

КРИТИКА И БИБЛИОГРАФИЯ

РЕЦЕНЗИИ

J. Gippert, N.P. Himmelmann, U. Mosel (eds.). Essentials of language documentation. Berlin; New York: Mouton de Gruyter, 2006. x + 424 p.

Рецензируемый сборник является введением в сравнительно новую область лингвистической науки (или, скорее, область лингвистической деятельности) – документирование языка, в первую очередь, документирование малых языков в полевых условиях. Актуальность книги определяется растущей обеспокоенностью в связи с ускоряющимися процессами вымирания языков и редукции языкового разнообразия. Многие фонды прилагают значительные усилия к тому, чтобы сохранить наследие языков и традиций, находящихся под угрозой исчезновения.

Документирование языка требует не только лингвистической компетенции, но и готовности решать разнообразные организационные, этические, правовые, технологические и другие, самые неожиданные вопросы. Поскольку область относительно молода и стремительно развивается, многие технические моменты в недалеком будущем могут существенно измениться. Сборник, однако, является прекрасным введением в проблематику, сочетая основы теории с практическими рекомендациями и разбором примеров. У книги есть веб-страница¹, на которой можно найти иллюстрации, аудиовидеоматериалы, подборку интернет-ссылок.

Два общих замечания, касающихся книги в целом. Читатель заметит, что ряд обсуждаемых в книге вопросов имеет к практике российских исследований лишь косвенное отношение и не представляет для нас особенного практического интереса – например, некоторые пункты Главы 2, посвященные правовым аспектам. Другие моменты, наоборот, крайне злободневны – например, вопрос о (не)допу-

стимости скрытой записи. В главах книги тесно перешлетены вопросы как значимые, так и второстепенные для российских ученых. Это неизбежное свойство книги, претендующей на разнородную целевую аудиторию и универсальность.

Второе замечание относится к понятию лингвистического документирования. Во многих фрагментах книги проявляется существенное концептуальное отличие этого рода деятельности от практики исследований, сложившейся в России. Наша полевая лингвистика традиционно теснее связана с лингвистической типологией, и многие исследователи во главу угла ставят разработку грамматического описания. Подход, излагаемый в рецензируемой книге, иной: главным результатом документирования являются собранные первичные материалы. Эти материалы должны сопровождаться минимальным объемом вспомогательной информации, которая позволит эффективно использовать их для дальнейших специальных исследований. Информацию, которую будущие исследователи могут относительно легко извлечь из аннотированного текста (например, надежные рамки глаголов), эксплицировать не обязательно. В грамматический очерк следует помещать информацию, которая из текстов извлекается с трудом (например, словоизменительные парадигмы). Несмотря на непривычность такой постановки вопроса, нельзя не признать, что прагматически она оправдана – грубо говоря, за то время, которое полевой исследователь потратит на описание грамматики одного вымирающего языка, он мог бы успеть задокументировать второй. По сути речь идет о разделении труда между полевыми лингвистами и авторами грамматик; эффективность такого разделения покажет время.

¹ <http://titus.uni-frankfurt.de/ld>; <http://titus.uni-frankfurt.de/ld/links.htm>

§ 1. Что такое документирование языка
и для чего оно нужно?
(N.P. Himmelmann)

Документирование языка² понимается как «долговременная многоцелевая фиксация языковых данных». В отличие от традиционной описательной лингвистики, для документирования языковые данные являются основным объектом, а не иллюстративным материалом. В идеале целью документирования является создание широкого корпуса первичных данных, который удовлетворит последующие поколения пользователей, какой бы аспект языка они ни захотели исследовать. Пользователями языковой документации могут быть как исследователи самых различных областей (лингвисты, этнографы, фольклористы), так и сами члены языкового сообщества и разнообразные организации, занимающиеся образованием и языковым планированием. Следует также иметь в виду неизвестных на данный момент потенциальных пользователей и такие задачи, которые на момент документирования еще никто не ставил. Поэтому, с одной стороны, нужно стараться избегать жесткой привязки к той или иной узкотеоретической парадигме; с другой, любой проект должен иметь теоретически обоснованный набор основных задач, чтобы не накапливать массы бесполезных данных.

Оптимальный корпус языковой документации должен содержать как традиционные «символьные», так и мультимедийные данные; пополняться из любых доступных источников; разрабатываться с расчетом на долгосрочное хранение и независимость от конкретной информационной среды (оборудования и программного обеспечения).

Новыми в данном подходе являются:

(i) Внимание к первичным данным, т.е. к (аудио- и видео-) записям языковой деятельности, и предоставление доступа к ним широкому кругу пользователей. Помимо собственно коммуникативных актов, документируются также суждения носителей языка о тех или иных лингвистических объектах, различного рода таксономии (системы родства, системы счета, морфологические парадигмы и пр.). Фиксации подлежат и сами сеансы работы, включая задаваемые вопросы и предъявляемый стимульный материал. (ii) Протоколирование сеансов записи и снабжение корпуса ме-

таданными, позволяющими ориентироваться в корпусе и оценивать качество данных (см. § 4). (iii) Забота о долгосрочном хранении первичных данных, в т.ч. о высоком качестве исходных записей и о целом ряде технологических вопросов при их обработке (см. § 4, 13–15). (iv) Создание междисциплинарных рабочих групп, включающих специалистов в смежных с лингвистикой гуманитарных областях. (v) Тесное взаимодействие с носителями языка и их активное вовлечение в работу (см. § 3).

§ 2. Этические и практические вопросы
реализации проектов полевых исследований
(A.M. Dwyer)

В главе обсуждаются этические и правовые проблемы, возникающие при планировании полевой работы и в ходе взаимодействия с консультантами и с изучаемым языковым сообществом. Автор указывает на двоякое положение исследователя, положение которого «на стыке» двух культур зачастую приводит к противонаправленным мотивациям. Так, интересы научных центров требуют скорейшего получения результата, а отношение консультантов зависит от их представлений о долгосрочности планов исследователя. Необходимо выбирать такую линию поведения, которая хотя бы отчасти уравнивает эти факторы.

Общие культурные установки и этические нормы ведения полевой работы должны быть адаптированы к изучаемому сообществу. Слепо применять универсальные правила столь же бессмысленно, как установить по всему миру единую почасовую оплату работы консультантов. Ярким примером является получение *информированного согласия* (informed consent) на запись и использование языковых материалов. Большинство западных научных учреждений продолжает требовать получения исследователем письменной расписки о том, что носитель не возражает против записи. Во многих культурах, однако, устный договор вызывает больше доверия и даже обладает большей связывающей силой, чем расписка; предложение дать расписку может привести к отказу от сотрудничества. Поэтому иногда правильнее получить устное согласие на запись, фиксируя его на аудио или видео. Указывая на конфликт между стремлением к фиксации спонтанной речи и этическими нормами, исключаящими скрытую запись, автор признает допустимым получение информированного согласия *сразу после* скрытой записи. В идеале такое согласие должно не просто подтверждать право записи на существование, но и в как можно более явной фор-

² В тексте рецензии мы будем стараться различать термины «документирование языка» (как процесс) и «языковая документация» (как совокупность его конкретных результатов).

ме оговаривать возможные способы ее использования.

Среди этических принципов упоминается широко понимаемый принцип принесения пользы исследуемому языковому сообществу – речь идет не просто об оплате труда консультанта, но, шире, об ответном даре исследователя сообществу, который он приносит в частичную компенсацию получаемого им дара языка. Этот ответный дар может принимать разные формы, главное, что направлен он не только на консультантов исследователя, но на все сообщество в целом. Здесь автор настаивает на необходимости использования собранных языковых материалов – лучше осуществить хотя бы первичную обработку данных и сделать их доступными, чем не обработать вообще; последнее тоже является неблагодарностью по отношению к языковому сообществу.

§ 3. Полевая работа и языковое сообщество (U. Mosel)

В этой главе речь идет о том, как организовать работу в языковом сообществе и привлечь к ней самих носителей языка. Для успеха проекта необходимо соблюдать баланс между академическими интересами и той отдачей, которую получают от проекта сами местные жители. Кроме того, различные социокультурные и психологические факторы могут ограничивать виды собираемого материала, способы работы и даже форма и содержание конечного продукта. Иногда целесообразно идти на компромисс, фиксируя материал в двух формах – в той, на которой настаивают носители, и в той, которая необходима лингвисту. Это касается фиксации текстов (в точном соответствии с тем, что было произнесено vs. в редакторской версии носителей), использования орфографии (более точной vs. более удобной), вносимой в словарь информации (грамматически необходимой vs. культурно значимой). В долгосрочном плане документирование может стать намного эффективнее, если в него будут вовлечены члены языкового сообщества: они могут обучать исследователя языку, записывать, расшифровывать и переводить тексты, работать над словарем. Для этого, очевидно, необходимо передать им определенные навыки: обращение с записывающей аппаратурой, ведение и организация записей на бумаге, владение базовыми лингвистическими понятиями, организация рабочего цикла и др. Обсуждаются практические приемы, полезные в начале работы с неизвестным языком.

§ 4. Языковые данные и документирование языков (P.K. Austin)

В главе описываются основные манипуляции с данными, необходимые в процессе документирования, и требования, предъявляемые к полученному корпусу документации, а также необходимый для этого инструментарий. Современный корпус должен быть цифровым, в идеале материал должен уже собираться в цифровом формате. Необходимо регулярно создавать резервные копии, чтобы в случае сбоя оборудования данные восстанавливались в полном объеме.

Основные процедуры должны быть продуманы уже на самых ранних стадиях: 1) запись (звука, видео и т.п.); 2) оцифровка аналоговых данных; 3) анализ – транскрипция, перевод, аннотация, метаданные; 4) архивирование; 5) мобилизация – публикация данных в различных формах.

При записи первостепенное значение имеет качество оборудования, в особенности микрофонов и сменных носителей. При аудиозаписи нельзя использовать сжатие с потерями (в частности, формат MP3). Перед записью важно устранить возможные источники шума, в том числе такого, который человеком обычно не осознается (шум ветра, тиканье часов, эхо от стен, шелест бумаги). Для видеозаписи могут потребоваться специальные осветительные приборы.

В последние годы сокращается жизненный цикл форматов данных. Замена старых форматов на новые, переход к новым типам носителей происходит все с большей частотой. Следует постоянно помнить о переводе имеющихся записей в новые форматы, чтобы в будущем они не оказались недоступными из-за отсутствия необходимого оборудования.

Отдельно отметим необходимость сопровождающих собранный материал метаданных (§ 4.2.2), важность которых еще не вполне осознается отечественными исследователями. Метаданные содержат сведения об источниках материала, участниках и обстоятельствах записей, содержании и структуре материала, технических параметрах. Только наличие метаданных обеспечит эффективный поиск в больших архивах³.

Автор различает три формы существования собранного материала, каждая из которых может требовать особого технического

³ Ср. поиск по метаданным в архивах программы документирования исчезающих языков DOBES (http://corpus1.mpi.nl/ds/imdi_browser/).

формата – исследование, архив, демонстрация (публикация). Задача разработчиков – обеспечить удобные способы преобразования данных из одной формы в другую.

§ 5. Язык в свете этнографии речи (J.H. Hill)

Это одна из самых коротких, но и самых увлекательных глав тома. По сути, текст является цепочкой иллюстративных примеров из практики полевых исследований и настолько захватывает, что было бы жаль лишать читателя возможности ознакомиться с ним самостоятельно. Поэтому мы не будем объяснять, почему носительницы юто-ацтекского языка науатль при встрече с Хилл пищали фальцетом, как получилось, что Билл Грейвз, исследователь пима, другого языка той же семьи, отказался от работы с самым знающим носителем языка, и что именно в выпущенном университетом Аризоны рекламном буклете словаря хопи заставило сообщество хопи так сильно изменить отношение к проекту, что словарь едва удалось опубликовать. А если вам когда-нибудь хотелось понять, почему переводчик, уже имея в кармане билет на восьмичасовой автобус, обещает встретиться с вами в девять, вам просто необходимо обратиться к тексту оригинала. Хилл выдвигает несколько принципов, опирающихся на хаймсовскую концепцию этнографии речи и имеющих непосредственное значение для эффективной полевой работы. Во-первых, следует всегда исходить из предположения, что необычные свойства речевого поведения носителей имеют определенную семантику или прагматику, а не мотивированы внеязыковыми факторами (особой физиологией, средой проживания и т.п.). Во-вторых, если вам кажется, что носители ведут себя по отношению к вам неадекватно или грубо, скорее всего вы не отдаете себе отчет в специфических для данной культуры особенностях речевого этикета и коммуникативных установок. В-третьих, необходимо четко понимать, в каких ситуациях используется тот или иной регистр или, для многоязычных сообществ, тот или иной языковой код – это поможет выбрать эффективный метод получения языкового материала. Наконец, важнейший элемент этнографического исследования речи – это выяснение отношения языкового коллектива к своему языку, модели языка как объекта.

§ 6. Документирование лексических данных (J.V. Naviland)

В главе освещаются избранные вопросы создания лексикографических баз данных, хо-

рошо известные в теоретической лексикографии, но забываемые в азарте полевой работы. Приводятся примеры разнообразных семантических областей в языках Мексики и Австралии, в т.ч. именных классификаторов, глаголов каузативного перемещения, терминов родства, междометий и восклицаний и др. Подчеркивается важность учета контекста и условий уместности употребления лексических единиц, коннотаций, сочетаемости и пр., обсуждаются различные семантические метаязыки и методы накопления лексического материала. В общем случае особое внимание следует уделять более разработанным областям лексики, сочетая опрос и наглядный стимульный материал с систематическим наблюдением за языковым поведением носителей, поскольку некоторые аспекты семантики не сводятся к различению референтов и с трудом могут быть вербализованы информантами.

§ 7. Просодия в документировании языков (N.P. Himmelmann)

В главе рассматриваются требования к собранному материалу, обеспечивающие последующий просодический анализ. Вследствие высокой вариативности просодических параметров важно иметь по несколько произнесений одних и тех же высказываний, желательно от разных дикторов (в идеале не менее двух мужчин и двух женщин). Сопоставимые высказывания получают путем предъявления одинаковых стимулов, разыгрывания одинаковых микродиалогов или произнесения клишированных высказываний.

При изучении лексической просодии следует иметь в виду, что изолированные слова являются элементарными высказываниями, поэтому необходимо отделять эффекты словесной просодии от просодии фразовой. Надежно (но трудоемко) помещать слово в рамочные предложения, чтобы получить употребления в начале, середине и конце высказывания (напр., «Он не знает слова “X”»). На практике проще группировать слова по три или четыре и записывать эти группы в разном порядке.

Одних только акустических данных недостаточно для выделения параметров, которые носители воспринимают как значимые (например, для определения места ударения). Необходимы данные о восприятии, собранные в ходе перцептивных экспериментов.

§ 8. Этнографический компонент лингвистического документирования (B. Franchetto)

Документирование языка тесно переплетено с документированием культуры; во всяком

случае лингвистическая документация должна быть пригодна для этнографического анализа. Что больше всего интересует антрополога в лингвистическом описании? Очевидно, это те элементы языка, которые напрямую связаны с окружающим миром, в первую очередь лексическая информация – значение слов, взаимосвязь различных значений слова. Лексемы полезно объединять не только по семантическому, но и по тематическому принципу (материальная культура, традиционное хозяйство и т.п.). Приводится краткий список стандартных тем, интересующих современного этнографа (мифы о происхождении, разрешение от беременности, видение языка и др.); отметим, что такой подход страдает известным недостатком – он привязан к современной парадигме и не в состоянии предсказать интересов будущих исследователей. Важнее другой принцип, не сформулированный автором, по крайней мере эксплицитно: корпус собираемых текстов должен быть максимально диверсифицированным и с лингвистической, и с социолингвистической и этнографической точек зрения.

По-настоящему захватывающим становится чтение (если употребление этого слова в данном случае правильно) главы начиная с третьего раздела. Дело в том, что все обсуждение проекта документации амазонского языка куикуро снабжено мультимедийными материалами, размещенными на интернет-сайте книги. Можно не только узнать о высоком уровне фонетической рефлексии этноса, рассуждающего о сравнительных артикуляционных и просодических свойствах своего языка в сравнении с языком соседей, но и *услышать*, как «танцуют», «подпрыгивают» или «колеблются» слова этих языков, *увидеть*, как вождь приветствует посланников, которые пришли пригласить жителей деревни на праздник, специальным напевным ритуальным речевым жанром. В целом, даже если теоретические предпосылки главы покажутся читателю несколько упрощенными, его не могут не захватить перспективы, которые открывает использование мультимедийных средств.

§ 9. Лингвистическое аннотирование (E. Schultze-Berndt)

Глава содержит обзор основных типов лингвистической аннотации, насыщена материалом и плохо поддается реферированию. Мы ограничимся тем, что наметим ее общую идеологию. Лингвистическая аннотация должна ориентироваться в первую очередь на требования компьютерной обработки текстов и общий

принцип «многослойного» представления. Центральным как для главы, так и для книги в целом является принцип *минимально достаточного* аннотирования; аннотация не столько должна содержать все разнообразие информации о языке, сколько давать будущим исследователям возможность получать эту информацию из аннотированного текста. Кроме принципа минимальной достаточности, объем разметки определяется в первую очередь временем и человеческими ресурсами, которыми располагает конкретный проект. Аннотированные материалы должны быть пригодны для использования внутри описываемого языкового сообщества, а также эксплицировать существенные элементы ситуативного контекста записи. Минимальный набор слоев аннотации – практическая орфография, перевод на языки-посредники (как минимум один из них должен быть понятен посетителям описываемого языка) и комментарии, необходимые для понимания аннотированных материалов не включенному в контекст читателю. Если фонетическая транскрипция не выводится из практической орфографии, следует ввести слой фонетической транскрипции. Необходимо, чтобы аннотация обладала формальными средствами для отражения диалога и переключения кода; желательно, чтобы она отражала элементарные супrasegmentные характеристики (интонационный контур конца интонационно-просодической единицы, резкое изменение тона, а также некоторые нелингвистические элементы – жесты, шепот, смех, паузы и т.п.).

Интересно отметить, что поморфемное глоссирование отнесено к желательным, но необязательным элементам аннотации, что находится в резком противоречии с другой распространенной точкой зрения – морфологическое глоссирование является важнейшим и обязательным, даже центральным элементом документации – точкой зрения, которой придерживаются и авторы обзора. Это отчасти объясняется различием целевой аудитории и методологического бэкграунда российской и западной полевой традиции. Для нас аннотированные тексты являются в первую очередь источником грамматической информации, должны быть ориентированы в том числе и на неспециалиста по документируемому языку, в то время как зарубежные полевые лингвисты уже давно делают акцент на предоставлении собранных материалов языковому коллективу. Позволим себе в этой связи предположить, что морфологическое глоссирование в ряде случаев позволит выявить ошибки записи даже исследователю, свободно говорящему на описываемом языке.

§ 10. Проблемы сегментирования устной речи (N.P. Himmelmann)

Здесь обсуждаются проблемы, связанные с выделением единиц трех уровней: слов, интонационных единиц и единиц следующего, более высокого уровня. При членении текста на слова чаще всего можно опираться на интуицию носителя (попросить «надиктовать» предложение). Спорные случаи носят универсальный характер (клитики, сложные слова); носители не дают однозначного ответа, и приходится формулировать независимую от интуиции систему критериев.

По мнению автора, большинство публикаций страдает недостаточной эксплицитностью принципов выделения интонационных групп. Перечислены основные признаки их границ – пограничный тон, возвращение к базовому значению частоты основного тона, паузация, замедление артикуляции в конце и ускорение в начале и др. В целом, по мнению автора, от 80 до 90% интонационных единиц отождествляются однозначно. В одной такой единице обычно вводится один новый участник или одно новое событие; отмечается, однако, что носитель или исследователь, владеющий описываемым языком, склонен приписывать паузы на основании информационной структуры и часто слышит их там, где их нет – и наоборот.

Третий тип единиц, которые кратко обсуждаются в главе – это единства, в которые объединяются интонационно-просодические единицы. Природа их плохо изучена, само существование или универсальность вызывает сомнения.

В характерном для книги прагматическом духе Химмельман рекомендует для проектов по документации малых языков использовать редуцированную транскрипцию (без инструментального измерения длительности пауз, с упрощенной классификацией пограничных тонов и др.). Глава является прекрасным практическим введением в проблематику и суммирует основные знания в этой области; читатель сможет приступить к практической сегментации записей исследуемого языка.

§ 11. Создание орфографии (F. Seifart)

Автор обсуждает принципы, проблемы и примеры разработки практической (ориентированной на носителей) орфографии малого языка. Доказывается, что в целом ряде случаев приоритет при принятии решений имеют не структурные характеристики фонетики и фонологии языка, а социолингвистические,

идеологические и даже технологические факторы (грамотность на доминирующем языке, стремление противопоставить себя иноязычной этнической группе, сложность освоения особых раскладок клавиатуры). Центральным для автора является противопоставление *deep* (глубокой) и *shallow* (поверхностной) типов орфографий, которые, по-видимому, можно обозначить как морфонологический и фонетический. Орфография должна отражать лингвистически значимые признаки лексических единиц, однако разработчик орфографии всегда стоит перед проблемой выбора уровня, на котором эти признаки действуют. Выбор более глубокого уровня может мотивироваться тем, что он может помочь избежать междиалектного орфографического конфликта (проблема, актуальная для многих регионов РФ). Однако этот выбор часто вступает в противоречие с использованием орфографической системы доминирующего языка. Нам кажется, что практически перевешивают плюсы фонетической системы. В российской реальности, во всяком случае, приоритет использования орфографии доминирующего языка преобладает над остальными соображениями, а стремление носителей языка миранья противопоставить себя более многочисленному этносу бора кажется в целом далеким от малых народов России и стран СНГ.

Существенным является понятие функциональной нагрузки фонологического признака и положение о том, что не все, а только активно эксплуатируемые противопоставления нуждаются в орфографическом отображении. Если противопоставление редко и если оно легко восстанавливается из контекста, им можно пренебречь в целях упрощения графики. Основным критерием здесь является, с точки зрения автора, количество и контекстуальность возникающих в результате фонологической неполноты орфографии омографов.

§ 12. Грамматическое описание в документировании малых языков (U. Mosel)

В главе обсуждается место кратких грамматических очерков среди других типов кратких грамматических описаний (пилотная грамматика на материале малого корпуса; вводные замечания к статье; резюме подробной грамматики, введение к словарю и пр.). Для Мозель очерк не только отражает лишь текущее, неполное состояние знаний исследователя о языке, но вообще является не грамматическим описанием языка, а инструментом для работы с текстами. Из этого следует сразу несколько важных выводов. Например,

в очерке не должны содержаться такие привычные сведения, как класс глаголов, образующих каузативы, а также падежная рамка каузативного глагола – эти сведения могут быть почерпнуты из словаря и аннотированных текстов, соответственно.

Еще одно требование, которое Мозель предъявляет к грамматическому очерку – это точность и удобство в использовании. Под точностью здесь понимается не предсказательная сила описания, а соответствие языковым фактам, которые представлены в собранных материалах, и адекватность конвенций грамматической аннотации: под удобством – пояснение по поводу неочевидных категорий, расшифровка всех используемых в аннотации сокращений, а также четкая структурированность текста. Вообще, в этой главе как ни в какой другой проявляется пока не очень привычное для нас понимание документирования как сбора материалов, не обязательно предполагающего написание грамматики. Тем не менее, условие прозрачности структуры приводит к таким минимальным требованиям, которые все же делают очерк похожим на модель обычного грамматического описания, пусть в миниатюре (примечательны, однако, уточнения в духе «для упрощения понимания глосс» и «для более быстрой интерпретации высказываний»).

§ 13. Трудности архивного дела (P. Trilsbeek, P. Wittenburg)

По оценкам специалистов, около 80% материалов, собранных по исчезающим языкам и культурам, находится в руках частных лиц или рабочих групп по краткосрочным проектам, хранятся в плохих условиях и на низкокачественных носителях. Не случайно, что в масштабных проектах по документированию языков (например, DoBeS), создание централизованного архива считается одной из главных задач. В главе рассказывается о принципах и трудностях ведения архивов.

По сравнению с традиционными архивами, цифровые архивы в меньшей степени привязаны к сохранению физических носителей – хранить требуется саму информацию. Однако возникают и новые задачи: пополнение, удобный доступ с поиском, редактирование. Кроме противоречия между задачами долгосрочного хранения и быстрого доступа возможны столкновения интересов трех групп – собирателей данных, хранителей (архивариусов) и пользователей. Так, полевым исследователям необходимо компактное оборудование, тогда как для архивации нужны записи очень высокого качества, что требует громоздких технических инструментов.

Важно обеспечивать не только сохранность файлов, но и возможность без потерь восстановить хранящуюся в них информацию. Это связано с целым рядом вопросов, касающихся применяемых форматов и кодировок (как текстовых, так и аудио- и видеоформатов), включая перевод (миграцию) в более современные форматы.

§ 14. Документирование языков и представление текстовых материалов (J. Gippert)

В главе рассказывается о проблемах электронного представления текстов – кодировки символов, разметка структуры текста и т.п. Современным стандартом кодирования символов, в котором каждому символу каждой письменности (включая слоговые азбуки, иероглифические системы, транскрипцию МФА и др.) сопоставляется уникальный числовой код, является Юникод (Unicode). Во избежание проблем, вызванных смешением предшествующих Юникоду 8-битных «кодовых страниц», в которых одни и те же 256 кодов использовались для разных наборов символов, все лингвистические архивы должны быть переведены из устаревших кодировок в Юникод. Автор допускает неточность, называя Юникод 16-битной кодировкой и сетуя на недостаточно большое количество символов в нем ($2^{16} = 65\,536$); а стандарт ISO 10646, по его словам, является 32-битным и содержит коды, более чем 4 миллиардов кодов символов. Эта информация устарела задолго до выхода книги: начиная с Unicode 2.0 (1996 г.), объемы Юникода и ISO 10646 идентичны и включают 1 114 112 символьных кодов (текущая версия – 5.1. 2008 г.).

Существенным компонентом содержания документа является его структура – членение на слова, предложения, абзацы и т.п. Обычные средства визуального оформления – пробелы, интервалы, отступы, шрифтовые выделения – крайне ненадежны, поскольку применяются от текста к тексту непоследовательно, а машинное их представление привязано к конкретному текстовому редактору. Перспективным решением представляется миграция в формат XML (см. также § 5) – язык разметки, позволяющий разделить описание структуры данных и описание их визуального оформления: одно может храниться независимо от другого. Преимуществами являются разработанные для XML средства преобразования (XSLT) и наличие бесплатных программ для работы с ним.

§ 15. «Тяжелые» интерфейсы:
мультимедийная мобилизация языковой
документации
(D. Nathan)

Заключительная глава написана директором Архива исчезающих языков SOAS Дэвидом Натаном и обсуждает роль информационных технологий в мобилизации собранных материалов. Если первоначально документирование рассматривалось как чисто академическая деятельность, то теперь члены языкового сообщества вовлекаются в него в качестве активных и полноправных участников. Новая задача, которая пока только поставлена: сделать результаты работы доступными и полезными для языкового коллектива. Для этого необходимо учитывать предпочтения членов сообщества и доступные им технические ресурсы, работать в тесном сотрудничестве с ними. Многие мультимедийные аспекты (дизайн ин-

терфейса, оформление) следует разрабатывать при их участии.

Существующие программные средства, которыми пользуются полевые лингвисты, по большей части представляют собой тонкий интерфейс: их структура определяется структурой данных, и пользование ими требует не только навыков работы с компьютером, но и познаний в предметной области. Такие интерфейсы не годятся для языкового сообщества. Необходимы мультимедийные продукты, ориентированные не на исходные данные, а на задачи, которые будут решать неспециалисты, они должны стимулировать пользователей к самообучению и исследованию ресурса; явным образом отражать участие членов сообщества; в полной мере использовать возможности передачи звука и изображения; иметь целостный, рациональный и адекватный с точки зрения культуры дизайн.

А.В. Архипов, М.А. Даниэль