

---

## 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
<b>ДОСТУП К ИКТ</b>	<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>– <b>Наличие инфраструктуры</b></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>– <b>Наличие инфраструктуры</b></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><b>– Приемлемый доступ к инфраструктуре</b></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><b>– Приемлемый доступ к инфраструктуре</b></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 цен<sup>3</sup>, 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 должно<sup>4</sup> соответствовать тому, что МСЭ и регуляторы телекоммуникаций называют «<b>всеобщим доступом</b>» (МСЭ, 2003). Но если и считать всеобщий доступ обязательным условием для преодоления цифрового разрыва, то его совсем недостаточно.</p> <p>– <b>Функциональная грамотность</b></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>– <b>Функциональная грамотность</b></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>– <b>Компьютеризация алфавита</b></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>– <b>Компьютеризация алфавита</b></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><b>Цифровая грамотность</b></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 <b>цифровой грамотности</b>.</p> <p>Изучение киберпространства, которое должно включать не только простую подготовку к работе с определенными компьютерными программами, но также и целостное видение социальных последствий<sup>5</sup> 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>
<b>ТЕХНОЛОГИЧЕСКОЕ ОСВОЕНИЕ ИКТ</b>	<p><i>Уровень, на котором пользователь настолько квалифицирован, что технология становится <b>незаметной в процессе его индивидуальной работы</b></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 <b>опыт</b>.</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>
<b>СМЫСЛОВОЕ СОДЕРЖАНИЕ</b>	<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 <b>создателя контента</b> или <b>виртуальных сообществ</b>.</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>
<b>ОБЩЕСТВЕННОЕ ОСВОЕНИЕ ИКТ</b>	<p><i>Уровень, на котором пользователь настолько квалифицирован, что технология становится <b>незаметной в процессе общественного использования</b></i></p>	
	<p>Этот уровень требует четкого понимания общественного значения использования ИКТ для развития, а также <b>культурных и этических последствий</b> 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
<b>ВООРУЖЕНИЕ ЗНАНИЯМИ</b> <sup>6</sup>	<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>	
<b>СОЦИАЛЬНЫЕ ИННОВАЦИИ</b>	<p>Этот уровень требует не только наличия способностей, но и их <b>практического применения</b>, 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?
<b>РАЗВИТИЕ ЧЕЛОВЕЧЕСКОГО ПОТЕНЦИАЛА</b>	<i>Уровень, на котором человек или сообщество свободны в своем выборе, и этот выбор существует в форме «потенциальных возможностей»<sup>7</sup></i>	
	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. Именно поэтому варианты, предоставляемые свободой, нельзя претворить в жизнь, если <b>участие</b> людей и сообществ не было подлинным в ходе всего процесса.	<b>Свобода выбора</b> в форме « <b>потенциальных возможностей</b> ». 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?

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.<sup>8</sup> 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 колониационного 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»<sup>9</sup>, состоявшейся в Бамако и организованной ЮНЕСКО совместно с Академией африканских языков (ACALAN) и Межправительственным агентством по франкоговорящим странам и регионам (AIF).

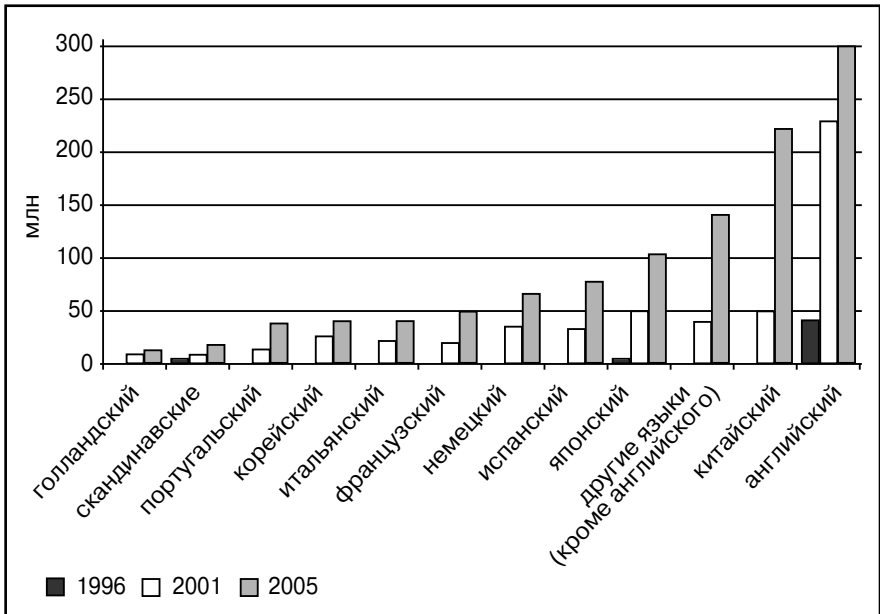
Однако если просто ждать, пока в ходе этого процесса будут разработаны аккуратно задокументированные показатели, обновляемые с учетом скорости развития новых технических средств, будет очень и очень трудно получить четкое представление о ситуации и имеющихся тенденциях.

**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](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

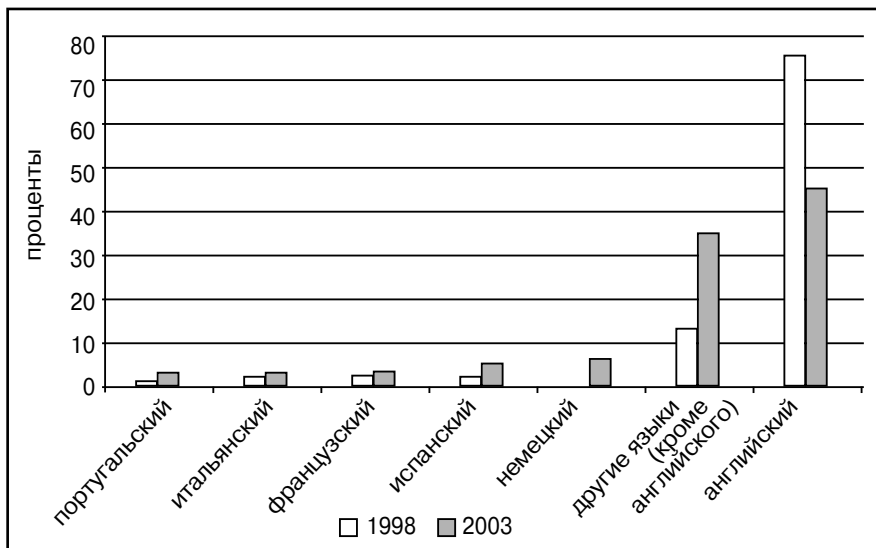
организациями, в частности, «Онлайновым компьютерным библиотечным центром» (OCLC). Центр проводил исследование, ставшее источником цитирования для многих авторов и изданий, которые упорно пишут о существовании более чем 70 % веб-страниц на английском языке (O'Neill, 2003). Суть метода заключается в создании 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 машин<sup>10</sup> и может быть недостаточно надежна, поскольку они выдают большие расхождения при проведении поиска по словам<sup>11</sup>.

Положительной стороной метода является то, что он предоставляет возможность осуществлять последовательный мониторинг в течение длительно времени, изучать сектора киберпространства, а не только Сеть<sup>12</sup> 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](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](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](http://funredes.org/mistica/francais/cyberotheque/thematique/fra_doc_wsis1.html)
- Pimienta, D. & Lamey B. 2001. «Lengua Española y Culturas Hispanicas 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](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](http://en.wikipedia.org/wiki/GNU_General_Public_License)

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