

UNESCO



Организация Объединенных Наций
по вопросам образования, науки и культуры



Всемирный саммит по информационному обществу

Измерение языкового разнообразия в Интернете

Измерение языкового разнообразия в Интернете

UNESCO Publications for the World Summit on the Information Society



United Nations
Educational,
Scientific and
Cultural Organization

2005

Measuring Linguistic Diversity on the Internet

A collection of papers by:

**John Paolillo, Daniel Pimienta,
Daniel Prado, et al.,**

Edited with an introduction by
UNESCO Institute for Statistics
Montreal, Canada

Комиссия Российской Федерации по делам ЮНЕСКО
Федеральное агентство по культуре и кинематографии
Российский комитет Программы ЮНЕСКО «Информация для всех»
Межрегиональный центр библиотечного сотрудничества

Измерение языкового разнообразия в Интернете

Сборник статей

**Авторы: Джон Паолилло, Даниэль Пимьента,
Даниэль Прадо и др.**

Москва
2007

УДК 81'272:004738.5

ББК 81.21(0)

И 33

Перевод с английского: *Малявская Е.В.*

Редактор русского издания: *Мурована Т.А.*

Введение и общая редакция на англ. яз.:
Институт статистики ЮНЕСКО, Монреаль, Канада

И 33 **Измерение языкового разнообразия в Интернете:** Сборник статей. – М.: МЦБС, 2007. – 118 с.

Издание на русском языке подготовлено по инициативе Российского комитета Программы ЮНЕСКО «Информация для всех» и Межрегионального центра библиотечного сотрудничества (МЦБС) при финансовой поддержке Федерального агентства по культуре и кинематографии. Публикуется с разрешения ЮНЕСКО.

УДК 81'272:004738.5

ББК 81.21(0)

Организация Объединенных Наций
по вопросам образования, науки и культуры
7, place de Fontenoy, 75352 Paris 07 SP

(CI-2005/WS/06 CLD 24821)

ISBN 978-5-91515-011-8

© ЮНЕСКО, 2005

© МЦБС, перевод и издание на русском языке, 2007

Содержание

1.	Введение – Институт статистики ЮНЕСКО	7
2.	Модели и подходы	
	а) <i>Языковое разнообразие в киберпространстве – модели развития и измерения – Даниэль Пимьента</i>	15
	б) <i>Политический и юридический контекст – Даниэль Прадо</i>	37
3.	Языковое разнообразие в Интернете: анализ языкового сдвига. Джон Паолилло	45
4.	Альтернативные точки зрения	
	а) <i>Языковое разнообразие в Интернете: ситуация в Азии – Йошики Миками и др.</i>	99
	б) <i>Несколько слов об африканских языках в Интернете – Ксавьер Фантоньян</i>	113
	Об авторах	117

Введение

ЮНЕСКО всегда признавала значение концепции «обществ знаний» в деле преодоления цифрового разрыва и формирования инклюзивного информационного общества. В основе этой концепции лежит не глобальное единообразие, а плюрализм и разнообразие, а ее важной составляющей является многоязычие, необходимое для обеспечения культурного разнообразия и присутствия всех языков в киберпространстве. Следует, однако, помнить, что в попытке преодолеть цифровой разрыв «на обочину дороги» могут попасть сотни малых языков и наречий. Этим соображением объясняется то значение, которое придается языковому разнообразию и локальному контенту в Плане действий Всемирного саммита по информационному обществу (ВСИО), ответственность за координацию которого лежит на ЮНЕСКО.¹

Языковое разнообразие в Интернете стало основным вопросом в ходе обсуждения различных аспектов информационного общества, а сами дискуссии принимали подчас неожиданные формы. На первый взгляд, вопрос языкового разнообразия связан с разными сообществами людей, использующими Интернет, который позволяет этим сообществам общаться друг с другом на родном языке. Однако тут же возникает и другой вопрос: по каким каналам осуществляется коммуникация в Интернете? Всемирная сеть – это набор источников информации с невысоким уровнем интерактивности. Пользовательские форумы и электронная почта обеспечивают более высокий уровень обмена информацией, однако данных о языках, используемых в общении по электронной почте или на форумах, явно недостаточно (более подробно об этом – в работах Паолилло и Сью Райт, указанных в списке литературы).

1 См. Коитиро Мацуура (Koichiro Matsuura), Генеральный директор, Обращение ЮНЕСКО к Постоянным делегациям на ВСИО, 8 июля 2005 г.

Для проведения языкового анализа ученые обращаются к веб-страницам, но здесь, как и во всех остальных видах коммуникации, нельзя не учитывать характер аудитории. Веб-страницы читают только те, кто имеет доступ к Интернету. Тогда получается, что, если даже и можно обеспечить языковое разнообразие за счет создания веб-страниц на «исчезающем» языке какого-то уединенно живущего племени, то очень немногие прочитают эти страницы, поскольку маловероятно, что члены племени имеют выход в Интернет. Веб-страницы о языке данного племени, написанные на одном из интернациональных языков, могут сыграть более важную роль, привлекая внимание к культурной ценности данного языка и, возможно, помогая его носителям. Кроме того, они могут способствовать сохранению вымирающих языков.

Статьи, представленные в данном сборнике, посвящены различным техническим проблемам измерения языкового разнообразия в Интернете. Легко получить приблизительные данные о количестве веб-страниц, используя коммерческие поисковые машины, но полученные данные не позволят понять, как часто обращаются люди к этим веб-страницам и принесло ли им чтение этих страниц какую-либо пользу. Следует также обратить особое внимание на то, что результаты поиска, представленные на разных языках, не должны отличаться друг от друга по ценности, значению и использованию (см. статью Даниэля Пимьенты).

Языки и информационное общество

Институт статистики ЮНЕСКО придерживается такого подхода к измерению информационного общества, который выходит за рамки технических проблем и включает вопросы социальной роли Интернета и других каналов распространения информации. Такой подход предполагает решение проблем огромной важности, а именно:

- стандартизации терминологии для достижения международной совместимости;
- идентификации показателей по экономически развитым и развивающимся странам;

– создания необходимых мощностей на национальном и международном уровнях для обеспечения постоянного сбора качественных данных.

Язык – это средство, с помощью которого происходит обмен информацией в обществе. Язык – это базовое средство коммуникации, основа самовыражения каждого индивида и общества как в устной, так и в письменной форме. ЮНЕСКО считает обучение на родном языке правом каждого ребенка. ЮНЕСКО поддерживает языковое разнообразие, чтобы сохранить для каждой страны и мира в целом богатство культур, основанное на этом разнообразии.

Культурный аспект представления языков в Интернете отражает значение тех вопросов информационного общества, которые связаны с информационными технологиями и их влиянием. Институт статистики ЮНЕСКО хотел бы продемонстрировать свою точку зрения, которая поднимает вопросы важности контента и условий его развития, указывая в то же время на технические проблемы измерения культуры и контента в информационном обществе.

Статьи в данном сборнике отражают разные взгляды на обозначенную проблему. Статья профессора Джона Паолилло представляет точку зрения профессионального лингвиста, работающего в англоязычном мире. Статья состоит из четырех основных разделов. Раздел I посвящен этическим вопросам оценки сдвига в компьютерных системах и соотносит эти вопросы со статусом мировых языков в Интернете. В разделе II анализируется ранее существовавший сдвиг в Интернете и его проявления в настоящее время; для анализа сложившейся ситуации используются статистические данные по развитию Интернета и их корреляция с общемировым языковым разнообразием. Раздел III изучает вопросы, связанные с языковыми сдвигами, которые появляются в Интернете. В разделе IV анализируются сдвиги в технических системах Интернета.

Статья Паолилло дополняется серией более коротких статей, которые освещают ситуацию в контексте неанглоязычных стран. Эти статьи подобрал Даниэль Пимьента из FUNREDES – неправительственного проекта, разработавшего систему подсчета языков с позиции романских языков. Пимьента освещает ситуацию с точки зрения неправительственной организации как части гражданского общества, рассказывает о барьерах, стоящих на пути локальных групп населения при

организации доступа к Интернету, и приводит краткий перечень имеющихся на сегодняшний день показателей. После его статьи приведена работа Даниэля Прадо, представляющего отношение романоязычного сообщества к ощутимому преобладанию английского языка. В числе более коротких статей – очень интересная работа Миками и др., представляющая взгляд стран Азии на сложившуюся ситуацию, а также работа Ксавьера Фантоньяна, посвященная анализу африканских языков.

Статьи, приведенные в данном сборнике, не дают окончательного ответа на вопрос, как следует измерять языки в Интернете, но они призваны развеять многие мифы, связанные с ранее опубликованными данными. Все они говорят о том, что простой подсчет веб-страниц недостаточен и что провайдеры Интернет-услуг, как и государства, должны приложить немало усилий для сохранения и развития мировых языков в киберпространстве. Каждый из представленных авторов высказывает ряд предложений по изменению сложившейся ситуации.

Обзор ситуации, связанной с языковым разнообразием в Интернете

Языковое разнообразие можно анализировать с разных позиций. На английском языке говорят практически одинаково в тех странах, в которых он является основным, а в Папуа – Новой Гвинее говорят на более чем 830 языках. Жители англоязычных стран могут говорить и на других языках, но вряд ли найдется еще страна, которая могла бы сравниться с Папуа по языковому разнообразию. Число носителей романских языков, включая граждан США, в два раза превышает количество людей, для которых родным языком является английский (см. статью Даниэля Прадо), но контроль большей части технических средств Всемирной сети осуществляется США (см. статьи Джона Паолитто, Миками и др.). Взаимоотношения между языками в Интернете и разнообразие языков внутри стран говорят о том, что даже при наличии глобальной сети национальные государства играют определенную роль в стимулировании языкового разнообразия в киберпространстве. Языковое разнообразие можно анализировать по ситуации как в отдельной стране, так и в Интернете в целом.

Общеизвестно, что преобладающим языком Интернета является английский язык. Статьи, представленные в настоящем сборнике, вы-

ражают разные точки зрения на эту ситуацию. Джон Паолилло принимает ее и считает, что это – проблема, как полагают и многие другие, признающие доминирование английского языка. По оценкам Даниэля Пимьенты английский охватывает около половины всех веб-страниц и его преобладание снижается по мере того, как другие страны и языковые группы увеличивают свое присутствие в Сети. Паолилло говорит о доминировании США в системах, поддерживающих Всемирную сеть, как коммерческих, так и регулятивных, и подчеркивает, что эти системы обеспечивают существование Сети. Миками поддерживает Паолилло в этом вопросе и делает акцент на сложностях, возникающих в процессе синхронизации американских и западноевропейских информационно-коммуникационных технологий и лингвистических преобразований иероглифов стран Азии. Правда, в статьях Миками и Пимьента есть намек на грядущие изменения в связи с экономическим ростом Индии и Китая. Такое разделение мнений по вопросу доминирования английского языка и будущего языков в Сети можно рассматривать как иллюстрацию разных вариантов оценки использования языков в Интернете, для которой пока еще не выработали надежных статистических показателей, несмотря на мириады платформ ИКТ. Другими причинами отсутствия показателей являются недостаток управления и феноменально быстрый рост Интернета. Пимьента высказывает мнение, что область Интернет-показателей охватывает, в основном, сферу торговли и потому нуждается в качественном научном изучении.

Паолилло возражает, говоря, что телекоммуникационные компании, получающие прибыль от спроса на коммуникационные и технические услуги, несут особую ответственность за языковое разнообразие в тех странах, на рынках которых они работают. Компании-производители технических и программных средств также влияют на языковой состав Интернета, производя компьютеры с клавиатурами, мониторами и операционными системами, работающими на строго определенных языках. Действия компьютерных компаний, зажатых конкуренцией за доминирование на рынке, оказывают разрушительное действие на развитие многоязычия и онлайн-языкового разнообразия. В таких условиях этнолингвистическое сознание телекоммуникационных и компьютерных компаний и организаций, отвечающих за работу Интернета, начинает развиваться только тогда, когда на рынке создается критическая масса недостаточно представленных этнолингвистических групп. Следовательно, основной вопрос о появляющемся языко-

вом сдвиге нуждается во внимательном изучении на глобальном, региональном и местном уровнях.

Измерение языков в Интернете можно использовать как парадигму для различных вопросов измерения контента. Проще говоря, если мы не в состоянии измерить этот кажущийся несложным параметр контента Всемирной сети, то что вообще мы можем измерить? Однако не стоит поддаваться пессимизму. Проект Миками обладает большим потенциалом для решения многих технических проблем, обозначенных в ранее публиковавшихся материалах, и открыто провозглашает точкой отсчета другой язык, не английский.

Нам необходимо заняться разработкой более интеллектуально сложных показателей. Оценка языков по общему количеству представленных в Сети страниц становится все более трудной задачей в связи с огромным объемом информации, но само по себе присутствие веб-страницы в Сети еще не говорит о ее использовании или даже о ее «посещении». Если мы действительно хотим измерить истинное влияние информационного общества, нам нужна статистика о том, кто и как пользуется Интернетом. С этой точки зрения веб-страницы выступают как ресурс, лингвистически однородный или разнообразный, но не отражающий спрос и потребление. На рынке, пересыщенном, скажем, англоязычными веб-страницами, и предлагающем самые разнообразные услуги, много низкосортных сайтов, которые посещаются редко или не посещаются совсем. Также не секрет, что большое число сайтов годами не обновляется и не подвергается никаким изменениям.

С экономической точки зрения Всемирная сеть обладает некоторыми чертами и недостатками свободного рынка (Паолилло). Веб-сайты разрабатываются с учетом потребностей определенной аудитории. При низком уровне доступа к Интернету внутри страны бизнес-сайты будут ориентироваться на внешний рынок и поэтому создаваться на одном из международных языков, например, английском. С другой стороны, низкое использование Интернет-сайта и низкая стоимость его технического обслуживания означают, что он будет существовать и регистрироваться поисковыми машинами еще долгое время после последнего посещения пользователем. В идеале нам нужен анализ «полезных» сайтов и их посетителей.

Даже с учетом всех недостатков имеющихся исследований указанные проблемы говорят о том, как мало информации о фундаментальных

изменениях в информационном обмене, вызванном информационным обществом, несут статистические данные о процентном соотношении людей, имеющих компьютеры, или о количестве зарегистрированных пользователей Интернета (оба параметра входят в число показателей Целей развития тысячелетия). Если оставить в стороне аргументы «за» и «против» доминирования английского языка, то мы увидим быстрый рост использования Интернета в Азии и, как следствие, рост числа веб-сайтов на языках стран Азии (Миками), а также то, как расширение Всемирной сети объединило романоязычные сообщества и заставило их пересмотреть свое место в глобальном информационном обществе (Прадо). Важно подчеркнуть, что цифровой мир предоставляет стимулирующую среду максимально возможному числу языков, обеспечивая тем самым для них подлинное цифровое участие.

Следующие действия

Мы надеемся, что наш сборник подчеркнет необходимость объединения усилий всех участников как на национальном, так и на международном уровне. Всемирный саммит по информационному обществу предоставляет хорошую платформу для обсуждения технических стандартов, языковой политики и ее целей, достичь которых можно за счет стимулирования более свободного обмена информацией.

Исследования показывают, как важно понимать роль культурного контекста для информационного обмена. Поэтому нам кажется маловероятным, что какой-либо глобальный проект в состоянии предоставить сопоставимые или политически релевантные данные, которые бы в достаточной степени соответствовали стоящим перед нами техническим и политическим задачам. Предпочтительнее поручить мониторинг каким-нибудь региональным организациям, исследования которых можно было бы впоследствии объединить на глобальном уровне. FUNREDES и Обсерватория Миками являются двумя потенциальными проектами, показывающими, как может функционировать подобная региональная сеть.

В заключение, как правильно отмечает в своей статье Джон Паоллилло, со стороны многих лингвистически разнообразных и развивающихся стран мира могут потребоваться действия в поддержку цифрового доступа и цифровой грамотности.

ЮНЕСКО рекомендует национальным, региональным и международным организациям работать в тесном сотрудничестве для создания необходимых ресурсов, принятия мер по устранению языковых барьеров и развития взаимодействия в Интернете путем создания и обработки образовательного, культурного и научного контента в цифровом формате и организации доступа к нему с целью предоставления всем культурам возможностей для самовыражения и получения доступа к киберпространству на всех языках, включая языки малочисленных народов.²

2 См. Рекомендацию ЮНЕСКО по развитию и использованию многоязычия и всеобщему доступу к киберпространству и документ ЮНЕСКО 32 С/27 (2003 г.) «Всеобщая декларация о культурном разнообразии».

Прим. ред. рус. изд.: на русском языке эти документы ЮНЕСКО размещены на веб-сайте Российского комитета Программы ЮНЕСКО «Информация для всех» <http://www.ifarcom.ru/ru/387>

Модели и подходы

а) Языковое разнообразие в киберпространстве – модели развития и измерения

Даниэль Пимьента, FUNREDES

Введение

Есть слово, используемое субъектами гражданского общества, в особенности теми, кто считает, что суть новых парадигм обществ знаний в демократической системе заключается в этике процессов, способное передать наше видение цели. Слово это – **соответствие**.

Соответствие между словом и делом – вот то, что позволяет нам верить обещаниям и прощать ошибки. С точки зрения процесса – это возможность учиться на ошибках, извлекать из них уроки и продолжать верить. Этот процесс, имеющий свою специфику для экспериментальных исследований, адаптированный к изучению вопросов развития, стал движущей силой настоящего документа. Цель документа – не столько в том, чтобы найти решение такого сложного вопроса как языковое разнообразие в Интернете, сколько в том, чтобы усомниться в недостоверных данных, представить провокационные точки зрения и обозначить дальнейшие идеи и действия, берущие свое начало на проторенных путях и в известных постулатах и способные охватить весь комплекс сложных вопросов. Такое возможно при наличии смирения у исследователя, действующего методом проб и ошибок, в сочетании с уверенностью человека действия, активно работающего в конкретной области.

В настоящем документе соответствие будет проявляться разными путями:

- в выборе общения на родном языке, что является неотъемлемым правом;
- в желании содействовать разнообразию, воплотившемся в выборе людей, известных в конкретной области, приглашенных для того, чтобы выразить свое мнение. Мы стремились максимально охватить географические регионы, культуры, языки, историю, сектора, возрастные группы и представителей обоих полов. Мы уже поняли, что не сумели достичь полного успеха: так, мы сожалеем, что пространство для женщин-авторов не стало больше, но соответствие проявляется, прежде всего, в искренности наших намерений;
- в решении не использовать язык дипломатии, отважиться на провокационные заявления, никогда не быть необоснованными, иногда – благодарными, всегда опираться на опыт и желание расширить кругозор людей.

Структурированный подход к интеграции ИКТ и развития человеческого потенциала

Концепция «цифрового разрыва» стала очень популярной. Она стимулирует к размышлениям и проведению большого числа международных встреч. Однако согласованная позиция гражданского общества (Пимьента, 2002, Communauté MISTICA, 2002) состоит в том, что нельзя игнорировать другие аспекты разрыва и следует избегать упрощенческого подхода, который считает технологии источником всех бед. Ниже мы представляем достаточно интересную таблицу использования ИКТ в интересах развития. Эта таблица показывает, что решение проблемы цифрового разрыва кроется не только в доступе к технологиям, но и в вопросах языка.

Эта таблица составлялась для того, чтобы показать последовательность преград, стоящих на пути использования ИКТ для развития человеческого потенциала. Таблица предполагает дальнейшее продолжение списка преград – от инфраструктуры до информационной культуры через инфраструктуру. Вполне возможно, что такое продолжение не будет

четко соответствовать реалиям жизни отдельного человека или социальной группы и что последовательность факторов зависит от индивидуального контекста. Тем не менее, по практическим и педагогическим соображениям мы упростили эту сложную реальность и представили ее как последовательность преград, преодолеть которые необходимо для выхода на следующие уровни.

**Таблица 1. ИКТ для развития – длинный и тернистый путь к сред-
ствам развития человеческого потенциала**

Уровень использования	Виды использования и преграды	Языковые вопросы
ДОСТУП К ИКТ	<i>Возможность для человека или группы обладать физическими средствами пользования ИКТ</i>	
	<p>Для получения доступа нужно преодолеть множество преград, и этот процесс происходит в несколько этапов:</p> <p>– Наличие инфраструктуры</p> <p>Услуги: провайдеры доступа к ИКТ и телекоммуникационным сетям, емкость которых позволяет обслуживать пользователей, поддерживая приемлемые показатели по времени ответа и скорости передачи при перегрузке сети.</p> <p>Пользователи: компьютерное оборудование, необходимое для доступа и обладающее требуемыми характеристиками для обеспечения приемлемой работы. Это касается индивидуальной (ПК) и коллек-</p>	<p>– Наличие инфраструктуры</p> <p>Интерфейсы должны обеспечивать доступ на родном языке пользователя и в форме, которая адаптирована к его культуре.</p> <p>В отношении технических средств вопрос языка касается клавиатур и программ, позволяющих управлять символами, ассоциированными с языком и кодифицированными для компьютерной обработки.</p> <p>Однако операционный аспект программного обеспечения, относящийся к язы-</p>

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>тивной базы (телецентры или Интернет-кафе).</p> <p>– Приемлемый доступ к инфраструктуре</p> <p>Стоимость доступа к инфраструктуре должна быть приемлемой для пользова-</p>	<p>ку, не ограничивается кодировкой. Для оптимальной работы на определенном языке издательские программы должны иметь базу и словари для проверки правописания и грамматики. Более амбициозная и долгосрочная цель предполагает использование программ машинного перевода на операционном, а не на прикладном уровне. Нужно проделать огромную работу по совершенствованию программ перевода и включить в них не только так называемые доминирующие языки. Это вполне выполнимая задача для программ с открытым исходным кодом, но, к сожалению, практически никакая деятельность в этой области не ведется, и разработчикам нужен стимул, чтобы заполнить этот пробел.</p> <p>ICANN (Webopedia, 2005b) рассматривает сейчас возможность введения доменных имен в Интернете на всех языках (Wikipedia, 2005a).</p> <p>– Приемлемый доступ к инфраструктуре</p> <p>Принцип всеобщего доступа должен учитывать стоимость доступа, которая</p>

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>теля. Существует много прямых и косвенных факторов в уравнении цен³, и индивидуальный или коллективный доступ обладают разными параметрами.</p> <p>Достаточно, например, сравнить стоимость доступа ADSL (Webopedia, 2005a) (от 10 до 50 долл. США/мес.) и зарплаты в социальной пирамиде, чтобы понять, что это больше, чем годовая зарплата значительной части человечества (тех, кто живет за чертой бедности), месячная зарплата для другой значительной части населения (живущей на юге), около 10 % месячной зарплаты среднего класса в развивающихся странах и порядка 1 % месячной зарплаты в развитых странах.</p> <p>Итак, первый разрыв – не цифровой, а экономический и социальный.</p>	<p>должна соответствовать уровню дохода определенной группы населения.</p>

- 3 Прямые расходы, такие как стоимость компьютера, расходы на поставщика Интернет-услуг, в отдельных случаях расходы на информационного провайдера, хостинг и Интернет-домен (поскольку доступ означает также создание контента); или не прямые расходы, такие как сбережения, позволяющие иметь доступ – например, IP-телефонию, или расходы на поездки, на техническую поддержку и подготовку персонала.

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>Решение двух первых вышеперечисленных проблем должно⁴ соответствовать тому, что МСЭ и регуляторы телекоммуникаций называют «всеобщим доступом» (МСЭ, 2003). Но если и считать всеобщий доступ обязательным условием для преодоления цифрового разрыва, то его совсем недостаточно.</p> <p>– Функциональная грамотность</p> <p>Предполагается, что человек, пользующийся инфраструктурой, обладает функциональной способностью читать и писать на родном языке. Это, пожалуй, второй разрыв, который необходимо преодолеть, когда мы утверждаем, что у нас есть «Интернет для всех».</p>	<p>– Функциональная грамотность</p> <p>Мультимедийный компонент ИКТ следует, конечно, принимать во внимание при адаптации интерфейсов в помощь неграмотным людям. Однако мы должны опираться на факты, когда речь идет о доступе к знаниям, а не просто о доступе к технологиям. Для неграмотного населения функциональная грамотность – задача более приоритетная, чем доступ к технологиям.</p> <p>При этом следует помнить о существовании языков,</p>

4 Мы используем слово «должно», потому что экономический аспект слишком часто игнорируется в планах всеобщей доступности, а концепция понимается как тотальный физический охват доступа к инфраструктуре, что, естественно, является задачей для поставщиков технических средств, но не обязательно для пользователей.

Уровень использования	Виды использования и преграды	Языковые вопросы
ПОЛЬЗОВАНИЕ ИКТ	<p>– Компьютеризация алфавита</p> <p>Родной язык человека, пользующегося инфраструктурой, должен быть переведен в цифровую форму. Чтобы это стало возможным, язык должен иметь письменную форму, а буквы его алфавита должны легко кодироваться. К сожалению, это не такая простая задача для большинства языков, которыми мы пользуемся.</p> <p><i>Потенциальная возможность сделать пользование ИКТ эффективным (т. е. позволяющим достигать поставленных целей) и экономным по времени.</i></p>	<p>имеющих только устную форму и не имеющих письменности. Для таких языков цифровой мир может стать фатальной преградой, если только их письменная форма не будет кодифицирована.</p> <p>– Компьютеризация алфавита</p> <p>Это все еще основная преграда для значительного числа языков и ее следует считать главным приоритетом с самого начала работы. Работа осуществляется в рамках Юникод (Wikipedia, 2005b), и ее следует продолжать и развивать.</p>
	<p>Для этого человек должен обладать разнообразными возможностями по управлению цифровыми средствами, понимать концепции, методы и культуру, ассоциируемую с киберпространством. Нельзя недооценивать спектр компетенций, необходимых</p>	<p>Цифровая грамотность</p> <p>Широкомасштабная деятельность, необходимая для достижения цифровой грамотности, должна быть продумана и реализована на родном языке и в контексте культуры определенной группы населения. Важно отметить,</p>

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>для достижения цифровой грамотности.</p> <p>Изучение киберпространства, которое должно включать не только простую подготовку к работе с определенными компьютерными программами, но также и целостное видение социальных последствий⁵ использования ИКТ и их влияния на развитие. Эта задача – крепкий орешек, самый сложный и самый игнорируемый аспект процесса достижения консенсуса при сокращении цифрового разрыва.</p> <p>В отличие от широко распространенного взгляда, три столпа информационного общества, над построением которого мы работаем, – не телекоммуникации, технические и программные средства, а этика информации, образование и участие.</p>	<p>что данное обязательное требование следует применять и к интерфейсу компьютерных приложений для государственных учреждений и организаций.</p>
ТЕХНОЛОГИЧЕСКОЕ ОСВОЕНИЕ ИКТ	<p><i>Уровень, на котором пользователь настолько квалифицирован, что технология становится незаметной в процессе его индивидуальной работы</i></p>	

5 Политические, экономические, социальные, культурные, лингвистические, организационные, этические, биологические и психологические последствия.

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>Хорошим примером в этом смысле являются очки – оптический прибор, который одевают на нос утром и забывают до вечера; применительно к ИКТ – это разговор по телефону, когда человек не думает постоянно о нем, как о средстве, позволяющем вести диалог на расстоянии.</p> <p>В отношении ИКТ такое освоение требует более развитых умений работы с ПК и компьютерными приложениями, которые составляют часть процесса. Поэтому очевидно, что для достижения данного уровня требуется квалификация, позволяющая анализировать информацию, общаться по электронной почте и просто правильно вести себя в виртуальном сообществе.</p> <p>В дополнение к хорошему компьютерному образованию для данного уровня необходим хотя бы минимальный практический опыт.</p>	<p>Как сделать технологию прозрачной, если доступ к ней требует использования языка, который отличается от родного языка человека? Постановка этого вопроса усиливает значение аргументов в пользу освоения предыдущих уровней.</p>
СМЫСЛОВОЕ СОДЕРЖАНИЕ	<p><i>Возможность пользоваться ИКТ имеет социальное значение для человека в личном, рабочем и общественном контексте</i></p>	

Уровень использования	Виды использования и преграды	Языковые вопросы
	<p>Для человека это означает выйти за рамки развлекательного использования ИКТ в качестве только средства коммуникации и нацелить их использование на собственное развитие.</p> <p>Именно на этом уровне фундаментальные возможности должны выйти на первый план, чтобы человек из простого потребителя превратился в создателя контента или виртуальных сообществ.</p> <p>Понятно, что для достижения этого уровня развития человеку требуется специальная подготовка.</p>	<p>На этом уровне язык имеет очень большое значение, т. к. предоставляет возможность и мотивирует человека к созданию контента и локальных виртуальных сообществ. На этом этапе встает вопрос «многоязычия и «наведения мостов» между языками.</p>
ОБЩЕСТВЕННОЕ ОСВОЕНИЕ ИКТ	<p><i>Уровень, на котором пользователь настолько квалифицирован, что технология становится незаметной в процессе общественного использования</i></p>	
	<p>Этот уровень требует четкого понимания общественного значения использования ИКТ для развития, а также культурных и этических последствий такого использования ИКТ (культуры/этики сети и информации, знания методических аспектов, связанных с продуктивным использованием ИКТ для развития).</p>	<p>Этический и культурный аспекты сетей не всегда нейтральны и должны быть пропущены через некий кросс-культурный фильтр (определенного вида синкретизм) на основе локальной культуры и этических норм. Язык, как часть культуры, не свободен от сложных и болезненных вопросов.</p>

Уровень использования	Виды использования и преграды	Языковые вопросы
ВООРУЖЕНИЕ ЗНАНИЯМИ ⁶	<p>Помимо хорошего компьютерного образования на данном уровне необходим еще практический опыт, направленный на развитие.</p> <p><i>Уровень, на котором человек и/или сообщество способен/способно трансформировать общественные реалии, благодаря общественному освоению ИКТ для целей развития</i></p>	
СОЦИАЛЬНЫЕ ИННОВАЦИИ	<p>Этот уровень требует не только наличия способностей, но и их практического применения, как индивидуального, так и коллективного. Осуществить практическое применение невозможно без опоры на ценности, которые ассоциируются с культурой Интернета и культурой информации, на направленность на коллективную работу, на достигнутый уровень «прозрачности» ИКТ и на активное участие в коммуникации.</p> <p><i>Уровень, на котором в деятельности по трансформации общественной реальности заложены</i></p>	<p>Чем ближе мы подходим к последнему звену в цепочке, протянувшейся от доступа к развитию, тем яснее становится значение как культуры, так и языка.</p> <p>Что такое «вооружение знаниями» и как оно проявляется в каждой отдельной культуре?</p>

6 Это словосочетание означает получение возможностей и их освоение, а также получение преимуществ от умения пользоваться этими возможностями.

Уровень использования	Виды использования и преграды	Языковые вопросы
	<i>оригинальные решения, автором которых является человек или сообщество</i>	
	В новой парадигме работы в сети заложены основы инноваций, социальных инноваций в частности (в виде новых организационных форм, новых решений известных проблем и т. п.).	Что такое «инновации» и как они проявляются в каждой отдельной культуре?
РАЗВИТИЕ ЧЕЛОВЕЧЕСКОГО ПОТЕНЦИАЛА	<i>Уровень, на котором человек или сообщество свободны в своем выборе, и этот выбор существует в форме «потенциальных возможностей»⁷</i>	
	<p>Это завершающий этап процесса, но не следует забывать, что в конце каждого социального процесса стоит только то, чего добились в ходе осуществления этого процесса, с самого его начала.</p> <p>Именно поэтому варианты, предоставляемые свободой, нельзя претворить в жизнь, если участие людей и сообществ не было подлинным в ходе всего процесса.</p>	<p>Свобода выбора в форме «потенциальных возможностей».</p> <p>Что такое «участие» и как оно проявляется в каждой культуре? Возможно ли подлинное участие в общественных процессах, особенно когда в обиход внедряется язык, отличный от родного языка?</p>

7 Развитие можно рассматривать как процесс расширения реальных свобод человека. Уважение свобод человека или потенциальных возможностей – другой взгляд на развитие, отличный от более узких представлений, которые связывают развитие с ростом ВВП, личными доходами, индустриализацией, техническим прогрессом или модернизацией общества (Sen, 2005).

Информационное общество – вопросы на перекрестке языков и культур

Информационная этика как дисциплина появилась в последние годы при значительном участии ЮНЕСКО. Связывая эту дисциплину с разнообразием культур и языков, мы открываем простор для размышлений, полностью соответствующих рассматриваемой нами теме. Этой теме была посвящена конференция, организованная в 2004 г.⁸ Международным центром по информационной этике (ICIE). Сборник материалов данной конференции опубликован в конце 2005 г. (Caruro, 2005). Эти материалы также имеют прямое отношение к обсуждаемой нами теме.

В их числе – доклад Чарльза Эсса (Charles Ess, 2004), в котором он высказал мнение, что вопреки общеизвестной гипотезе о культурной нейтральности ИКТ данные многих исследований показывают, что ИКТ, произведенные в странах западной культуры, в основном, в Северной Америке, несут в себе и, в определенной степени, рекламируют «западные» культурные ценности и коммуникационные предпочтения. По мнению Эсса это проявляется тогда, когда такие ценности и предпочтения вступают в конфликт с культурами стран, получающими технологии, – стран Азии, Латинской Америки и арабских стран. Возникающие в результате конфликты проявляются подчас в оглушительных провалах благих намерений преодолеть бедность и изоляцию (Postma, 2001). Эсс идет дальше и указывает на опасность «компьютерной колонизации», которая может стать продуктом бесхитростного плана «объединения мира», но игнорирует риски, доказавшие на практике негативные последствия для местных культур и ценностей небрежного внедрения ИКТ.

Однако Эсс успокаивает нас, говоря, что можно избежать таких конфликтов, если, в первую очередь, занять осознанную позицию в отношении вопросов культуры. Он показывает направления, по которым можно пойти в построении взаимодействий типа «человек-машина», отвечающих этому культурному критерию (Hall, 1976).

8 Локализация Интернета: этические вопросы в межкультурной перспективе. 4–6 октября 2004 г. – Карлсруэ. – <http://icie.zkm.de/congress2004>

Если мы согласны с тем, что компьютерное (цифровое) образование является важной составляющей перехода к инклюзивному информационному обществу, то понятно, что такое образование должно соответствовать фундаментальному этическому критерию уважения культурного и языкового разнообразия и поэтому избегать проявлений национализма и колониционного подтекста в технологиях.

В числе вопросов, связанных с информационным обществом, есть еще один существенный и многоаспектный вопрос – о знаниях, являющихся всеобщим достоянием, которые должны быть свободны от рынка и брать свое начало в открытом контенте и программном обеспечении. Этот вопрос также непосредственно связан с языковым разнообразием в информационном обществе.

Хосе Антонио Миллан (Jose Antonio Millán, 2001), испанский специалист по языкам и Интернету, напоминает нам, что самым сложным из всех существующих способов взаимодействия человека и компьютера остаются наши языки. В устной или письменной форме они все больше используются для работы с самыми разными программами, например, поисковыми машинами. Лингвистическое знание, инкорпорированное в программы, – автокоррекция, аутлайнинг, преобразование текст-речь – невидимо для пользователя, но его экономическое значение огромно. Базовые ресурсы, лежащие в основе программ, зачастую создаются на государственные средства, выделяемые на научные исследования. Однако внедряются эти программы зачастую в рамках коммерческого программного обеспечения, которое не относится к разряду программ с открытым исходным кодом и поэтому не подлежит улучшению или расширению (например, для того, чтобы включать миноритарные языковые вариации наиболее используемых языков). Более того, их нельзя использовать в качестве основы, на которой люди, говорящие, допустим, на миноритарных языках, могли бы составить собственные программы. По мнению Миллана, демократизация программного обеспечения на миноритарных языках произойдет после «освобождения» языковых ресурсов (на основании общедоступных лицензий (GPL) или аналогичных соглашений – Wikipedia, 2005с), произведенных на государственные средства или просто находящихся во всеобщем доступе.

Так или иначе программы с открытым исходным кодом, которым надлежит по своей природе сыграть важную роль в языковом секторе, представлены очень скромно, и разработчики программ нуждаются в дополнительной мотивации для продолжения работы над ними.

Тема открытого контента естественным образом приводит нас к рассмотрению изменений, необходимых в системе научного издательского дела, которая признается специалистами в области информационного общества (Guedon, 1998) как устаревшая, поскольку препятствует обмену научными знаниями, в особенности в странах Юга. Она подвергается критике в таких проектах, как Public Library of Science и Berlin Declaration on Open Access to Knowledge in the Sciences (ZIM, 20). Преобразование этой системы в направлении моделей, взятых в значительной степени из ИКТ и базирующихся на понятии открытого контента, могло бы внести значительный вклад в развитие языкового разнообразия.

Причина сложившейся ситуации кроется в отсутствии языковой политики и в определенном бездействии заинтересованных стран. Миллан подчеркивает, что для закрытия этого вопиющего пробела необходимо выработать настоящую политику в области цифрового контента, включающую языковую политику для цифрового мира. В этом отношении роль международных организаций, таких как ЮНЕСКО, должна состоять в разъяснении государствам-членам важности добровольно принятой политики, способствующей развитию многоязычия.

Критерии и показатели

Разумно ли определять и руководить языковой политикой в цифровом пространстве, не имея адекватных, аккуратных и точных показателей текущей языковой ситуации и изменений, происходящих с языками?

Парадоксально, но мир сетей, рожденный и развивающийся в университетах, отдал измерение языковой ситуации маркетинговым компаниям, намерения которых отличаются от намерений научных издательств, и которые по этой причине не слишком озабочены составлением документации по своим методикам. Результатом стало

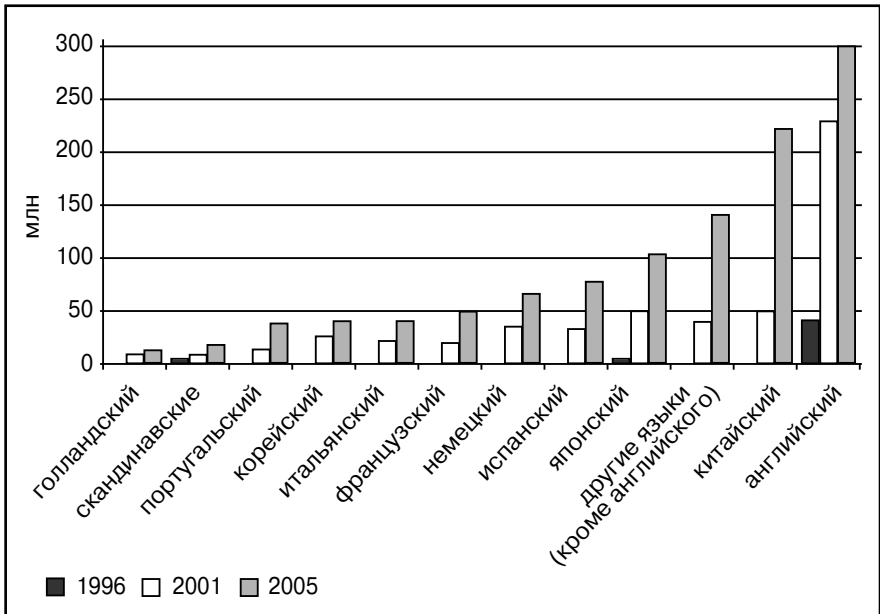
отсутствие порядка и полная неразбериха в отношении состояния языков в Интернете, а это чревато дезинформацией. Так, несмотря на то, что доля англоязычных пользователей Интернета упала с 80 % в год создания Всемирной сети до 35 % на сегодняшний день, пресса стабильно, наперекор всем фактам, приводит данные в пределах 70–80 %!

Крайне важно, чтобы научный мир восстановил свою роль в этой области вместе с государственными организациями, как национальными, так и международными. И сейчас есть четкие основания говорить, что такие изменения происходят! Для получения актуальной информации можно обратиться к онлайн-материалам встречи «Многоязычие в киберпространстве»⁹, состоявшейся в Бамако и организованной ЮНЕСКО совместно с Академией африканских языков (ACALAN) и Межправительственным агентством по франкоговорящим странам и регионам (AIF).

Однако если просто ждать, пока в ходе этого процесса будут разработаны аккуратно задокументированные показатели, обновляемые с учетом скорости развития новых технических средств, будет очень и очень трудно получить четкое представление о ситуации и имеющихся тенденциях.

I – Данные по числу пользователей Интернета в каждой языковой группе. В этой категории есть один источник, хорошо зарекомендовавший себя по итогам многолетней работы – Global Reach. Он с высокой регулярностью поставлял данные, которые, хотя и были взяты из разных источников и не были объединены общей методикой, все же стали достоянием общественности (Рис. 1). Цифры не абсолютно точны, но все же они есть, хотя и нуждаются в частом обновлении. Даже если мы будем относиться к ним с сомнением (погрешность составляет 20 %), они позволяют нам видеть перспективу роста пользователей Интернета по языковым группам.

9 http://portal.UNESCO.org/ci/en/ev.php-URL_DO=DO_TOPIC&URL_SECTION=465.html или <http://www.UNESCO.org/webworld/multilingualism>

Рис. 1 Количество пользователей Интернета по языкам

Источник: Global Reach 2005 (<http://global-reach.biz/globalstats/index.php3>).

II – Данные по языкам во Всемирной сети. В этой категории существует несколько параллельных подходов:

1) Можно экстраполировать данные по языкам, полученные с помощью поисковых машин. Это самый простой способ, дающий приемлемый порядок величин, но получаемые данные не настолько надежны, чтобы их можно было использовать для мониторинга состояния языков, учитывая слабость алгоритмов распознавания языков и непредсказуемое поведение поисковых машин при подсчете итоговых величин.

2) Второй подход был использован в одном из первых исследований данного вопроса, которое было проведено в июне 1997 г. компанией «Alis Technologies» при содействии неправительственной организации «Internet Society». Позднее этот метод был взят на вооружение другими

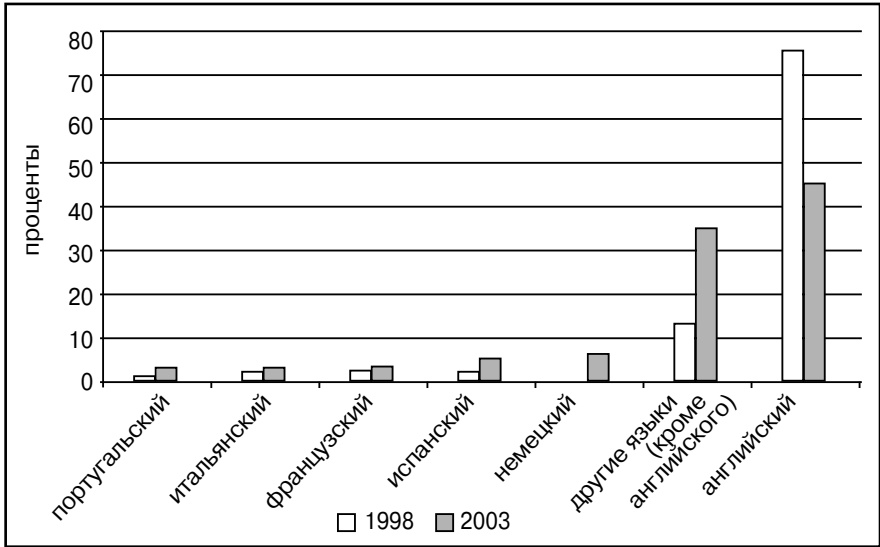
организациями, в частности, «Онлайновым компьютерным библиотечным центром» (OCLC). Центр проводил исследование, ставшее источником цитирования для многих авторов и изданий, которые упорно пишут о существовании более чем 70 % веб-страниц на английском языке (O'Neill, 2003). Суть метода заключается в создании случайной выборки IP-адресов, включающей несколько тысяч веб-сайтов (Wikipedia, 2005d), пропуска ее через машину, распознающую языки, и обобщении полученных результатов.

Этот подход имеет такой же недостаток, как и первый, – слабость алгоритмов распознавания языков. Однако с 1997 г. был достигнут большой прогресс, и новые методы должны коренным образом повысить надежность получаемых результатов.

Другой недостаток этого подхода значительно серьезнее, т. к. определяется недостатками используемого статистического инструментария. Математическая обработка, составленная для случайной переменной (т. е. в нашем случае для случайной выборки веб-сайтов, подвергнутых процедуре распознавания языков), должна анализировать статистическое распределение для выведения среднего значения, дисперсии и доверительного интервала. Данные по одной случайной выборке не могут дать надежный результат. Что такое 8000 веб-сайтов, когда существуют 8 миллиардов веб-страниц, индексированных Google? Однако, если основываться на скудных публикациях на данную тему, оказывается, что цифры по языкам, представленные OCLC, получены именно этим методом.

3) Существует еще один подход, при котором полученные данные публикуются, но методика их получения не раскрывается. Как следствие, оценить результаты не представляется возможным. Речь идет об исследовании Inktomi, проведенном в 2001 г. и широко анонсированном. Это исследование содержит грубейшие ошибки, например, представление соотношения веб-страниц по ограниченному числу языков, общее число которых принято за 100 %!

4) Последнюю группу составляют редкие, но документально оформленные методики. В их числе оригинальный подход, который использовали научные работники Херох в 2001 г. (Grefenstette & Nioche, 2001), и особый метод, который применялся в FUNREDES и Латинском союзе с 1996 г. (Рис. 2).

Рис. 2. Соотношение веб-страниц по языкам

Источник: FUNREDES 2003, <http://funredes.org/lc>

Данный метод основан на получении числа встречаемости определенного слова в заданном секторе киберпространства, таком как веб-страницы или дискуссионные группы, с помощью поисковых машин. Для каждого из исследуемых языков создается набор ключевых слов, наилучшим образом отражающий семантические и синтаксические особенности этих языков. Данные по встречаемости каждого слова, предоставленные поисковой машиной, группируются в соответствии с встречаемостью слова в каждом конкретном языке. Затем эти значения обрабатываются как случайные переменные, математическое распределение которых анализируется с помощью традиционных статистических инструментов, таких как среднее значение, дисперсия, доверительный интервал и закон распределения Фишера. В результате получается весовая оценка присутствия каждого языка в сравнении с английским, который принят в качестве отправной точки. Полученная оценка анализируется количественно с помощью статистических инструментов (доверительный интервал). Повторение этого процесса измерения через последовательные промежутки времени позволяет видеть динамику присутствия языков в исследуемых секторах Сети. Ценность метода заключается

в том, что он обеспечивает получение совместимых результатов, позволяющих отслеживать тенденции изменений.

Несмотря на то, что эта методика ни разу не подвергалась научной критике с момента своего появления, у нее есть определенные недостатки:

- она дает не абсолютное, а относительное значение (относительно английского языка) присутствия веб-страниц на разных языках (немецком, испанском, французском, итальянском, португальском или румынском). Для получения абсолютного значения по каждому языку должна существовать оценка абсолютного значения присутствия английского языка, учитывающая возрастающую сложность и неопределенность проверки встречаемости ключевых слов с учетом роста числа языков в Интернете;
- применение этого метода для новых языков сложно с лингвистической точки зрения и дорого стоит;
- она дает оценку, соответствующую киберпространству страниц, индексированных поисковыми машинами, но не принимает во внимание «невидимую Сеть» (Bergman, 2001). Но «существуют» ли в действительности неиндексированные страницы?
- кроме того, методика сильно зависит от точности счетчиков поисковых машин¹⁰ и может быть недостаточно надежна, поскольку они выдают большие расхождения при проведении поиска по словам¹¹.

Положительной стороной метода является то, что он предоставляет возможность осуществлять последовательный мониторинг в течение длительно времени, изучать сектора киберпространства, а не только Сеть¹² в целом, и, помимо этого, позволяет создавать серии уникальных

10 На сегодняшний день основная работа по проведению измерений состоит в проверке поведения поисковых машин, отборе наиболее точных из них и компенсации их непредвиденного поведения, особенно при обработке диакритических знаков.

11 Вполне вероятно, что вскоре поисковые машины будут предоставлять результаты поиска, включающие документы с переводом поисковых слов на разные языки.

12 Он также предоставил данные в первом приближении, которые, конечно, очень неточны, но интересны в изучении развития культур в Интернете.

и очень значимых показателей, основанных на исследованиях по странам и доменам (Pimienta, 2001).

Перспективы новых подходов

Проект «Обсерватория языков» (см. статью Йошики Миками) обещает заполнить имеющийся пробел и дает ответы, необходимые разработчикам политики для выработки стратегий и измерения их воздействия.

Наш опыт подсказывает, что перспективный подход, который, похоже, еще никем не применяется, будет аналогичен методу, которым пользуется «Алеха» для составления портрета наиболее посещаемых сайтов и предоставления другой ценной информации. «Алеха» компилирует данные по поведению значительного числа пользователей, которые согласились поставить на свои компьютеры spyware (программы-шпионы), позволяющие получать очень подробную статистику. Приняв этот подход за основу, можно представить себе программу, которая могла бы количественно оценивать языки, используемые в разнообразных контекстах, соответствующих таким показателям, как языки написания и чтения электронных писем, языки доступных сайтов и т. д.

Список литературы

- Bergman, M.K. 2001. The Deep Web: Surfacing Hidden Value. *Bright Planet – Deep Web*. <http://www.brightplanet.com/technology/deepweb.asp>
- Capurro, R. & al. (Eds.) 2005. Localizing the Internet. Ethical Issues in Intercultural Perspective. *Schriftenreihe des ICIE* Bd. 4, München: Fink Verlag.
- Communauté MISTICA. 2002. Travailler l'Internet avec une vision sociale. http://funredes.org/mistica/francais/cyberotheque/thematique/fra_doc_olis_t2.html
- Ess, C. 2004. Moral Imperatives for Life in an Intercultural Global Village in *The Internet and Our Moral Lives*, ed. R. Cavalier, State University of New York Press, Albany, pp. 161–193.
- Ess, C. 2005. Can the Local Reshape the Global? Ethical Imperatives for Human Intercultural Communication Online, in Capurro, 2005.

- Ess, C. 2006. From Computer-Mediated Colonization to Culturally-Aware ICT Usage and Design, In P. Zaphiris and S. Kurniawan (eds.), *Human Computer Interaction Research in Web Design and Evaluation*. Hershey, PA: Idea Publishing.
- Ess, C. & Fay S. 2005. Introduction: Culture and Computer-Mediated Communication – Toward New Understandings, *Journal of Computer-Mediated Communication* Vol. 11, No. 1. <http://jcmc.indiana.edu/>
- Grefenstette, G. & Nioche, J. 2001. Estimation of English and non-English Language Use on the WWW. Xerox Research Centre Europe, Meylan.
- Guédon, J.C. 1998. «La bibliothèque virtuelle: une antinomie? «conférence prononcée à la National Library of Medicine. conférence prononcée à la *National Library of Medicine*.
- Washington. <http://sophia.univ-lyon2.fr/francophonie/doc/nlm-fr.html>
- Hall, E.T. 1976. *Beyond Culture*. Anchor Books, New York.
- ITU. 2003. *Competitive Markets Required to Bridge Digital Divide: Regulators map 'Universal Access' route to Information and Communication Technology*. http://www.itu.int/newsarchive/press_releases/2003/33.html
- Millán, J.A. How much is a language worth: A Quantification of the Digital Industry for the Spanish Language. *Language Diversity in the Information Society International Colloquium*. Paris, France. <http://jamillan.com/worth.htm>
- O'Neill & al. 2003. Trends in the Evolution of the Public Web: 1998–2002 <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>
- Pimienta, D. 2002. La fracture numérique, un concept boiteux». *Communauté Virtuelle MISTICA*. http://funredes.org/mistica/francais/cyberotheque/thematique/fra_doc_wsis1.html
- Pimienta, D. & Lamey B. 2001. «Lengua Española y Culturas Hispánicas en la Internet: Comparación con el inglés y el francés. *II Congreso Internacional de la Lengua*. Valladolid. <http://www.funredes.org/LC/L5/valladolid.html>
- Postma, L. 2001. A Theoretical Argumentation and Evaluation of South African Learners. Orientation towards and Perceptions of the Empowering Use of Information. *New Media and Society*. Vol. 3, No. 3, pp. 315–28.
- Sen, A. 2005. *Human Development and Capability Association*. <http://www.fas.harvard.edu/~freedom/>
- UNESCO. 2000. «Infoethics». UNESCO *WebWorld News*. <http://www.UNESCO.org/webworld/news/infoethics.shtm>

UNESCO. 2005. Multilinguisme pour la diversité culturelle et la participation de tous dans le cyberspace. http://portal.unesco.org/ci/fr/ev.php-URL_ID=17688&URL_DO=DO_TOPIC&URL_SECTION=201.html

ZIM. 2003. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. *Conference on Open Access to Knowledge in the Sciences and Humanities*. Berlin. <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

Глоссарий

Webopedia. 2005a. ADSL. <http://www.webopedia.com/TERM/A/ADSL.html>

Webopedia. 2005b. ICANN. <http://www.webopedia.com/TERM/I/icann.html>

Wikipedia. 2005a. Internationalized Domain Name.
<http://en.wikipedia.org/wiki/IDNA>

Wikipedia. 2005b. Unicode. <http://en.wikipedia.org/wiki/Unicode>

Wikipedia. 2005c. GNU General Public License.
http://en.wikipedia.org/wiki/GNU_General_Public_License

Wikipedia. 2005d. IP Address. http://en.wikipedia.org/wiki/IP_address

б) Политический и юридический контекст

Даниэль Прадо, Латинский союз

Факты говорят о том, что самые известные европейские языки переживают период серьезного спада по использованию в области научно-технической коммуникации, уступая место английскому языку. Это влияние испытали на себе такие великие европейские языки как немецкий, испанский, французский, итальянский, португальский, русский и скандинавские языки, и исключение составляют только отдельные языки ограниченного распространения, которые в последние годы переживают некоторый подъем (Hamel, 2002).

Среди европейских языков наибольшему влиянию этого фактора подверглись романские языки, что проявилось в области издания тех-

нической литературы, научных конференциях, международных организациях, СМИ, обучении и др.

В ноябре 2002 г. была проведена первая международная конференция, посвященная анализу места романских языков в научно-технической коммуникации (UNILAT, 2002a). Она собрала специалистов по языковой политике из стран и регионов трех основных языковых групп – французской, португальской и испанской.

В ходе этой конференции были представлены статистические данные и доклады, показавшие резкое снижение роли романских языков во многих областях, связанных с наукой и техникой. По данным Calvet (2002), основанным на наиболее значительных международных базах данных¹³, на таких языках как французский, испанский, португальский, итальянский, румынский и каталонский, а также на 20 других менее распространенных языках создается только 1/10 часть научных публикаций от числа написанных на английском языке – и это несмотря на официальный статус этих языков в более четверти всех стран мира (27,53 %) и почти одного миллиарда романоязычного населения. Действительно, по мнению Namel, на английском языке пишется от 80 до 90 % публикаций по естественным наукам и от 74 до 82 % публикаций по общественным и гуманитарным наукам, в то время как на три наиболее распространенных романских языка приходится 12 % публикаций по общественным наукам и 18 % по гуманитарным. Правда, Namel уточняет свои наблюдения, говоря, что эта статистика взята из баз данных научных статей, и что книгоиздание развивается так же энергично, как научные журналы. Интересно отметить, что издательское дело на романских языках идет сейчас вполне неплохо, на него приходится 18,9 % всех мировых публикаций (Rousseau, 2002), но цифра эта включает, в основном, художественную литературу (Leñez Aristimuño, 2002).

Вполне понятно, что в сравнении с большинством языков на планете, по критерию распространения знаний романские языки находятся еще не в самом плохом положении. И действительно, на каждые

13 Часто говорят, что научные журналы на английском языке избыточно представлены в этих международных базах данных и что, как следствие, журналы в странах, не входящих в ОЭСР, представлены недостаточно (UIS).

100 страниц на английском языке приходится почти 38 страниц (UNILAT, 2005) на романских языках¹⁴. Французский язык занимает 2-е место по использованию в мире, испанский язык «удобно устроился» на 3-м, а преподавание испанского языка растет во всем мире. Португальский отличается хорошей демографической ситуацией, да и распространен он не на одном континенте; а итальянский язык остается престижным языком, несмотря на низкий демографический уровень и ограниченное географическое распространение (Италия, Швейцария и Сан-Марино).

Не следует забывать, что на английском языке, на котором говорит в 2,5 раза больше людей, чем на всей совокупности романских языков, создано в 2,5 раза больше Интернет-страниц, чем на всех романских языках вместе взятых. Научные публикации на английском языке представляют 2/3 всех научных публикаций в мире, в то время как научные публикации на всех романских языках составляют 1/10 часть.

Мы также совершенно не намерены игнорировать спад, который происходит в наши дни в области научно-технического использования других языков, таких как языки Северной Европы и скандинавские в особенности: ведь из этих языков по причине засилья английского языка как средства научного общения в определенных областях знаний исчезает огромная доля научного вокабуляра (Nilsson, 2005). Также в наши намерения не входит драматизировать ситуацию с европейскими языками, если помнить, что по данным Леанез (Leáñez) 98 % языков планеты вообще не имеют базовых терминологических словарей – административных, научных, технических, юридических или коммерческих. Надо «бить в колокола» в связи с этой тревожной ситуацией, поскольку это касается всех языков, кроме английского.

Вернемся, однако, к языкам в Интернете. Статистика FUNREDES/ Латинского союза показывает, что в 2003 г. почти 14 % веб-страниц были опубликованы на одном из романских языков и около 45 % веб-страниц – на английском. Даже на немецком, на котором говорит в 10 раз меньше людей, количество веб-страниц было только в 2 раза меньше, чем на всех романских языках вместе взятых. Однако наи-

14 Исследование было проведено по пяти самым распространенным романским языкам: испанскому, французскому, итальянскому, португальскому и румынскому.

большую тревогу в отношении присутствия романских языков в Интернете вызывают неопубликованные данные, «невидимый» Интернет – интрасети, базы данных, списки рассылки, форумы и т. п. Мы не располагаем статистикой по ним, но повседневная практика показывает, что английский язык начинает доминировать, как только в электронной дискуссионной группе начинается международное обсуждение технических вопросов, или когда научная база данных приобретает международное значение, или когда группа молодых людей начинает обсуждать в чате своего кумира. Феномен этот объяснить легко, т. к. телематические сети обслуживают исследователей международного уровня, и было бы бессмысленно повторять, что именно английский считается в научном мире основным языком общения. Огорчение вызывает тот факт, что эта ситуация не меняется, и поэтому население и группы людей, не владеющие английским языком, оказываются в изоляции.

Леанез напоминает нам о том, что «язык, имеющий невысокую ценность, используется мало, а малоиспользуемый язык обладает невысокой ценностью». Тем самым он говорит, что если наш родной язык не соответствует нашим потребностям, мы выучим другой язык и научим других говорить на нем.

С учетом сказанного План действий ЮНЕСКО (UNESCO's Action Plan, 2005) оказался для ВСИО своевременным. Смотрите сами: в главе I одна из рекомендаций касается культурного и языкового разнообразия. Рекомендуются «формировать политику, которая поддерживает уважение, сохранность, развитие и совершенствование культурного и языкового разнообразия и культурного наследия в рамках информационного общества». В настоящее время нет ни одной романоязычной страны, в которой была бы принята политика, позволяющая полностью использовать романские языки, особенно в информационном обществе и при обмене знаниями.

Более того, языковая политика романоязычных стран, за очень редким исключением, излишне сосредоточена на административных вопросах с одной стороны и на защите эндогенных языков с другой (в редких случаях, на защите прав потребителей). Не создавая необходимый механизм контроля и инструменты, позволяющие претворить в жизнь то, что превозносится законами о книгах, они не имеют достаточных средств для развития своего языка и, как следствие, оставляя

ют пустое место, которое быстро заполняется английским языком, в особенности в таких областях как научные дискуссии, техническая документация, высшее образование, Интернет и т. п.

За исключением Квебека, Каталонии и Франции ни одна государственная организация в провинциях или странах с населением, говорящим на романских языках, не взяла на себя ответственность за элементы, из которых складывается общая политика развития, благосостояния, модернизации и распространения языка. В Бельгии, Швейцарии, Испании и Португалии есть соответствующие институты, но такую работу они выполняют лишь частично. Более того, в регионах или странах, более развитых в отношении языковой политики, нет политики, поддерживающей цифровое многоязычие. Часто частные организации, ограниченные в средствах, или межправительственные организации, не наделенные четким мандатом, вынуждены подключаться и выполнять эту работу.

Хорошим моментом во всем этом является то, что многие миноритарные или миноритизированные языки в противовес всему, что происходит с более крупными языками, принимают участие в научно-технической коммуникации на невиданном ранее уровне. Именно это происходит с каталонским языком, а также с галицийским, баскским, сардинским и другими языками. Однако еще многое нужно сделать для того, чтобы эти языки были в состоянии охватить все сферы, необходимые для процветания людей, говорящих на этих языках.

Здесь есть еще один аспект: доступ к информации, созданной на языке, которым мы владеем не свободно. Перевод, как мы знаем, – дело дорогое. Перевод ряда процессов, таких как тендеры межправительственных организаций, занимает много времени. Машинный перевод никогда не заменит перевод, сделанный человеком, но может помочь в повышении производительности, скорости и доступности. Он является неотъемлемой частью трансформации, необходимой в мире цифрового и бумажного издательского дела. Ни одна из ныне существующих систем не в состоянии выполнить перевод на удовлетворительном уровне для наиболее распространенных языковых пар. Все машинные переводы требуют редактирования. Но самым серьезным моментом в этом деле является то, что системы машинного перевода или

переводы с помощью компьютера охватывают очень ограниченное число языковых пар¹⁵.

Нужно повышать качество существующих систем, и это произойдет, учитывая уровень их технологического развития, но ничто не заставит нас поверить в то, что можно совершить рывок от критически малого количества языков (1 %), соединенных в пары компьютерным переводом. Добровольные инициативы должны наметить пути перевода с разных языков, не представляющих рыночный интерес для коммерческих предприятий. Латинский союз, как и Университет ООН, инициировал ряд процессов в этом направлении¹⁶. Мы надеемся, что появятся и другие инициативы для наименее популярных языков.

Так что же нужно сделать, чтобы создать многоязычный цифровой мир? Недавно прошедшие во Франции дискуссии по европейскому Google, которые были подхвачены международной прессой, дали толчок некоторым новым идеям (Millán, 2005), да и ЮНЕСКО всегда подчеркивала роль библиотек и формата их фондов. Одна из идей состоит во внедрении крупномасштабных проектов по компьютеризации фондов силами разных стран, межправительственных и неправительственных организаций или частных Интернет-провайдеров, но только тех, кто обязуется уважать этические аспекты использования информации. Очевидно, что следует запретить как приобретение права собственности на эту цифровую информацию для коммерческих целей, так и передачу прав или лицензий на ее распространение. Цель состоит в широком бесплатном распространении цифрового контента как единственного средства, обеспечивающего подлинное языковое разнообразие.

В своем ежедневном использовании Интернет спонтанно порождает новые виды коммуникации: независимые и автономные пресс-службы, блоги, инициативы граждан. Это говорит о существовании аудитории помимо моноязычных монополий. Может быть, нам следует более

15 Действительно, было подсчитано, что из почти 6000 существующих в мире языков менее 100 подвергались обработке системами машинного перевода или переводились с помощью компьютерных программ.

16 Следует отметить включение румынского языка в проект Atamiri (<http://lux0.atamiri.cc/forum/init.do>).

внимательно понаблюдать за этой аудиторией, поддерживать ее и вдохновлять.

Не секрет, что страны, говорящие на романских языках, отстают в организации присутствия этих языков в цифровом мире. Это заставляет нас действовать – нужно выработать добровольную политику оцифровки информации и каталогов, которые в настоящее время существуют только в бумажной форме; нужно выработать актуальную политику в области научного издательского дела на национальных языках, или, по крайней мере, переводить издания, опубликованные на английском языке, и мгновенно распространять их через Интернет; нужно принять хартию прав граждан, включающую право получения информации на родном языке и, как следствие, обязательство уважать многоязычие на веб-сайтах международных организаций и международных компаний, а также обязательство со стороны национальных корпораций распространять материалы на языках местного населения; нужно также распространять машинный перевод на те языки, которые еще не охвачены соответствующим программным обеспечением.

В соответствии с рекомендациями, принятыми на первой встрече, Латинский союз занят подготовкой второй встречи, посвященной месту романских языков в профессиональной и технической коммуникации (UNILAT, 2002b). В наших планах – разработка механизмов консультирования, мониторинга, сбора и распространения статистических данных, а также поиск механизмов стимулирования развития издательского дела и проведения исследований на романских языках; развитие эффективных лингвистических инструментов. Эта встреча запланирована в 2006 г. в Испании в тесном сотрудничестве с «Trois Espaces Linguistiques» (представляющей франкоговорящие, испаноговорящие и португалоговорящие регионы земного шара). Мы надеемся, что в ходе этой встречи удастся найти решение вышеперечисленных проблем.

Список литературы

Calvet, L.J. 2002. *Le marché aux langues*. Plon, Paris.

Hamel, R.E. 2002. El español como lengua de las ciencias frente a la globalización del inglés. Diagnóstico y propuestas de acción para una política iberoamericana del lenguaje en las ciencias, au *Congrés international sur les langues néolatines*

dans la communication spécialisée. Mexique.

http://unilat.org/dtil/cong_com_esp/comunicaciones_es/hamel.htm#a

Leáñez Aristimuco, C. 2002. Español, francés, portugués: ¿equipamiento omerma? au Congrès international sur les langues néolatines dans la communication spécialisée. Mexique. http://unilat.org/dtil/cong_com_esp/comunicaciones_es/leanez.htm#a

Millán, J.A. 2005. A quoi bon un projet européen concurrent? Courrier International. http://www.courrierint.com/article.asp?obj_id=51004&provenance=hebdo

Nilsson, H. 2005. Perte de domaine, perte de fonctionnalité : indicateurs et enjeux, au Lexipraxis. <http://www.aifl.asso.fr/presentation.htm>

Rousseau, L.-J. 2002. Le français dans la communication scientifique et technique au Congrès international sur les langues néolatines dans la communication spécialisée. Mexique. http://unilat.org/dtil/cong_com_esp/comunicaciones_es/rousseau.htm#a

UNESCO. 2005. Plan d'action du SMSI. http://portal.UNESCO.org/ci/fr/ev.php-URL_ID=15897&URL_DO=DO_TOPIC&URL_SECTION=201.html

UNILAT. 2002a. Congrès international sur les langues néolatines dans la communication spécialisée. http://www.unilat.org/dtil/cong_com_esp/es/index.htm

UNILAT. 2002b. Recommandations. Congrès international sur les langues néolatines dans la communication spécialisée. http://www.unilat.org/dtil/cong_com_esp/es/index.htm

UNILAT. 2005. Etude sur La place des langues latines sur l'Internet. (http://www.unilat.org/dtil/LI/2003_2005.htm)

Языковое разнообразие в Интернете

Джон Паолилло, Институт информатики,
Университет Индианы

Прошло уже более 20 лет с момента появления Интернета в англоязычном мире, а представленность различных языков в этом пространстве по-прежнему сильно смещена в направлении преобладания английского языка. Английский остается доминирующим языком в Интернете, а это значит, что языки, на которых говорит множество людей, представлены либо плохо, либо не представлены совсем. В какой степени эта ситуация отражает предпочтение, отдаваемое английскому языку, и отказ от других языков? В данном докладе¹⁷ мы рассмотрим этот вопрос с позиций этической системы Фридмана и Ниссенбаума (Friedman and Nissenbaum, 1997), разработанной для оценки сдвига в компьютерных системах и сравнения полученных данных со статусом мировых языков в Интернете. Эта система поможет нам интерпретировать возможные причины и способы корректировки потенциальных сдвигов. Мы также обратим внимание на звучащее в настоящее время недовольство статусом языков в Интернете и рассмотрим эту ситуацию

17 Помощь в написании настоящего доклада оказали: Elijah Wright и Hong Zhang (Университет Индианы), S., Baskaran, G.V., Ramanan, S.V., Rameshkumar, S., Shoba Nair, I., Vinoshbaru James, Viswanathan (Университет Св. Анны, Ченнай, Индия). Полная версия доклада представлена на <http://ella.slis.indiana.edu/paolillo/paolillo.divesity041027.pdf>

в рамках указанной системы, которая требует от нас анализа не только распределения и использования языков в Интернете, но и анализа общественных институтов, осуществляющих управление развитием Интернета, которые, по выражению Фридмана и Ниссенбаума, могут вызвать «возникающие сдвиги». Наконец, мы рассмотрим вопросы языковых сдвигов в технических системах Интернета.

Сдвиг, многоязычие и компьютерные системы

Цифровой разрыв, т. е. неравное распределение доступа к источникам и сервисам цифровой информации, стал одним из ключевых моментов в наше время цифровой информации. Правительства, международные агентства, группы граждан, корпорации и другие объединения стремятся овладеть преимуществами, которые сулят обещания низкой стоимости и мгновенного доступа к информации за счет перевода многих коммуникационных функций на сетевые компьютерные носители. Однако, если традиционные социальные барьеры, такие как социально-экономический статус, образование, этническая принадлежность, пол и пр. затрудняют доступ к цифровой информации, то политика должна быть направлена на выравнивание доступа для реализации обещанных преимуществ.

Вопрос статуса мировых языков можно рассматривать в свете цифрового контента. Некоторые языки имеют огромные объемы доступного цифрового контента. Пользователи Интернета, говорящие, читающие и пишущие на таких языках, имеют намного меньше проблем, связанных с доступом и обменом полезной информацией, чем люди, говорящие на языках, хуже представленных в информационных системах. Анализ этой ситуации наводит на мысль о том, что, сдвиг, возможно, создается самими цифровыми информационными системами, их конфигурацией или использованием. Стало ли языковое различие барьером на пути доступа к информации, который предоставляет несправедливые преимущества одним и ставит в невыгодное положение других? Анализ подобных вопросов этического и морального порядка требует создания специальной системы.

ЮНЕСКО и культурное разнообразие

В 2001 г. государства – члены ЮНЕСКО приняли Всеобщую декларацию о культурном разнообразии¹⁸. Статья 6 «К культурному разнообразию, доступному для всех» гласит:

«Обеспечивая свободное распространение идей словесным и изобразительным путем, следует добиваться, чтобы все культуры могли быть объектом самовыражения и распространения. Залогом культурного разнообразия является свобода выражения мнения, плюрализм средств информации, многоязычие, равный доступ к возможностям для художественного творчества, к научно-техническим знаниям, в том числе в цифровой форме, и обеспечение всем культурам доступа к средствам выражения и распространения идей».

Опираясь на данную Декларацию, ЮНЕСКО недвусмысленно поддерживает предоставление всем языковым и культурным группам равного доступа к цифровой информации, как в сфере производства, так и в сфере потребления. Декларация конкретизирует эту позицию, перечисляя направления действий, которые необходимы для претворения ее в жизнь. Три из них имеют непосредственное отношение к вопросам цифровых носителей и информационных технологий:

9. Стимулирование «цифровой грамотности» и обеспечение более высокого уровня владения новыми информационными и коммуникационными технологиями, которые следует рассматривать и как учебную дисциплину, и как педагогические средства, способные повысить эффективность образовательных услуг.
10. Продвижение языкового разнообразия в киберпространстве и стимулирование всеобщего доступа через глобальную сеть к информации, являющейся общественным достоянием.
11. Преодоление цифрового разрыва в тесном сотрудничестве с соответствующими организациями системы ООН путем по-

18 <http://unesdoc.UNESCO.org/images/0012/001271/127160m.pdf>

Прим. ред. рус. изд.: Всеобщая декларация о культурном разнообразии на русском языке размещена на веб-сайте Российского комитета Программы ЮНЕСКО «Информация для всех»: <http://www.ifapcom.ru/ru/387/425/>

ощрения доступа развивающихся стран к новым технологиям и облегчения распространения местного контента в цифровом формате и доступа этих стран к образовательным, культурным и научным цифровым ресурсам, доступным во всем мире (ЮНЕСКО, 2001, стр. 6).

Эти принципы и направления действий определяют критерии оценки признаков информационного общества в терминах этики и задач развития такого общества. Однако они не предоставляют достаточно глубокого анализа возможных причин любых сдвигов, которые могут проявиться. Без такого анализа достаточно трудно дать правильные рекомендации по необходимым действиям.

Так, например, электронные библиотеки не встретили одобрения у народа маори в Новой Зеландии. Внимательное изучение показало, что это не просто вопрос цифровой грамотности, а отражение культурных традиций, препятствующих успешному использованию электронных ресурсов. Выяснилось, что метафора библиотеки относится к форме организаций, принадлежащих «Pakeha» (белым, выходцам из Западной Европы), предполагающих доступ к информации, чуждой для народа маори (Dunker, 2002). Ключевое место конфликта для маори – открытый доступ к информации, которая традиционно защищается в их культуре. Такой информацией являются для них генеалогические данные. Концепция библиотек, применяющих принцип открытого доступа к информации независимо от ее содержания, вступила в конфликт с ценностями маори. Следовательно, модель информационного доступа для электронных библиотек необходимо пересмотреть для того, чтобы создать электронную библиотеку, признанную народом маори¹⁹.

Этическая система

Фридман и Ниссенбаум (Friedman and Nissenbaum, 1995, 1997) создали полезную основу для анализа сдвигов в компьютерных системах, помогающую сосредоточить внимание на причинах существующих предпочтений. Они определили 3 основные категории сдвигов: исторические, технические и возникающие. Исторические сдвиги берут начало

19 Эта ситуация напоминает проблемы, связанные со случайным опубликованием в Интернете личных медицинских карт людей.

в общественных институтах, практиках и отношениях, и существуют независимо от компьютерных систем. Технические сдвиги возникают тогда, когда технические свойства используемых систем не учитывают всех особенностей той сферы, в которой они применяются. Возникающие сдвиги появляются в процессе работы реальных пользователей; они не являются осознанной частью проекта или системы, или неотъемлемой частью социального контекста, а возникают из взаимодействия того и другого в определенной ситуации.

Проиллюстрируем все три формы сдвигов на языковых примерах. Исторические сдвиги проявляются тогда, когда правительство, отрасль промышленности или мощная корпорация отказываются предоставлять носителям одного или нескольких языков информацию, технологии или продукты. Так, в середине 1990-х годов Microsoft Inc. отказалась сделать версии своих продуктов, которые могли бы работать с алфавитами, отличными от латинского, несмотря на наличие уже имеющихся на рынке технологических решений, таких как WorldScript, созданный Apple Computer Inc. Microsoft Inc. объявила, что отказывается сделать это потому, что рынки языков, не использующих латиницу, слишком малы, чтобы оправдать создание новой версии их продукта; так, мы видим, что за этим примером исторического сдвига стояла экономическая мотивация²⁰.

Технические сдвиги возникают в системах кодирования текстов, таких как Юникод UTF-8, и приводят к тому, что текст не на латинице требует в 2–3 раза больше места, чем аналогичный текст на латинице. В этом случае мотивация заключается в вопросах совместимости старых систем, основанных на латинице, с более молодыми системами, использующими Юникод.

Наконец, возникающие сдвиги появляются тогда, когда компьютерные системы, разработанные для одной цели, используются для другой цели, например, когда система электронных библиотек, разработанная для городских условий, в которых проживает белое население Новой Зеландии, воспринимается с неодобрением народом маори, проживающим в сельской местности.

20 За прошедшее время Microsoft изменила свою позицию и создала версии своих продуктов для рынков других языков.

Каждый из трех типов сдвигов требует отдельного подхода. Исторические сдвиги следует рассматривать в свете образовательных, юридических и институциональных ресурсов стран, отраслей промышленности или корпораций. Технические сдвиги следует рассматривать при разработке основополагающих принципов создаваемых компьютерных систем. Возникающие сдвиги следует рассматривать через образование и создание проектов, основанных на наблюдениях за реальным использованием компьютерных систем.

Поскольку развитие Интернета предполагает взаимодействие технологий, предварительных условий, целей, отраслей промышленности и действующих лиц, все три типа сдвигов отражаются на развитии языков в Интернете, в разных местах и в разное время.

Интернационализация и Интернет: популярные концепции

Дискуссии в СМИ о потенциальном языковом сдвиге в Интернете развиваются в двух противоположных направлениях. Вассерман (Wasserman) так описывает эту ситуацию:

«Поскольку Интернет способствует... пониманию мира как взаимосвязанной и взаимозависимой среды, его можно рассматривать как один из факторов ускорения глобализации... Поскольку глобализация рассматривается как сила, источником которой является так называемый экономически развитый мир, некоторые критики предсказывают разрушение отличительных местных и культурных особенностей в миноритарных странах и сообществах. С другой стороны, некоторые критики говорят, что глобальные и локальные силы взаимодействуют в процессе глобализации, превращая ее в многосторонний процесс, приносящий пользу и даже стимулирующий развитие местных культур и языков. (Wasserman, 2002:2)».

Те, кто придерживается первого взгляда, выступают защитниками прав меньшинств, а те, кто придерживается второго взгляда, выступают сторонниками новых сетевых информационных технологий. Первый взгляд – это до некоторой степени реакция на быстрые и далеко идущие изменения, вызванные распространением Интернета, в то время как второй взгляд был с самого начала в значительной степени спровоцирован создателями этой технологии.

Достаточно просто найти популярные отчеты, отражающие работу команд инженеров над ранее существовавшей сетью ARPANET (первой компьютерной сетью) как над идеализированной децентрализованной демократической организацией (см. Hafner and Lyon, 1996), или над сетью Whole Earth 'Lectronic Link (известной также как WELL), способствующей созданию виртуальных Интернет-сообществ (Rheingold, 2000). От этого взгляда совсем недалеко до теории, которая считает языковое доминирование одной из форм неравенства, которую технология Интернета сможет быстро устранить. Во-первых (как говорят сторонники этой теории), Интернет – это глобальная и децентрализованная сеть; никакой пользователь или группа пользователей не может осуществлять иерархический контроль над другим пользователем или группой пользователей, потому что Интернет допускает полную свободу общения. Как следствие, каждый желающий может пользоваться любым языком, если есть еще хотя бы один человек, который также хочет пользоваться этим языком. Во-вторых, рост числа неанглоязычных пользователей, в особенности пользователей, пишущих и читающих на китайском языке, должен, по прогнозам, превысить текущий показатель роста англоязычных пользователей. Иными словами, в конце концов, английский язык не будет доминировать в Интернете, потому что на других языках будет говорить гораздо больше людей. Вопрос о том, какой язык является доминирующим в сети, на самом деле относится к разряду вопросов демографического распределения. И, наконец, сторонники Интернета говорят, что такие технические возможности Интернета, как Юникод для многоязычных текстов и системы типа BabelFish для моментального перевода веб-документов, способны разрешить любые проблемы, связанные с использованием информации в Интернете, с которыми могут столкнуться люди, говорящие на других языках. В частности, эта перспектива в значительной степени характеризует позицию, отраженную в издании «Культурное и языковое разнообразие в информационном обществе», подготовленном ЮНЕСКО для Всемирного саммита по информационному обществу (UNESCO, 2003)*.

* *Прим. ред. рус. изд.:* эта книга издана на русском языке Российским комитетом Программы ЮНЕСКО «Информация для всех» и Российской национальной библиотекой: <http://www.ifapcom.ru/365/432/>

На каждый из этих аргументов есть контраргумент, представляющий альтернативный взгляд, согласно которому английский язык и, в меньшей степени, другие европейские языки являются доминирующими в коммуникации посредством Интернета. Причины этого явления частично социальные, частично технические. Во-первых, приводится аргумент в пользу того, что Интернет использует телекоммуникационную инфраструктуру, экономическое доминирование в которой принадлежит компаниям США. Географическим центром связи в глобальной телекоммуникационной сети являются США, так что любое действие, способствующее этой ситуации, будет непропорционально выгодно для США за счет более низкой стоимости связи и большого числа абонентов. Во-вторых, несмотря на наметившиеся в последнее время тенденции, англоязычные пользователи по-прежнему составляют самую многочисленную группу пользователей Интернета. В-третьих, представительство англоязычных пользователей Интернета диспропорционально по отношению к представительству народов, говорящих на других языках. Наконец, большинство технологий, используемых в Интернете, лучше всего адаптировано к английскому языку. Интерфейсы для нелатинских алфавитов громоздки, а для ряда языков не существуют вообще. Даже такие системы, как Юникод, инкорпорируют технические сдвиги, отражающие предпочтение английского языка, а системы перевода недостаточно совершенны, чтобы соответствовать требуемому уровню²¹.

Эти точки зрения отличаются подходами, которые составляют основу трех типов сдвигов, определенных Фридманом и Ниссенбаумом (Friedman and Nissenbaum, 1997). Языковая демография пользователей Интернета заставляет задуматься об исторических сдвигах. Вопрос доступности технических возможностей для разных языков поднимает тему технических сдвигов. Кроме того, вопросы децентрализации вместо существующего де-факто централизованного контроля заставляют говорить о возникающих сдвигах в системе, которая вышла в своем развитии за рамки первоначальных, национальных границ.

21 Варианты этих двух точек зрения и их отношение к аналогичным последствиям глобализации обсуждаются в Block (2004).

Несмотря на различные мнения и периодически появляющиеся жаркие дебаты, мы ощущаем явный дефицит эмпирических исследований, которые непосредственно занимались бы вопросами исторических, технических и возникающих сдвигов в Интернете. Причина этого частично заключается в том, что Интернет огромен по масштабам и подвержен быстрым изменениям. Оба эти обстоятельства сильно затрудняют получение надежных данных. И, хотя лингвистические обзоры периодически проводятся маркетинговыми организациями, такими как Jupiter Research (<http://jupiterresearch.com>) и Global Reach (<http://www.greach.com/>), ценность их данных по оценке вопросов языковых сдвигов сомнительна по причине наличия экономических интересов у специалистов по маркетингу и их клиентов. Более того, надежный крупномасштабный анализ онлайн-многоязычия будет стоить дорого и потому недоступен научным организациям, имеющим ограниченный бюджет или вообще не получающим финансирование.

Источники исторических сдвигов

Исторические сдвиги затрагивают общественные институты, практику и отношения, независимо от технологий. В числе источников такого типа сдвигов – историческое распределение населения по языкам, экономические меры, способствующие процветанию более многочисленных языков, и институциональная политика государств. Что касается языкового разнообразия в Интернете, то исторические сдвиги имеют место в отношении правительств, институтов и компаний к людям, говорящим на разных языках, при реализации политики в области информационных технологий. Понять эти сдвиги непросто, но чрезвычайно важно, и, поскольку Интернет – явление глобальное, понимать их нужно в контексте глобального языкового разнообразия.

Глобальное языковое разнообразие

Логически построенное обсуждение языкового разнообразия в глобальном или региональном масштабе требует использования количественного показателя разнообразия. К сожалению, количественное измерение языкового разнообразия редко используется в современных языковых исследованиях, а широко используемой общепринятой меры не существует. Существующие критерии по числу языков или языко-

вых групп представляются несколько упрощенными (Barrera-Brassols and Zenck, 2002; Smith, 2001). Более сложные показатели разнообразия предлагались (Greenberg, 1951; Lieberson, 1964), но не получили статистического обоснования и перестали использоваться. Подход, о котором мы расскажем в настоящей статье, основан на методе Неттла (Nettle, 1999) и заключается в применении меры вариативности в качестве показателя разнообразия.

Удовлетворительный индекс языкового разнообразия должен учитывать несколько факторов. Во-первых, он должен отражать некий объект анализа – страну, континент или Интернет. Во-вторых, языковое разнообразие должно учитывать вероятность нахождения людей, говорящих на любом выбранном языке. Он должен иметь естественный равный нулю минимум для абсолютно однородного населения и не иметь фиксированного максимума. Чем больше языков, тем больше должно быть значение индекса, но по мере уменьшения доли языковой группы ее вклад в разнообразие тоже должен уменьшаться. Страны с большим числом языковых групп примерно одинаковой численности (например, Танзания; Mafu, 2004) покажут относительно высокое языковое разнообразие, в то время как страны с примерно таким же числом языков, но с одним или двумя доминирующими языками (такие как США), продемонстрируют относительно низкое языковое разнообразие. Мера, обладающая такими характеристиками, – информационно-теоретическое образование под названием «энтропия», на которой мы строим нашу меру языкового разнообразия. В статистических терминах энтропия есть мера вариативности. Энтропия складывается из известного количества населения страны, говорящего на каждом языке, умноженного на его натуральный логарифм и суммированного со всеми записями для данной единицы – страны или региона. Окончательная величина индекса будет в 2 раза больше этой суммы.

Таблица 1 и рисунок 1 представляют данные по этой основанной на энтропии мере для разных регионов мира на базе 7639 языковых групп, представленных в Ethnologue (www.ethnologue.com) в порядке от самого низкого к самому высокому уровню языкового разнообразия. Данные США, родины Интернета, были размещены в первом ряду для сравнения. Регионы, известные своим языковым разнообразием (Африка, Океания), показывают наивысшее языковое разнообразие, а регионы с крупными национальными языками (Восточная Азия, Северная Америка) показывают самое низкое языковое разнообразие. Два по-

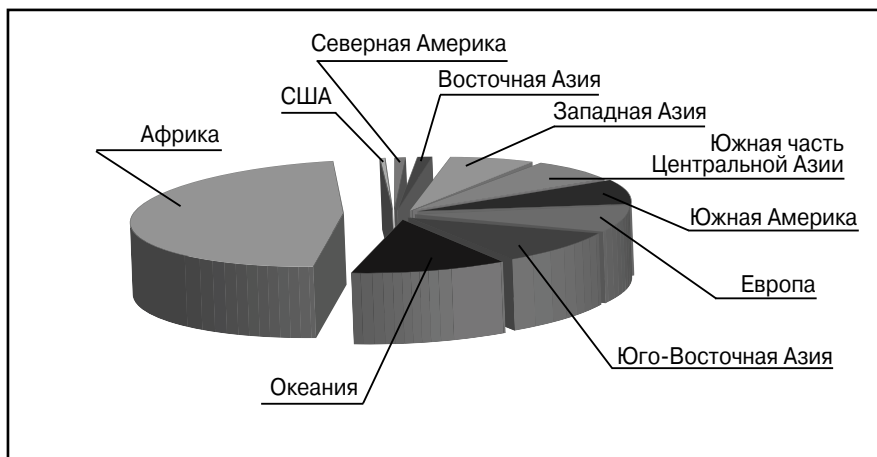
следних региона особенно важны для понимания языкового разнообразия в Интернете: США и Китай, вероятно, представляют собой двух крупнейших игроков в Интернете (по некоторым прогнозам в ближайшие несколько лет Китай обойдет США по числу пользователей); ни одна из этих стран не проявляет большего языкового разнообразия в сравнении со странами Океании или Африки. До тех пор, пока эти страны доминируют в Интернете (или, если говорить шире, в дискуссии по языковому разнообразию в Сети), Интернет не может отражать языковое разнообразие мира.

Таблица 1. Показатели индекса языкового разнообразия по регионам

Регион	Языки	Индекс разнообразия	Доля от общего числа языков в мире
США	170	0,7809	0,0020
Северная Америка (вкл. США)	248	3,3843	0,0086
Восточная Азия	200	4,4514	0,0112
Западная Азия	159	26,1539	0,0659
Южная часть Центральной Азии	661	29,8093	0,0752
Южная Америка	930	30,5007	0,0769
Европа	364	32,4369	0,0818
Юго-Восточная Азия	1317	37,6615	0,0949
Океания	1322	46,5653	0,1174
Африка	2390	185,6836	0,4681

Источник: Ethnologue.

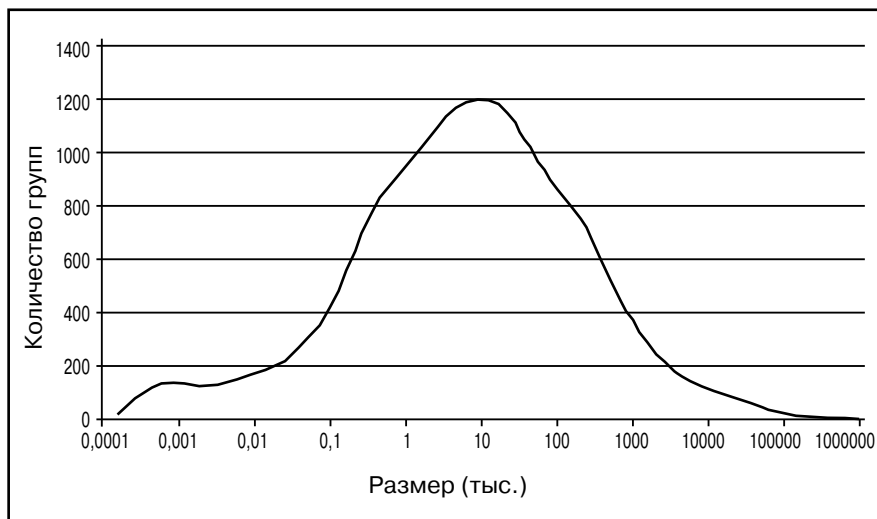
Рис. 1. Индекс языкового разнообразия по регионам



Источник: Ethnologue.

Эволюция языкового разнообразия

Общее представление о языковом разнообразии можно получить в результате изучения размеров языковых популяций. На рис. 2 показано количество языковых групп разной численности, цифры приводятся по данным Ethnologue. Горизонтальная ось представлена в логарифмическом масштабе, что означает, что колоколовидная кривая отражает логарифмически нормальное распределение (Grimes, 1986). Типичный размер языковой группы – несколько десятков тысяч человек, что примерно равно численности населения небольшого города. Языки с сотнями миллионов носителей, такие как английский, китайский, испанский, французский и пр., являются несколько нетипичными, так же как и более мелкие языковые группы с населением в сотни человек. С бытовой точки зрения ситуация выглядит иначе: почти половина людей на Земле говорит на языке, на котором говорят еще сотни миллионов других людей.

Рис. 2. Размеры языковых групп

Источник: Ethnologue, ПРООН.

Глобальное и региональное языковое разнообразие – явление не статическое, оно имеет тенденцию развития во времени. Это разнообразие находится под влиянием социально-исторических событий, таких как массовая миграция, колонизация, войны, эпидемии и пр. Глобальное языковое разнообразие в настоящее время сокращается, и процесс этот идет уже давно. Для лингвистов, изучающих разнообразие человеческой речи, это кризисная ситуация. Вымирание сотен языков за короткий исторический период означает, что большие массивы знаний о способности человека говорить пропадают навсегда, а с ними пропадают литература, история и культура людей, на этих языках говоривших. Для сообществ, чьи языки, история и культура исчезают, ситуация приравнивается к катастрофе.

По некоторым оценкам почти половина языков в мире исчезнет к концу 2050 г. (Dalby, 2003; Krauss, 1992; Nettle and Romaine, 2000). Когда языковое разнообразие перестанет существовать вследствие исчезновения малых языковых групп, возрастет доля населения, относящаяся к большим языковым группам.

Потеря языкового разнообразия не ограничивается каким-то одним регионом мира: языки исчезали в большом количестве в Европе по причине расцвета национальных государств; в Северной и Южной Америке и Австралии драматические потери шли следом за европейской колонизацией и продолжаются до сих пор; на островах Тихого океана и в Индонезии английский и индонезийский вытесняют языки аборигенов; в Азии крупные языки в Китае, Японии, Индии и России веками расширяли сферу своего влияния за счет других языков (Crystal, 2000; Muhlhausler, 1996).

Некоторые причины вымирания языков очевидны. Так, случайная или намеренная ликвидация группы людей может привести к смерти языка (Wurm, 1991). Большая часть языкового разнообразия Северной Америки была потеряна именно таким путем: войны с европейскими поселенцами и болезни, распространявшиеся в результате контактов с европейцами, приводили к сокращению аборигенного населения до такой степени, что некому было поддерживать их родные языки. Другие причины вымирания языков менее очевидны, особенно когда они связаны с изменениями в экологии культуры.

Глобальное языковое разнообразие и Интернет

Низкий уровень языкового разнообразия, характерный для Северной Америки, Латинской Америки и Карибского бассейна, Европы и Восточной Азии, способствует предоставлению доступа к Интернету с использованием ограниченного числа стандартизированных технологических решений, ориентированных на каждую из основных языковых групп. Регионы и страны с более высоким языковым разнообразием требуют, как правило, более сложных подготовительных работ для налаживания Интернет-доступа, которые могут включать адаптацию (кастомизацию) ресурсов для каждого из большого числа миноритарных языков. Исходя из этого, можно сказать, что Интернет изначально «предпочитал» более многочисленные языки. Однако даже крупные языковые группы часто не имеют устойчивых технических стандартов. Так, люди, говорящие на хинди, исчисляются сотнями миллионов, но при этом, по данным одного из исследователей Университета Южной Калифорнии, почти каждый веб-сайт на хинди имеет собственный набор шрифтов, которые несовместимы со всеми остальными наборами шрифтов для этого языка. Люди, желающие прочитать матери-

алы на хинди, представленные на этих веб-сайтах, должны устанавливать шрифты, соответствующие требованиям каждого отдельного сайта. Поиск по этим сайтам проводить очень сложно, т. к. слова, отображенные разным образом, могут быть некорректно распознаны поисковыми машинами (Information Sciences Institute, 2003). Иными словами, не все крупные языковые группы находятся в Интернете в одинаково выгодном положении. Такие регионы как Африка, Океания и Юго-Восточная Азия сталкиваются с еще более серьезными проблемами по причине наличия большого числа языков, не представленных до настоящего времени в Интернете. Как следствие, до того, как включить этих членов языковых групп в число пользователей Интернета, надо проделать большую работу по техническому развитию этих стран.

При рассмотрении влияния Интернета важно не упустить из виду эволюционный взгляд на языковое разнообразие. И хотя Интернет вполне может оказывать долговременное влияние на языковое разнообразие, неясно, каким и насколько большим оно может быть в исторической перспективе. Расширяя сферу действия отдельных языков, Интернет потенциально усиливает их, но, поскольку он проделывает то же самое с большими языками, способствуя одновременно языковым контактам внутри сети, он потенциально ослабляет миноритарные языки. Это влияние может быть намного меньше, чем влияние других общественных факторов, таких как развитие сельского хозяйства, урбанизация населения, геополитические события и пр., которые вполне могут быть вне пределов влияния человека и коллективных организаций, например, ООН, чтобы такое воздействие можно было предотвратить. В то же время, мир является свидетелем настоящего сокращения языкового разнообразия, и под угрозой находится сохранение исторической и культурной традиции сотен сообществ во всем мире. Важно, чтобы эти проблемы были поняты и учтены при формировании любой политики, в задачи которой входят вопросы языкового разнообразия в Интернете.

Источники возникающих сдвигов

Возникающие сдвиги – это предпочтения, появляющиеся при применении Интернет-технологий. По отношению к языковому разнообразию в Интернете возникающие сдвиги появляются в результате работы пользователей информационных технологий, когда знание языков

пользователями начинает соответствовать их возможностям использования предоставленных технологий или информации. Такие сдвиги имеют два основных проявления: во-первых, в распространении языков в Интернете, во-вторых, в экономическом контроле над рынками телекоммуникационных и информационных технологий. В данном разделе мы рассмотрим источники этого вида сдвигов. Представленные здесь данные говорят о существенном предпочтении, отдаваемом английскому языку, и проявления этого сдвига мы наблюдаем в настоящее время.

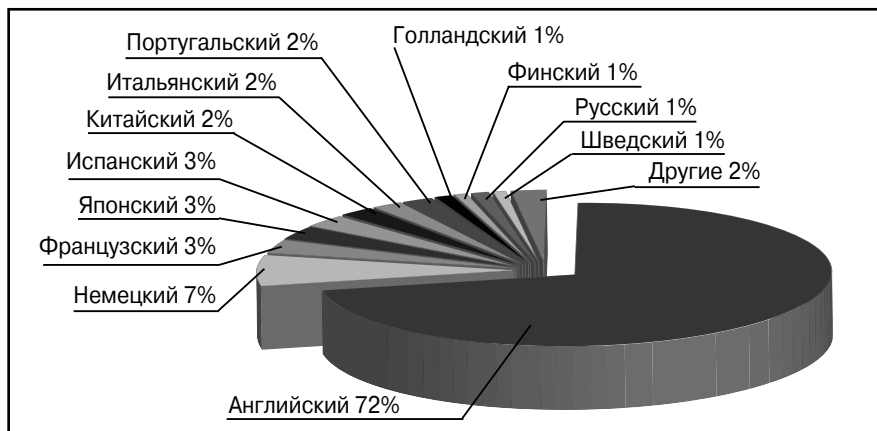
Языковое разнообразие информационных ресурсов Интернета

Не так много исследований посвящено крупномасштабному количественному анализу языков, используемых в Интернете. Как правило, такие исследования фокусируются на Всемирной сети, исключая другие коммуникационные модели, такие как электронная почта и чат, поскольку Сеть легче поддается наблюдению и анализу, чем другие формы Интернет-коммуникации. Два заслуживающих внимания исследования в этой области показали интересные результаты: серия исследований Лавуа, О'Нейлла и др. (Lavoie, O'Neill) из OCLC и исследование Нанберга (Nunberg, 1998) из PARC по неанглоязычным веб-сайтам.

Исследования OCLC (Lavoie and O'Neill, 1999; O'Neill, Lavoie and Bennett, 2003) используют произвольную выборку веб-сайтов в Интернете. Выборка была создана путем генерации произвольных IP адресов, а затем были предприняты попытки выйти на веб-сайты по каждому из этих адресов. Если веб-сервер отвечал, ученые загружали его домашнюю страницу и вводили ее в автоматизированную систему классификации языков (O'Neill, McClain and Lavoie, 1997). Этот метод хорош тем, что исследуется произвольная выборка сайтов. Все другие методы выборки косвенно или напрямую зависят от поисковых машин или «веб-спайдеров» – программ, осуществляющих поиск новых веб-страниц путем следования по ссылкам в известном наборе веб-страниц. Спайдеры создают так называемую выборку по методу «снежного кома», которая определяется близостью к произвольной точке отсчета. Поисковые машины зависят от индексов, создаваемых спайдерами, и такие выборки также будут иметь сдвиги. Если мы хотим получить надежную оценку доминирования различных языков в Интернете, выборку с наличием сдвигов следует избегать.

Начальное исследование тенденций в использовании разных языков проводилось в два разных этапа с интервалом в один год. Более позднее исследование, проведенное в 2002 г., должно было подтвердить полученные наблюдения. Исследование 1998–1999 гг. исходило из предположения, что имела место некоторая международная экспансия Сети и что использование разных языков было тесно связано с доменом, в котором был создан каждый веб-сайт. Выборка, сделанная в 1999 г. и включавшая 2229 произвольно полученных веб-сайтов, содержала 29 идентифицируемых языков, распределение которых представлено на рис. 3. Как и ожидалось, английский язык явно доминировал, и именно на нем было создано 72 % всех исследованных веб-сайтов. Индекс разнообразия для данной выборки веб-страниц составил 2,47, что меньше, чем для типичной страны Юго-Восточной Азии, и больше, чем для типичной страны южной части Центральной Азии. Он также в сотни раз меньше глобального языкового разнообразия. Следовательно, несмотря на то, что языковое разнообразие во Всемирной сети не сильно отличается от ситуации в большинстве многоязычных стран, оно является плохим отражением языкового разнообразия в мире в целом.

Рис. 3. Соотношение языков в Сети по произвольной выборке веб-страниц



Источник: O'Neill, Lavoir and Bennett, 2003.

Второй этап исследования, проведенный в 2002 г., показывает достаточно постоянную долю английского в Сети в сравнении с пре-

дыдущим исследованием, хотя небольшие отличия появляются среди других языков (O'Neill, Lavoie and Bennett, 2003). Индекс разнообразия в 2002 г. оказался равен 2,44, т. е. немного изменился по сравнению с предыдущим исследованием. Частично это могло быть связано с методологией исследования. 29 языков, которые идентифицируются в выборке веб-страниц, представляют собой предел той программы идентификации языков, которую они используют (<http://www-rali.umontreal.ca/SILC/SILC.en.cgi>), и этот метод не позволяет обнаружить новые языки, появляющиеся в Сети. Даже если бы программа идентификации могла охватить большее количество языков, их процент был бы незначителен и поэтому существенно не изменил бы полученный индекс разнообразия языков во Всемирной сети.

Исследование, проведенное OCLC в 1999 г., также определило долю многоязычных веб-страниц каждого домена и языковые пары, используемые на каждом веб-сайте. Если на веб-сайте использовалось более одного языка, английский всегда был одним из них: 100 % из 156 идентифицированных многоязычных сайтов содержали страницы на английском языке. Страницы на французском, немецком, итальянском и испанском языках были выявлены на 30 % многоязычных сайтов, а другие языки составляли еще меньшую долю. Более того, 87 % многоязычных веб-сайтов происходили из доменов вне крупнейших англоязычных стран (Австралии, Канады, Великобритании и США). В рамках каждого из этих доменов уровни многоязычия варьировались от 6 из 13 (42 %) на российских сайтах и до 16 из 1103 (1,5 %) на сайтах США. Следовательно, Всемирная сеть имеет сильную тенденцию к моноязычию, и многие проявления многоязычия – просто «дань вежливости» доминирующему английскому. Это открытие прямо противоречит распространенному мнению, что Сеть, так или иначе, стимулирует разнообразие.

Тенденции, выявленные исследованием OCLC, нашли подтверждение в работе Нанберга (Nunberg, 1998), который использовал другую методику. В его работе была проведена проверка 2,5 млн веб-страниц, отобранных в 1997 г. компанией Alexa, предоставляющей Интернет-услуги. Проверка проводилась с использованием автоматического идентификатора языков, созданного Гейнрихом Шютце (Heinrich Schuetze) – коллегой Нанберга. Несмотря на то, что полученная выборка относится к категории «снежного кома», она в тысячу раз больше вы-

борки OCLC. Основным результатом, полученным Нанбергом, стало следующее: страны с низким уровнем использования Интернета применяли на своих веб-сайтах, в основном, английский язык, а страны с более высоким уровнем использования Интернета, применяли большее количество других языков. Латинская Америка продемонстрировала контрастные показатели по причине очень низкого уровня использования Интернета в 1997 г. и колоссального числа неанглоязычных сайтов. Поэтому степень использования английского как второго языка в неанглоязычных странах может повлиять на языковое многообразие на веб-сайтах этих стран.

В стороне от перечисленных исследований стоят еще несколько попыток изучить распределение языков на основе статистических данных, полученных поисковыми машинами. По разным причинам эти попытки не дали достаточно информации для изучения. Например, FUNREDES – неправительственная организация, занимающаяся вопросами внедрения информационных и коммуникационных технологий в странах Латинской Америки, с 1995 г. проводила серию исследований для оценки распространения языков и национального влияния на Интернет (Pimienta and Lamey, 2001; Pimienta et al., 1995–2003). В ходе этих исследований был осуществлен подсчет веб-страниц, индексируемых наиболее распространенными поисковыми машинами и содержащих определенные слова из разных языков и национальных групп. Интересно, что в результате был получен значительно меньший процент англоязычных страниц (52 % в 2001 г. и 45 % в 2003 г.), чем результат, приведенный в исследованиях Лавуа, О’Нейлла и Нанберга.

Однако подсчет страниц, отобранных поисковыми машинами, представляет собой ненадежную методику определения репрезентативности языков в Сети. Помимо необъективных выборок, в результате которых определенные страницы попадают в поле зрения поисковых машин, существуют и другие схемы, искажающие результаты. Как правило, поисковые машины используют набор методов индексирования индивидуальной разработки, закрытый для проверки, а это может приводить к необъективности при подсчете страниц, результаты которого вы не можете исправить или просто проверить. Ключевое слово не обязательно должно присутствовать на странице, чтобы та попала в подсчет, а страницы, содержащие данное слово, могут выпасть из подсчета. Кроме того, этот метод предполагает, что

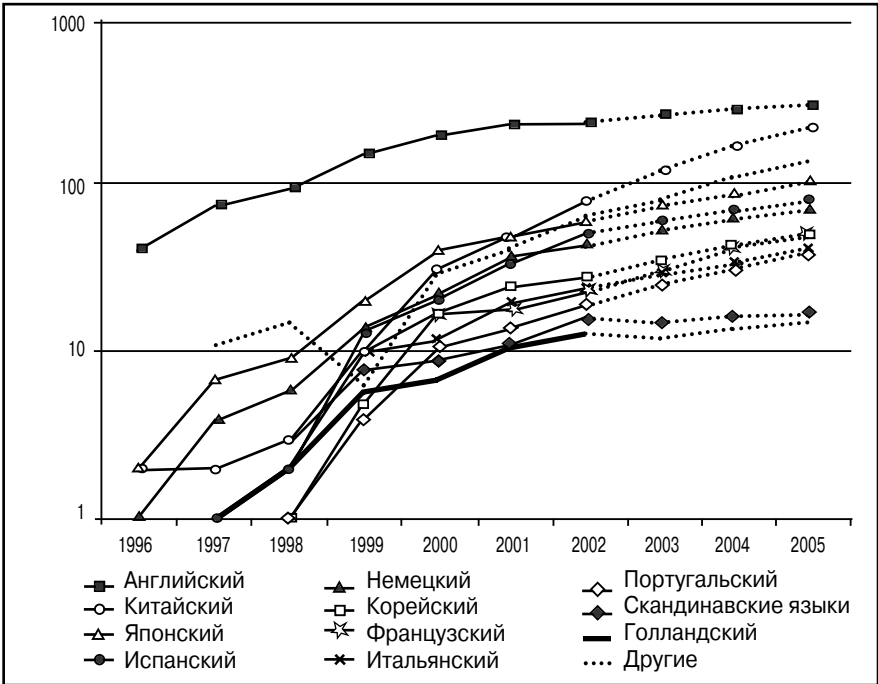
частота встречаемости слов соответствующих «культурно-нейтральных» концептов одинакова в разных языках. Однако культурный нейтралитет недостижим. Многие слова, чью частоту учитывают, представляют культурно-связанные концепты, например, «сыр» (cheese): американская англоязычная культура и культура континентальной Франции придают совершенно разное значение словам, обозначающим этот продукт – cheese и fromage. Данный факт найдет отражение в частоте употребления соответствующих терминов. Более того, если в результате мы получаем число страниц, а не слов, то результат, учитывающий разные формы слова в языке, может включать двуязычные или многоязычные страницы, которые подсчитываются многократно.

Языковое разнообразие пользователей Интернета

Наиболее эффективная попытка оценить языковое разнообразие пользователей Интернета была сделана переводческой компанией Global Reach. Оценки, которые она составляла ежегодно с 1996 по 2002 гг., широко цитируются в качестве иллюстрации мнения об Интернете как средстве развития языкового разнообразия²². Эти оценки базируются на данных Международного союза электросвязи (МСЭ) о количестве пользователей в каждой стране: пользователь определяется как человек, который пользовался Интернетом последние 3 месяца. Группы пользователей делятся по языкам, данные по которым взяты из Ethnologue и сверены с данными ООН по численности населения – т. е. практически так же, как поступали мы, подсчитывая языковое разнообразие (см. выше). В ряде случаев Global Reach дополняла эти источники маркетинговыми данными, полученными от компании Nielsen Net Ratings и некоторых других. В этом исследовании отсутствуют актуальные данные по пользователям Интернета, поэтому данные Global Reach не представляют языки, на которых говорят сегодня пользователи Интернета. Именно поэтому, что эти цифры часто цитируются как свидетельство языкового разнообразия пользователей Интернета, стоит посмотреть на них внимательнее.

22 Эти данные доступны на <http://global-reach.biz/globstats/evol.html>

Рис. 4. Количество пользователей Интернета, говорящих на различных языках (ось Y дана в логарифмическом масштабе)



Источник: Global Reach.

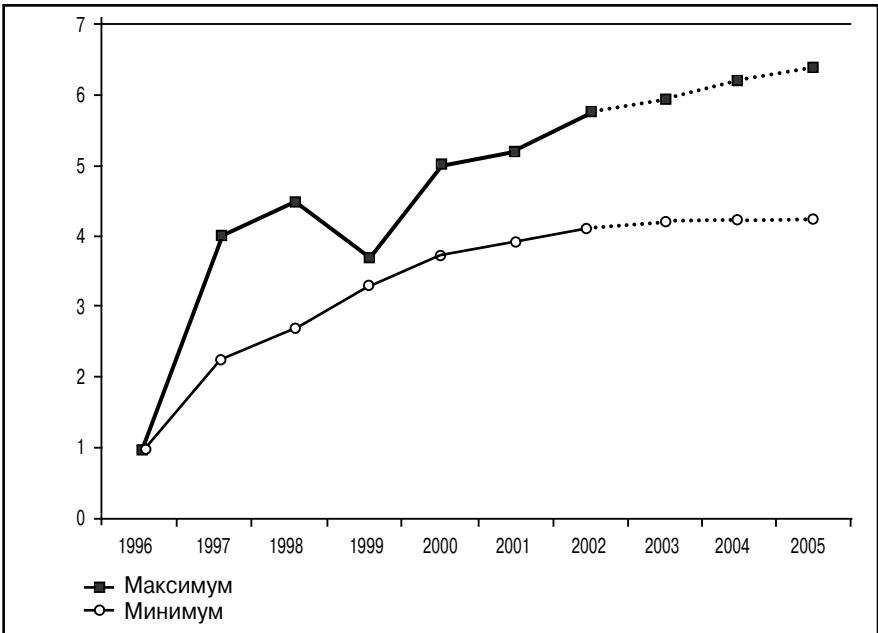
На Рис. 4 представлены данные Global Reach о распределении пользователей Интернета по языкам, на которых они говорят. Период с 2003 по 2005 г. показан пунктирной линией, поскольку это были прогнозируемые оценки. Установленные исследованием языки пользователей совпадают с языками, определенными OCLC. Как и ожидалось, в 2001 г. количество пользователей, работающих и общающихся в Интернете на английском языке (230 млн человек), в три раза превосходило количество пользователей, читающих и пишущих на китайском языке (60 млн пользователей)²³. Рис. 4 показывает, что все

23 Эти оценки рассматривают все варианты китайского как один язык, несмотря на то, что лингвисты считают китайский язык семьей из 9 разных языков (часто называемых «диалектами» среди нелингвистов).

группы пользователей переживают экспоненциальный рост, за исключением пользователей с английским и японским языками, рост численности которых замедляется. По полученным данным, 50 % от общего количества возможных пользователей Интернета, говорящих на этих двух языках, уже являются реальными пользователями Интернета.

Из оценочных данных, представленных Global Reach, можно вычислить индекс языкового разнообразия для общего числа пользователей Интернета; эти данные представлены на Рис. 5. Поскольку состав языков в группе «Другие» не расшифровывается в данных Global Reach, мы подсчитали минимальные и максимальные значения для индекса, приняв за «другие» один язык (минимальное разнообразие) или равное распределение по 6000 языков (максимальное разнообразие). Удивительно, что, несмотря на значительное увеличение индекса с 1996 по 1999 г., языковое разнообразие, похоже, стабилизируется после 2000 г., хотя и наблюдается экспоненциальный рост многих языков. Кроме того, прогнозы на 2003–2005 г. тоже показывают эту тенденцию выравнивания; прогнозируемый рост числа людей, говорящих на китайском языке, в силу его многочисленности, практически сдерживает рост разнообразия. Конечный результат – это индекс языкового разнообразия, находящийся где-то между типичной африканской страной и суммарного индекса Северной Америки и региональных индексов стран Европы. Ничего удивительного в этом нет, если учесть, что Интернет-хостинг по-прежнему сосредоточен в Северной Америке и Европе. Тем не менее, языковое разнообразие Интернета нигде не превышает индекса какого-то одного региона или мира в целом. С учетом сказанного делаем вывод, что, вопреки популярному убеждению, Интернет в этом смысле не может считаться олицетворением языкового разнообразия.

Итак, глобальный характер и способность соединять огромное количество людей не сделали Интернет лингвистически разнообразным. Чтобы гарантировать представление в сети языков, на которых говорят пользователи Интернета, надо решать и другие вопросы, что, как мы уже видели выше, очень зависит от конкретных сообществ, которые соединяет Интернет.

Рис. 5. Оценка разнообразия пользователей Интернета

Источник: Global Reach

Интернет и практика многоязычия

Доступ к Интернету является обязательным условием использования информации, представленной в Сети. До настоящего момента мы старались понять, что может означать этот доступ в глобальном масштабе. Однако наши усилия не будут успешными, если люди, говорящие на многочисленных языках мира, просто не выберут один из нескольких доминирующих языков. Что определяет языковой выбор пользователей Интернета?

Языки – это не просто средства передачи информации, а сложные системы символов, несущие богатые и тонкие оттенки смысла. Социолингвистические исследования многоязычия говорят о тонкой и неспокойной ситуации существования контактирующих языков, а недавнее исследование многоязычия в Интернете показывает, что такая же ситуация характерна и для Интернета. Более того, глубокий интерес к

Интернету во всем мире основан на экономических преимуществах, которые он предлагает. А способствует ли Интернет развитию языков?

Пользуясь общими терминами, довольно сложно идентифицировать, какие языки и как используются в Сети. Требуется рассмотреть целый комплекс вопросов, начиная с индивидуализации языковых сообществ и заканчивая дифференцированным доступом к Интернету, разными системами письма и компьютерными кодировками, разными коммуникационными режимами. Большинство работ, посвященных возможному влиянию Интернета на язык и культуру, изучают конкретные языковые группы, использующие Интернет в своем специфическом контексте, а не занимаются исследованиями с макросоциальных позиций. Такие конкретные исследования показывают, что языковые контакты в Интернете способствуют развитию больших языков, равно как и контакты вне Сети. Так, Райт (Wright, 2004) и Холмс (Holmes, 2004) приводят данные исследования, посвященного языковому поведению в Сети студентов колледжей, проживающих в восьми разных странах. Результаты показывают, что степень использования родных языков в Сети сильно изменяется в зависимости от исследуемого контекста. В то же время, ни одна из исследованных групп студентов не показала, что использует в Сети свой полный языковой репертуар. Менее распространенные языки вообще не используются в Интернете. Как следствие, вопросы, связанные с этой темой, сложны и туманны.

В своем первом исследовании Паолилло (Paolillo, 1996) открыл, что дискуссионные группы в Usenet, аудитория которой, в основном, говорит на пенджаби, предпочли английский своему родному языку. Такое поведение частично предсказуемо в аудитории, состоящей, в основном, из эмигрантов и людей, получивших образование на английском языке, но наблюдаемые тенденции говорят о том, что пенджаби используется в Сети практически исключительно в ритуальных целях или при общении по национальным вопросам, и служит больше как знак принадлежности, чем как средство информативной коммуникации. В одной из своих следующих работ Паолилло (2006) провел сравнение интерактивности и языковой однородности жителей Южной Азии в ходе их общения в чате и дискуссионных группах в Интернете и обнаружил, что и в том, и другом случае предпочтение отдавалось миноритарному языку (хинди или пенджаби, в зависимости от форума). Эти тенденции получили подтверждение у Пила (Peel, 2004), который показал, что пред-

почтительным языком интерактивных чатов в ОАЭ является арабский язык, а электронные сообщения составляются на английском языке. В другой работе Паолилло (2001) писал, что участники чата, живущие в центре, чаще используют миноритарные языки, чем люди с периферии. Поскольку система чатов позволяет людям включаться в их работу и выходить из нее по собственному желанию, участники с периферии и их языковые предпочтения оказываются доминирующими. Как мы видим, технологические и социальные аспекты Интернет-коммуникации взаимодействуют сложным образом, но все же предпочтение отдается мажоритарным, а не миноритарным языкам. С помощью технических средств можно сгладить последствия языкового доминирования, но неизвестно, насколько это окажется эффективным.

Исследованиями греческого языка в Интернет-коммуникации занимались Кутсогианнис и Митсакопулу (Koutsogiannis and Mitsakopoulou, 2004), Георгакопулу (Georgakopoulou, 2004) и Андросополус (Androtsopolous, 1998). Тематика их исследований охватывает многие вышеперечисленные вопросы. Греческий алфавит, так же как шрифт гурмухи в пенджаби, нелегко использовать в Интернете, а латинизированная форма греческого, адаптированная для внесетевого общения и известная как «Greeklish» (текст на греческом языке, написанный на латинице), встречает поддержку у эмигрантов в многоязычных контекстах, отдающих предпочтение английскому (Georgakopoulou, 2004) или немецкому (Androtsopolous, 1998) языкам. Это, в свою очередь, разрушает греческую норму диглоссии (Ferguson, 1959), при которой говорящие используют отчетливую разговорную форму в неофициальной речи и классический язык в письменной речи. В прошлом правительство Греции прикладывало немало усилий для поддержания грамотности на катаревуса – классическом официальном письменном языке; разрушение греческой диглоссии в Интернете подрывает эти усилия. В другом контексте диглоссии – арабском – Варшауер и др. авторы (Warschauer et al., 2002) отмечали, что разговорный египетский, арабский и английский языки «покушаются» на традиционные функции классического арабского. Такое «покушение» имеет тенденцию дестабилизировать ситуацию с диглоссией, приводя, в конце концов, к предпочтению внешнего, доминирующего языка. В результате, когда в Интернете происходит разрушение лингвистических норм, универсальный доступ к Интернету может иметь разрушительные последствия для языкового разнообразия.

Влияние английского языка одновременно широко и изоциренно. Резко контрастные ситуации с электронной почтой в Швейцарии (Durham, 2004) и использованием Интернета в Танзании (Mafu, 2004) показывают, что люди, говорящие на двух языках, отдают предпочтение английскому, а не более близким им местным языкам. Если в употреблении английского языка среди элиты Танзании есть колониальный подтекст, то его нет и не может быть в Швейцарии. Объяснение этому феномену мы находим только в международном статусе английского языка (Crystal, 2003; Phillipson, 1992, 2003). Другим примером влияния английского в Интернете является распространение ряда функций устных языков на письменные языки через SMS, IM (instant messages) и Интернет-чат на шведском языке (Hard af Segerstad, 2002). Точно так же Торрес (Torres, 1991, 2001) отмечал прагматические функции эмоциональных иконок (смайликов) в чате на каталонском языке. Эти формы произошли из англоязычных контекстов и поэтому указывают на контактное влияние английского языка на каталонский через Интернет.

Эти и другие исследования освещают богатство и сложность факторов, влияющих на использование миноритарных языков многоязычными пользователями Интернета. Тема, постоянно появляющаяся во многих исследованиях, касается хрупкости использования недоминирующих языков в коммуникационных контекстах Интернета.

Организации и интересы, управляющие Интернетом

Наперекор популярным убеждениям, Интернет – не открытая и не демократическая (или анархическая) организация. Скорее, это организация со сложной системой мощных интересов, многие из которых сильно централизованы. Эти мощные интересы зачастую не принимают в расчет действия индивидуальных пользователей и, тем самым, создается впечатление, что Интернет свободен от ограничений со стороны гражданских, государственных или корпоративных интересов. Тем не менее, каждый уровень интересов – это возможность определить по языковым сдвигам, какие языки используются в Интернете.

Регулированием Интернета занимаются несколько основных действующих лиц. Во-первых, это телекоммуникационные монополии и олигополии разных регионов мира. Эти компании осуществляют под-

держку инфраструктуры, которая позволяет индивидуальным пользователям подключаться к Интернету, а Интернет-сайтам – соединяться друг с другом. Во-вторых, это компании, производящие компьютеры и программное обеспечение. В их числе Intel, IBM, Hewlett-Packard, Cisco Systems, Sun Microsystems, Microsoft, Adobe. Эти компании создают и продают технические и программные средства, составляющие инфраструктуру Интернета. Кроме того, это управляющие органы Интернета, такие как Корпорация Интернета для специализированных адресов и номеров (ICANN), и Информационный центр сети Интернет (NIC), Американское бюро регистрации адресов (ARIN), Координационный Центр распределения ресурсов сети Интернет в Европейском регионе (RIPE) и Азиатско-тихоокеанский информационный центр сети Интернет (APNIC), которые принимают решения по вопросам подключения к Интернету (см. глоссарий). Национальные правительства также играют определенную роль как в руководстве Интернет-ресурсами на государственном уровне, так и во внедрении других форм информационной политики. И, наконец, есть организации и консорциумы, такие как Консорциум Всемирной сети (W3C), консорциум Юникод (Unicode Consortium) и Международная организация по стандартизации (ISO), которые разрабатывают стандарты для применения технологий Интернета.

Телефонные сети с самого начала были очень важны для Интернета. Когда Интернет-хост соединяется с другим хостом, модемы, выделенные линии, цифровые абонентские линии, оптоволоконные опорные магистрали и геосинхронные спутники – все могут быть задействованы на некотором этапе цифровой коммуникации, физически передавая данные по телефонной сети. За последнее время для передачи Интернет-трафика были адаптированы другие формы телекоммуникационных сетей, такие как телевизионные кабельные сети. Исторически и сегодня экономический контроль над этими ресурсами находится в руках крупных компаний – частных или государственных монополий. Эти концерны наиболее развиты в США. Так, MCI через свое дочернее предприятие UUNET управляет сетью, на которую приходится большая часть международного интернет-трафика (см. Mapnet, <http://www.caida.org/tools/visualization/mapnet>). Оптоволоконная опорная магистраль, введенная MCI несколько лет назад, является главной в этой сети. В то время как компании, аналогичные MCI, практически не интересуются языками, на которых работают Интернет-

пользователи их линий, центральное место, занимаемое США в распределении трафика данных, гарантирует, что административные задачи высокого уровня относительно трафика на опорных магистралях будут решаться на английском языке. Как следствие, региональные сети, связанные с этими центральными сетями, просто обязаны привлекать людей, хорошо владеющих английским языком.

Если на первый взгляд это не представляет собой проблему, поскольку специалисты в области компьютеров во всем мире хорошо владеют английским языком, обе эти тенденции подпитывают и усиливают друг друга. Если руководство региональных сетей не может общаться со своими провайдерами на предпочитаемом ими языке, тогда английский по умолчанию остается доминирующим языком сетевой администрации. Телекоммуникационные компании, собирающие урожай огромных прибылей благодаря спросу на коммуникационные и технологические услуги, несут особую ответственность за языковое разнообразие на рынках, которые они обслуживают.

Компании, производящие компьютерное оборудование и программное обеспечение, оказывают аналогичное влияние на языковой состав Интернета, создавая компьютеры с клавиатурами, мониторами и операционными системами, настроенными на конкретные языки. Эти изделия производятся по низкой цене за счет масштаба производства, и в результате самые крупные рынки мира оказываются заполненными стандартизированной продукцией. Компьютерные технологии с их офшорным производством чипов, разработкой программ за счет аутсорсинга (включая даже менеджмент) и товарными рынками представляют собой один из первых глобальных секторов промышленности. По этой причине и в силу лидирующего положения компаний США в области разработки новых систем и стандартов, компьютерные системы, прокладывающие путь в такие разнообразные в языковом отношении регионы, как Африка, в подавляющем большинстве настроены на использование английского или европейских языков, и почти или совсем не приспособлены к настройке на локальные языки. Это выражается в еще одной форме возникающего сдвига в Интернете в направлении европейских языков, уводящего все дальше от языков тех стран, которые менее развиты в экономическом отношении. Как и телекоммуникационные компании, фирмы, производящие технические средства и программы, несут особую ответственность за языковое разнообразие в тех странах, рынки которых они обслуживают.

Таким образом, действия компьютерных компаний, зажатых конкуренцией за доминирование на рынке, оказывают отрицательное влияние на климат многоязычия в области вычислительной техники и онлайн-языкового разнообразия. Чтобы стимулировать многоязычие в вычислительной технике, нужны меры, предусматривающие приоритет международных интересов над конкурентными целями частных компаний. Некоторые из этих тенденций подкрепляются действиями международных организаций и консорциумов, таких как ISO, Консорциум Юникод, W3C, которые в состоянии предвидеть различные аспекты развития Интернет-технологий. Многие крупнейшие компьютерные компании, включая Apple и Microsoft, работают через эти организации. И неважно, что кто-то из специалистов по технологиям жалуется, что данные организации тормозят инновации, ведь их международный статус помогает принимать во внимание интересы различных национальных и языковых групп. С другой стороны, эти организации по стандартизации не имеют реальных механизмов принудительного применения принятых ими решений. Как следствие, ряд интернет-технологий имеет стандарты, редко соблюдаемые на практике. В их числе использование HTML для веб-страниц и язык программирования ECMAScript для интерактивности веб-браузеров. Несовместимость, возникающая в результате несоблюдения стандартов, разрушительна для развития многоязычия в вычислительной технике. Если мы хотим, чтобы эти организации стимулировали и защищали языковое разнообразие, то нам следует усилить их правоприменительные возможности.

Есть еще одно действующее лицо, управляющее Интернетом и оказывающее большое влияние на его языковое разнообразие, – это ICANN, корпорация, осуществляющая по контракту с Министерством торговли США администрирование протокола, известного под названием «Система доменных имен» (Domain Name System, DNS). DNS осуществляет функцию ассоциирования уникальных мнемонических имен со всеми хостами в Интернете – функцию, которая, в основе своей, является лингвистической. К сожалению, DNS с трудом совмещается с каким-то языком, если только он не является американским английским, и более того, с трудом уживается с функцией присвоения названий на любом естественном языке. DNS глубоко интегрирована в Интернет, поскольку большинство других прикладных протоколов зависят от ее способности находить Интернет-хосты. Это также единственный действительно

контролируемый организацией, а не просто кодифицируемый, протокол. ICANN контролирует DNS, в основном, путем делегирования, но его административная структура, система контрактов с правительством США и другими сторонами, методы работы – все было направлено на ограничение многоязычия в названиях Интернет-хостов. В результате этого DNS не в состоянии выполнить свою изначальную задачу предоставления мнемонических схем для Интернет-хостов. Для изменения этой ситуации требуется изменить ICANN, DNS и политику, проводимую администрацией доменных имен.

Пользователи Интернета воспринимают имена Интернет-хостов почти как обычные имена, а на самом деле, они сильно разнятся. DNS требует, чтобы имена Интернет-хостов были уникальными во всем мире, в то время как в естественном языке метафора, система условных обозначений и аббревиатуры исключают возможность уникальности какого-то конкретного имени. После присвоения домена «acl.org» Ассоциации библиотекарей-христиан (Association of Christian Librarians), это имя становится недоступным для Ассоциации по вычислительной лингвистике (Association for Computational Linguistics) или для любой другой организации в мире, которая хотела бы иметь эту же аббревиатуру в качестве своего названия.

Для обеспечения уникальности в условиях ограниченной гибкости DNS использует иерархически-структурированные имена: индивидуальные имена хостов состоят из цепочек имен, ранжированных от специфичного к более общему. Верхний уровень иерархии занимает последнее поле имени; это будет родовой домен верхнего уровня или код страны (gTLD или ccTLD), который функционирует как общий классификатор. Однако часто остается непонятным, какой классификатор нужно использовать для конкретной цели. TLD, на основании соглашений с ICANN, присваиваются с учетом различных функций: .com предназначен для коммерческих сайтов, .net – для сетей, .org – для неприбыльных и некоммерческих организаций, а коды стран должны присваиваться самими странами с учетом их собственных целей. Доменные имена в gTLD более привлекательны, поскольку, как правило, короче и легче для запоминания. Однако, поскольку количество gTLD незначительно в сравнении с сотнями миллионов хостов, конфликты при присвоении доменных имен неизбежны.

Отношение ICANN к таким конфликтам и к их разрешению строится на поддержке владельцев официально зарегистрированных товарных знаков. В иных случаях первая сторона, регистрирующая свое доменное имя, сохраняет его до тех пор, пока регистрация остается в силе. Это не помогает регистрантам, не имеющим товарного знака или происходящим из маленькой области, или носителям миноритарного языка. Это особенно мешает иностранным заявителям, если именем их национальной идентификации являются омографы ранее зарегистрированного доменного имени. Чтобы изменить имя зарегистрированного домена, требуется провести дорогостоящие переговоры и/или совершить необходимые юридические действия. Регистрация сотен миллионов хостов на английском языке привела к явному сдвигу против интересов неанглийских имен хостов, поскольку тысячи выбранных имен хостов на других языках будут омографами ранее зарегистрированных хостов в gTLD. Как следствие, в DNS присвоение товарного знака как юридически обязательное в США торговое действие получает преимущество над прозрачным многоязычным наименованием, относящимся к сфере языка и международной коммуникации. Такой ассиметричный порядок приоритетов не изменится, пока DNS не будет находиться под управлением полностью международной, а не частной организации, имеющей договорные отношения с правительством США или любой другой страны.

Оригинальная разработка DNS отличается явным техническим сдвигом в направлении предпочтения английского языка, потому что может использовать только 7-битный US-ASCII-код. Поэтому даже европейские языки, такие как французский, испанский и немецкий, использующие диакритику, не отображаемую в US-ASCII, оказываются в затруднении, когда речь заходит о выборе подходящих имен для Интернет-хостов. Ряд организаций, в числе которых Консорциум многоязычных Интернет-имен (MINC), New.net and RealNames долго боролись, стараясь убедить ICANN разработать альтернативы современной системе DNS, обладающие более совершенной поддержкой многоязычия. Несмотря на то, что эти группы внесли много конструктивных, заслуживающих внимания предложений, их усилия встретили большое сопротивление со стороны ICANN. Только недавно ICANN приняла вариант Юникода, известный как «пуникод» (punycode), допускающий существование многоязычных доменных имен, но внедрение этого варианта

проходило неприемлемо медленно, и политически было обречено на провал.

Вопрос доменных имен, в принципе, достаточно символичен. Тем не менее, символизм – это сила, и нетерпимость ICANN к многоязычным доменным именам привела к тому, что мир понял, что этой организации нет дела до интернационализма или языкового разнообразия. Несмотря на то, что в последнее время ICANN была сильно реформирована и сейчас претендует на более интернациональное звучание, она потеряла общественное доверие по вопросу о многоязычных доменных именах, и не ясно, приведут ли эти изменения к справедливой, функциональной и международной системе доменных имен, и можно ли восстановить утраченное доверие.

Роль организаций ARIN, RIPE и APNIC (равно как и других NIC'ов – сетевых информационных центров) в появлении возникающего языкового сдвига не столь очевидна, как роль ICANN. Эти организации с относительно открытым членством осуществляют руководство физическим взаимодействием региональных и локальных сетей. Одной из их основных функций является поддержка пространства IP-адресов. IP-адреса – это 32-битные номера, используемые для уникальной идентификации хостов. Как и доменные имена, IP-адреса присваиваются в процессе делегирования посредникам, которые могут передавать полномочия дальше. В отличие от доменных имен, каждый присвоенный диапазон соответствует физической ветви Сети, соответствующее оборудование которой управляется одним провайдером. IP-адреса присваиваются в диапазонах, а поскольку адресное пространство, в конечном итоге, лимитировано, каждое такое присвоение имеет цену возможности – такие же номера не могут быть присвоены позднее где-то в другом месте, если только эта часть Сети не прекращает своего существования.

Сетевые информационные центры (NIC) вынуждены заниматься вопросами языкового разнообразия, выполняя возложенные на них функции региональной администрации. Сетевые ресурсы, доступные конкретной стране или языковой группе, зависят от диапазона IP-адресов, предоставленных соответствующему региональному администратору, и их передачи от него другим группам и странам. Плохая передача адресов или малый диапазон, с которого начинается работа, представляют собой два препятствия, которые могут привести к недостатку

адресов для новых хостов. Разгорелась полемика, имеет ли APNIC, региональные обязанности которой охватывают Океанию, Восточную и Юго-Восточную Азию, достаточно места для адресов, чтобы продолжать присвоение IP-диапазонов с необходимой скоростью. APNIC отрицает существование такой проблемы, но призрак надвигающегося кризиса вызывает тревогу. Как ожидается, проблемы, связанные с адресным пространством, могут усугубиться в результате модернизации текущей 4-й версии IP (IPv4) и появления 6-й версии IP (IPv6), которая использует более широкий диапазон адресов; но эта конверсия была отложена на несколько лет по причине технической несовместимости с IPv4.

Тем не менее, присвоение адресного пространства IPv4 очень неэффективно. Большие диапазоны адресного пространства выделяются для специальных целей или остаются полностью неиспользованными; они известны как «bogons» и находятся под строгим контролем для того, чтобы системные администраторы могли осуществлять их мониторинг в целях безопасности (см. <http://www.cymru.com.Bogons/>). Даже когда диапазоны «bogon» были выведены из области использования, произвольная выборка из 1107 IP-адресов возвратила 203 IP-адреса (18 %), которые, очевидно, были отданы под тестирование редко используемого протокола многоадресной передачи. Иными словами, 18 % глобально доступного пространства IP-адресов были заблокированы и стали неиспользуемыми вследствие неэффективности передачи этого пространства. Именно потому, что такая неэффективность может иметь место, и потому, что она может оказывать влияние на адресное пространство, выделяемое региональным администрациям, локальные языковые группы могут не получить доступ к Интернет-ресурсам. Чтобы разные языки имели равный шанс на использование в Сети, администрирование и выделение адресного пространства в Интернете должно осуществляться справедливо.

Национальные правительства могут играть как положительную, так и отрицательную роль при оказании влияния на языковые сдвиги в Интернете. Когда национальные правительства принимают внутри своих стран политику, защищающую и развивающую языковые права своих многоязычных граждан (Skutnabb-Kanngas and Phillipson, 1995), исторические языковые сдвиги в этих странах находятся под контролем. Поскольку эта языковая политика переносится на соответствующие области информационной политики, она

способствует развитию языкового разнообразия в Интернете. Однако, как правило, правительства больше озабочены административной стороной деятельности и опасностями сепаратизма, и значительная часть населения мира живет без гарантий обеспечения своих базовых языковых прав. Когда страны контактируют с глобальным Интернетом и требуют приспособить его к своим национальным языкам, они стимулируют появление возникающего сдвига, в ущерб своим собственным этнолингвистическим меньшинствам, принося, в конечном итоге, мало содействия языковому разнообразию в сети. Если национальные языковые группы надеются обеспечить себе нишу в глобальной телекоммуникационной этносфере, они должны признать языковое разнообразие и заниматься им в пределах своих национальных границ. Они могут также стремиться к обучению компьютерной грамотности граждан всех своих языковых групп, поскольку это необходимо для полноценного использования Интернета. Этнолингвистическое понимание телекоммуникационных компаний, компьютерных компаний и руководящих органов Интернета начнет расширяться только тогда, когда критическая масса недостаточно представленных этнолингвистических групп сможет привлечь к себе внимание. Это вряд ли произойдет, пока скрыт истинный размах международного языкового разнообразия.

Возникающие языковые сдвиги – важный аспект, требующий к себе особого внимания при рассмотрении вопросов языкового разнообразия в Интернете. Проблемы, рассмотренные нами в настоящей статье, носят чисто иллюстративный характер и не претендуют на представление исчерпывающего списка возможных возникающих сдвигов. По мере изменения рынка телекоммуникаций, компьютерной техники и программного обеспечения, по мере развития организаций, осуществляющих руководство Интернетом, могут появляться новые языковые сдвиги. Возникающие в разных технических и лингвистических контекстах языковые сдвиги могут быть глубоко локальными, поскольку проявляются особым образом внутри конкретной страны. Именно поэтому общая проблема возникающих языковых сдвигов требует тщательного мониторинга на глобальном, региональном и локальном уровнях.

Источники технических сдвигов

Три области технических сдвигов, имеющих разное отношение к языковому разнообразию, заслуживают внимания в свете усилий по интернационализации в рамках трех рассмотренных выше направлений действий ЮНЕСКО. Во-первых, это вопрос стандартов кодирования, который имеет прямое отношение к действиям по п. 10, стимулируя языковое и культурное разнообразие в Интернете. Кодирование текстов – основное техническое средство достижения языкового разнообразия в этой базовой, основанной на текстах, коммуникационной среде. Во-вторых, это вопрос языков разметки и программирования, используемых для создания и поддержания Интернет-приложений и контента. Эти технические системы напрямую связаны с п. 9 – продвижением компьютерной грамотности. Если компьютерная грамотность в качестве обязательного условия требует грамотности в другом языке, открытость и универсальный доступ гарантировать невозможно. Наконец, это вопросы, связанные с техническими языковыми сдвигами в протоколах Интернет-приложений – пп. 9 и 10. Для содействия доступу развивающихся стран к информационным технологиям основные Интернет-приложения (электронная почта, гипертекстовые браузеры, мгновенный обмен сообщениями в сети и пр.) должны разрешать использование языков этих стран. Без этого барьеры на пути принятия технологий станут непреодолимыми. Рассмотрим подробнее три названные области технических сдвигов.

Кодирование

Кодировки определяют произвольное присвоение чисел символам письменных языков. Два разных вида кодировок могут быть несовместимыми в случае присвоения одного числа двум разным символам и наоборот. Чтобы воспользоваться возможностями, которые предоставляет нам компьютер при работе с текстом (по его отображению, редактированию, сортировке, поиску и быстрой передаче), коммуникация на определенном языке должна быть представлена в некой кодировке. Поэтому большая часть из того, что предлагает нам Интернет в отношении языкового разнообразия, сводится к кодировкам конкретного текста.

Наиболее распространенной кодировкой является ASCII код (American Standard Code for Information Interchange), разработанный в 1950–1960-е гг. под эгидой ANSI – Американского института нацио-

нальных стандартов – для стандартизации технологии работы по теле-тайпу. Этот код состоит из 128 символов и пригоден, в основном, для североамериканского английского. По причине его раннего появления и широкого распространения все коды, появившиеся позже, строились с ориентацией на ASCII, например, ISO-8859-1 или Latin-1 Международной организации по стандартизации. Этот код содержит 256 символов, первые 128 из которых совпадают с символами ASCII. Юникод – попытка предоставить совместимые кодировки для всех языков мира (Unicode Consortium 1991, 1996, 2000, 2003) – принял на вооружение такую же стратегию, сделав первые 256 из 65 536 символов в BMP (Базовой многоязыковой плоскости) такими же, как в ISO-8859-1. Большинство технологий, поддерживающих Интернет, основано на ASCII или его производных. DNS, Usenet news и Internet Relay Chat, разрешают использование только поднабора символов ASCII. Операционные системы, такие как Linux, используют исключительно «текстовые ASCII-файлы» для ряда своих базовых функций. Все эти системы усиливают технический сдвиг в направлении английского языка.

Надежды на интернационализацию инфраструктуры Интернета связаны с окончательным принятием Юникода – попыткой стандартизации, предпринятой Консорциумом Юникод в сотрудничестве в ISO. Членами Консорциума являются крупнейшие поставщики программного обеспечения, представители мировых религий, региональные организации и учреждения образования, правительственные организации разных стран. Стандарт Юникод (в нынешней 4-й версии) предоставляет более миллиона возможных кодов символов, что позволяет использовать в едином тексте все современные и исторические письменности. 65 000 символов составляют Базовую многоязыковую плоскость, которая должна обеспечить большую часть письменной коммуникации. Однако такое разнообразие стоит денег. В своей базовой форме UTF-32 текст в Юникод занимает в 4 раза больше места в сравнении с аналогичным текстом в ASCII. Многие разработчики программ предполагали, что пользователи не захотят получить такое «наказание» за пользование многоязычным текстом, особенно если работа на компьютере происходит в монологичном контексте²⁴. Юникод предлагает другие виды кодирования с символами переменной длины, которые бо-

24 Так это или нет – важный вопрос, не получивший достаточного рассмотрения в научной литературе.

лее эффективны, но ценовое бремя переносится на нелатинские языки, которые вынуждены занимать больше места. И хотя за последние 10 лет цены на хранение данных сильно упали, позволяя снизить проблемы Юникода, работа с этой кодировкой по-прежнему сильно усложняет задачу разработчикам программ, поскольку большинство приложений требуют взаимодействия с ASCII. Кроме того, большой размер документов в Юникоде требует больших затрат на пересылку, сжатие и распаковку, и эти расходы – достаточное основание для того, чтобы в ряде случаев отказаться от использования Юникода.

Несмотря на крупные изменения в деле интернационализации обработки данных с помощью Юникода, проблемы с использованием многоязычного текста в Интернете еще далеко не решены. По ряду технических, экономических и организационных причин разработка приемлемого технического стандарта идет медленнее, чем развитие самого Интернета. Как следствие, международное использование Интернета поддерживало языки, основанные на латинице (английский язык в особенности), чему способствовало наличие широко распространенного стандарта кодировки, принятого до распространения Интернета. Чтобы в Интернете можно было одинаково использовать все языки мира, Юникод должен получить более широкое распространение. Как и в случае с DNS, может потребоваться модернизация ряда Интернет-протоколов и приведение их в состояние, совместимое с Юникодом.

Языки разметки и программирования

Есть и другой инструмент, благодаря которому в Интернете сохраняются технические сдвиги в пользу английского языка, – это компьютерные «коды», т. е. языки разметки и программирования, которые используются для конфигурации Интернет-контента и услуг. Первым и наиболее очевидным способом создания технических сдвигов является поддержка этими языками многоязычного контента. Языки разметки, такие как Hypertext Markup Language (HTML) и eXtensible Markup Language (XML), должны уметь описывать текст на всех возможных человеческих языках. Консорциум Всемирной сети предпринял для этого ряд шагов, потребовав поддержать Юникод как элемент своих стандартов. Это означает, что там, где поддержки Юникода недостаточно, поддержки HTML и XML тоже недостаточно. Именно это происходит сейчас с большинством языков Западной и Юго-Восточной Азии, а также южной части

Центральной Азии. Так создается общая основа для предпочтения определенных языков. Языки программирования также должны быть совместимы с многоязычным текстом. К сожалению, многие распространенные языки программирования, такие как С, пока еще не обеспечивают стандартной поддержки Юникода²⁵. Растущее число языков, созданных для веб-приложений (в их числе широко используемые Java, JavaScript, Perl, PHP, Python, Ruby), способны на поддержку Юникода, но другие системы, как, например, программное обеспечение для баз данных, непостоянны в своей поддержке Юникода. Организация электронной коммерции на различных языках, а не только на английском, предполагает, что совместимые с Юникодом базы данных получат широкое распространение.

Следующий путь создания сдвига в сторону английского языка – дизайн самих языков разметки и программирования. Языки программирования предлагают самый простой из имеющихся интерфейс для управления компьютером, который является посредником между когнитивными процессами программистов и логическими возможностями самих компьютеров. Число языков программирования огромно – от 2500 до количества, превышающего число человеческих языков. Несмотря на такое очевидное разнообразие, подавляющее большинство этих языков берет свое начало в FORTRAN – самом первом из языков программирования высокого уровня, созданного IBM в 1957 г. (Levenez, 2003). Эти языки широко используют английские ключевые слова для определения важных компонентов программирования, таких как условные операторы (if, then, else, case и др.) и итерационные циклы (while, for, until и др). Несмотря на то, что во многих человеческих языках есть эквиваленты этих слов, они никогда не используются вместо этих английских слов в выполняемой программе. Так, Руби, работающий в соавторстве с японским программистом Юкихио Матсумото и обращающий особое внимание на интернационализацию, также использует английские ключевые слова.

HTML и XML в этом отношении очень похожи. Тэги HTML представляют собой, как правило, мнемонические аббревиатуры английских

25 Веб-сайт «International Components for Unicode» предлагает C-библиотеку с открытым кодом, поддерживающую использование Юникода (<http://oss.software.ibm.com/icu/>).

слов (например, **b** для **bold** – полужирный шрифт, **ul** для **unordered list** – неупорядоченный список, **li** для **list item** – элемент списка и др.). И, хотя XML не является языком разметки, это – синтаксис для определения языков разметки, все языки разметки, базирующиеся на XML, используют английский (например, MathML для математических выражений и XML: FO для форматирования текстовых документов), несмотря на стандарт XML, использующий Юникод. Эта тенденция усугубляется в связи с проектом развития Семантического веба («Semantic Web»), цель которого в том, чтобы привнести во Всемирную сеть идеологию «общеизвестного знания». Предполагается, что крупные базы данных искусственного интеллекта, такие как Cyc (Reed and Lenat, 2002) и WordNet (Fellbaum and Miller, 1998), будут использоваться для разработки новых языков разметки, и эти языки будут способствовать тому, что Интернет-программы смогут находить и обрабатывать информацию для пользователей. Такие базы данных уже подверглись критике с позиций культуры стран Северного полушария за сексистские, патриархальные предпочтения (Adam, 1998). Именно поэтому проекты типа Семантического веба, которые обещают дать жизнь «новому поколению» Интернет-услуг, несут в себе опасность углубления существующих языковых и культурных сдвигов.

Потенциальную опасность языковых сдвигов, заложенную в языках программирования и разметке, следует анализировать параллельно с культурной природой вычислений. Современная система вычислений основывается на многовековой истории математических знаний, и распространение этой системы на современном этапе можно сравнить по характеру и по значению с распространением системы десятичного исчисления. Десятичные числа были изобретены в Северной Индии около VII века н.э. и распространились по всему миру, вытесняя другие системы исчислений. При этом культурная передача десятичных чисел не требовала импортирования словарного состава; многие языки модифицировали свои названия чисел для восприятия нового метода. Компьютер развивает принципы десятичных чисел путем автоматизации работы с ними. Однако в отличие от десятичных чисел распространение компьютеров принесло с собой объемные и сложные английские словари – языки программирования.

Нет сомнения, что компьютер, как физический артефакт, привязывая символы к действиям, играет определенную роль в их взаимосвязи. Точное сочетание символа и действия имеет произвольный характер,

и поэтому адаптировать можно любой язык, но это – довольно сложная задача, выполнение которой далеко не тривиально. Как следствие, один важный вопрос, касающийся языкового разнообразия, не получил должного внимания в научной литературе, а именно: до какой степени различные характеристики языков программирования способствуют тому, что их воспринимают и используют люди, говорящие на разных языках?²⁶ Хорошо известны эффекты переноса у людей, говорящих на одном языке и изучающие другой. Вполне естественно, что языки программирования, как формальные лингвистические системы, могли проявлять существующий в естественных языках перенос, создавая систематические проблемы или ошибки у людей, говорящих на определенных языках. По своим свойствам языки программирования сильно отличаются друг от друга. Возможно ли, что люди, говорящие на каком-то языке, получают более качественные услуги от того языка программирования, который больше соответствует их родному языку? Тогда, может быть, языки программирования могут быть составлены с учетом различных культурных и языковых традиций? Поможет ли такая адаптация людям, говорящим на каком-то определенном языке, освоить ресурсы информационных технологий?

ЮНЕСКО и другие агентства ООН испытывают настоятельную потребность увидеть ответы на эти вопросы, особенно тогда, когда для продвижения языкового разнообразия необходимо решать задачи в области образования. Компьютерное программирование сделало язык живым и сильным, способным видоизменять культуры. Жаль, что сегодня такую поддержку получил, в основном, английский язык. Если цифровая грамотность в области языков программирования требует языковых и/или культурных знаний на английском, тогда люди, говорящие на других языках, должны нести большие расходы на образование и, возможно, на культуру, чтобы заявить о своем праве на информационные ресурсы Интернета.

Коммуникационные режимы

Несмотря на то, что Интернет стал известен большинству людей благодаря Всемирной сети (а многие считают, что это одно и то же), он пред-

26 См. Anis (1997) для ознакомления с предложениями по данному вопросу.

ставляет собой более разнородную среду, предлагающую различные варианты коммуникации. Более того, дизайн Интернета подразумевает возможность создания и недорогого разворачивания новых вариантов. Сегодня мы широко используем электронную почту, Сеть и мгновенный обмен сообщениями в Интернете и не знаем, что мы будем использовать в будущем. Определенные коммуникационные режимы стали очень популярны и периодически инкорпорируют технические формы языковых сдвигов.

Одним из таких коммуникационных режимов является Usenet News, изначально созданная в 1978 г. как сеть трех университетских компьютерных систем (Spencer and Lawrence, 1998). Usenet – это коллекция тысяч «новостных групп», пространство для публичных сообщений с именами, позволяющими отнести их к определенной теме. Сервисное и клиентское программное обеспечение Usenet имеется в свободном доступе, и его администрация относительно открыта. Администраторы Usenet могут в индивидуальном порядке устанавливать объем, скорость и частоту обмена сообщениями с другими серверами, так что они могут быстро оптимизировать пользование сетью в районах с плохой связью. Таким образом, барьеры для доступа в Usenet относительно невысоки. Usenet – чрезвычайно важный международный ресурс. По состоянию на 1999 г. доступ в Usenet имели 205 стран (Smith, 1999).

Как техническая система, Usenet – это микрокосм Интернета. Система имен для новостных групп носит иерархический характер и использует подсистему ASCII, почти так же, как DNS. Она имеет иерархии верхнего уровня и локальные, региональные и национальные иерархии²⁷. Текст сообщений должен быть совместим с ASCII. Тексты на китайском и японском языках используют в Usenet особые кодировки. Как и в остальной части Интернета, имеет место преобладание английского языка в родовых иерархиях верхнего уровня. Так, в иерархии comp. – родовой категории записей, касающихся компьютерных систем, почти нет записей на японском языке, даже на comp.lang.ruby. Только в иерархии fj.comp. можно найти дискуссии по вычислительной технике на японском языке. В ветви иерархии soc.culture. есть место для многоязычного трафика, но, в основном, на европейских языках.

27 Пространство имен в Usenet, как и пространство имен в DNS, было использовано с серьезными нарушениями.

Таким образом, несмотря на низкую стоимость для стран с ограниченными ресурсами, Usenet плохо приспособлена для международного использования и обладает множеством технических сдвигов в направлении английского языка. Некоторые из этих сдвигов приводят к дополнительным возникающим сдвигам.

Второй коммуникационный режим, ставший популярным в начале 1990-х годов, – ретранслируемый Интернет-чат (Internet Relay Chat, IRC), многосторонний синхронный режим коммуникации в реальном времени. Участники этого канала коммуникации общаются между собой в реальном времени почти так же, как если бы они участвовали в телефонной конференции, с той лишь разницей, что разговоры записываются. Соединенные в общую сеть, серверы IRC могут объединять тысячи каналов, и эта сеть, как и EFNet или UnderNet, открывает культурные, региональные или национальные темы и привлекает участников со всего мира (Paolillo, 2001). IRC был создан в Северной Европе, так что некоторые его характеристики, такие как допустимые символы в текстовых сообщениях и имена участников, отличаются от принятых в Usenet. Однако поддержка многоязычных текстов в IRC несколько не лучше, чем в Usenet. Фактически, отображение текстов на американском варианте английского языка на компьютерах в странах Северной Европы обнаруживает очевидные проблемы, в числе которых замена знаками препинания символов диакритических гласных в скандинавских именах и словах.

Итак, несмотря на притязания этих двух систем на международное использование, они обладают недостатками, которые берут свое начало в языковых сдвигах, составляющих часть дизайна этих систем. Понятно, что новые коммуникационные режимы, такие как Instant Messaging, Web-logging, Web-chat и другие, появляются постоянно. Несмотря на то, что часть из них изначально инкорпорирует XML и Юникод, состояние этих стандартов таково, что только небольшая часть населения мира и мировых языков может пользоваться услугами этих технологий. Многие сторонники технического развития возлагают надежды на новые коммуникационные протоколы, такие как voice-over-internet (голос через Интернет) или multi-modal interfaces (мультимодальные интерфейсы). Но даже если они смогут решить какие-то языковые вопросы, останутся другие проблемы, и в их числе поддержка людей с дефектами слуха и зрения. Более того, существующие технические сдвиги усиливают сдвиги, появляющиеся по демографическим, экономическим

и другим причинам. Чтобы минимизировать эти явления в Интернете, новые коммуникационные режимы должны быть внимательно изучены на предмет наличия латентных технических сдвигов до того, как они будут предложены для массового использования.

Многие «технофилы» возлагали надежды на машинный перевод как на решение проблем многоязычной коммуникации в Интернете. Уже сейчас существует большой спрос на переводческие услуги, предлагаемые компаниями типа Systran – поставщиком системы перевода BabelFish, – и в определенных ситуациях, например, в случае перевода с каталонского языка на испанский и обратно, машинный перевод решил серьезные коммуникационные проблемы (Climent et al., 2004). Так смогут ли люди получать доступ в Интернет на своем родном языке, просто используя одну из онлайн-овых систем перевода? Этот вопрос носит слишком оптимистический характер – и причин тому несколько.

Во-первых, наличие системы машинного перевода предполагает, что «мирские» проблемы представления и понимания текста на определенном языке уже решены, в то время как для большинства языков мира это не так. Во-вторых, создание систем машинного перевода слишком трудоемко. При переводе любой пары языков могут возникнуть специфические проблемы, которые решаются только для этой конкретной пары языков. Так что перевод с любого языка на любой другой язык мира – это задача, которая вряд ли будет решена в ближайшем будущем. В-третьих, создание систем машинного перевода требует огромного объема материалов на языках, с которых и на которые будет осуществляться перевод²⁸; на сегодняшний день эти материалы собирают с веб-сайтов на тех языках, на которые будет реализован перевод (Grefenstette, 1999; Resnik, 1999), и поэтому они должны быть составлены носителями этих языков. Такая работа не будет сделана, пока не будет подготовлена адекватная техническая поддержка для каждого конкретного языка. Наконец, машинный перевод никогда не будет так же хорош, как перевод, выполненный человеком (Kau et al., 1993). Пользователи систем машинного перевода должны приспосабливаться к странным обрывкам слов и их порядку, которые просто представляют

28 Для этого нельзя использовать любой текст. Как правило, требуются двуязычные тексты, сопоставленные по каждому предложению. Их создание стоит довольно дорого, и для всех пар языков таких текстов нет.

собой скрытую форму языкового сдвига, собственно, и приведшего к необходимости перевода. Следовательно, не приходится ожидать, что такие технологические подходы, как машинный перевод, сумеют существенным образом сократить проблемы языковых сдвигов в Интернете.

Выводы

Изучение потенциальных источников сдвигов, рассмотренных в данной статье, вскрыло большое количество источников языковых сдвигов в Интернете, как исторических и технологических, так и возникающих. Следовательно, можно ответить утвердительно на вопрос, существуют ли языковые предпочтения в Интернете. Основные последствия сдвигов заключаются в благоприятствовании большим языкам с хорошо развитыми техническими стандартами, самым значительным из которых является английский язык, и это не случайно, т. к. именно английский был языком основателей Интернета и предшествующих проектов. Однако также очевидно, что причины и следствия сдвигов разнообразны, неявны и, во многих случаях, непредсказуемы. Если ЮНЕСКО серьезно надеется решить проблему языковых сдвигов в Интернете, нам надо больше работать над собственным образованием и образованием отвечающих за развитие Интернета основных действующих лиц в двух сферах: проявления языковых сдвигов, и значение и ценность языкового разнообразия.

Глоссарий

Возникающий сдвиг (Emergent bias) – по Фридману и Ниссенбауму (Friedman and Nissenbaum, 1997) – это предпочтение, которое является следствием взаимодействия технических систем в разных социальных контекстах.

Всемирная сеть (Сеть, World-Wide Web, Web) – приложение для обмена форматированными документами, программами и мультимедийным контентом по Интернету. Это также весь набор документов и контента, доступных в Сети. Сеть – самое популярное приложение Интернета, благодаря простоте работы веб-браузеров при обработке запросов на документы и другого контента.

Интернет (Internet) – глобальная компьютерная сеть, возникшая в результате слияния ARPANET и других региональных компьютерных сетей.

Исторический сдвиг (Pre-existing bias) – по Фридману и Ниссенбауму (Friedman and Nissenbaum, 1997) – любые предпочтения, вызванные социальными причинами и возникшие до появления конкретного технологического приложения, в котором они проявились.

Консорциум Юникод (Unicode Consortium) – организация, осуществляющая контроль над развитием Юникода.

Протокол (Protocol) – стандартизированный набор сообщений и правил для обмена сообщениями между компьютерами в Сети. Протоколы – это сложные системы, часто обсуждаемые в терминах т. н. «уровней»: прикладной уровень, канальный уровень и т. п.

Регистратор доменных имен (Domain-name registry) – организация, которая по контракту с ICANN осуществляет руководство некой частью пространства имен DNS. Как правило, регистратор отвечает за ведение одного или более TLD. Примерами регистраторов доменных имен являются Verisign и Educause.

Технический сдвиг (Technical bias) – по Фридману и Ниссенбауму (Friedman and Nissenbaum, 1997) – любые предпочтения, являющиеся составной частью технической системы. Предпочтение в ASCII американского варианта английского языка является типичным примером технического сдвига.

Техно-лингвистическая жизнеспособность (Techno-linguistic vitality) – возможность этно-лингвистического сообщества воспользоваться преимуществами технологий, в особенности информационных технологий, и возможность использовать родной язык при работе с этими технологиями. По аналогии с этно-лингвистической жизнеспособностью.

Хост (Host, Internet host) – любой компьютер, подключенный к Интернету.

Юникод (Unicode) – 64-битный стандарт кодирования символов, находящийся на стадии развития и позволяющий представить знаки практически всех письменных языков. Разработкой Юникод занимается Консорциум Юникод совместно с ISO и W3C, что должно обеспечить совместимость стандартов всех трех организаций.

Язык разметки (Markup Language) – система, внедряющая форматирование или другие коды («разметку») в текстовые документы, позволяющие форматировать текст или интерпретировать его с помощью устройства, понимающего разметку. Примером языка разметки является HTML (язык гипертекстовой разметки); другие языки, такие как SVG (Scalable Vector Graphics – масштабируемая векторная графика), работают по сходной схеме, но выполняют совершенно разные функции. См.: XML.

АСМ (Association for Computing Machinery) – Ассоциация по вычислительной технике – крупнейшая международная профессиональная организация в области вычислительной техники. Ассоциация образует множество специальных групп, которые занимаются техническими, социальными и политическими вопросами, имеющими отношение к компьютерам и компьютерным сетям.

APNIC (Asia-Pacific Network Information Center) – Сетевой информационный центр, осуществляющий контроль за работой Интернета в Азии и Тихоокеанском регионе. Под его контролем находятся Австралия, Китай, Япония, Корея, Индонезия, Малайзия и все страны островов Тихого океана.

ARIN (American Registry for Internet Numbers) – американское бюро регистрации адресов в Интернете. Это сетевой информационный центр, который осуществляет контроль за работой Интернета в Северной Америке.

ASCII (American Standard Code for Information Interchange) – один из первых 7-битных стандартов кодирования компьютерных текстов, который получил повсеместную поддержку в большинстве компьютерных приложений. Большинство современных текстовых кодов, включая Юникод, совместимы с ASCII, сеть битов которого позволяла кодировать 128 разных символов. Расширенный ASCII – это 8-битный вариант прежнего ASCII, который не имеет стандарта; разные поставщики поддерживают разные версии этого расширенного ASCII, которые друг с другом не совместимы.

BMP (Basic Multilingual Plane) – Базовая многоязыковая плоскость, область значений кодов Юникод, содержащая символы практически для всех современных письменностей и большое число специальных символов.

ccTLD (Country-Code Top-Level Domain) – домены верхнего уровня, которые ассоциируются с определенными странами. ccTLD идентичны ко-

дам стран ISO-3166. В качестве примеров можно привести .uk (Великобритания) и .za (ЮАР).

СМС (Computer-Mediated Communication) – компьютерно-опосредованная коммуникация – любая форма коммуникации «человек-человек», которая происходит при помощи сетевых компьютеров как средства коммуникации.

CNNIC (China Network Information Center) – китайский сетевой информационный центр, осуществляющий контроль за работой Интернета в Китае.

DNS (Domain-Name System) – система доменных имен – техническая система, которая находится под управлением ICANN, что позволяет присваивать мнемонические коды компьютерам, объединенным в Сеть.

Ethnologue – база данных, ведение которой осуществляет SIL International (редактор – Barbara Grimes). База содержит записи, представляющие общую дескриптивную информацию всех известных языковых популяций мира.

GPL (Gnu Public License) – юридическое лицензирующее соглашение, которое используется в ряде программ с открытым кодом и предназначено для защиты авторского права, предоставляя при этом для разработчиков открытый доступ к исходному коду программ.

gTLD (Generic Top-Level Domain) – родовой домен верхнего уровня. Домены верхнего уровня присваиваются «родовым» категориям, без обязательной ссылки на страну. Общие gTLD – это .com (коммерческие), .edu (высшие учебные заведения США), .mil (организации военного ведомства США), .net (сетевые провайдеры), .org (неприбыльные и некоммерческие организации) и т. д.

HTML (Hypertext Markup Language) – гипертекстовый язык разметки, который используется для форматирования страниц во Всемирной сети. Это – простой язык разметки, понимаемый широким спектром веб-браузеров и другим программным обеспечением, со стандартом, который теперь поддерживается W3C.

IANA (Internet Assigned Numbers Authority) – уполномоченная организация по распределению нумерации в Интернете.

ICANN (Internet Corporation for Assigned Names and Numbers) – Корпорация Интернета для специализированных адресов и номеров, контролирующая выдачу доменных имен верхнего уровня и отвечающая за адресное пространство Интернета.

ICT (Information and Communications Technology) – информационно-коммуникационные технологии.

IP (Internet protocol) – Интернет-протокол. См.: TCP/IP.

IPv4 – IP версия 4. Это версия IP, имеющая сегодня самое широкое применение. Она характеризуется 32-битными номерами адресов для каждого хост-компьютера Интернета. Адресное пространство под IPv4 ограничено и, как следствие, Интернет сегодня переживает переход от IPv4 к IPv6.

IPv6 – IP версия 6. Это – версия «следующего поколения» Интернет-протокола с использованием 128-битных адресов. Поддержка IPv6 расширяется в целом ряде сетевых приложений, но широкого применения она пока еще не нашла, поскольку приложения IPv4 не совместимы с хостами IPv6.

IRC (Internet Relay Chat) – протокол приложений для одновременной многосторонней коммуникации в Интернете в режиме реального времени. В программах «чатов», включая и частные программы, есть много заимствований из IRC. Существует множество сетей IRC, которыми пользуются миллионы людей во всем мире, в основном, в досуговых целях.

ISO-8859-1, Latin-1 – 8-битный стандартный текстовый код, поддерживающий большинство европейских языков, использующих латинский алфавит.

NIC (Network Information Center) – сетевой информационный центр, техническая организация, контролирующая работу Интернета на региональном или местном уровне. Существует 3 основных региональных центра: ARIN, RIPE, APNIC – для Северной Америки, Европы и Азиатско-Тихоокеанского региона соответственно.

RIPE (Réseaux IP Européens) – Координационный Центр распределения ресурсов сети Интернет в Европейском регионе, сетевой информационный центр, контролирующий работу Интернета в Европе.

SGML (Standard Generalized Markup Language) – стандартный общий язык разметки, широко использующийся в печатной и издательской сфере. Язык HTML был изначально разработан как приложение SGML.

TCP/IP (Transmission Control Protocol/Internet Protocol) – набор сетевых протоколов, на которых базируется Интернет. TCP и IP – это практически независимые «уровни» сетевых Интернет-протоколов, поскольку отвечают за разные аспекты функционирования сети, но часто используются вместе.

TLD (Top-Level Domain) – домен верхнего уровня. Доменное имя, присваиваемое ICANN регистратору доменных имен, который объединяет взаимосвязанные хосты, как правило, в рамках одной страны или в организационных целях.

Usenet (news) – компьютерная сеть, используемая для обмена открытыми опубликованными и широко распространяемыми сообщениями («новостями») между сетевыми компьютерами. Это также вся совокупность новостей, которыми обмениваются подобным образом. Usenet – важный элемент Интернета, т. к. этот недорогой и легко устанавливаемый протокол обеспечивает работу почты и не требует выделенного канала. Как следствие, именно Usenet является первым приложением Интернета при подключении нового пользователя.

UTF-8, UTF-16, UTF-32 – кодировки Юникода, использующие 8-, 16- и 32-битное кодирование текста соответственно. UTF-8 и UTF-16 – кодировки переменной длины, означающие, что некоторые символы требуют больше, чем 8 или 16 битов. UTF-32 – кодировка фиксированной длины, означающая, что все символы требуют для кодировки 32 бита.

W3C (World-Wide Web Consortium) – Консорциум Всемирной сети, осуществляющий контроль за развитием протоколов, языков разметки и других технических стандартов Всемирной сети.

XML (Extensible Markup Language) – расширяемый язык разметки – определение языка разметки, который является упрощенной версией SGML и создан как более гибкий, чем HTML, формат доставки информации по Сети, т. к. позволяет определить различные типы разметки. На сегодняшний день языки разметки, определенные в XML, включают языки для веб-контента (XHTML), графики (Scalable Vector Graphics, SVG), математических уравнений (MathML), музыки (MML, MusicML) и многих других приложений.

Список литературы

- Adam, A. 1998. *Artificial Knowing: Gender & the Thinking Machine*. London: Routledge.
- Androutsopoulos, J. 1998. Orthographic variation in Greek e-mails: a first approach. *Glossa* 46, pp. 49–67.
- Anis, J. 1997. A Linguistic Approach to Programming. Arob@se, 1.2. <http://www.liane.net/arobase>
- Barrera-Bassols, N. and Zinck, J.A. 2002. Ethnopedological research: a worldwide review. In 17th World congress of soil science CD-ROM proceedings: Confronting new realities in the 21st century. 590.1-590.12. Bangkok: Kasetsart University. (http://www.itc.nl/library/Papers/arti_conf_pr/barrera.pdf)
- Block, D. 2004. Globalization, transnational communication and the Internet. *International Journal on Multicultural Societies*, Vol. 6, No.1, pp.13–28.
- Climent, S., J. Moré, A. Oliver, M Salvatierra, I Sanchez, M. Taulé and L. Vallmanya. 2004. Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. *Journal of Computer-Mediated Communication* Vol. 9, No. 1. <http://www.ascusc.org/jcmc/>
- Crystal, D. 2000. *Language Death*. Cambridge: Cambridge University Press.
- . 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- . 2003. *English as a Global Language, Second Edition*. Cambridge: Cambridge University Press.
- Dalby, A. 2003. *Language in Danger*. New York: Columbia University Press.
- Durham, M. 2004. Language Choice on a Swiss Mailing List. *Journal of Computer-Mediated Communication* 9.1. <http://www.ascusc.org/jcmc/>
- Dunker, E. 2002. Cross-cultural usability of the library metaphor. Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries. Portland, OR.
- Fellbaum, C., and G. Miller. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Ferguson, C. A. 1959. Diglossia. *Word*, 15, pp. 325–340.
- Friedman, B. and H. Nissenbaum. 1995. Minimizing bias in computer systems. Conference companion on Human factors in computing systems, 444. ACM Press.
- Friedman, B. and H. Nissenbaum. 1997. Bias in computer systems. In Friedman, B., ed. *Human Values and the Design of Computer Technology*, pp. 21–40. Stanford, California. Cambridge ; New York, CSLI Publications; Cambridge University Press.

- . 1997. Self-presentation and interactional alliances in e-mail discourse: the style- and codeswitches of Greek messages, *International Journal of Applied Linguistics* 7: pp. 141–164.
- Georgakopolou, A. (Forthcoming). On for drinkies? E-mail cues of participant alignments. In S. Herring (ed.), *Computer-Mediated Conversation*.
- Global Reach. 1999-2005. Global internet statistics by language. Online marketing information. <http://global-reach.biz/globstats/index.php3>
- Greenberg, J. 1956. The measurement of linguistic diversity. *Language*, Vol. 32, No. 2, pp. 109–115.
- Grefenstette, Gregory. 1999. The WWW as a resource for example-based MT tasks. Paper presented at ASLIB “Translating and the Computer” conference, London.
- Grimes, J. E. 1986. Area norms of language size. In B.F. Elson, ed., *Language in global perspective: Papers in honor of the 50th anniversary of the Summer Institute of Linguistics, 1935–1985*, pp. 5–19. Dallas: Summer Institute of Linguistics.
- Hafner, K., and Lyon, M. 1996. *Where Wizards Stay Up Late: The Origins of the Internet*. New York: Simon and Schuster.
- Hård af Segerstad, Y. 2002. Effects of Mobile Text Messaging on Swedish Written Language — human adaptability made visible. *International Conference on Cultural Attitudes towards Technology and Communication, The Net(s) of Power: Language, Culture and Technology*, Montréal.
- Holmes, H. K. 2004. An analysis of the language repertoires of students in higher education and their language choices on the Internet (Ukraine, Poland, Macedonia, Italy, France, Tanzania, Oman and Indonesia). *International Journal on Multicultural Societies*, Vol. 6, No. 1, pp. 29–52.
- Ifrah, G. 1999. *The Universal History of Numbers: From Prehistory to the Invention of the Computer*. New York: John Wiley and Sons.
- Information Sciences Institute. 2003. USC Researchers Build Machine Translation System — and More — For Hindi in Less Than a Month. <http://www.usc.edu/isinews/stories/98.html>
- Kay, Martin, Jean-Mark Gawron, and Peter Norvig. 1993. *Verbmobil: A Translation System for Face-to-Face Dialog*. Stanford, CA: CSLI Publications.
- Krauss, Michael. 1992. The world’s languages in crisis. *Language* Vol. 68, No. 1, pp. 4–10.
- Koutsogiannis, D., and B. Mitsikopolou. 2004. Greeklish and Greekness: Trends and Discourses of “Glocalness”. *Journal of Computer-Mediated Communication* 9.1. <http://www.ascusc.org/jcmc/>

- Lavoie, B. F. and E. T. O'Neill. 1999. How "World Wide" is the Web? Annual Review of OCLC Research 1999. 2003.
- Lévénéz, Eric. 2003. Computer languages timeline. <http://www.levenez.com/lang/>
- Lieberson, S. 1964. An extension of Greenberg's linguistic diversity measures. *Language*, 40, pp. 526-531.
- Mafu, S. 2004. From oral tradition to the information era: The case of Tanzania. *International Journal on Multicultural Societies*, Vol.6, No.1, pp. 53-78.
- Muhlhäusler, P. 1996. *Linguistic Ecology: Language Change & Linguistic Imperialism in the Pacific Rim*. London: Routledge.
- Nettle, D. 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Nettle, D., and S. Romaine. 2000. *Vanishing Voices: The Extinction of the World's Languages*. Oxford: Oxford University Press.
- Nunberg, Geoffrey. 1998. Languages in the Wired World. Paper presented at La politique de la langue et la formation des nations modernes, Centre d'Etudes et Recherches Internationales de Paris.
- O'Neill, Edward T, Brian F. Lavoie, and Rick Bennett. 2003. Trends in the Evolution of the Public Web: 1998 - 2002. *D-Lib Magazine*, 9.4. <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>
- O'Neil, E.T.; P.D. McClain; and B.F. Lavoie 1997. A methodology for sampling the World-Wide Web. Technical report, OCLC Annual Review of Research. <http://www.oclc.org/oclc/research/publications/review97/oneill/o'neilla%r980213.html>
- Paolillo, J. C. 1996. Language Choice on soc.culture.Punjab. *Electronic Journal of Communication/Revue Electronique de Communication*, 6(3). <http://www.cios.org/>
- Paolillo, J. C. 2001. Language Variation in the Virtual Speech Community: A Social Network Approach. *Journal of Sociolinguistics*, 5.2.
- Paolillo, J. C. 2002. Finite-state transliteration of South Asian text encodings. In *Recent Advances in Natural Language Processing: Proceedings of the ICON International Conference on Natural Language Processing*. New Delhi: Vikas Publishing House, Ltd.
- Paolillo, J. C. To appear, 2006. 'Conversational' code switching on Usenet and Internet Relay Chat. To appear in S. Herring, ed., *Computer-Mediated Conversation*. Cresskill, NJ: Hampton Press.

- Peel, R. 2004. The Internet and language use: A case study in the United Arab Emirates. *International Journal on Multicultural Societies*, Vol. 6, No. 1, pp. 79–91.
- Phillipson, R. 1992. *Linguistic Imperialism*. Oxford: Oxford University Press.
- Phillipson, R. 2003. *English-Only Europe?* London: Routledge.
- Pimienta, D.; and B. Lamey. 2001. Lengua española y cultura hispanicas en la Internet: Comparació con el inglés y el francés. II Congreso Internacional de la Lengua Española, Valladolid, 16–19 October 2001.
- Pimienta, D.; et al. 2001. L5: The fifth study of languages on the Internet. <http://funredes.org/LC/english/L5/L5tendencies.html>
- Reed, S. L., and D. B. Lenat. 2002. Mapping Ontologies onto Cyc. American Association for Artificial Intelligence. <http://www.aaai.org/>
- Resnik, P. 1999. Mining the Web for Bilingual Text. 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), College Park, Maryland.
- Rheingold, H. 2000. *The Virtual Community: Homesteading on the Electronic Frontier*, revised edition. Cambridge, MA: MIT Press.
- Skutnabb-Kangas, T., and R. Phillipson. 1995. *Linguistic Human Rights: Overcoming Linguistic Discrimination*. Berlin: Mouton de Gruyter.
- Smith, E. A. 2001. On the co-evolution of linguistic, cultural and biological diversity. In L. Maffi, ed. *On Biocultural Diversity*, 95–117. Washington DC: Smithsonian Institution Press.
- Smith, M. 1999. Invisible Crowds in Cyberspace: Measuring and Mapping the Social Structure of USENET. In M. Smith and P. Kollock, eds., *Communities in Cyberspace*. London: Routledge Press.
- Spencer, H. and Lawrence, D. 1998. *Managing Usenet*. Sebastopol, CA: O'Reilly.
- Su, H.-Y. 2004. The Multilingual and Multi-Orthographic Taiwan-Based Internet: Creative Uses of Writing Systems on College-Affiliated BBSs. *Journal of Computer-mediated Communication* 9.1. <http://www.ascusc.org/jcmc/>
- Torres i Vilatarsana, Marta. 2001. Funciones pragmáticas de los emoticonos en los chats. *Interlingüística* 11.
- Torres i Vilatarsana, Marta. 1999. Els xats: entre l'oralitat i l'escriptura. Article publicat a la revista *Els Marges*, 65 (desembre, 1999). Publicat a Internet (gener, 2001) amb el consentiment d'aquesta revista.

- UNESCO. 2003. Cultural and Linguistic Diversity in the Information Society. UNESCO publications for the World Summit on the Information Society. CI.2003/WS/07 <http://unesdoc>
- [UNESCO.org/images/0013/001329/132965e.pdf](http://unesco.org/images/0013/001329/132965e.pdf)
- Unicode Consortium. 1991. The Unicode Standard: Worldwide Character Encoding. Reading, Mass., Addison-Wesley Pub.
- Unicode Consortium. 1996. The Unicode Standard, Version 2.0. Reading, Mass., Addison-Wesley Developers Press.
- Unicode Consortium. 2000. The Unicode Standard, Version 3.0. Reading, Mass., Addison-Wesley.
- Unicode Consortium. 2003. The Unicode Standard, Version 4.0. Reading, Mass., Addison-Wesley.
- Warschauer, M., G. R. El Said and A. Zohry. 2002. Language Choice Online: Globalization and Identity in Egypt. *Journal of Computer-Mediated Communication (JCMC)*, 7.4. <http://www.ascusc.org/jcmc/>
- Wasserman, Herman. 2002. Between the local and the global: South African languages and the Internet. Litnet Seminar Room. <http://www.litnet.co.za/seminarroom/11wasserman.asp>
- Wright, S. 2004. Introduction. *International Journal on Multicultural Societies*, Vol. 6, No. 1, pp. 3–11.
- Wurm, S. A.. 1991. Language death and disappearance: causes and circumstances. In R. H. Robbins and E. M. Uhlenbeck, eds., *Endangered Languages*, 1–18. Oxford: Berg.
- Wurm, S. A., ed. 1996. *Atlas of the World's Languages in Danger of Disappearing*. Paris: UNESCO Publishing/Pacific Linguistics.

Альтернативные точки зрения

Языковое разнообразие в Интернете: ситуация в Азии

Йошики Миками*, Ахамед Заки абу Бакар[•],
Вираш Сонлерт-ламваниш[○], Ом Викас[■],
Заварски Павол*, Мохд Заиди Абдул Розан*,
Гендри Надь Янош[▲], Томоз Такахаша*

(Члены Проекта «Обсерватория языков» (Language Observatory Project, LOP), Японское агентство по науке и технологии)

«Прежде чем закончить это письмо, я хочу довести до сведения Вашего Преосвященства тот факт, что в течение многих лет я жаждал увидеть в данной Провинции какие-нибудь книги, напечатанные на языке этой страны и на ее алфавите, какие видел я в Малабаре к большой чести тамошней христианской общины. Сделать это мне не удалось по двум причинам: первая из них в том, что казалось

* Технологический университет г. Нагаока (Nagaoka University of Technology), Япония;

• Малазийский технологический университет (Universiti Teknologi Malaysia), Малайзия; ○ Тайская лаборатория вычислительной лингвистики (Thai Computational Linguistic Laboratory), Таиланд; ■ Технологический департамент индийских языков, Министерство информационных технологий (Technology Department of Indian Languages, Ministry of Information Technology), Индия; ▲ Университет г. Мишкольц (Miskolc University), Венгрия.

Адрес для контактов с авторами: mi-kami@kjs.nagaokaut.ac.jp

невозможным составить текст из такого количества форм, число которых доходило до шести сотен против наших двадцати четырех в Европе...» Письмо отца-иезуита Фриара в Рим, 1608 (Priolkar, 1958).

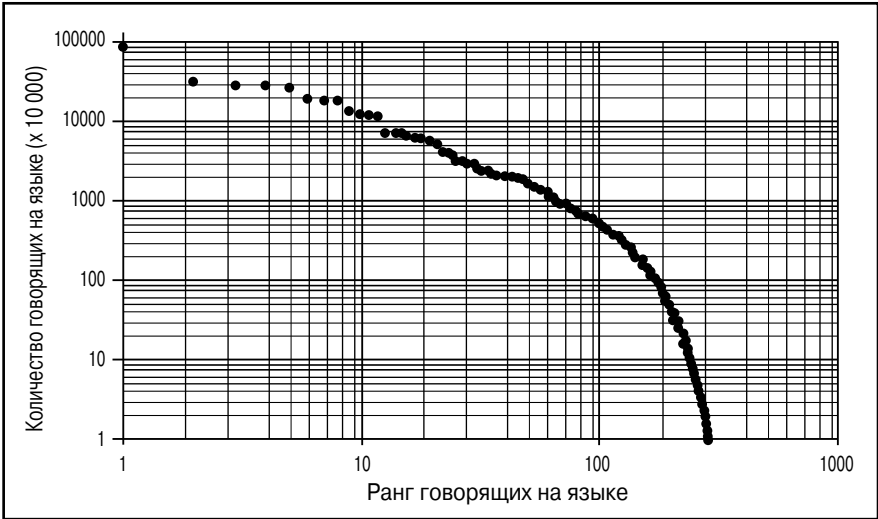
«Когда 500 лет назад в Майнце Гуттенберг напечатал свою знаменитую Библию, ему нужно было только одно основное клише для каждой буквы алфавита, а в 1849 году, когда издательство American Mission Press напечатало Библию в Бейруте на арабском, то использовало не менее 900 знаков, и даже такого большого числа оказалось недостаточно...» Джон М. Мунро (John M. Munro), 1981 (Lunde, 1981).

Разнообразие языков и алфавитов в Азии

По оценкам специалистов на сегодняшний день в мире существует 7000 устных языков (Gordon, 2005). Что касается официальных языков, то их по-прежнему много и, может быть, их количество превышает три сотни. Текст универсального значения – Всеобщая декларация прав человека – был переведен Управлением Верховного комиссара ООН по правам человека (United Nations Higher Commission for Human Rights, UNHCHR) на 328 языков (UNHCHR, 2005).

Из всех языков, представленных на сайте этой организации, самую большую аудиторию имеет китайский язык – почти миллиард человек, за ним идут английский, русский, арабский, испанский, бенгали, хинди, португальский, индонезийский и японский. В конце этого списка стоят языки, на которых говорит менее ста тысяч человек. Азиатские языки занимают шесть из 10 верхних позиций и почти половину (48) из первых по распространенности 100 языков.

На сайте Управлением Верховного комиссара ООН по правам человека (УВКПЧ) представлены также оценочные данные по количеству людей, говорящих на каждом языке. Если мы ранжируем языки по количеству говорящих на них людей и нанесем языки на график, построенный в логарифмическом масштабе, то увидим, что соотношение количества говорящих на языке и их ранга (позиции) среди говорящих на других языках мира практически соответствует кривой закона Ципфа (Рис. 1) (по крайней мере, в интервале от десятков до сотен).

Рис. 1 Кривая квази-закона Ципфа

Языковое разнообразие в Азии становится более явным, если посмотреть на разнообразие алфавитов, используемых для представления языков. С позиции сложности локализации разнообразие алфавитов – это проблема. Трудно ответить на вопрос, сколько алфавитов существует в мире, поскольку ответ зависит от единицы измерения. В данной статье для простоты мы объединяем в одну категорию все латинские алфавиты, алфавиты и их расширения для европейских языков, вьетнамского, филиппинского и др. Мы принимаем за одну категорию кириллические и арабские языки. Точно так же, в рамках одной категории мы рассматриваем китайские иероглифы, японское силлабическое письмо и корейский хангыль. Остальные алфавиты включают индийские письменности, которые составляют пятую категорию. В нее входят не только индийские алфавиты типа деванагари, бенгали, тамильский, гуджарати и другие, но и 4 крупнейших языка Юго-Восточной Азии: тайский, лаосский, камбоджийский (кхмерский) и мьянмский. Независимо от разницы в написании, эти алфавиты имеют общее происхождение (древний язык брахми) и ведут себя одинаково при кодировании. Если сложить число людей, говорящих на каждом языке, в соответствии с данной группировкой по алфавитам, то мы получим картину, представленную в Таблице 1. Тогда алфавиты, используемые в Азии, распространятся на все пять категорий, в то время как алфавиты, ис-

пользуемые в других частях мира, представляют собой, в основном, латинский, кириллический, арабский и некоторые другие.

Таблица 1. Распределение групп пользователей по основным категориям алфавитов

Алфавит	Латинский	Кириллический	Арабский	Ханьцзы	Индийская группа	Другие*
Кол-во пользователей (млн)	2 238	451	462	1 085	807	129
% от общего числа	43,28	8,71	8,93	20,98	15,61	2,49

* Другие: греческий, грузинский, армянский, амхарский, дивехи, иврит и пр.

Современное состояние охвата языков – ситуация с Windows

По сравнению с прошедшим десятилетием современные продукты информационно-телекоммуникационных технологий (ИКТ) способны до известной степени оперировать с многоязычием. Благодаря появлению стандарта кода многоязычных знаков в виде ISO/IEC 10646, который используется и для стандарта Юникод, а также благодаря сложной интернационализации программного обеспечения в течение 10 лет выросло количество языков, поддерживаемых основными настольными платформами ИКТ. Однако охват языков данными платформами все еще ограничен. Самая последняя версия Windows XP (Professional SP2)* способна работать с 123 языками. Однако, если мы внимательнее посмотрим на этот список, то увидим, что большинство представленных в нем языков – европейские и лишь немного азиатских и африканских языков. Охват языков показан в Таблице 2. Здесь языки разбиты по группам алфавитов, как это было описано в первой части статьи. Так, по подсчетам, на тех языках, с которыми работает Windows XP, говорят 83,72 % от общего числа населения Земли. Хотя данная цифра

* По состоянию на год издания оригинальной версии данного сборника.

может быть истолкована как довольно неплохая, нам она кажется завышенной, и плохо соотносится с реальностью, что мы покажем ниже.

Таблица 2. Охват языков Windows XP SP 2 с разбивкой по основным категориям алфавитов

Регион алфавита	Латинский	Кириллица	Арабский	Ханьцзы	Индийская группа	Прочие
Европа	Европейские* и славянские** языки	Русский, македонский и славянские*** языки	–	–	–	Греция Грузия Армения
Азия	Азербайджанский, вьетнамский, малайский, индонезийский, узбекский, турецкий	Монгольский, азербайджанский, казахский, киргизский, узбекский	Арабский, урду, персидский	Китайский, японский, корейский	Гуджарати, тамильский, телугу, каннада, бенгальский, малайялам, пенджаби, хинди, маратхи, санскрит, конкани, ория, тайский	Ассирийский, дживехи, иврит

* Включают: албанский, баскский, каталанский, датский, голландский, английский, эстонский, фарерский, финский, французский, галисийский, немецкий, венгерский, исландский, итальянский, латвийский, литовский, мальтийский, норвежский, португальский, румынский, саами, испанский, шведский и валлийский языки.

** Включают: сербский, чешский, хорватский, словацкий, боснийский, польский и словенский языки.

*** Включают: белорусский, болгарский, сербский, боснийский и украинский языки.

Ситуация с Google

Поисковые машины стали неотъемлемой составляющей глобального информационного общества. Их работа делает доступным огромный массив знаний. Когда мы изучаем охват языков наиболее распространенными поисковыми машинами, мы видим, что ситуация здесь гораздо хуже, чем в случае с Windows. Одна из самых распространенных многоязычных поисковых машин – Google – проиндексировала по состоянию на апрель 2005 г. свыше 8 млрд страниц на разных языках мира. Однако оказалось, что эти страницы представляют всего лишь 35 языков. Среди них только 7 азиатских языков: индонезийский, арабский, китайский традиционный, китайский упрощенный, японский, корейский и иврит (Таблица 3). Если подсчитать численность охваченного населения, то она сократится до 61,37 % именно потому, что материалы, написанные на азиатских и африканских языках, недоступны для поиска.

Таблица 3. Охват языков Google с разбивкой по основным категориям алфавитов

Регион алфавита	Латинский	Кириллица	Арабский	Ханьцзы	Индийская группа	Прочие
Европа	Европейские* и славянские** языки	Русский, болгарский, сербский	–	–	–	Греция
Азия	Индонезийский	–	Арабский	Китайский традиционный и упрощенный, японский, корейский	–	Иврит, турецкий

*Включает: каталанский, датский, голландский, английский, эстонский, финский, французский, немецкий, венгерский, исландский, итальянский, латвийский, литовский, норвежский, португальский, румынский, испанский и шведский языки.

**Включает: хорватский, чешский, польский, словацкий и словенский языки.

Многоязычный характер Всеобщей декларации прав человека

Приведем еще один пример. Как мы упомянули в начале нашей статьи, на веб-сайте Управления Верховного комиссара ООН по правам человека Всеобщая декларация прав человека представлена на более чем 300 языках мира, начиная с абхазского и заканчивая зулу. К сожалению, там также можно найти и много переводов на разные языки, особенно, на языки с нелатинским алфавитом, и эти переводы даны в виде GIF- или PDF-файлов, а не в виде кодированных текстов. И снова, как и в предыдущих случаях, представим эту ситуацию в виде таблицы (Таблица 4). Из нее понятно, что в виде кодированных текстов лучше всего представлены языки, использующие латинский алфавит. Языки, использующие другие алфавиты, индийские в особенности, с трудом поддаются кодировке. Если алфавит не удастся представить в какой-то одной из имеющихся трех форм, он попадает в категорию «недоступных». Более того, не так просто загрузить специальные шрифты для надлежащего просмотра этих алфавитов. Сложность ситуации можно назвать цифровым разрывом между языками или «языковым цифровым разрывом».

Таблица 4. Представление Всеобщей декларации прав человека с разбивкой по основным категориям алфавитов

Регион алфавита	Латинский	Кириллица	Арабский	Ханьцзы	Индийская группа	Прочие
Европа	Европейские и славянские языки	Русский, болгарский, сербский	–	–	–	Греция
Азия	Индонезийский	–	Арабский	Китайский традиционный и упрощенный, Японский, корейский	–	Иврит, турецкий

Регион алфавита	Латинский	Кириллица	Арабский	Ханьцзы	Индийская группа	Прочие
В каком алфавите представлен	Латинский	Кириллица	Арабский	Ханьцзы	Индийский	Другие
Кодировка	253	10	1	3	0	1
PDF	2	4	2	0	7	10
GIF	1	3	7	0	12	7
Недоступны	0	0	0	0	1*	1*

* Недоступными языками являются магади и бходжпури.

Локализация информационных технологий – взгляд в прошлое

Давайте посмотрим, что было 500 лет назад, когда была изобретена эпохальная технология книгопечатания. Буквопечатающая технология была независимо изобретена и на Востоке, и на Западе. На Востоке эта технология была впервые создана в XIII веке корейскими ремесленниками и затем подхвачена китайцами. Но технология эта не получила развития и впоследствии была вытеснена ксилографией. Буквопечатающая технология, распространенная сегодня повсеместно в Азии, уходит своими корнями в изобретение, сделанное Гуттенбергом в середине XV века.

Первый печатный пресс был привезен на Гоа в 1556 г. Считается, что это – первая печатная машина, привезенная в Азию. Вслед за ней другие машины были привезены в Манилу, Малакку, Макао и другие города Азии. Поначалу эти машины использовались для печати переводных или транслитерированных священных текстов с применением латинского алфавита, но позднее на них стали печатать разные тексты с отпечатками букв местных алфавитов. По мнению одного индийского историка первым печатным текстом в Азии с использованием местного алфавита стала книга на тамильском языке «Христианская доктрина».

На второй странице этого текста содержится рассказ о том, какой подход был использован при локализации печатной технологии на тамильском языке. Несмотря на то, что в тамильском языке всего 246 слогов, на второй странице представлено более 150 знаков в комплекте шрифта. Отец-иезуит, проживавший в XVII веке где-то на берегу Малабара, писал в Рим: «...в течение многих лет я жаждал увидеть в этой Провинции какие-нибудь книги, напечатанные на языке этой страны и на ее алфавите, ...но сделать это мне пока не удалось. Главная причина в том, что мы должны составлять текст из более чем 600 отпечатков против 24 в Риме...» (Priolkar, 1958).

В Маниле, в то время центре испанских колоний, «Доктрина» была переведена на тагалогский язык в 1593 г. Однако так случилось, что перевод сопроводили транслитерированным текстом. Тагалогская версия «Доктрины» была составлена в трех вариантах: на тагалогском языке с использованием тагалогского алфавита; на тагалогском языке с использованием латинского алфавита и на испанском языке с использованием латинского алфавита. За последующие 100 лет после того, как буквопечатающая технология была привезена в Манилу, два вторые варианта полностью вытеснили первый. В итоге тагалогский алфавит был полностью забыт даже местным населением (Hernandez, 1996). Изображение тагалогского шрифта на почтовой марке, выпущенной почтой Филиппин в 1995 г., показывает нам, как выглядел этот шрифт, и служит напоминанием об утерянном культурном наследии.

Эти два исторических события дают нам урок: когда локализация реализуется неудачно, появление новой технологии может разрушить систему письменности или даже саму культуру.

Стандарты кодировки как краеугольный камень локализации

За цифровым языковым разрывом стоит множество причин: экономических, политических, социальных и пр. Однако с технических позиций локализация должна оставаться главным фактором. Как явствует из письма отца-иезуита, отправленного 400 лет назад в Рим, отрывок из которого мы привели в самом начале нашей статьи, даже во времена технологии книгопечатания пионеры информационных технологий были вынуждены преодолевать аналогичные сложности, локализуя технологии в другую языковую среду, почти так же, как это делают сего-

дня инженеры-компьютерщики. Особым препятствием для нелатинских алфавитов является отсутствие или недоступность соответствующих стандартов кодировки. По этой причине разработчикам веб-сайта с текстом Всемирной декларации прав человека пришлось поместить файлы, не поддающиеся кодированию, в виде изображений или в формате PDF. Если мы посмотрим на международно-признанные справочники схем кодирования, например, IANA Registry of Character Codes (IANA, 2005) или ISO International Registry of Escape Sequences (IPSJ/ITSCJ, 2004), то не сможем найти в них схемы кодирования для таких языков, которые считаются «упущенными сквозь ячейки сети». Следует отметить, что многие стандарты кодировки, принятые на национальном уровне, используются для нескольких языков и имеют национальный статус. Для семи индийских письменностей первый национальный индийский стандарт был принят в 1983 г. Он получил название Indian Standard Script Code for the Information Interchange (ISSCII). Позже, в 1991 г., он претерпел изменения и вышел во втором издании под названием «National Standard IS 13194», который и используется в Индии в настоящий момент. Однако, несмотря на существование национальных стандартов, поставщики технических средств, разработчики шрифтов и даже конечные пользователи сами создавали собственные таблицы кодирования, что приводило к неразберихе. Стимулом для создания так называемых экзотических схем кодирования или локальных внутренних кодировок послужило внедрение дружественных для пользователя средств создания шрифтов. Несмотря на то, что прикладные системы для этой области не являются автономными и широко распространяются через Сеть, необходимость в стандартизации не была осознана пользователями, поставщиками или разработчиками шрифтов. Отсутствие профессиональных ассоциаций и соответствующих государственных учреждений – еще одна причина сложившейся неконтролируемой ситуации. Интересное исследование по всему многообразию индийских языков провела компания Aruna Rohra and Ananda of Saora Inc. (www.gse.uci.edu/markw/languages.html): на 49 тамильских веб-сайтах она обнаружила существование 15 различных схем кодирования (Aruna & Ananda, 2005).

UCS/Unicode

Первая версия Универсального многооктетного набора кодированных символов (Universal Multiple-Octet Coded Character Set, UCS, ISO/IEC

10646) была выпущена в 1993 г. Юникод, разработанный изначально промышленным консорциумом, приведен сегодня в соответствие последней версии UCS и мог бы устранить неразбериху. Но он не стал доминирующим, по крайней мере, в азиатской части мира. Наше последнее исследование показало, что код UTF-8 охватывает только 8,35 % всех веб-страниц под азиатским ccTLD (Mikami, et al., 2005). Первые и последние десять ccTLD показаны в Таблице 5. Несмотря на то, что ожидается высокая скорость миграции, процесс этот следует тщательно отслеживать.

Таблица 5. Доля веб-страниц, использующих UTF-8, по ccTLD

ccTLD	Название	Доля	ccTLD	Название	Доля
tj	Tajikistan Таджикистан	92,75 %	uz	Uzbekistan Узбекистан	0,00 %
vn	Viet Nam Вьетнам	72,58 %	tm	Turkmenistan Туркменистан	0,00 %
np	Nepal Непал	70,33 %	sy	Syria Сирия	0,00 %
ir	Iran Иран	51,30 %	mv	Maldives Мальдивы	0,00 %
tp	Timor East Восточный Тимор	49,40 %	la	Lao Лаос	0,01 %
bd	Bangladesh Бангладеш	46,54 %	ye	Yemen Йемен	0,05 %
kw	Kuwait Кувейт	36,82 %	mm	Myanmar Мьянма	0,07 %
ae	UAE ОАЭ	35,66 %	ps	Palestine Палестина	0,12 %
lk	Sri Lanka Шри Ланка	34,79 %	bn	Brunei Бруней	0,36 %
ph	Philippines Филиппины	20,72 %	kg	Kyrgyzstan Киргизстан	0,37 %

Источник: Language Observatory Project.

Цели проекта «Обсерватория языков»

Проект «Обсерватория языков» был запущен в 2003 г. как признание растущего значения мониторинга уровня языковой активности в киберпространстве (Language Observatory Project, LOP; UNESCO, 2004). Проект призван создать средства для оценки уровня использования каждого языка в киберпространстве. Если говорить точнее, то от проекта ждут периодического предоставления статистики по языкам, алфавитам и кодировкам в киберпространстве. После полного запуска проект должен будет дать ответы на следующие вопросы:

- Сколько разных языков существует в виртуальной вселенной?
- Какие языки отсутствуют в виртуальной вселенной?
- Сколько веб-страниц написано на определенном языке, скажем, на пушту?
- Сколько веб-страниц написано с использованием тамильского варианта письма?
- Какие виды схем кодирования символов используются для кодирования какого-то определенного языка, скажем, берберского?
- Как быстро Юникод замещает традиционные и локальные схемы кодирования в сети?

Наряду с таким анализом проект будет заниматься подготовкой предложений по преодолению сложившейся ситуации, как на техническом, так и на политическом уровнях.

Альянс проектов

В настоящее время несколько групп экспертов работают в тесном сотрудничестве в рамках обсерватории языков. Организациями-учредителями проекта являются: Технологический университет г. Нагаока (Nagaoka University of Technology), Япония; Токийский университет зарубежных исследований (Tokyo University of Foreign Studies), Япония; Университет Кейо (Keio University), Япония; Технологический университет Малайзии (Universiti Teknologi Malaysia), Малайзия; Университет г. Мишкольц (Miskolc University), Венгрия; проект «Технологическое

развитие индийских языков» (Technology Development of Indian Languages) под руководством Министерства информационных технологий Индии; Лаборатория исследований в области коммуникации (Communication Research Laboratory), Таиланд. Финансирование проекта осуществляется Агентством по науке и технологиям Японии (Japan Science and Technology Agency) в рамках программы RISTEX (RISTEX, 2005). ЮНЕСКО выразила официальную поддержку данному проекту с самого начала его создания. Основные технические компоненты «Обсерватории языков» включают мощную технологию поиска в Сети (кролер-технология) и технологию идентификации особенностей языков (Suzuki, et al., 2002). В проекте используется UbiCrawler (Boldi, et al., 2004) – масштабируемый, полностью распределенный веб-кролер, разработанный совместно Отделом информационных наук Исследовательского университета Милана (Dipartimento di Scienze dell'Informazione, Universita degli Studi di Milano) и Институтом информатики и телематики Итальянского национального совета по исследованиям (Istituto di Informatica e Telematica). Это мощная машина по сбору данных для «Обсерватории языков». Краткое описание совместных усилий проекта и команды UbiCrawler можно найти в публикации ЮНЕСКО (UNESCO, 2004).

Заключение

В данной статье мы стремились подчеркнуть значение мониторинга поведения и активности мировых языков в киберпространстве. Проект «Обсерватория языков» предусматривает использование сложных научных методов для понимания и мониторинга мировых языков. Консорциум проекта надеется, что ему удастся сделать так, чтобы мир больше знал о живущих и умирающих языках. В этом случае можно будет предпринять шаги, чтобы предотвратить исчезновение языков, оказавшихся в тяжелой ситуации. Чтобы эти усилия принесли плоды, «Обсерватория» должна стать центром развития человеческого капитала и депозитарием языковых ресурсов. Накопление цифровых языковых ресурсов в результате проведенных научно-исследовательских работ позволит развивающимся странам и региональным сообществам вывести свои языки в киберпространство и, тем самым, спасти национальное наследие от исчезновения.

Список литературы

- Aruna, R. & Ananda, P. 2005. Collecting Language Corpora: Indian Languages. The Second Language Observatory Work Shop Proceedings. Tokyo University of Foreign Studies, Tokyo.
- Boldi, P., Codenotti, B., Santini, M., & Vigna, S. 2004. UbiCrawler: A scalable fully distributed Web crawler. *Software: Practice & Experience*, Vol. 34, No. 8, pp. 711–726.
- Gordon, R. 2005. *Ethnologue: Languages of the World 15th Edition*. (<http://www.ethnologue.com/>).
- Hernandez, Vincente S. 1996. *History of Books and Libraries in the Philippines: Manila*, The National Commission for Culture and the Arts, pp. 24–31.
- IANA. 2005. Character Sets. (<http://www.iana.org/assignments/character-sets>).
- IPSC/ITSCJ. 2004. International Register of Coded Character Sets to be used with Escape Sequences. (<http://www.itscj.ipsj.or.jp/ISO-IR/>).
- Mikami, Y., Zavorsky, P., Zaidi, M., Rozan, A., Suzuki, I., Takahashi, M., Maki, T., Ayob, I. N., Boldi, P., Santini, M. & Vigna, S. 2005. The Language Observatory Project (LOP). *Proceedings of the Fourteenth International World Wide Web Conference*, May 2005. Chiba, Japan., pp. 990–991.
- Lunde, P. 1981. *Arabic and the Art of Printing*, Saudi, Aramco World.
- Priolkar, A. K. 1958. *The Printing Press in India – Its Beginning and Early Development*. Bombay, Marathi Samshodhana Mandala, pp. 13–14.
- RISTEX. 2005. (http://www.ristex.jp/english/top_e.html).
- Suzuki, I., Mikami, Y., Ohsato, A. & Chubachi, Y. 2002. A language and character set determination method based on N-gram statistics, *ACM Transactions on Asian Language Information Processing*, Vol. 1, No. 3, pp. 270–279.
- UNESCO. 2004. *Parcourir le cyberspace à la recherche de la diversité linguistique*. UNESCO WebWorld News, 23rd Feb. 2004. (http://portal.UNESCO.org/ci/en/ev.php-URL_ID=14480&URL_DO=DO_TOPIC&URL_SECTION=201.html).
- UNHCHR. 2005. *Universal Declaration of Human Rights*. (<http://www.unhchr.ch/udhr/navigate/alpha.htm>).

Несколько слов об африканских языках во Всемирной сети

Ксавьер Фантоньян

Резюме

По данным *Cahiers du RIFAL No. 23* «Компьютеризация африканских языков» («*Traitement informatique des langues africaines*»), на африканском континенте говорят на 2000 языков, что составляет 2/3 от общего количества языков в мире. Это богатство наследия и культуры заслуживает серьезного к себе отношения. Сегодня киберпространство предоставляет всем языкам инструмент, с помощью которого они могут стать полноправными участниками широкомасштабной коммуникации. Однако не все языки мира используются в киберпространстве, и не все могут воспользоваться теми возможностями, которые оно предоставляет. Понятно, что для подлинного многоязычия необходим процесс компьютеризации языков, который начинается с их кодирования. Первый вопрос, который следует задать, касается степени использования африканских языков в киберпространстве. В своей работе «Африканские языки в Сети» («*Les langues africaines sur la Toile*») Марсель Дики-Кидири и Эдема Атибаква (Marcel Diki-Kidiri, Edema Atibakwa) рассказали об исследовании 3000 веб-сайтов, из которых были отобраны содержавшие информацию на африканских языках. Их анализ показывает, что в Сети есть множество материалов об африканских языках, но лишь немногие сайты используют какой-либо африканский язык в качестве средства коммуникации. Есть множество факторов, объясняющих эту ситуацию, но двумя основными являются следующие: отсутствие кибер-сообществ, способных передавать информацию на своих родных языках, и отсутствие соответствующих инструментов для компьютеризации и обработки языков.

Однако исследование, проведенное Жилем Морисом де Шрайвером и Аннелин ван де Векен «Африканские языки в Сети: исследование положения дел с языками хауса, сомали, лингала и исикоса» (Gilles Maurice de Schryver, Anneleen Van de Veken, «*Les langues africaines sur la Toile: etude des cas haoussa, somali, lingala et isixhosa*»), приходит к более

сбалансированным выводам, уточняет и даже исправляет результаты вышеупомянутого исследования. Авторы исследовали дискуссионные форумы и увидели, что для них характерен вполне удовлетворительный уровень использования широко распространенных африканских языков – суахили, хауса и лингала.

Перечислим основные результаты исследования RIFAL:

- африканские языки появляются в Сети, скорее не как средства коммуникации, а как объекты исследований (в справочных материалах, документации, описаниях, примерах, текстах, курсах);
- языком коммуникации при обсуждениях африканских языков является английский, даже для языков во франкоговорящих регионах;
- курсы изучения африканских языков крайне редко встречаются в Сети, что открывает возможности для развития кибер-сообществ людей, говорящих на африканских языках, которые будут осуществлять коммуникацию в Интернете на своем родном языке;
- программные продукты или технологические компьютерные решения, стандартизирующие шрифты для всех африканских языков, редко предлагаются в Сети.

Для изменения этой ситуации предлагаются следующие меры:

- увеличение числа двуязычных или многоязычных сайтов, использующих в качестве языков коммуникации французский или английский языки и, по крайней мере, один африканский язык;
- расширение распространения документации по африканским языкам, которая существует, но пока что не распространяется в Сети регулярно;
- открытие в Сети качественных языковых курсов африканских языков;
- разработка и распространение программных и технологических компьютерных решений, поддерживающих письменные формы африканских языков и их регулярное использование в киберпространстве.

Нельзя сказать, что сегодня африканские языки не представлены в Сети. Существует много материалов об африканских языках, но лишь немногие из них написаны на каком-либо африканском языке. Одной из причин этого является отсутствие мотивации для африканцев писать на своем родном языке, и эта причина объясняет относительный неуспех африканских языков в Сети. Интернет-пользователи, осуществляющие коммуникацию в Сети, хотят, чтобы их читали и понимали, и поэтому пишут на том языке, который использует большее количество людей.

Есть и другая причина: множество материалов об Африке, найденных в Сети, были написаны не африканцами. Это – религиозные документы или материалы для обучения. Форумы, на которых африканцы общаются с африканцами на африканских языках, скорее исключение, чем правило.

Microsoft объявил, что в скором времени Windows и Office будут переведены на суахили, который, несомненно, является самым распространенным языком Африки. На нем говорит около 100 миллионов человек на континенте и островах Индийского океана. До начала перевода специалисты-лингвисты Microsoft должны будут создать общий словарь различных диалектов суахили. Microsoft также планирует перевести свои программы на другие африканские языки, такие как хауса и йоруба.

И, хотя намерения Microsoft заслуживают одобрения, приходится с горечью констатировать, что это будет единственная альтернатива для людей, говорящих на суахили и не знающих никакого другого языка. Да, компьютерные программы с открытым кодом, переведенные на суахили, немногочисленны. Будем надеяться, что усилия Microsoft стандартизировать африканские языки, будут выгодны и Linux, и другим программам с открытым кодом.

В настоящее время в Африке ведется большая работа над программами с открытым исходным кодом. В Буркина-Фасо осуществляются попытки локализации Open Office на языки морэ и диула. Аналогичная работа проводится в Мали с языком бамбара, в Бенине с фонгбе, йоруба, мина и денди. Огромная работа, проделанная с амхарским языком и его алфавитом, наглядно демонстрирует результаты более эффективных исследований по компьютеризации африканских языков. Шаги, предпринятые Юникод по стандартизации н'ко, были с энтузиазмом восприняты многими людьми.

Однако остаются вопросы, ожидающие своего решения, и касаются они орфографии и стандартизации африканских языков. Многие языки по-прежнему транскрибируются фонетически, и нельзя игнорировать опасность потери языками своего алфавита.

Таким образом, несмотря на то, что мы знаем о существовании около 2000 африканских языков, изучены всего 400 из них. Остается еще 1600 языков, не удостоившихся серьезного внимания исследователей. Ни один из них не имеет на сегодня веб-аудитории, и даже те, изученные, 400 языков не достигли того уровня технологической адаптации, который позволял бы считать их живыми языками Всемирной сети.

Список литературы

- Diki-Kidiri M., Don D. Dimo-Lexis, Dictionnaires monolingues et Lexiques spécialisés, Outils logiciels pour linguiste, CNRS-LACITO, Paris.
- Meloni H., 1996. Fondements et Perspectives en traitement automatique de la parole. AUPELF/UREF. Morvan P., 2000. Dictionnaire de l'Informatique: Acteurs concepts, réseaux, Larousse, Paris.
- Peek J., Lui C., et al ; 1997. Système d'information sur Internet: Installation et mise en œuvre, Editions O'Reilly International Thomson.
- Rint-Riofi I, C., Chanard, et Diki-Kidiri, M. (undated) Stage de formation niveau 1 et 3, Document de travail: Introduction aux inforoutes par le développement de la terminologie et des contenus textuels pour le français et les langues partenaires, Lumigny, Marseilles.

Об авторах

Ксавьер Фантоньян (Xavier Fantognon) – студент, занимающийся тоголезской лингвистикой в Университете Бенина (xavier@bj.refer.org) и решивший посвятить себя продвижению африканских языков в Интернете. Он перевел на язык фонгбе (Fongbe) интерфейс платформы SPIP (<http://www.spip.net/fon>), а также принимает активное участие в деятельности, связанной с культурой, в ее традиционной и мультимедийной форме.

Йошики Миками (Yoshiki Mikami) – профессор менеджмента и информационных наук в Университете г. Нагаока. Занимал руководящие должности в Министерстве внешней торговли и промышленности. Координатор проекта «Обсерватория языков» (<http://www.language-observatory.org/>; <http://gii.nagaokaut.ac.jp/gii/>; <http://kjs.nagaokaut.ac.jp/mikami/>)

Джон Паолилло (John Paolillo) – доцент информационных наук и информатики; адъюнкт-доцент лингвистики, Школа библиотечных и информационных наук; доктор лингвистики, Стэнфордский университет, 1992; бакалавр гуманитарных наук, лингвистика, Корнеллский университет, 1986. Сфера научных интересов: компьютерная лингвистика, информационный поиск, связь с использованием компьютеров, статистические модели и методы количественных исследований, социолингвистика и обучение языку, обучение второму языку, языки Южной Азии.

Даниэль Пимьента (Daniel Pimienta) – француз марокканского происхождения, проживающий в Санто-Доминго, Президент Ассоциации «Сеть и развитие» (Association Réseaux & Développement, FUNREDES – <http://www.funredes.org>), неправительственной организации, работающей в сфере ИКТ и развития с 1988 г. FUNREDES провел несколько экспериментов в области языков и культуры, некоторые из них – в сотрудничестве с Латинским союзом и/или при поддержке Агентства по франкофонии (Agence de la Francophonie) (<http://funredes.org/tradauto/index.htm/bamaktxt> – <http://funredes.org/lc>).

Даниэль Прадо (Daniel Prado) – уроженец Аргентины, живущий в Париже, Директор Программы терминологии и промышленности языков Латинского союза (Programme de Terminologie et Industries de la Langue de l'Union Latine) (<http://unilat.org/dtil/>), межправительственной организации, продвигающей латинские языки. Отвечает за подготовку статистических данных по состоянию языков в современном обществе и информации по языковой и терминологической политике.

ЯЗЫКОВОЕ РАЗНООБРАЗИЕ В ИНТЕРНЕТЕ

Сборник статей

Джон Паолилло, Даниэль Пимьента, Даниэль Прадо и др.

Редактор *Т.А. Мурована*

Ответственные за выпуск *С. Д. Бакейкин, Е.И. Кузьмин*

Технический редактор *Ю.Ю. Таранова*

Корректор *О.И. Церковская*

Издатель:

Межрегиональный центр библиотечного сотрудничества (МЦБС)

105066, г. Москва, 1-й Басманный пер., д. 2а, стр. 1

Тел.: (495) 267 33 34, факс: (495) 657 96 20

www.mcbs.ru

Художественное оформление:

Издательство «Права человека»

119992, Москва, Зубовский бульвар, 17

веб-сайт: www.hrpublishers.org

ИД № 02184 от 30.06.2000. Подписано в печать 01.12.2007.

Формат 60х90 1/16. Бумага офсетная. Гарнитура FranklinGothicBook.

Печ. л. 7,5. Печать офсетная. Тираж 1000 экз. Заказ № 359.

Отпечатано на Фабрике офсетной печати,

249039, г. Обнинск, ул. Королева, 6

Межрегиональный центр библиотечного сотрудничества представляет:



«Языковое разнообразие в киберпространстве: российский и зарубежный опыт»

В сборник включены аналитические материалы и статьи экспертов и специалистов из различных областей: культуры, языкознания, библиотечного дела, информационно-коммуникационных технологий, – которых объединил интерес к проблеме сохранения языкового разнообразия в киберпространстве.

Сборник подготовлен Российским комитетом Программы ЮНЕСКО «Информация для всех» и Межрегиональным центром библиотечного сотрудничества в рамках нового направления их со-

вместной деятельности – «Развитие многоязычия в Интернете». Это направление представляет особую важность в плане сохранения богатейшего и многообразного письменного и устного культурного наследия народов России.

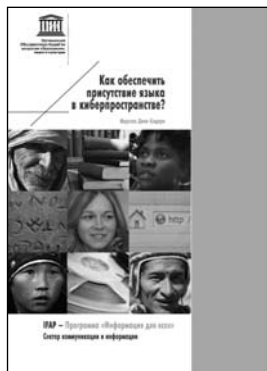
Сборник был подготовлен накануне объявленного ООН Международного года языков и рассказывает о мерах, предпринимаемых в Российской Федерации по развитию многоязычия в киберпространстве, а также о зарубежном опыте представления языков в Интернете. В нем содержится также подборка документов международных организаций (ООН, ЮНЕСКО, МСЭ), а также документы Всемирного саммита по информационному обществу, касающиеся этой тематики.

Издание адресовано руководителям и сотрудникам органов государственной власти, местного самоуправления, специалистам информационной сферы, ученым-филологам, этнографам, социологам, работникам учреждений культуры, преподавателям высших и средних учебных заведений, а также всем, кого волнует судьба культурного наследия народов России в эпоху информационного общества.

Распространяется бесплатно.

Обращаться по адресу:
105066, Москва, 1-й Басманный пер., д. 2а, стр. 1
Тел.: (495) 267-33-34

Электронная версия:
<http://www.ifapcom.ru/ru/365/>



«Как обеспечить присутствие языка в киберпространстве?»

В современном мире наблюдается тенденция исчезновения языков и унификации культур. Как этому противостоять?

Книга «Как обеспечить присутствие языка в киберпространстве?», подготовленная ЮНЕСКО при содействии Латинского союза и Лаборатории «Речь, языки и культуры Африки южнее Сахары», рассказывает о том, какие шаги необходимо предпринять для того, чтобы язык, даже не имеющий письменности, мог занять свое место в киберпространстве и активно в нем использоваться.

Книга переведена на русский язык Российским комитетом Программы ЮНЕСКО «Информация для всех» и Межрегиональным центром библиотечного сотрудничества и представляет интерес для всех, кому не безразличны проблемы сохранения и развития богатейшего языкового и культурного наследия России и всего человечества.

Распространяется бесплатно.

Обращаться по адресу:

105066, Москва, 1-й Басманный пер., д. 2а, стр.1

Тел.: (495) 267-33-34

Электронная версия:

<http://www.ifapcom.ru/ru/365/>