	7	8
Экспериментальный материал	Исходный материал	Звук	Титры и изображение	Просодия и изображение	Титры	Просодия	Изображение	Только контекстный отрывок
Количество каналов	3	2	2	2	1	1	1	0
Средняя доля правильных ответов	87,4%	70,4%	73,9%	51,2%	72,0%	51,1%	61,7%	38,3%

Некоторые выводы из этого исследования состоят в следующем:

- каждый из каналов по отдельности (группы 5–7) очень информативен, так как в соответствующих группах процент правильных ответов гораздо выше, чем в 8-й группе;
- иерархия каналов по информативности выглядит так: вербальный > визуальный > просодический;
- комбинация 4 дает самый низкий результат – это редкое, непривычное сочетание, и, вероятно, испытуемые с трудом интегрируют информацию из этих двух каналов в отсутствие вербального;
- вербальный канал, по-видимому, является основным, а информация из двух других каналов интегрируется на основе вербально передаваемой информации.

Если условно предположить, что три информационных канала являются независимыми (что, конечно, является сильным огрублением), и нормировать вклад трех изолированных каналов (пятая, шестая и седьмая колонки таблицы 2), то вклады каналов могут быть оценены так: вербальный – 39%, просодический – 28%, визуальный – 33%. Очевидно, и традиционная лингвистическая точка зрения, и мнение, циркулирующее в практической психологии, являются весьма неточными.

Некоторые дополнительные результаты были получены в исследовании На. Молчановой (2009). В этой работе, в отличие от предыдущих, в качестве экспериментального материала был использован не фрагмент фильма, а бытовой диалог. Были изменены вербальная и просодическая модификации исходного материала: для первой использовались титры, полностью очищенные от маркирования просодии (таких, как пунктуационные знаки), для второй – иной вариант акустического фильтра. Также был оптимизирован вопросник, увеличен контраст между группами «исходный материал» и «только контекстный отрывок». Результаты оказались весьма близки к предшествующим исследованиям. Это позволяет заключить, что высокая значимость трех информационных каналов и их относительный вклад оценены достаточно близко к истине.

6 Жестовые языки

Научное лингвистическое изучение *жестовых языков* (ЖЯ) глухих – это особая и весьма разработанная область лингвистики. Исследования ЖЯ уже в течение ряда десятилетий ведутся в США и других западных странах, и в последнее время начаты и в России (см.: Прозорова, 2007; Кибрик, 2008б). В жестовых языках, в отличие от звучащих языков, визуальный канал является основным. (Пользователи ЖЯ могут иметь быть частично слышащими и поэтому использовать звуковой канал как вспомогательный.) В сущности, следует признать, что существует два основных типа человеческих языков – звучащие и жестовые.

Жестовые языки в принципе обладают теми же сущностными характеристиками, что и звучащие. В работе В. И. Киммельман (2008) подробно обсуждается вопрос о том, как и в какой степени в русском жестовом языке (РЖЯ) можно выделять основные части речи – имя и глагол.

В работе А. Кибрика и Е. Прозоровой (2007) описаны процессы референции в РЖЯ, в сопоставлении со звучащими языками. Основные явления, связанные с референцией, в РЖЯ те же, что и в звучащих языках. При дейксисе говорящий на РЖЯ может привлекать внимание адресата к референтам, представленным в перцептивном пространстве. При анафоре говорящий опирается на имеющуюся активацию референтов в когнитивной системе; в случаях анафоры в РЖЯ в большинстве случаев используется нулевое выражение – подобно таким звучащим языкам, как японский. Кроме того, в РЖЯ широко распространено виртуальное указание на элементы воображаемого конструктивного пространства – внешней проекции когнитивного пространства. По нашим данным, виртуальное указание использовалось в 9% всех случаев референции в корпусе. Как правило, этот способ референции применяется при промежуточном уровне активации референтов – не слишком высоким, но одновременно и не слишком низким. Явление, в зачаточной форме представленное при жестуляции, сопровождающей звучащую речь (см. параграф 4), в жестовых языках приобретает статус важного конвенционального референциального средства.

В примере (4) и на рисунке 5 можно видеть иллюстрации всех основных типов референциальных средств.

Данный фрагмент жестовой речи состоит из четырех ЭДЕ. В третьей ЭДЕ употреблена полная именная группа МАЛЬЧИК с адъективным указательным местоимением (ДЕМ) – указательным жестом, направленным влево (→). Во второй ЭДЕ использовано субстантивное указательное местоимение. В той же ЭДЕ, а также в четвертой можно видеть примеры нулевых референциальных выражений. Наконец, в первой ЭДЕ используется особое референциальное средство, присущее жестовым языкам – так называемая конструкция с классификатором, кумулятивно указывающая на референта человека (посредством формы руки) и на характер действия.

В работах Е. В. Прозоровой (2008, 2009) были исследованы вопросы локальной структуры дискурса РЖЯ. Как было установлено (Прозорова, 2009),

- (4) 1. Человек_i подходит.
2. \emptyset_i отдает это ему_k.
3. Тот мальчик_k держит велосипед,
4. \emptyset_k благодарит.



ЧЕЛОВЕК-ДВИГАТЬСЯ^{→л}-----ДАТЬ^{→л}-----ДЕМ^{→л}



МАЛЬЧИК ДЕМ^{→л} ДЕРЖАТЬ-ВЕЛОС.-----БЛАГОДАРИТЬ

Рис. 5. Фрагмент жестовой речи, соответствующий примеру (4)

главным средством кодирования границ ЭДЕ в жестовой речи (т. е. аналогом просодических средств звучащих языков) являются движения головы. Кроме того, используются вспомогательные «артикуляторы», такие как моргание и паузация. Таким образом, визуальный канал в жестовых языках, как и звуковой канал в звучащих языках, содержит два особых компонента – вербальный и просодический.

В целом можно заключить, что учет ЖЯ добавляет исследованиям универсалий и типологии целое дополнительное измерение. В дальнейшем можно ожидать, что исследования жестовых языков будут играть все большую роль в общелингвистической проблематике. В центре внимания при этом будет вопрос о том, какие свойства человеческих языков зависимы, а какие независимы от доминирующей модальности.

7 Мультиmodalность и прикладные исследования

Окружающий нас мир мультиmodalен, и мы являемся мультиmodalными организмами. Языковая коммуникация по природе мультиmodalна, но узко специализированные научные традиции, фокусирующиеся лишь на каком-то из аспектов реальности, склонны эту мультиmodalность не замечать. Как это нередко бывает, важные и реальные явления раньше осознаются и исследуются в прикладных, практически ориентированных разработках, чем в теоретических моделях – ср. следующее высказывание из статьи о системах коммуникации для инвалидов: «Нормальная, ненарушенная коммуникация, конечно, является изначально мультиmodalной, речевое содержание в ней модифицируется просодией и передается параллельно

с использование мимики, жестов, поз, а также целого набора других невербальных коммуникативных средств» (Alm, 2006, p. 569). В данной статье я не буду останавливаться на собственно технологических и инженерных изобретениях – таких, как мультимодальные коммуникативные устройства, мультимодальные GPS-навигаторы, или системы, относящиеся к Web 2.0. Однако имеет смысл кратко упомянуть некоторые связанные с языком прикладные исследования, в которых изучаются и реализуются мультимодальные коммуникативные процессы. Следует отметить, что на западе работы такого рода активно ведутся уже в течение более десяти лет. Подобные прикладные исследования можно условно разделить на три группы: а) работы по компьютерному моделированию коммуникации; б) исследования мультимодальности в дискурсе, в особенности, в дискурсе средств массовой информации; в) корпусная лингвистика.

В настоящее время *человеко-машинное общение* развивается по пути все большего сближения с естественным межличностным общением, и этот процесс естественно затрагивает и создание мультимодальных компьютерных систем. Считается, что системы, позволяющие пользователям распределять контент между разными модальностями будут более удобными для пользователей, более адаптивными и в конечном счете более надежными. При этом мультимодальность означает не просто наличие нескольких каналов связи, а их взаимодействие и интеграцию. «В биологии, экспериментальной психологии и когнитивной нейронауке существует особая, быстро развивающаяся литература, показывающая, что мультисенсорное восприятие и интеграция не могут быть сведены к рассмотрению этих чувств в изоляции друг от друга» (Cohen, Oviatt, 2006, p. 378). Проектируемые в настоящее время системы включают как традиционную обработку письменного текста и звучащей речи, так и опознавание жестов, выражений лица, направления взгляда, а также учет прагматического контекста (о некоторых из таких систем см.: Gibbon et al. eds., 2000, Granström et al. eds., 2002; Bengio, Bourlard, eds., 2005; Minker et al. eds., 2005). Помимо пропозиционального содержания, большую роль играет передача и понимание эмоций (см.: Котов, 2009). Активно ведутся работы по созданию «воплощенных говорящих агентов» (*embodied conversational agents*), способных в какой-то мере пользоваться всеми информационными каналами (Cassell, 2006). Особое место среди мультимодальных компьютерных систем занимают те, которые предназначены для людей с ограниченными возможностями (см., напр.: Edwards, 2002; Alm, 2006).

Мультимодальность естественного дискурса часто отмечают исследователи *средств массовой коммуникации*, например: «Мультимодальный подход предполагает, что сообщение распространяется через все коммуникативные модусы. Если это так, то каждый модус – это лишь частичный носитель глобального значения сообщения» (Kress, 2002, p. 6). В исследованиях текстов СМИ большое внимание придается визуальным аспектам – использованию цветов, шрифтов, формата, иллюстрациям и т. д. (Schröder, 2006, p. 629); письменный текст, таким образом, по природе мультимодален. Тем более важен визуальный ряд в таком СМИ, как телевидение. Интересно, что кон-

серватизм мышления характерен не только для академического сообщества, но и для медийных специалистов. Так, при возникновении такой изначально мультимодальной технологии, как телевидение, в течение ряда десятилетий новости главным образом подавались в виде диктора, зачитывающего вслух заранее заготовленный письменный текст, и считалось, что движущаяся «картинка» может помешать восприятию и отвлечь внимание (Busch, 2006, p. 614). Максимальное разнообразие одновременно используемых модальностей характерно для современного интернета, сочетающего множество взаимодействующих средств – графических (как статических, так и движущихся) и звуковых. «Изучение текста, включающего изображения и звуки, бросает вызов обычному дискурсивному анализу <...> Массмедиа создают мультимодальные тексты <...> Несмотря на трудности, возникающие при попытке отразить такую мультимодальность, сосредотачиваться лишь на языке и игнорировать остальные модусы означает упустить многое из потенциальных смыслов современных медийных текстов» (McKay, 2006, p. 599). Существует уже целый ряд сборников и монографий, посвященных специально мультимодальности дискурса, в особенности медийного (Kress, van Leeuwen, 2001; LeVine, Scollon, eds., 2004, Norris 2004, O'Halloran, ed., 2004; Ventola et al. eds., 2004). Проводятся также международные конференции по мультимодальности в языке и культуре (см., напр.: <http://www.multimodality.it/site>).

Как известно, начиная с 1990-х годов набирает силу методологически важное направление – *корпусная лингвистика* (см., например, Biber et al., 1998; Баранов, 2003: 112–137). Корпусный метод – это разновидность общенаучного метода наблюдения, который состоит в систематическом рассмотрении большого объема естественных данных. Корпус того или иного языка – достаточно большая, репрезентативная коллекция дискурсов (текстов) на данном языке, доступная для компьютерного поиска и, как правило, обеспеченная специальной разметкой. В настоящее время существуют корпуса для многих крупных языков, в том числе русского, см. <http://corpora.yandex.ru>. Первоначально языковые корпуса были ориентированы на письменные тексты, позже начали появляться подкорпуса устной речи, с постепенно возрастающей сложностью просодической транскрипции; выше не раз упоминался просодически размеченный корпус, описанный в работе «Рассказы о сновидениях» (2009). В настоящее время ставится задача создания полноценных мультимодальных корпусов, включающих не только звуковые, но и визуальные каналы. На конференции LREC-2008 (Language Resources and Evaluation Conference) был представлен целый ряд докладов и даже секций, посвященных проектам по созданию мультимодальных корпусов (см., напр.: Blache et al., 2008; Gallo et al., 2008). Правда, авторы таких исследований, учитывая визуальные аспекты коммуникации, не обращают внимания на просодический компонент. Актуальная задача на ближайшее будущее – создание мультимодальных корпусов, включающих все информационные каналы – в частности, для русского языка. Группа японских исследователей (Kitazawa et al., 2008) соединяет проблематику мультимодальности с изучением усвоения языка

детьми – в этой работе рассматривается усвоение указательной референции, с учетом вербальных элементов, жестов и даже просодии. В рамках Национального корпуса русского языка создается мультимедийный подкорпус звучащей речи, включающий разметку жестов и других визуальных элементов (Гришина, Савчук, 2008).

8 Заключение

Вопреки сильным традициям фрагментации, характерным для современной науки, представляется несомненным, что просодия и визуальные средства коммуникации должны изучаться вместе с вербальными средствами в рамках единого подхода.

Адекватное представление о естественной языковой коммуникации может быть получено лишь при мультимодальном подходе, учитывающим все коммуникативные каналы и их внутреннюю организацию. Принципиальной разницы между вербально передаваемой информацией и информацией, передаваемой просодически или жестами, не существует. Только подход, признающий мультимодальность коммуникации, имеет шанс на создание реалистической лингвистики будущего.

Понятие (мульти) модальности в узком смысле слова относится к различию между человеческими органами чувств, в первую очередь различию между зрительным и слуховым каналами. Однако в рамках каждого из этих каналов есть дальнейшие, более дробные различия, которые также охватываются понятием мультимодальности. Так, в звучащей речи имеется сегментный (вербальный) компонент и множество несегментных (просодических) параметров. Визуальный канал включает жестикуляцию, направление взора, мимику и другие аспекты «языка тела». Письменный дискурс, также воспринимаемый визуально, помимо вербального компонента включает целый набор графических параметров, таких как шрифт, цвета, формат и т. д. Современное понятие мультимодальности включает все это разнообразие.

Уже сейчас следует отметить, что мультимодальный подход, констатирующий множественность информационных каналов, недостаточен. Различные каналы не просто аддитивно участвуют в передаче информации. Они являются взаимозависимыми, взаимодействуют между собой, и информация распределяется между ними сложным образом. Поэтому в перспективе необходим не только мультимодальный, но и *кросс-модальный* подход, учитывающий эти обстоятельства.

ЛИТЕРАТУРА

Баранов А. Н. Введение в прикладную лингвистику. М.: УРСС, 2003.
 Бутовская М. Л. Язык тела: природа и культура (эволюционные и кросс-культурные основы невербальной коммуникации человека). М.: Научный мир, 2004.
 Гришина Е. А., Савчук С. О. Корпус звучащей русской речи в составе Национального корпуса русского языка. Проект // Компьютерная лингвистика

и интеллектуальные технологии. Труды международной конференции «Диалог 2007» (Бекасово, 4–8 июня 2008 г.) М.: РГГУ, 2008. С. 125–132.
 Кибрик А. А. Есть ли предложение в устной речи? // Фонетика и нефонетика / Под ред. А. В. Архипова и др. М.: ЯСЛ, 2008а. С. 104–115.
 Кибрик А. А. О важности лингвистического изучения русского жестового языка // Лингвистические права глухих / Под ред. А. А. Комарова, Н. А. Чаушьян. М.: ОООИ ВОГ, 2008. С. 122–129.
 Кибрик А. А., Подлеская В. И. К созданию корпусов устной русской речи: принципы транскрибирования // Научно-техническая информация. Серия 2, 6. 2003. С. 5–11.
 Кибрик А. А., Подлеская В. И. Проблема сегментации устного дискурса и когнитивная система говорящего // Когнитивные исследования. Вып. 1 / Под ред. В. Д. Соловьева. М.: Институт психологии РАН, 2006. С. 138–158.
 Кибрик А. А., Прозорова Е. В. Референциальный выбор в русском жестовом языке // Диалог. Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции Диалог'2007 / Под ред. Л. Л. Иомдина, Н. И. Лауфер, А. С. Нариньяни, В. П. Селегея. М.: ИПИ РАН, 2007. С. 220–230.
 Киммельман В. И. Противопоставление имени и глагола в русском жестовом языке: конкретная лексика. Курсовая работа. М.: РГГУ, 2008.
 Кодзасов С. В. Комбинаторная модель фразовой просодии // Просодический строй русской речи. М.: Институт русского языка РАН, 1996. С. 85–123.
 Кодзасов С. В. Комбинаторные методы в фонологии: Дис. ... докт. филол. наук. М.: МГУ им. М. В. Ломоносова, 2001.
 Кодзасов С. В. Фазовая символика тона // Логический анализ языка. Семантика начала и конца / Под ред. Н. Д. Арутюновой. М.: Индрик, 2002.
 Котов А. А. Паттерны эмоциональных коммуникативных реакций: проблемы создания корпуса и перенос на компьютерных агентов // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2009». М.: РГГУ, 2009. С. 211–219.
 Крейдлин Г. Е. Невербальная семиотика. М.: НЛО, 2002.
 Молчанова Н. Вклад различных каналов передачи информации в понимание устного бытового диалогического дискурса: Курсовая работа. М.: МГУ, 2009.
 Николаева Ю. В. Иллюстративные жесты: семантика и особенности вербального контекста: Дипломная работа. М.: МГУ, 2003.
 Николаева Ю. В. Иллюстративные жесты в устном дискурсе (типы, функции, коммуникативное поведение) // Московский лингвистический журнал, 2004. Т. 8. № 2. С. 89–117.
 Николаева Т. М. Почему лингвисты не любят «про интонацию»? // Фонетика и нефонетика / Под ред. А. В. Архипова и др. М.: ЯСЛ, 2008. С. 43–52.
 Прозорова Е. В. Российский жестовый язык как предмет лингвистического исследования // Вопросы языкознания. 2007. № 1. С. 44–61

- Прозорова Е. В. Транскрипция как средство анализа пауз в русском жестовом дискурсе // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог» / Под ред. А. Е. Кибрика и др. М.: ИПИ РАН, 2008. С. 511–518.
- Прозорова Е. В. Маркеры локальной структуры дискурса в русском жестовом языке: Дисс.... канд. филол. наук. М.: МГУ, 2009.
- Рассказы о сновидениях: корпусное исследование устного дискурса / Под ред. А. А. Кибрика, В. И. Подлесской. М.: ЯСЛ, 2009.
- Сухова Н. В. Способы взаимодействия пауз колебания и кинетических фраз // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог» / Под ред. А. Е. Кибрик и др. М.: ИПИ РАН, 2008. С. 511–518.
- Федорова О. В. Методика регистрации движений глаз «Визуальный мир»: шанс для сближения психолингвистических традиций // Вопросы языкознания, 2008. № 6. С. 98–120
- Эльберт Е. М. Вклад вербального, просодического и визуального каналов в понимание устного дискурса. Дипломная работа. М.: МГУ, 2007.
- Alm N. Augmentative and alternative communication // Encyclopedia of language and linguistics, 2nd ed. / Увью Brown, Keith. Oxford: Elsevier, 2006. V. 1. P. 569–574.
- Bengio S., Bourlard H. (eds.) Machine learning for multimodal interaction. Berlin: Springer. 2005.
- Biber, D., Conrad, S., Reppen, R. Corpus linguistics: investigating language structure and use. Cambridge: Cambridge University Press. 1998.
- Blache P., Bertrand R., Ferré G. Creating and exploiting multimodal annotated corpora // Materials of the Language Resources and Evaluation Conference, Marrakech, May 2008. 2008.
- Busch B. Media, politics, and discourse: Interactions // Encyclopedia of language and linguistics, 2nd ed. / Ed. by Brown, Keith. Oxford: Elsevier, 2006. Vol. 8. P. 609–616.
- Butcher C., Goldin-Meadow S. Gesture and the transition from one- to two-word speech: when hand and mouth come together // D. McNeill (ed.) Language and gesture. Cambridge: CUP, 2000. 235–257.
- Cassell J. Conversational agents, synthetic // Encyclopedia of language and linguistics, 2nd ed. / Ed. by Brown, Keith. Oxford: Elsevier, 2006. Vol. 3. P. 163–169.
- Chafe Wallace. Discourse, consciousness, and time. Chicago: University of Chicago Press. 1994.
- Cohen P. R., Oviatt S. L. Multimodal interaction with computers // Encyclopedia of language and linguistics, 2nd ed. // Ed. by Brown, Keith. Oxford: Elsevier, 2006. V. 9. P. 376–386.
- Edwards A. D. N. Multimodal interaction and people with disabilities / Eds. Gransström et al. 2002. P. 73–92.

- Gallo C. G., Jaeger T. F., Allen J., Swift M. Production in a multimodal corpus: How speakers communicate complex actions // Materials of the Language Resources and Evaluation Conference, Marrakech, May 2008. 2008.
- Gransström Björn, House D., Karlsson I. (eds.). Multimodality in Language and Speech Systems. Norwell: Kluwer. 2002.
- Gibbon D., Mertins I., and Moore R. (eds.). Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation. Norwell: Kluwer, 2000.
- Kibrik A. A. and El'bert E. M. Understanding spoken discourse: the contribution of three information channels // Materials of the 3rd International Conference on Cognitive Science, Moscow, June 2008 / Eds. Yu. I. Alexandrov et al. 2008. P. 82–84.
- Kibrik A. A., Podlesskaya V. I. Is sentence viable? // Materials of the 3rd International Conference on Cognitive Science, Moscow, June 2008 / Eds. Yu. I. Alexandrov et al. 2008. P. 84–85.
- Kitazawa Sh., Kiriya Sh., Kasami T., Ishikawa Sh., Otani N., Horiuchi H., Takebayashi Y. A Multimodal infant behavior annotation for developmental analysis of demonstrative expressions // Materials of the Language Resources and Evaluation Conference, Marrakech, May 2008. 2008.
- Kress G. The multimodal landscape of communication. Medien Journal 2002. V. 4. P. 4–19.
- Kress G., van Leeuwen T. Multimodal discourse. The modes and media of contemporary communication. London: Arnold, 2001.
- Levy E., Fowler C. A. The role of gestures and other graded language forms in the grounding of reference in perception // Language and gesture / Ed. by D. McNeill. Cambridge: CUP, 2000. P. 215–234.
- LeVine P., Scollon, R. (eds.). Discourse and technology: multimodal discourse analysis. Washington, DC: Georgetown University Press, 2004.
- McKay S. Media and language: Overview // Encyclopedia of language and linguistics, 2nd ed. Eds. Yu. I. Alexandrov et al. Oxford: Elsevier, 2006. V. 8. P. 597–602.
- McNeill D. Hand and mind: what gestures reveal about thought. Chicago: University of Chicago Press, 1992.
- Mehrabian A. Silent messages. Belmont, CA: Wadsworth Publishing Company, 1971.
- Minker W., Bühler D., Dybkjær L. (eds.). Spoken multimodal human-computer dialogue in mobile environments. Dordrecht: Springer, 2005.
- Norris S. Analyzing multimodal interaction: A methodological framework. London: Routledge, 2004.
- O'Halloran K. L. (ed.). Multimodal discourse analysis: Systemic functional perspective. London: Continuum. 2004.
- O'Neill D. Two-year-old children's sensitivity to a parent's knowledge state when making requests // Child development. 1996. V. 67 (2). P. 659–677.

- Scollon R. Multimodality and the language of politics *Encyclopedia of language and linguistics*, 2nd ed. / Eds. Yu. I. Alexandrov et al. Oxford: Elsevier, 2006. V. 9. P. 386–387.
- Schröder K. C. Media: Pragmatics. *Encyclopedia of language and linguistics*, 2nd ed. Eds. Yu. I. Alexandrov et al. Oxford: Elsevier, 2006. V. 8. P. 623–631.
- Tomasello M., Carpenter M., Liszkowski U. A new look at infant pointing // *Child development*. 2007. V. 78 (3). P. 705–722.
- Ventola E., Charles C., Kaltenbacher M. (eds.). *Perspectives on multimodality*. Amsterdam: Benjamins, 2004.

Методы автоматического разрешения лексической многозначности с использованием знаний

В. Д. Соловьев, В. В. Иванов

Введение

Статья преследует следующие цели: представить обзор современных приложений лингвистических ресурсов, основанных на знаниях, при автоматическом разрешении лексической многозначности (Word Sense Disambiguation, WSD), указать новые задачи, для решения которых требуется использовать методы WSD.

Несмотря на то, что явление лексической многозначности и задача WSD относятся к лингвистике, статья имеет тесную связь с информационными технологиями и когнитивной наукой. Это обусловлено тем, что методы, описанные в статье, опираются на ресурсы, содержащие машиночитаемые знания о лексических единицах (ЛЕ) естественного языка (ЕЯ) или подъязыка некоторой предметной области. Таким образом, и задача автоматизированного WSD, и методы ее решения скорее относятся к области искусственного интеллекта и представления знаний.

В параграфе 1 описываются предпосылки и необходимость решения задачи WSD с помощью автоматизированных методов, а также общие подходы к решению на основе баз знаний, корпусов текстов и их комбинаций. Затрагиваются проблемы, возникающие при автоматическом WSD, различные постановки задачи WSD, открытые вопросы, дана классификация методов, основанных на знаниях. Параграф 2 посвящен обзору методов, основанных на знаниях, а также комбинированных подходах к решению задачи WSD. В параграфе 3 внимание уделяется особенностям использования баз знаний при решении задачи WSD в фиксированных предметных областях и структурированных массивах текстовых данных.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 09-07-97007-р_поволжье_a), и ФАО РФ, проект № 2.2.1.1/6944 «Развитие Российского научно-образовательного центра по лингвистике им. И. А. Бодуэна де Куртенэ».