

НЕИЗВЕСТНЫЕ ТЕКСТЫ

После появления речи, ставшей основным способом передачи сообщений, возникла потребность особой разновидностью зрительной сигнализации, фиксирующей сообщения на каком-либо материале для отсутствующего адресата. Передача сообщения о ситуации достигалась путем копирования того, что видит (или воображает) наблюдатель. Адресат, наблюдая копию ситуации, получает о ней приблизительно такое же представление, как индуктор-наблюдатель. Каждая ситуация может быть описана неопределенным количеством фраз.

При копировании чаще всего применялась контурная проекция на плоскость. При этом исключались объекты, не имеющие (по мнению изображающего) отношения к копируемой ситуации или несущественные, терялся ряд признаков и давалась приблизительно (в пределах, допускающих опознание) передача контура.

Так как передать реально наблюдаемые непрерывно меняющиеся ситуации с помощью статических копий невозможно, непрерывный ряд ситуаций преобразуется в дискретный путем квантования на интервалы неравной абсолютной продолжительности. Принимается, что в течение определенного интервала ситуация остается неизменной и, таким образом, может быть передана статической копией. При этом копируется все то, что остается неизменным на протяжении избранного интервала, и устраняется все меняющееся — это пиктографическое изображение.

В пиктографической сцене ситуация передается с помощью изображений и в меньшей степени условных (типа некоторых современных дорожных) знаков, расположенных, как правило, по масштабному принципу. При этом изображения предметов помещаются в соответствии с их действительным (или воображаемым) пространственным расположением на определенном отрезке времени. В связи с этим предметы изображаются так, как их можно видеть из различных точек, т.е. для каждого предмета или его детали может быть своя точка наблюдения.

Пиктографические сообщения в виде групп сцен возникли в верхнем палеолите и употреблялись у всех племен до появления государств. Для вспомогательных целей пиктография употребляется до сих пор.

В государствах, которые нуждались в разнообразных точных записях, возникла настоятельная потребность в графических сообщениях, фиксирующих словесные. В результате появилось письмо, передающее речь, т.е. являющееся вторичным кодом по отношению к языку.

При изобретении письма можно было графически кодировать основные единицы языка, т.е. фонемы, морфемы, словоформы, предложения. В двух последних случаях потребовалось бы настолько большое

количество знаков, что реализовать их практически было бы невозможно. Древние изобретатели письма независимо друг от друга остановились на графическом кодировании морфем — наименьших смысловых единиц в языке. При этом можно было широко использовать в качестве материала пиктографические знаки, закрепляя за ними определенное чтение и значение.

Однако многие понятия изобразить было по меньшей мере затруднительно. Изобретатели письма пользовались двумя приемами. Первый состоял в широком использовании омонимов, т. е. знаку придавался не первоначальный, а совсем другой смысл. Этот прием, в свою очередь, вызвал появление ключевых знаков (детерминативов), указывающих смысл. Таким образом, морфема стала передаваться знаком, указывающим чтение (но не смысл), и детерминативом, указывающим смысл, но не имеющим чтения. Второй прием состоял в использовании знаков, передающих самые короткие по фонетическому составу морфемы, для передачи частей других морфем. При этом, естественно, также могли употребляться детерминативы, поясняющие смысл.

Другая трудность состояла в передаче служебных морфем, изобразить которые было невозможно. В этом случае также использовались знаки, фактически ставшие фонетическими. Появление последних дало возможность уточнять чтение с помощью звуковых подтверждений, обычно передающих начало чтения знака.

Таким образом, в ранних системах письма (иероглифических) стали употребляться три вида знаков: имеющие только чтение (фонетические), имеющие чтение и смысл (обычно называемые идеограммами), имеющие только смысл (детерминативы).

В некоторых ранних вариантах иероглифики (андское, рапануйское, протошумерское письмо) широко употреблялось аббревиатурное написание, при котором фиксировались в основном знаменательные корневые морфемы, составляющие смысловый костяк предложения, а служебные морфемы не указывались. Так как в контексте знаки корней фактически передавали словоформы, они получили название логограмм.

В дальнейшем иероглифическое письмо (сохранившееся в Китае и Японии), слишком громоздкое по количеству знаков и орфографии, сменилось чисто фонетическим в нескольких разновидностях соответственно характеру языков (слоговое, фонемное, консонантное). Появились даже инфрафонемные алфавиты, в которых одной фонеме соответствовала группа знаков (огамическое и некоторые виды рунического письма).

Как правило, в фонемных алфавитах единый принцип полностью не выдерживается. Например, некоторые знаки передают две фонемы, и, наоборот, некоторые фонемы передаются группами знаков.

Количество знаков в разных системах письма резко различается. В иероглифическом (морфемно-силлабическом) оно доходит примерно до 400 синхронно употребляемых знаков, в силлабическом — до 100, в фонемном — до 40, в инфрафонемном — до 10. По количеству употребляемых знаков можно судить о характере письма, для чего, однако, требуется обширный текст, в котором представлена основная часть знаков.

При наличии только короткого текста о характере письма можно судить по убыванию появления новых знаков, наиболее медленному в иероглифической записи. Естественно, чем меньше знаков в письме, тем длиннее цепочка знаков, передающих слова, однако в иероглифике цепочки могут удлиняться за счет добавления детерминативов и звуковых подтверждений, т.е. в зависимости от правил орфографии.

Были найдены записи, сделанные полностью забытым письмом на неизвестном языке, в связи с чем возник вопрос об их изучении и по возможности дешифровке.

Термин «дешифровка» понимается весьма различно и требует уточнения. Прежде всего следует отметить, что дешифровка исторических систем письма и дешифровка секретных шифров не имеют почти ничего общего. В древних текстах знаки стоят в обычном порядке, но чтение их забыто, в шифрованных записях известные знаки заменены другими и порядок их смешан. В первом случае язык либо неизвестен, либо сильно изменился, во втором — известен.

Под дешифровкой в узком смысле следует понимать установление чтения забытых знаков. Однако чтение текста отнюдь не означает его понимания, так как язык мог полностью исчезнуть или же сохраниться в виде языков-потомков, отличающихся по грамматике и лексике. Некоторые древние тексты (например, этрусские) написаны известным письмом, но на вымершем языке. Таким образом, наряду с дешифровкой письма, необходимо изучение языка неизвестных текстов. Наконец, если достаточно известны и чтение знаков, и язык, необходимо дать чтение, перевод и интерпретацию каждого конкретного текста со всеми его особенностями, что, собственно, относится уже к области филологии, но часто называется дешифровкой текста.

Для того чтобы прочесть текст, необходимо знание данного кода (т.е. чтения знаков и правил их употребления) и знание языка, на котором написан текст. Предполагается, что текст, подлежащий дешифровке, является записью человеческой речи. Человеческие языки сильно изменялись с течением времени, но во всех языках всех времен используются сходные способы передачи информации. Кроме того, предполагается, что подлежащий дешифровке текст не зашифрован умышленно. При наличии умышленной зашифровки необходимо предварительно восстановить нормальный порядок знаков, а затем уже вести дальнейшие исследования.

При дешифровке неизвестных текстов основным и решающим источником информации являются сами тексты. Однако о любом тексте всегда имеется дополнительная информация, которую можно использовать для целей дешифровки. Так, сведения о времени, к которому относятся тексты, дают возможность ограничить круг поисков лингвистических и графических аналогий и определить хронологический разрыв между изучаемым языком и языком-потомком. Сведения о месте находки и объекте, на котором начертан текст, могут дать указания о его содержании. Иногда тексты сопровождаются изображениями, которые могут оказаться важным источником дополнительной информации.

Неизвестный текст может сопровождаться параллельным известным текстом. Последний может оказаться либо независимым текстом на ту же тему (псевдобилингва), либо переводом неизвестного текста (билингва). Значение билингвы для дешифровки трудно переоценить. Дешифровщик как бы получает специально подготовленное учебное пособие (как известно, такие пособия — параллельное издание иностранного текста и перевода — широко используются и при изучении иностранных языков). Однако отсутствие билингвы вовсе не означает невозможности дешифровки.

Изучение текста требует его формализации. Прежде всего текст должен быть транскрибирован стандартными знаками. В качестве последних могут быть использованы стандартизированные знаки изучаемого письма, а также (для удобства обработки и публикации) общепринятые знаки (цифры, буквы). Эта работа требует не только большой точности, но и приобретения специальных навыков — овладения данным шрифтом и индивидуальным почерком. Составление транскрипции предусматривает опознание всех вариаций написания, а также полустертых и искаженных графем, восстановление утраченных мест, обнаружение ошибок и внесение конъектур. Эта работа обычно не бывает закончена к моменту дешифровки и продолжается по ходу дешифровки и после нее. Ошибки при опознании графем представляют значительную опасность, так как исправление их затруднительно.

По ходу составления унифицированной транскрипции возникает необходимость в составлении каталога графем (т.е. знаков и их аллографов). Составление каталога дает возможность начать планомерную работу по выявлению аллографов, выделить особые группы знаков (например, цифры) и «модель порождения» знаков, развернуть работу по сопоставлению изучаемого алфавита с известными, а иногда и по опознанию изображаемых знаками предметов, что может дать важную дополнительную информацию.

При формальном изучении текстов исследователь временно игнорирует всю дополнительную информацию, сосредоточиваясь исклю-

чительно на той, которую несут сами тексты. Дополнительная информация оказывается необходимой на более поздних этапах изучения текстов.

Для удобства исследования текст целесообразно рассматривать как ряд морфем, расположенных в последовательности, свойственной данному языку. Общее количество морфем в любом языке не зависит от количества фонем и не превышает синхронно 1500. Стабильность количества морфем определяется свойствами человеческого мозга. Превышение критического количества создает трудности для запоминания (оперативной памяти). В свою очередь, значительное уменьшение числа морфем повлечет за собой удлинение словоформ и создаст трудности для их распознавания (т.е. для восприятия устной речи). Возможное число сочетаний фонем резко ограничено законами образования морфем в данном языке (фиксированные ограничения). Морфема — наименьшая семантическая единица языка, и поэтому она обычно является предельным референтом знака письма. Каждая группа тождественных морфем характеризуется позициями этих морфем в ряду (адресами) и частотой.

Все морфемы можно подразделить на корневые и служебные. Такое подразделение имеет относительный характер, так как в ряде случаев корневые морфемы могут употребляться в качестве служебных. Однако морфема, занимающая конкретную позицию в ряду, является альтернативно либо корневой, либо служебной.

С помощью служебных морфем образуются словоформы и осуществляется связь между словами в предложении. Количество морфем в словоформе обычно практически не превышает пяти. Общее количество служебных морфем, естественно, значительно меньше, чем количество корневых. Так как одна и та же служебная морфема соединяется с различными корневыми морфемами, частота наиболее употребительных служебных морфем должна намного превышать частоту корневых морфем. Однако в специфических текстах (где часто повторяются некоторые слова) рекордную частоту могут иметь и корневые морфемы. Чтобы избежать влияния специфики текста, целесообразно учитывать помимо абсолютной частоты также относительную, исключая повторяющиеся блоки. При этом рекордные по абсолютной частоте корневые морфемы займут соответствующее место независимо от специфики текста.

Ряд морфем, составляющий текст, может быть разделен на отдельные цепочки, соответствующие морфемам, словоформам и предложениям данного языка. Для целей дешифровки имеет смысл разделить текст на цепочки, соответствующие словоформам.

Общепринятой формой изучения языка является составление словарей и грамматик. Сведения по лексике сосредоточиваются именно

в словарях. Поэтому при изучении языка неизвестного текста также целесообразно иметь упорядоченный набор цепочек, представляющих лексику данного языка.

В словоформах служебные морфемы располагаются обычно в начале и в конце, а в некоторых языках и в середине слова (внутренняя флексия). При разбивке текста на цепочки практически целесообразно включать в состав словоформы не только корневые, словообразующие и словоизменяющие морфемы, но и все остальные служебные морфемы (например, предлоги и послелогии, частицы, союзы). Дело в том, что дешифровщик, ставящий задачу не допускать присоединения к словоформе служебных морфем (выделяемых в грамматиках в качестве самостоятельных частей речи), должен иметь критерии, позволяющие отделить словообразующие и словоизменяющие морфемы от остальных служебных морфем. Такие критерии вообще дать очень трудно, а до начала исследования текста просто невозможно. Наоборот, именно изучение текста может дать основания для классификации служебных морфем. Однако если даже и удалось бы отделить словообразующие и словоизменяющие морфемы от остальных служебных, то изучение последних от этого только затруднилось бы. В самом деле, грамматические функции служебных морфем можно определить, только изучая их сочетания со знаменательными словами, а в этом случае служебные морфемы оказались бы от них изолированными.

Иногда к словоформе приходится присоединять и определение. Это целесообразно в тех случаях, когда определение выражено неизменяемыми словами, лишенными своих служебных морфем и именно поэтому практически неотличимыми от них в неизвестных текстах.

Таким образом, неизвестный текст целесообразно разделить на отдельные цепочки, в общем соответствующие словоформам, к которым присоединены также служебные морфемы непосредственного окружения и неизменяемое определение. Такие цепочки получили название **б л о к о в**.

Технически разделение текста на блоки может вестись следующим образом.

Регистрируются все цепочки, повторяющиеся в тексте дважды и чаще. Такие цепочки могут соответствовать словоформе (если она встретилась более одного раза), основе словоформы (если ненулевой словоизменяющий показатель встретился один раз), корню словоформы (если ненулевые словообразующий и словоизменяющий показатели встретились один раз), сочетанию части одной словоформы с частью другой (случайно повторенному не менее двух раз), сочетанию двух или нескольких словоформ (повторенному не менее двух раз).

Ориентируясь на расчетную длину блока и на практическую длину, полученную в результате выделения блоков, легко устранить слишком

длинные повторяющиеся цепочки, сильно превышающие среднюю длину блока.

Все зарегистрированные цепочки упорядочиваются в порядке убывания частоты. Кроме этого их целесообразно упорядочить по составляющим знакам на основе принятого каталога (например, в порядке нарастания номеров знаков при числовой транскрипции). Упорядоченный набор зарегистрированных цепочек может рассматриваться как исходный материал для составления словаря блоков.

Многие тексты имеют уже в оригинале разделение на цепочки знаков. Такие цепочки во всех текстах, имеющих разбивку, соответствуют, как правило, словоформам, к которым могут присоединяться различные служебные морфемы и неизменяемое определение.

Ближайшей задачей после установления последовательности знаков в ряду и составления стандартной транскрипции является изучение состава блоков с целью выяснить морфологию изучаемого языка.

Состав блоков целесообразнее всего изучать с помощью словарей, построенных в порядке возрастания номеров знаков. Эти словари могут быть упорядочены по первому знаку (прямой словарь), по последнему знаку (обратный словарь) и по внутренним знакам (глубинные словари). Прямой и обратный словари необходимы во всех случаях. Необходимость составления глубинных словарей определяется спецификой языка.

Комбинируя данные прямого и обратного словарей, можно легко получить набор микропарадигм (грамматических показателей, употребляемых с корнями), а затем свести их в парадигмы.

Имеется опасность спутать знаки, входящие в состав устойчивой (корневой) группы блока, с переменными, передающими грамматические показатели. Однако если сомнительные знаки действительно входят в состав устойчивой группы, то они не должны встречаться в качестве переменных перед другими заведомо устойчивыми группами и не должны заменяться заведомо переменными знаками.

Выделение переменных знаков (инициальных и финальных) дает возможность исключить из словаря блоков многие случайные цепочки (случайно повторенные сочетания конца одной словоформы и начала другой), а также приступить к расчленению на блоки оставшихся нерасчлененными частей текста.

Характеристика блоков должна включать также сведения о позициях, которые они могут занимать в предложениях. Как известно, во многих случаях связь между словами обеспечивается только порядком слов, без морфологических показателей. Учет позиций блоков особенно важен для языков, которым свойствен твердый порядок слов. Впрочем, и в тех языках, синтаксис которых не требует твердого порядка слов, обычно имеется предпочтительный порядок (зависящий от литературного стиля).

Определение порядка слов значительно облегчает установление функций грамматических показателей. С одной стороны, появляется возможность расположить последние в синтаксическом порядке («синтаксическая сетка») и дать их классификацию (например, выделить словообразующие, словоизменяющие и иные служебные морфемы).

С другой стороны, определение порядка слов дает возможность развернуть классификацию блоков по частям речи (условным или фактическим) или по иным группам, удобным для изучаемого материала.

Выявление служебных морфем и их функций открывает возможность развернуть сопоставление изучаемого языка с известными языками. Первой задачей является отыскание ближайшего языка-потомка (или группы таких языков). При установлении генетического родства решающую роль играет, конечно, общая характеристика морфологии и синтаксиса. Для детального сопоставления грамматических показателей необходимо подвергнуть тексты на языке-потомке такой же обработке, какой подвергались древние тексты. Кроме того, в зависимости от величины хронологического разрыва между изучаемым языком и языком-потомком в последнем должны быть восстановлены старые грамматические формы методами внутренней реконструкции и сравнительно-историческими (эта работа может быть выполнена только профессиональными лингвистами). Сопоставление служебных морфем изучаемого языка и языка-потомка дает возможность приписать соответствующим знакам условное чтение (которое не следует смешивать с фактическим чтением).

Изучение морфологии и синтаксиса и классификация блоков дают возможность развернуть изучение лексики неизвестных текстов. При переходе к фонетическому чтению решающую роль могут сыграть условные чтения знаков, установленные при сопоставлении грамматических показателей изучаемого языка и языка-потомка. Однако фонетическое чтение слов во многих случаях не дает возможности определить их смысл. Для успешного изучения неизвестной лексики необходимы специальные морфемные словари языка-потомка и детальное изучение фонетических изменений. Кроме того, даже в тех случаях, когда перевод вполне возможен, текст остается непонятным по причине полной невразумительности. Для того чтобы придать древним текстам смысл, кроме грамматического перевода необходим широкий и всесторонний комментарий. Составление такого комментированного перевода уже выходит за рамки формального изучения текстов и, несомненно, требует привлечения всей возможной дополнительной информации.