

Борис Путько

**Введение
в структурную иероглифику**

2020

"角之字形，乃刀下用也。今头上用刀，其凶甚矣！"

– Потому что иероглиф, обозначающий рог, состоит из двух частей: вверху пишется "меч" – "дао", а внизу "юн" – "применять", – объяснил Чжао Чжи.
– Вэй Яню приснились рога на голове – это не к добру: над его головой занесен меч.

Ло Гуань-Чжун. Троецарствие.
(перевод с китайского
В.А.Панасюка)

Я не думаю, что открою кому-нибудь "страшную тайну", если скажу, что чтение иероглифических текстов есть распознавание сложных графических образов. Тогда, изучение иероглифики есть одновременное решение трех различных по сути задач: распознавание образов, фиксация этих образов в памяти и создание графико-семантических ассоциаций между этими образами и их значением или графико-звуковых ассоциаций между образом и его произношением, или его "именем". Структурная иероглифика, на первый взгляд, имеет отношение к решению первой из этих трех задач, но, как мы увидим далее, способна оказать действенную помощь и в решении двух других задач, а также в некоторых утилитарных приложениях, крайне необходимых при изучении иероглифики и в овладении китайским языком вообще, ибо невозможно считать знающим китайский язык человека, который не может читать китайские тексты и писать по-китайски. Кроме помощи при распознавании и запоминании иероглифов, структурная иероглифика, как оказалось, способна помочь и в поиске незнакомых иероглифов в особым образом организованном словаре, и (о, боже!) в святой святынях классической иероглифики — в письме иероглифов. Конечно, не в письме кисточкой на бумаге или на шёлке, а во вводе иероглифов на цифровых устройствах от компьютера до телефона.

Иероглифическим словарям уже больше 2-х тысяч лет, и знающие люди давно понимали значение этих "устройств" для передачи традиции. Но в Китае (возможно, и поэтому) всегда существовала тесная связь, если не сказать "спайка" между учеником и учителем. В рамках именно этой связи и возможна была

передача традиции с помощью словарей с минимальными искажениями. Пользоваться словарем мог только человек, хорошо знакомый с основами иероглифики, имеющий хорошие навыки, полученные от учителя. В те стародавние времена иероглифы в словаре (в "лексиконе") располагались либо "по темам", либо "по ключам", а затем появились и "тонические, т.е. такие, в которых иероглифы расположены по произношению" [1, стр II]. Как отмечал В.П. Васильев "понятно, что ни тот ни другой [способ расположения] не могут годиться для начинающего учиться; надобно сперва выучиться языку, чтобы приискивать [нужные иероглифы] в этих лексиконах. Китайцы так и делают, потому что они учатся всегда по несколько лет у учителей, при помощи которых и запоминают чтение и значение всех главнейших иероглифов. Без учителя, с помощью одного лексикона, для китайца непонятно изучение своего собственного языка." [там же, стр. III]. С появлением различных систем записи произношения иероглифов ситуация вряд ли стала проще. Чтобы найти иероглиф в словаре по произношению, его, произношение, нужно уже знать.

Русская графическая система, основоположником которой и был В.П. Васильев, не сказать, чтобы прервала эту традицию, но показала, что изощренный аналитический ум, "европейский гений", как говорил Василий Павлович, способен предложить достаточно удобные методы структурирования графических образов. По собственному опыту могу сказать, что лупы и полу-часа времени мне было достаточно, чтобы самостоятельно, без какой-либо посторонней помощи, найти первый в моей жизни иероглиф в Мудровском словаре, построенном по принципам русской графической системы [2]. Эта система оказала заметное влияние на кодировку иероглифов "по четырем углам", принятую и в Китае [3], но, тем не менее, и в 20-ом веке и в начале 21-го, словари, даже электронные, остались по сути "плоскими", простыми эмуляторами бумажных словарей. Простая идея поиска иероглифа не по одному ключу, а по комбинации его элементов, по четырем углам, по последнему элементу, не важно, до сих пор не может получить достаточного распространения при организации электрон-

ных словарей. Идея такого поиска иероглифа по набору наперед заданных компонентов не то чтобы не оригинальна, она тривиальна, и реализации подобного механизма появились достаточно давно. Стоит упомянуть текстовый процессор австралийской фирмы NJStar Software Company [4], где, по крайней мере, с начала 2000-ых уже был реализован "выбор по ключам" (Radical Lookup). В 2015-ом году на сетевом ресурсе "MDBG English to Chinese dictionary" [5] появилась новая функция "Show entries containing character components: (汉字部件)", позволяющая найти иероглиф, указав один или несколько его компонентов из списка в 300 штук. Pleco Software, производитель широко известного электронного словаря Pleco, по словам её основателя Майкла Лава (Michael Love) еще в 2017 году вел собственную разработку подобного инструмента [6], но я не знаком с результатами этой разработки, и в последней версии Pleco её следов не заметно. Есть разделение иероглифов на компоненты, но реализовано оно как "исключение" указанного компонента из состава иероглифа и последовательное деление "остатка" по тому же принципу. Обратное действие возможно по той же цепочке, когда указанный компонент отмечается в составе других иероглифов, в том числе и в составе первоначального. Но никакого механизма для "сборки" иероглифа из произвольных компонентов пока нет.

Понимание структуры иероглифов, закономерностей их формирования, и, главное, закономерностей их восприятия, распознавания и запоминания, по моему, способны необычайно облегчить освоение иероглифики, способны снизить порог вхождения в иероглифику для новичков и предоставить дополнительные возможности изучения иероглифики для специалистов.

Структурный подход к иероглифике вовсе не нов, явные преимущества его использования были понятны синологам уже в 19 веке. Анализ графического состава, сходство и различие иероглифов безотносительно их значения и произношения предоставляют дополнительные преимущества при запоминании иероглифов. "Притом вы имеете перед вашими глазами сконцентрированную картину постепенного разрастания китайской письменности; если два сходных

иероглифа поставлены один подле другого, то вы их, конечно, запомните всего скорее; если вы отыскиваете неизвестный иероглиф, то вам невольно припомнятся те, которые сходны, и главное, при этой системе разрастания, кажется только и возможно выучиться писать по-китайски. Не легко, напр[имер], написать с первого разу иер[оглиф] 嬌, но когда перед вами поставили сперва 冫, потом 冫, потом 冫, потом 冫, далее 冫, и, наконец, 冫, то вы не только запомните самый иероглиф, но перед вашими глазами пройдут целые группы других иероглифов." [1, стр. VIII]. Но слишком ранние попытки использовать анализ структуры иероглифов для машинной обработки, в частности, для ввода иероглифов (методы *цанцзе* и *уби*), кажется, лишь дискредитировали саму идею структурного подхода к иероглифике. Недостаточно четко сформулированная задача, произвольное, не связанное с естественными законами восприятия выделение элементов иероглифов для их описания, желание авторов "объять необъятное", отсутствие инструментальных возможностей обработки комбинаций элементов, родили странных монстров, которые могут жить только в голове разработчика и среди убежденных адептов этих систем, и, как показала практика, с трудом воспринимаются и с легкостью отвергаются пользователями.

Мне кажется, структурная иероглифика способна избежать такой ловушки необоснованных претензий. Простота и наглядность реализации композиционного ввода, основанного на закономерностях структурной иероглифики обнадеживает. Мы не можем утверждать, что создание иероглифов было неким "проектом", что существовал некий "стандарт" добавления иероглифа в общую копилку иероглифики, но нельзя утверждать и то, что создатели иероглифов делали это бездумно. Без сомнения у них были какие-то соображения по этому поводу, без сомнения они руководствовались и здравым смыслом, и традицией, и какими-то эстетическими представлениями. И многовековая "обкатка и подгонка", казалось бы, произвольно созданных иероглифов, привела к тому, что сейчас мы имеем то, что имеем, может быть и не до конца понимая причины и истоки нынешнего положения дел. И мне представляется допустимым сослаться

здесь на авторитет Исаака Ньютона, который в "Математических началах натуральной философии", говоря о причинах свойств силы тяготения сказал, что "довольно того, что тяготение на самом деле существует и действует согласно изложенным нами законам" [7, стр. 662]. Механизмы восприятия графических образов могут отличаться от описанных в данной работе, могут быть совершенно иными, но, в данном случае, нам довольно того, что эти механизмы существуют и действуют уже в течение многих столетий, создавая, поддерживая, и обеспечивая будущие перспективы этого удивительного культурного феномена — иероглифического письма. И именно эти механизмы позволяют структурной иероглифике добиваться того, на что она пока лишь претендует, но претендует весьма обоснованно. Надеюсь, читателю из дальнейшего изложения это будет совершенно понятно.

1. Структурная иероглифика. Прологомены.

1.1. Элементы иероглифов. Графы.

Структурная иероглифика относится к тому удивительному разряду предметов, которые стоят на простых и очевидных основаниях, но результативность или эвристическая ценность целого предмета становится очевидной лишь тогда, когда эти, сами по себе вполне очевидные, основания собраны в одном месте, причем в определенном порядке, строго определены и поставлены в соответствие друг другу для достижения вполне конкретной цели, т.е. тогда, когда предмет достаточно формализован. Претензия на построение формализованной системы иероглифики, это одна из тех самых ловушек необоснованных претензий, которые я хотел бы избежать, но попытка построить в первом приближении нечто полезное, конечно же, связана с необходимостью более или менее точно представить себе задачу. Другими словами, структурную иероглифику в нынешнем ее варианте можно представить как *постановку задачи*, а именно как попытку "организовать" обширное поле иероглифики в целях упрощения распознавания, запоминания, и использования (ввода) иероглифов.

На первый взгляд, и это правильно, все иероглифы состоят из отдельных элементов. Эта сентенция сама по себе никого не удивляет, но и ничего не проясняет. В классической иероглифике, например, считается, что иероглиф состоит из набора черт, коих выделяют от 32-х до 38-ми. В общепринятом понимании черта есть элемент иероглифа, который можно изобразить не отрывая кисти от бумаги. Но, это очевидное высказывание, как и другие самоочевидные высказывания, ничего не может дать "ни уму, ни сердцу". Иероглифы и возникли как некое изображение посредством нанесения на носитель последовательности черт, но знание этого само по себе ничего не дает кроме возможности воспроизведения иероглифов, своего рода древняя функция "copy-paste": хо-

чешь научиться читать и писать иероглифические тексты, копируй то, что видишь в нужном порядке. И нынешние компьютерные системы рукописного ввода, по сути, никуда не ушли от черчения иероглифов на бараньих лопатках или на панцире черепахи. Изоощренные машинные алгоритмы, используемые при рукописном вводе, предназначены для того, чтобы заставить машину понять, что нарисовал человек своими руками, идентифицировать написанное, сохранить в приемлемом для машины виде. Это огромный скачок в обучении машины, но к человеку, пишущему иероглифы, это всё не имеет никакого отношения.

На второй взгляд, и это тоже правильно, разделение иероглифов на отдельные черты мало информативно. Последовательность черт относительно простого иероглифа легко запомнить и воспроизвести, но для графически сложного иероглифа запомнить и воспроизвести последовательность большого количества черт просто так, без дополнительных ухищрений, уже невозможно. Собственно говоря, так никогда и не делали. Сложные иероглифы изначально создавались как компаунды, как модификации уже существующих иероглифов добавлением к ним либо дополнительных черт, либо целых "блоков" таких черт, отдельных элементов, которые сами по себе могли иметь какое-то значение и могли выступать в качестве самостоятельных иероглифов. Это также широко использовалось в мнемотехнике, при письме иероглифов и даже в их толковании (см. эпиграф к этому тексту). Другими словами, выделение и использование таких блоков было простым и *естественным* путем развития иероглифики. Ударение следует делать именно на слове "естественный", и, как мы увидим далее, именно естественные законы восприятия сложных графических образов создают основу структурной иероглифики.

Кажется, теперь мы можем позволить себе утверждать, что все иероглифы состоят из некоторой последовательности черт, но при этом должны добавить, что большинство иероглифов состоят из неких *блоков* черт, неких элементов, которые можно воспринимать как отдельные, легко распознаваемые сущности,

при этом в традиции часто имеющие свои собственные имена: 手 — "рука", 戈 — "копье", 鸟 — "птица". Объединение этих сущностей создает новую сущность с новым именем. "Рука" и "копье" создают новую сущность 我 с именем "я". "Я" и "птица" дают иероглиф 鹅 с именем "гусь". Обратите внимание, что даже сливаясь друг с другом, как "рука" и "копье" в иероглифе "я" (我), и даже имея общую черту, эти сущности легко вычлняются нашим вниманием, остаются *инвариантными* для него.

Фиксация этих легко распознаваемых сущностей, по сути, была началом структурной иероглифики. О их существовании знали, использовали в мнемотехнике и при воспроизведении сложных иероглифов, но зафиксировали впервые при составлении словарей, когда иероглифы в словаре стали располагать в соответствии с одним из таких элементов, с первым, чаще всего, при написании иероглифа, и основным или главным при его запоминании или толковании. Тогда эти элементы стали *ключами* в словарях, и это их имя сохранилось и используется до сих пор. Каждый иероглиф имеет свой ключ. Точнее говоря, ключ имеет словарь, а каждый иероглиф имеет некий элемент, который считается главным и выступает в качестве ключа в данном словаре. Стоит отметить, что традиционная иероглифика весьма свободно обращается с терминологией. Один и тот же компонент иероглифа в разных ситуациях может быть чем угодно: графемой, ключом, радикалом, детерминативом, фонетиком, — по воле автора, полагающего, что все всё равно понимают, что он имеет ввиду в той или иной ситуации. Чаще всего в традиционной иероглифике ключ, а иногда это радикал, есть некая часть иероглифа, про которую вам сказали — "это ключ". Так мы и будем поступать: сказано "ключ", значит ключ. Иногда, ключом может быть и сам иероглиф как целое. Обычно это достаточно простой иероглиф, который не то чтобы не поддается разделению на элементы (на отдельные черты можно разделить любой иероглиф, разве, кроме иероглифа 一 или 乙), но разделение которого на элементы почему-то "не нужно". Это может быть и достаточно сложный иероглиф, например, 鬼 или 黑. Потом я скажу, что такого

рода элементы *атомарны*, т.е. неделимы.

Таким образом, мы можем утверждать, что структурная иероглифика началась с выделения отдельной сущности, которую до сих пор называют ключом или радикалом. Но технологические возможности того времени не позволяли реализовать очевидные возможности комбинаторики ключей. Комбинации двух элементов легко реализуются в двумерной таблице, но что делать при увеличении их количества? Уже третий элемент делает кросс-таблицы неоперабельными. "Плоские" словари позволяли просто зафиксировать лишь индекс по ключу и "остаток", много или мало чего есть "после ключа". Иероглифы, имеющие данный ключ, внутри группы могли располагаться по тематическому принципу или по произношению, однако в нынешних словарях закрепился вариант по количеству черт "в остатке". Это позволяло как-то решить задачу классификации иероглифов, помогало запомнить их, помогало и воспроизвести нечетко зафиксированный в памяти иероглиф, когда вспоминался первый его элемент и была надежда на то, что "рука сама вспомнит" остальное в процессе письма. Количество этих ключей зависело от набора иероглифов, их количества и, в большой степени, от произвола составителя словаря. Из общих соображений понятно, что при отсутствии стандарта при составлении словаря, который сам становился стандартом для последующих поколений пользователей, количество ключей напрямую зависело от количества иероглифов в данном словаре — чем больше иероглифов, тем больше и ключей, и количество ключей в ранних словарях доходило до 540 (как в известном словаре "Шовэнь цзецзы" 说文解字 эпохи Хань). В таком состоянии иероглифика пребывала столетиями. До тех пор, пока не возникла настоятельная необходимость стандартизации начертания иероглифов. Тогда же стало ясно, что начинать нужно со стандартизации основных идентификаторов иероглифов, с их ключей или радикалов, и количество их впервые в словаре 字彙 (Zìhuì) 1615-го года было ограничено 214-ю. Этот набор ключей использовался уже почти 100 лет, когда тогдашний император повелел за 5 лет издать "единственно правильный" словарь, что и было сде-

лано в 1716 г. Этого толкового императора, урожденного Сюанье (玄燁), больше знают по девизу его правления как Канси (康熙). И словарь, изданный при его правлении до сих пор называется Словарь Канси (康熙字典 – Kāngxī Zìdiǎn). И набор ключей этого словаря называют "ключами Канси" (Kāngxī radicals). К этому набору элементов иероглифов мы и обратимся в последствии уже из отдельных и вполне определенных соображений структурной иероглифики.

Кроме роли идентификаторов в словарях, ключи издавна выполняют еще несколько задач. Во-первых, ключи несут мощную и проверенную веками смысловую нагрузку. Как правило, иероглифы, обозначающие некие понятия, предметы, действия, относящиеся к одному классу имеют одинаковый ключ в своем составе. Примеров тому тьма в учебниках. Ключи, как правило, выполняют роль смыслового детерминатива, учебники любят делить иероглифы на детерминативы и фонетики [8]. Детерминатив "служит для передачи примерного значения" знака, фонетик "передает точное или приблизительное чтение" знака [там же, стр. 161]. Что значит "примерное значение иероглифа"? Что значит "приблизительное чтение"? Пока вы пользуетесь тысячей иероглифов, специально отобранных для вас преподавателем, все кажется простым и логичным. Но стоит уйти за пределы учебного словаря, все выглядит несколько иначе. Простой пример: традиционный иероглиф 義 (yì), детерминатив 羊 (yáng), фонетик 我 (wǒ)! А тут грянула реформа по упрощению иероглифов. И в упрощенном начертании этот иероглиф выглядит так – 义. Где тут детерминатив? Где фонетик? Дело заключается в том, что разделение иероглифов на детерминатив и фонетик не имеет никакого отношения к структуре иероглифа. Звучание иероглифа в тот момент, когда он создавался, это отдельная песня. Как его изображают и озвучивают сейчас, отнюдь не обязано совпадать с тем, что имело место в момент рождения его. Детерминативы и фонетики имеют отношение лишь к мнемотехнике. И зависят они только от остроты ума обучающего. Какую мнемоническую схему он придумает для того, чтобы вам было проще запомнить

(внимание!) три совершенно различные вещи: графический образ, его произношение и его значение. Простой иероглиф 椅 (yǐ) - кресло. Детерминатив - 木 (дерево), 奇 - фонетик (jī? или qí?). А теперь попробуйте понять, какое отношение имеет дерево к тому устройству, на котором вы сидите, которое регулируется в трех плоскостях, целиком состоит из металла, пластика и ткани? Ах, ведь нужно добавить, что подобные (только по назначению) устройства раньше делали из дерева. Вот и результат - эта ассоциативная информация, будучи связанной к конкретному графическому образу, позволяет вам легче запомнить его. Все, остальное — от лукавого.

То же можно сказать о выделении группы "графем" при описании иероглифов, о разделении иероглифов на "монограммы", "гетерограммы", "идеограммы", "фоноидеограммы". Никакого структурного смысла подобные классификации не несут, ибо в структуре иероглифа нет никаких критериев, признаков, параметров или свойств, которые определяют принадлежность иероглифа к той или иной группе. Такие произвольные классификации позволяют лишь структурировать учебный материал, облегчить его передачу т.е несут чисто дидактическую нагрузку. Пока авторитетный учитель не скажет ученику, что иероглиф 吃 есть идеограмма, а иероглиф 吧 есть фоноидеограмма, нет никакой возможности отнести иероглиф к той или иной группе. Почему 玉 — графема, а 王 — монограмма, почему 男 — идеограмма, а 气 — фоноидеограмма, почему 果 — монограмма, а 栗 — гетерограмма? (Все примеры взяты из [8]).

Забудем о такой псевдоклассификации иероглифов. Еще раз переформулируем основное положение иероглифики: каждый иероглиф («двумерную картинку») можно представить как набор неких элементарных единиц (при этом, некоторые иероглифы могут состоять из одной лишь такой единицы, т.е. из самого себя). Эти элементарные единицы в разных иероглифах могут иметь различные относительные размеры, занимать различное положение относительно других таких же единиц. Т.е., условно говоря, в некоем сложном иероглифе

единица А может быть меньше или больше единицы Б, она может располагаться выше или ниже, левее или правее единицы Б, при этом оставаясь самой собой, некой элементарной единицей А, легко заметной в составе иероглифа и легко опознаваемой, т.е. некоторой *инвариантой*. В таком случае мы можем спокойно абстрагироваться от взаимных размеров этих единиц.

Чтобы избавиться и от указания их взаимного расположения в составе иероглифа, используем понятие *направления декомпозиции*, будем "разрезать" иероглиф на компоненты и описывать его состав начиная с верхнего левого угла и последовательно продвигаться до нижнего правого — слева сверху вправо вниз. Обратите внимание, традиция сделала все, что смогла, для того, чтобы мы пришли к этому. Во первых, традиционная иероглифика требует, чтобы независимо от взаимных размеров его элементов, каждый иероглиф вписывался в условный квадрат, размер которого должен сохраняться во всем тексте, во-вторых, элементы иероглифа, в основном, пишутся слева сверху вправо вниз. Традиция — хорошая вещь, и нарушать ее можно лишь тогда, когда тебе это очень нужно. При этом, нужно понимать, зачем ты это делаешь. Таким образом, пользуясь подсказками традиции, мы можем описать двумерную картинку иероглифа как линейную последовательность его элементов, в которой номер элемента отражает его положение относительно других элементов. При этом иногда возникают некоторые особые ситуации, но о них мы поговорим чуть позже.

Обратите внимание, я только что сказал очень важную вещь: структурная иероглифика *описывает* иероглифы с помощью линейной последовательности элементарных графических единиц. И дальше нам следует поговорить об этих графических единицах. А именно, установить, какие графические единицы можно использовать для описания структуры иероглифов, какими свойствами они должны обладать, сколько их должно быть, какими свойствами должен обладать весь набор этих графических единиц, какой пул иероглифов можно описать с помощью данного набора графических единиц.

В первом приближении эти графические единицы можно считать компонентами иероглифов, и тогда, можно сказать, что иероглиф можно описать просто указав последовательность его компонентов. И опять я никого не удивлю, если скажу, что необходимое для описания иероглифов количество таких графических единиц зависит от того количества иероглифов, которые мы хотим описать с помощью этих компонентов. Принято думать, что количество иероглифов в китайском языке невообразимо. Возможно это и так, но работать с такими псевдо-количествами как "велико", "огромно", "невообразимо" и т.п. совершенно невозможно. Невозможно объять необъятное и нужно отталкиваться от чего-то более приземленного и точного. Например, стандарт UNICODE [9] на данный момент описывает во всех блоках, относящихся к китайскому, японскому, корейскому и вьетнамскому языкам 87870 иероглифов. При этом только блок CJK Unified Ideographs, который стандарт определяет как общеупотребительный (common) [там же, стр. 683], содержит 20971 иероглиф.

Набор IICore (International Ideograph Core), представленный Ideographic Rapporteur Group в мае 2004 г. (Last Updated: 2004.06.16) [10] содержит 9810 иероглифов, о которых источник говорит, что это набор постоянно используемых иероглифов по всей Юго-Восточной Азии, и что он "способен удовлетворить все потребности почти всех пользователей" [9, стр. 685]. 10 тысяч это уже не 87 тысяч. Еще один любопытный источник: набор иероглифов современного китайского языка в упрощенном начертании с указанием встречаемости иероглифов: The Lancaster Corpus of Mandarin Chinese, - составленный Ричардом Сяо (R. Xiao) и Тони МакЭнери (T. McEneary) [11]. Весь список состоит из 6839 иероглифов. Это уже вполне обозримое количество. Если из этого списка исключить те иероглифы, которые встречаются реже, чем 1 раз на миллион, останется всего 4344 иероглифов. На базе такого корпуса иероглифов уже можно попытаться сформировать набор необходимых графических элементов. Но для этого нужно представлять себе, что мы хотим от структурной иероглифики вообще и от этого набора в частности.

Первая и основная задача структурной иероглифики — идентификация иероглифов. При этом мы можем, и должны, усилить требование к идентификации, сделать ее *однозначной*. Что бы мы не делали, наши действия должны иметь на выходе однозначный результат, а именно, один и только один иероглиф. Структурная иероглифика описывает графически уникальные объекты, следовательно, описание каждого объекта должно сохранить эту уникальность, должно быть установлено взаимно-однозначное соответствие между иероглифом и последовательностью его компонентов, как существует взаимно-однозначное соответствие между иероглифом и последовательностью, взаимным расположением и взаимными размерами составляющих его черт. Если нарушить это соответствие при письме, например, то ясно, что "на выходе" мы получим совершенно другой объект. Достаточно расположить откидную вправо с небольшим отступом от откидной влево и вместо человека (人) мы получим цифру восемь (八). Модель взаимно-однозначного соответствия между иероглифом и способом его "рукописного ввода на бумаге" является хорошим примером для структурной иероглифики, но эта модель излишне сложна для реализации, требует соблюдения большого количества параметров: последовательности черт, взаимного их расположения, взаимных размеров. Структурная иероглифика претендует на то, что указания простой последовательности компонентов должно быть достаточно для однозначной идентификации иероглифа. Другими словами, должным образом составленная последовательность компонентов способна однозначно идентифицировать каждый иероглиф в рамках рассматриваемого корпуса таковых. Понятно, что уже одно это требование накладывает некие ограничения на список искомых компонентов. Этот список должен быть *необходимым*, т.е. с одной стороны, он не должен содержать компонентов, без которых можно обойтись, с другой стороны, удаление любого компонента из этого списка может сделать описание некоторых иероглифов с помощью оставшихся компонентов недостаточно эффективным, а некоторых и вообще невозможным. При этом, весь набор компонентов должен быть *достаточным* для

однозначного описания любого иероглифа из рассматриваемого корпуса, и любой иероглиф должен "делиться на компоненты нацело", т.е. в нем и не должно содержаться ничего иного кроме компонентов из вышеуказанного списка.

Из общих соображений понятно, что чем проще выбранные компоненты, тем длиннее будет описывающая иероглиф последовательность, особенно для графически сложных иероглифов. Поэтому использование давно и хорошо разработанной системы черт для идентификации иероглифов представляется слабо эффективным. Стоит отметить, что ввиду особенностей "двумерной" технологии письма, разделение иероглифа на черты до появления печати было единственно возможным способом их "ввода". Также стоит отметить, что запоминание иероглифа как последовательности черт (именно последовательности, а не набора) является одним из самых трудных моментов в освоении иероглифики и не представляется возможным без привлечения посторонних для распознавания техник, главной из которых, безусловно, является использование мышечной памяти.

Наиболее интересным с точки зрения "кандидатов в компоненты" представляется хорошо известный набор ключей Канси или радикалов, давно используемый традицией из других соображений и для других целей. Уже беглый взгляд на традиционный набор ключей говорит о хорошем графическом "покрытии" иероглифов, т.е. подавляющее большинство иероглифов естественным образом раскладываются на ключи-компоненты. В этом есть прямой резон следовать за традицией. Поэтому в качестве основы набора компонентов был выбран один из многочисленных (и не совсем совпадающих друг с другом) наборов иероглифических ключей [12]. При этом некоторые компоненты были удалены из набора, некоторые из отсутствующих в списке ключей компоненты были добавлены. Некоторые из ключей встречаются крайне редко, например, ключ , одновременно являющийся и отдельным иероглифом, в составе ланкастерского корпуса иероглифов встречается с уничтожающе малой частотой 2 раза на более чем 58 млн. иероглифов как отдельный иероглиф и ни разу в

составе других иероглифов. Кроме традиционных ключей Канси, в составе иероглифов встречаются "не ключевые" компоненты, т.е. такие, которые не имеют графического сходства с традиционными ключами Канси, и которые, тем не менее, достаточно точно и однозначно дискриминируют иероглифы. Для примера достаточно взглянуть на иероглифы 乙 и 乞. Верхний элемент второго иероглифа не представлен в списке ключей Канси, но именно он отличает данный иероглиф от предыдущего, являющегося как раз ключом в традиционном понимании.

В результате визуального анализа вышеуказанного ланкастерского корпуса иероглифов [11] был выделен набор компонентов в количестве 259 шт. С помощью этого набора компонентов удалось описать 6839 иероглифов (за исключением небольшого количества единичных включений "мусора", присутствующего в оригинале, который мне не удалось распознать). При этом каждый иероглиф однозначно описывался своей последовательностью компонентов, уникальной в границах рассматриваемого корпуса иероглифов. Этот первоначальный набор компонентов включал в себя все ключи Канси, все их варианты и небольшой набор не ключевых компонентов. Но такое количество компонентов неоперативно в конкретных приложениях. Матрица выбора слишком велика для ее запоминания или даже для расположения на экране цифрового устройства.

И снова обратимся к традиции. Веками работая с ключами, постоянно сталкиваясь с проблемами отображения ограниченного количества ключей на значительно большее поле иероглифов, вынужденная постоянно передавать свои знания следующему поколению, пишущая и читающая братия, конечно же нашла выход. "Явочным порядком", по сути, была проведена некая "начальная формализация" графических образов. Совершенно понятно, "очевидно", что все три элемента иероглифа 众, различающиеся один от другого, являются вариантами одного компонента – 人. Ограниченность пространства ввода иероглифа привела к необходимости выработки неких каллиграфических приемов изоб-

ражения компонентов, сохраняющих их общность с первоисточником. Большинство таких каллиграфических приемов отражены в приложении 1, о котором речь еще впереди. И если в примере с иероглифом 众 это очевидно, то в других случаях встречаются не очевидные графические варианты тех же компонентов. Верхние элементы иероглифов 少, 当 и 兴, например, традиция считает вариантами изображения ключа 小 от "усечения" до переворота "вверх ногами". Идея "вариантов" и есть первичная неосознанная и "не сформулированная формализация" графических образов: видим одно, а в уме держим другое. Ничто не мешает и нам использовать такого рода формализацию, но не в целях запоминания, толкования или индексации иероглифов в словаре, а в целях формального описания последовательности компонентов иероглифа. Несомненно, мы погрешим против идеи наглядности отображения компонентов иероглифа в его описании, но этот веками проверенный прием позволит нам существенно уменьшить количество компонентов для формального описания последовательности компонентов иероглифов. Еще одно преимущество идеи формального отражения, или *кодирования* одних графических единиц с помощью других графических единиц, поможет нам в устранении некоторых коллизий декомпозиции. Отказавшись от наглядного и более или менее очевидного соответствия частей или элементов иероглифа и описывающих их графических единиц, мы можем для своих целей ввести некие "полезные фикции", которые не существуют в начертании иероглифа, но по тем или иным соображениям могут присутствовать в описании иероглифа в качестве одного из равноправных с остальными компонентами.

Когда я говорил о направлении декомпозиции как панацее от необходимости указывать взаимное положение компонентов иероглифа, я был не точен. Так происходит в большинстве случаев, но не всегда. Существуют ситуации, когда правило направления декомпозиции оказывается недостаточным. В качестве примера рассмотрим иероглифы 员 и 呗, последовательность компонентов которых по правилу направления декомпозиции одинакова: 口 и 贝. Для того, чтобы

исключить неопределенность в таких случаях, в конец полной последовательности компонентов более редкого иероглифа (частота использования иероглифов может быть взята из ланкастерского корпуса Р. Сяо [11]) добавляется "пустой", не существующий в составе графического образа иероглифа псевдокомпонент-модификатор "*". Этот пустой псевдокомпонент добавляется в описание иероглифа для того, чтобы обеспечить уникальность описания последовательности его компонентов. Последовательность компонентов чаще встречающегося иероглифа 员 (yuán) будет представлена строкой "口贝", а более редкого иероглифа 呗 (bei) строкой "口贝*" (обратим внимание: *более редкого*, - это важно для последующей разработки системы ввода иероглифов в реальных текстах, таким образом чаще употребляемые иероглифы и вводятся быстрее, и запоминаются проще). Этот же компонент-модификатор "*" можно использовать и для решения других коллизий, например, для различения иероглифов 鸟 (niǎo) и 乌 (wū), последовательность графов которых, соответственно, будет выглядеть как "鸟" и "鸟*", поскольку иероглиф 乌 (wū) используется реже. Компонент-модификатор не имеет собственного графического образа в составе иероглифа, но фиксирует при этом с минимальными затратами визуальное отличие двух иероглифов, тогда лишь, когда это действительно нужно, как и показано в обоих примерах.

Идея использования вариантов начертания ключей была взята напрямую из традиционной иероглифики, и, после некоторой подгонки распределения этих вариантов по компонентам, позволила достаточно заметно уменьшить число компонентов, используемых структурной иероглификой для описания иероглифов. С учетом пустого псевдокомпонента-модификатора "*" достаточное число компонентов было редуцировано до 200. (В дальнейшем, при знакомстве с композиционной системой ввода, основанной на закономерностях структурной иероглифики, станет понятно, почему именно 200.) Поскольку эти 200 компонентов в данном контексте не являются ни ключами, ни радикалами, ни детерминативами, ни фонетиками, т.е. ни чем иным кроме как графическими

образами, из соображений терминологической определенности в дальнейшем будем называть их *графами*. Более строгое определение графа я приведу позже, а пока будем просто считать, что иероглифы описываются последовательностью графов. Графы это не ключи и не иероглифы, хотя графически могут совпадать и с теми и с другими. Графы могут совпадать с отдельными чертами, как например графы J или \backslash , иметь "странное" с точки зрения традиционной иероглифики начертание как графы \wedge или $||$, или не иметь графического представления вообще, как пустой граф-модификатор "*". Графы — *формальные, графически представленные коды* неких компонентов иероглифа. Такого рода формализация графических образов, произведенная нашим привычным к распознаванию сложных графических образов умом, с одной стороны, позволяет "нагрузить" компьютер, заставить его быстро и эффективно обрабатывать нужную нам информацию по идентификации (а потом и вводу) иероглифов, с другой стороны, позволяет не потерять наглядность процесса, простоту его освоения и доступность для любого человека, изучающего иероглифику независимо от учителя (или с его минимальным участием, о чем ратовал В.П. Васильев уже больше 150-ти лет назад [1]). Существует устойчивое мнение, что иероглифика есть настолько сложная и особенная область, что освоить ее можно лишь проникшись корнями, влияниями и вообще духом китайской культуры, лишь сделавшись в некотором смысле китайцем. Структурная иероглифика же позволяет изучать иероглифику, научиться читать и писать по-китайски без всякой необходимости "окитаиваться".

1.2. Постановка задачи.

Структурная иероглифика есть изучение иероглифов как сложных графических образов, восприятие, распознавание и запоминание которых, имеет свои естественные закономерности. Попытка выявить эти закономерности и использовать их при изучении иероглифики может существенно облегчить проникновение в эту загадочную область китайской культуры, в китайскую письмен-

ность. Понимание этих закономерностей способно упростить и облегчить как распознавание, так и запоминание иероглифов, а также, что самое главное, предоставить простой и эффективный метод ввода иероглифов независимо от их произношения, простой и наглядный способ идентификации иероглифов для поиска их в словаре, когда не только не известно "звучание" иероглифа, но даже и сам графический образ которого может представляться не четко или распознаваться приблизительно, как "что-то очень похожее на нечто уже известное", но "слегка не то", т.е. распознаваться по подобию, узнаваться только в ряду ему подобных иероглифов. Более формально задачи, стоящие перед структурной иероглификой, можно изложить следующим образом.

1. На основе визуального анализа базового набора иероглифов, включающего в себя наиболее употребляемые в современном китайском языке иероглифы упрощенного начертания, встречающиеся в современных текстах не реже одного раза на миллион, разработать набор формальных графических единиц для описания полного состава иероглифов, которое позволяло бы:

- однозначно идентифицировать каждый иероглиф из вышеуказанного корпуса таковых;
- максимально свободно выбирать иероглифы из указанного корпуса по любому наперед заданному составу его компонентов;
- формировать группы иероглифов по подобию их компонентов или целых блоков компонентов.

2. Имея в распоряжении описание полного состава всех иероглифов указанного набора, найти механизм простого, наглядного и эффективного ввода иероглифов на любых цифровых устройствах, предназначенных для ввода, обработки, хранения и передачи текстовой информации.

3. Разработать прототипы программных продуктов, использующие установленные принципы, механизмы и алгоритмы и позволяющие не только освоить методы ввода, но и использовать их в реальной жизни: инструменты для запоминания иероглифов, для поиска их в словарях и для ввода иероглифи-

ческих текстов.

Имея уже достаточно точное представление о том, что мы хотим получить, и зачем это нечто нам нужно, можно переходить к более точным определениям.

1.3. Определения.

Структурная иероглифика исходит из того, что:

(1) каждый иероглиф можно описать с помощью некой линейной последовательности отдельных *неделимых* графических единиц из наперед заданного набора таковых, каждая из которых может быть поставлена в соответствие отдельному компоненту в составе иероглифа или отражать модификации или взаимное расположение компонентов в составе данного иероглифа. Эти графические единицы далее будем называть *графами*. Также, структурная иероглифика утверждает, что (2) графы *исчерпывающе принадлежат множеству из 200 единиц* (см. приложение 1), что (3) это множество графов *необходимо и достаточно* для однозначного описания каждого иероглифа из рассматриваемого корпуса таковых, что (4) графы описывают иероглиф *полностью*, т.е. графы полностью охватывают все графическое многообразие компонентов иероглифов, иероглифы описываться графами и только графами, и в описание иероглифов не входит ничего, не принадлежащего разработанному набору графов, что (5) описание каждого иероглифа *однозначно*, и это значит, что последовательность графов, описывающая каждый отдельный иероглиф, *уникальна*.

Правила описания иероглифов (то, что в традиционной иероглифике обозначают как правила декомпозиции) предельно просты и состоят в следующем.

1. Последовательность указания графов в описании иероглифа соответствует составу его компонентов в направлении *слева-сверху вправо-вниз*, для охватывающих компонентов (т.е. таких, внутри которых располагаются другие компоненты) — *снаружи внутрь*. Например, иероглиф  описывается последовательностью графов  и .

2. В случае неоднозначности разделения иероглифа на компоненты, а такое

может случиться для графически сложных иероглифов, приоритет отдается варианту с *меньшим количеством* графов. Например, иероглиф 喜 может быть описан двумя последовательностями графов: "十豆口" и "士口丶一口", - по этому правилу предпочтителен первый вариант. Это правило может быть подтверждено также первым свойством графов, их неделимостью. Графы *атомарны*, поэтому попытка разделить компонент, который полностью описывается графом 豆 на меньшие графические элементы приведет к ошибке.

3. Все *коллизии* описания разрешаются добавлением в описание одного из иероглифов участников коллизии "пустого" графа-модификатора, не имеющему соответствующего ему компонента в составе иероглифа, обозначаемого знаком "*". Коллизиями будем называть ситуации, когда по двум первым правилам не удастся получить уникальную последовательность графов, т.е. два или более иероглифов имеют одну и ту же последовательность графов (эти ситуации исчерпывающе отражены в приложении 2).

Этих простых правил достаточно для построения таблицы описания заданного количества иероглифов (приложение 3). Как я уже говорил, базой для разработки метода анализа состава иероглифов послужило множество в 4344 иероглифов. Простое расширение этого множества до полного ланкастерского корпуса в 6839 иероглифов не привело к изменению ни принципов кодирования компонентов с помощью графов, ни самого набора графов. Оказалось, что набор графов и принципы описания иероглифов с их помощью остались теми же самыми, несмотря на увеличение множества иероглифов более чем на половину. И это обнадеживает, ибо говорит о наличии неких скрытых закономерностей восприятия сложных графических образов, которыми должны были, скорее всего неосознанно, руководствоваться люди, веками использующие подобные системы письма. Это может быть свидетельством того, что и дальнейшее расширение этого подхода на более мощные множества иероглифов не приведет к принципиальным изменениям механизмов и алгоритмов структурной иероглифики. Но, как бы то ни было, мы можем с уверенностью утверждать,

что на пространстве наиболее часто употребляемых иероглифов эти закономерности работают и приносят ощутимую пользу.

Основным результатом структурного подхода к иероглифике является построение таблицы описания иероглифов, являющейся, в свою очередь, базой приложений, использующих формальное описание графической структуры иероглифов для прикладных задач, в первую очередь для ввода иероглифов, для построения электронных словарей с возможностями многократно превышающими таковые традиционных плоских словарей, для создания приложений, облегчающих процесс запоминания иероглифов.

1.4. Таблица описания иероглифов.

Имея полную таблицу описания иероглифов рассматриваемого множества, можно уже заняться "занимательной арифметикой". При этом выясняются занятые вещи (см. приложение 4).

Количество графов в описании каждого иероглифа варьирует от одного до одиннадцати. Максимальное количество иероглифов (1574 шт. или 36,23 %) состоит из трех компонентов (точнее, описывается последовательностью из трех графов). Доля иероглифов, имеющая от одного до трех графов в их описании, в сумме составляет уже 61,74 %. Как мы помним, последовательность графов сохраняет уникальность, следовательно, для однозначной идентификации более половины иероглифов достаточно лишь трех графов. А если учесть частоту встречаемости иероглифов в реальных текстах, то оказывается, что в 82,25 % случаев трех графов достаточно. Собственно говоря, это и послужило отправной точкой в попытке создать простую, эффективную и наглядную систему ввода иероглифов, для освоения которой ничего не нужно, кроме толики внимания, нормальной памяти и некоторого упорства.

Но что делать с остальными иероглифами, которые состоят из большего набора компонентов? Анализ полной последовательности графов для всех иероглифов показывает, что на самом деле, используемое нами описание иероглифов

несет в себе некоторую "информационную избыточность", и мы вправе предположить, что хотя бы для некоторых иероглифов существуют наборы графов меньшие, чем полная последовательность, но способные однозначно идентифицировать данный иероглиф. Выяснилось, например, что иероглиф 驟, который описывается последовательностью из 9-ти графов (马耳又丿丨丶丿丿丿) во всем корпусе рассматриваемых иероглифов однозначно идентифицируется последовательностью из двух только графов: 𠂇 и 耳. Другими словами, в данном наборе иероглифов нет ни одного другого иероглифа, в котором бы встречались вместе графы 𠂇 и 耳. В действительности для всех 4344 иероглифов были найдены такие последовательности графов, равные или меньшие общей последовательности графов для каждого иероглифа, причем по своему размеру эти последовательности не превышают трех графов (см приложение 3). Назовем такие последовательности графов *маркерами*. Теоретически, для ввода любого наперед заданного иероглифа достаточно ввести его маркер, состоящий, как мы видим, максимально из трех графов. При этом, такие последовательности из одного или двух графов имеют 72% иероглифов. Если привести долю этих иероглифов к общей частоте встречаемости иероглифов, то окажется, что в 82-х случаях из 100 при вводе текста достаточно указать 2 только графа. В остальных случаях понадобится три графа, и никогда больше. На самом деле, это достаточно неожиданный результат. В последующем изложении при анализе ныне существующих методов ввода мы увидим, что пока такой эффективностью не обладает ни один метод.

2. Структурная иероглифика. Область применения.

Для того, чтобы понять, почему это работает, полезно рассмотреть сам процесс распознавания сложных графических образов. Здесь необходимо сделать одно очень важное пояснение относительно содержания данного раздела. Мне далеко до строгого научного описания процессов нервной деятельности при распознавании графических образов. Я могу основываться лишь на общебиологических представлениях о восприятии и на собственном опыте. И всё изложенное в данном разделе не есть результат научного эксперимента, а всего лишь результат внимательного наблюдения за собой и собственными ощущениями в процессе распознавания образов — узнавания иероглифов. Для человека далеко не детского возраста, когда способность к образному восприятию давно потеряна, когда любой образ давно привязан к имени, и сознание с большим удовлетворением оперирует именами, возврат к распознаванию и запоминанию образов составляет достаточно трудную задачу. Настолько трудную, что она не решается сама собой, требует каких-то усилий и времени. Самое главное, что этот процесс растянут во времени. А значит, за ним можно наблюдать. И неизбежная многократность повторов позволяет наблюдать за процессом в деталях.

Я полагаю, это допустимо *во введении* в новую область, которая, естественно, ожидает участия в ее разработке профессиональных исследователей в области физиологии развития человека, психофизиологии, педагогики, лингвистики и других областей, для которых владение материалом и методами их науки гораздо полезнее не предвзятого, но поверхностного взгляда дилетанта.

2.1. Распознавание графических образов.

Самая первая и самая трудная для анализа часть восприятия сложных графических образов есть как раз распознавание, фиксация графического

образа как понятого, включение его "в обработку" как отдельного объекта, некой инвариантной сущности, способной вступать в ассоциативные связи, обрабатываться как часть графико-звуковых ассоциаций, возникающих при заучивании иероглифов и использующихся при чтении иероглифов вслух, когда после возникновения графико-звуковых ассоциаций возникают звуко-семантические ассоциации, или как часть графико-семантических ассоциаций при чтении про себя. Дело осложняется тем, что обработка графических образов это настолько древний механизм обработки информации, что он эволюционно запрятан в самую глубину подсознания. Другими словами, восприятие графических объектов происходит в подсознании, и мы не можем сознательно влиять на процессы распознавания образов. Это, конечно, недостаток, но в этом есть и одно весьма важное преимущество. Оно заключается в том, что процессы распознавания графических образов, поскольку происходят в подсознании, если идут, то проходят катастрофически быстро. Было бы не рационально не использовать это преимущество.

К нашему счастью, мы представляем в общих словах как мозг обрабатывает сложные графические образы. К нашему же счастью, за многие века изучения и использования иероглифов люди, неосознанно используя некий культурный аналог естественного отбора, когда нежизнеспособные структуры просто отмирали, подспудно привели структуру иероглифов в соответствие с естественными законами восприятия графических образов, в соответствие с теми способами, которые позволяли наиболее просто и естественно запомнить, воспроизвести и передать смысл графических образов.

Первое, что стоит отметить при этом, это то, что мозг хранит графический образ как *единое целое*. И это важно. На этом основано использование карточек при запоминании иероглифов. Учащийся видит какое-то время иероглиф как картинку, и еще через какое-то время ему предъявляется его значение или произношение. Все, что нужно, для того, чтобы запомнить образ, это видеть "картинку" необходимое время. В этом случае активность нашего мозга заклю-

чается лишь в попытках запомнить данный образ. И подсказка его значения или произношения создает лишь графико-семантическую ассоциацию, не более того, и нисколько не способствует запоминанию самого образа. Фиксация образа в памяти происходит спонтанно, сама собой безотносительно нашей судорожной активности : "запомни, запомни, запомни". Но мы знаем, что мы помним образ как целое, и, поэтому мы достаточно просто *узнаём* иероглиф, хотя первое время не в состоянии не то чтобы воспроизвести его на бумаге, но даже не можем представить его во всех деталях в уме. Но увидев его, узнаем сразу, мгновенно. Именно поэтому нам проще запомнить иероглиф и узнать его в ряду ему подобных. Именно поэтому фонетические системы ввода ускоряют ввод иероглифов по сравнению с рукописным вводом, но нисколько не способствуют запоминанию иероглифов, поскольку в ответ на ввод слога отвечают списком иероглифов, которые *звучат* подобно искомому, а не *выглядят* подобно ему.

Второе, что следует здесь отметить, это то, что для подсознания *сходство есть отсутствие различий*. И это позволяет нам понять, как происходит распознавание образов. Наш мозг находит сходство, отмечая различия. Где и как он ищет различия? Взгляните на эти картинки: 猴, 侯, 喉. Вы сразу заметите, что эти картинки различны и отличаются они своими первыми элементами. И уже потом вы замечаете, что они очень похожи своими последними элементами. Мы видим здесь, что наш мозг воспринимает картинки разделяя их на некие элементы и затем анализирует их последовательно, один за другим. Это направление анализа, конечно, может зависеть от наших культурных предпочтений (я полагаю, арабы, например, могут видеть все с точностью наоборот, но это не важно). Все последовательно и очень просто: 什, 化, 代, 位, 们.

А что происходит, когда образ содержит больше элементов и они не укладываются в простую линейную последовательность? Если они располагаются сразу в двух измерениях: сверху и снизу, слева и справа? 媛 или 暖, или 暖. Наш мозг не склонен усложнять себе жизнь без необходимости. На пер-

вой картинке мы сразу видим 女 и еще "нечто", не зная ещё имени ни для 女, ни для последующего "нечто". На второй картинке мы видим 日 и снова то же самое "нечто". Этого нам достаточно, чтобы различить эти образы. Но на третьей картинке мы видим то же 日, и уже другое "нечто", похожее на предыдущее "нечто", но отличное от него небольшим элементом внутри: 爰 и 爰. Мы помним, что сходство есть отсутствие различий, и мозг ищет различия, и соглашается на сходство лишь тогда, когда не находит различий. Где? Сначала снаружи, в начале и в конце привычной последовательности, и лишь потом внутри образа, двигаясь от периферии внутрь. Мы видели это на примере. Затем мозг фокусируется на обнаруженных различиях, запоминает их и использует эти различия как *отличительный признак*, как своего рода метку этого образа. Но, обратите внимание, как метку, отличающую только эти два конкретных образа. И тогда мы понимаем, почему нам проще отличить и запомнить именно эти два образа в сравнении друг с другом, потому что это естественный для нашего мозга путь распознавания образов. Таким образом, первый элемент (日 или 女) отличает 暖 от 媛, и внутренний элемент (一 или 冫) отличает 暖 от 暖. Но если существуют некие элементы, отличающие один иероглиф от другого, то мы вправе предположить, что для каждого иероглифа может существовать некий набор элементов, который отличает данный иероглиф от всех остальных иероглифов в рамках рассматриваемого множества. В данном случае, для нашего набора из трех иероглифов мы можем утверждать, что для первого иероглифа (媛) таким элементом является 女, для второго (暖) – 日 и 一, для третьего (暖) 日 и 冫. Эти последовательности компонентов иероглифа и есть их маркеры, которые мы уже обнаружили ранее аналитически, не обращаясь к механизмам их появления. Памятуя о том, что последовательность компонентов иероглифа описывается последовательностью графов, можно утверждать, что *маркер есть уникальная последовательность графов, отличающая данный иероглиф от всех остальных иероглифов данного множества*. Сейчас для нас самое главное это то, что разделение сложного графического

образа на компоненты, последовательный их анализ и выбор маркеров для каждого графического образа есть естественный способ распознавания сложных графических образов. При этом нужно помнить, что графический образ хранится в памяти как целостный объект, а не как набор компонентов. Такого рода "графическая аналитика" нужна лишь в процессе распознавания образов.

2.2. Фиксация образа в памяти.

Следующий важный вопрос есть фиксация графических образов в памяти, отдельный и неконтролируемый сознанием процесс. Необходимо иметь в виду, что этот процесс происходит спонтанно, самопроизвольно, без участия сознания, и все наши усилия повлиять на этот процесс должны сводиться к созданию благоприятной обстановки для такой фиксации. К примеру, заметное влияние на скорость запоминания оказывает комфортная динамика появления и исчезновения образа. Эта динамика индивидуальна, она должна быть эмпирически "подогнана" к восприятию, должна быть синхронизирована с процессом восприятия. Чтобы запомнить графический образ, на него нужно "насмотреться". При этом его появление и исчезновение каким-то образом стимулирует запоминание. Вопрос только в оптимальном чередовании и комфортном времени рассматривания образа. Кому-то нужно дольше смотреть, кому-то меньше, кому-то нужно больше сеансов появления-исчезновения, кому-то меньше. Но в динамике появляющийся и исчезающий образ запоминается быстрее. Можно прочитывать иероглиф по-китайски, пока вы его видите, можно вспоминать его перевод, когда он исчезает, можно представлять его отдельные компоненты, но при этом нужно понимать, что построение графико-семантических или графико-звуковых ассоциаций есть совершенно другой процесс, и он происходит по своим законам, независимо от фиксации образа как такового в памяти. Сначала фиксируется образ независимо от его имени или звучания, потом возникают ассоциативные связи этого образа.

Привычный способ идентификации образов через имя: $\dot{\downarrow}$ — "говорить",

己 — "сам", — порождает и привычный способ контроля запоминания. Образов? Нет, графико-семантических ассоциаций. Сам образ в этом процессе обучения фиксируется в памяти *наряду* с выработкой графико-семантических ассоциаций, параллельно и независимо, фиксируется как отдельная сущность, образ отдельно от имени. Иероглиф 己 лишь на первом этапе обучения воспринимается аналитическим умом как совокупность имен "говорить" и "сам", но со временем, естественно и спонтанно, 己 становится совокупностью образов 讠 и 己. Образ "живет" в памяти отдельно от его имени, и объединяет их только ассоциативная связь. Освобождение старой ходульной ассоциации "запомнить — говорить самому" приводит к возникновению новой графико-графической ассоциации "讠 = 讠 + 己", и новой графико-семантической ассоциации "己 — запомнить".

Тренировка такой "безымянной" фиксации графики облегчается с помощью нескольких приемов. Когда мы говорили о естественных законах распознавания графических образов, мы обращали внимание на то, что сознание фиксирует сходство только как отсутствие отличий, поэтому и проще всего запоминать иероглифы в ряду себе подобных, явно фиксируя эти отличия подобных образов: 瓣, 辩, 辮, 辨. Это естественное внимание нашего сознания к мелким различиям подобных объектов проявляется в том, что проще угадать нужный иероглиф в ряду ему подобных, чем воспроизвести его не только на бумаге или на экране, но даже просто представить его в уме, так сказать "визуализировать" его в сознании. Потом придет способность "рассматривать" иероглиф в голове, как считать в уме, способность разложить его на компоненты из стандартного набора, на которые тоже нужно "насмотреться", уметь видеть их в составе новых иероглифов. Недаром при отсутствии тренировки через какое-то время при анализе состава нового иероглифа основная ошибка, приводящая к неудачному поиску такого иероглифа в словаре есть попытка разложения редко встречающихся и атомарных по определению компонентов, которые, ввиду редкости своего употребления, имеют тенденцию вылетать из памяти.

Поэтому кажется, что при изучении (особенно начальном) иероглифики есть смысл делать уклон в сторону графической, а не тематической организации учебного материала. Кстати, китайская педагогика это явно использовала в практике обучения. Достаточно взглянуть на оригинальные китайские материалы для начальной школы еще 80-х годов прошлого века [13]. К сожалению, новые "школы" изучения китайского языка, наводнившие в последние годы интернет, делают ставку на "новизну" материала, ориентируют учащихся на "современный китайский язык" и, под предлогом того, что "нынче так не говорят", вольно или невольно затрудняют и без того непростую задачу освоения иероглифики. Никакой диалог нынешних подростков не даст столько для освоения иероглифики как коротенькая сказка 小猴子下山 [там же].

При обучении композиционному вводу для стимулирования процесса "отделения образа от имени" в инструменты для запоминания графов и иероглифов (см. ниже раздел 3) специально был добавлен режим "картинки" (picture mode), когда "на опознание" предлагается не значение, не произношение, а именно образ, начертание (графа или иероглифа). И оказалось, что для поиска графа на матрице или для воспроизведения предложенного иероглифа тратится времени почти в 2(!) раза меньше, чем в традиционных уже режимах обучения, например, на флэш-картах, когда для проверки предлагается значение или произношение иероглифа. Попытки внести в процесс распознавания образов элементы "рационального осмысления" ("образ-имя-образ") замедляют процесс. Грубо говоря, рука (с мышью или с сенсорным экраном) работает быстрее головы. Образ фиксируется в сознании раньше его имени или значения, или произношения. И этот временной лаг явно заметен в процессе фиксации образа: он сразу узнается как отличный от других, но еще не имеет своего имени. Поначалу это обескураживает и раздражает: "Почему я не могу запомнить его?" Но это нормально. Это значит лишь то, что образ уже зафиксирован в подсознании и нужна лишь ассоциативная связь этого образа с его значением. А как мы уже знаем, это отдельная задача со своими особенностями и приемами решения. От-

существование в этой связке имени образа особенно показательны тогда, когда необходимо воспроизвести неизвестный иероглиф через какое-то время после того, как он впервые встретился. Опыт показывает, что запомнить иероглиф через имена его компонентов, как в примере с иероглифом "запомнить" (記), чрезвычайно трудно, почти невозможно, а графический образ сам всплывет в сознании, и оказывается, что увиденный 2 - 3 часа назад где-нибудь на вывеске иероглиф можно без труда найти в словаре.

Алгоритмика композиционного ввода организована так, что единственный иероглиф пользователь видит тогда, когда машина уже идентифицировала его. А на всех промежуточных этапах идентификации пользователь видит иероглифы выбранные из таблицы по тому или иному принципу сходства, т.е. почти всегда в окружении себе подобных. Отдельные алгоритмы выборки специально разработаны для того, чтобы можно было не случайно формировать списки иероглифов, а целенаправленно отбирать те из них, которые имеют определенную степень сходства: по начальным компонентам, по конечным, по любому порядку следования наперед заданных компонентов, выявлять часто встречающиеся блоки компонентов, обнаруживать вхождение одних иероглифов в состав других. Подобную работу делали всегда, но это была "штучная", кропотливая работа, связанная с огромными трудозатратами, с большими наборами карточек и изображений. Сейчас эта задача с помощью компьютерных систем управления базами данных выполняется структурной иероглификой "на лету". Теперь мы имеем неограниченные возможности выборки и сравнительного анализа иероглифов. Выявление схожих элементов иероглифов, вхождения одного иероглифа в состав других — все это значительно упрощает процесс запоминания иероглифов. Теперь даже для новичка никакого труда не составляет выделение таких рядов иероглифов. Чтобы получить ряды иероглифов с одинаковыми компонентами, например, 喪畏俛喂煨猥展輾碾 или 喝噶揭渴竭歇遏褐葛葛渴霽蝎, достаточно всего нескольких кликов мыши.

И эти возможности дают преимущество не только в изучении структуры

иероглифов. В конце концов, иероглиф не только "картинка", но еще и "звук". Работая со словарем, построенным по принципам структурной иероглифики, одним запросом можно выяснить произношение целой группы иероглифов. Например: 畏 — wèi, 偎 — wēi, 喂 — wèi, 煨 — wēi, 猥 — wěi. Чем 畏 не "фонетик"? Или: 曷 — hé, 喝 — hē, 褐 — hè, 遏 — è, 揭 — jiē, 竭 — jié, 渴 — kě, 葛 — gé, 噶 — gá, 藿 — ħi, 蝎 — xiē, 谒 — yè, 霭 — ǎi. 曷 в первом и втором случае, наверное, "фонетик", а далее? Такой подход дает возможность построить список *облигатных* фонетиков, таких, которые всегда определяют произношение иероглифов в рамках наперед заданных вариаций, например, в пределах одного слога с тоновыми только изменениями, как в первом примере, и исключить из списка таковых те случаи, когда в потугах притянуть к умозрительной схеме совсем не подобающий материал, преподаватель вынужден говорить о "приблизительном звучании". Без такого анализа, все разговоры о "фонетиках" могут оказаться, в лучшем случае, удачной мнемотехникой для запоминания некоторого количества иероглифов, а в худшем — пустым заполнением учебного времени и лишней нагрузкой для студента.

В дошкольном и раннем школьном возрасте дети спокойно и без усилий запоминают образы. Во взрослом состоянии человек, приобретая способности к аналитическому познанию, постепенно теряет способность непосредственного образного восприятия. Поначалу наш испорченный привычкой к аналитике ум бунтует, отказывается запоминать то, что не понимает. Но его можно обмануть. Стоит сказать ему, что он понимает, как он перестает мешать в этом деле, и даже начинает помогать в чем-то. Включаясь в сочинение всякого рода "ходульных ассоциаций", он думает, что "анализирует" образы. Пусть, ибо в это время образы и запоминаются, и организуются графико-семантические ассоциации, которые потом будут жить в подсознании. Пусть аналитический ум участвует в транспортировке образов в подсознание, ибо здесь огромное поле деятельности для аналитического ума, для уже сознательного запоминания значения "готового" графического образа. Традиционные мнемотехники, связанные с вы-

делением ключа в иероглифе, с разделением иероглифа на детерминативы и фонетики, вынужденно используют структуру иероглифа, неявно признают наличие сущностных характеристик его структуры. Изобретая всякого рода произвольные связки структуры с некими своими представлениями, которые зачастую не имеют никакого отношения к реальности, или имели их когда-то, но давно потеряли, они подспудно создают полезные ассоциативные связи. Нужно понимать, что графико-семантические ассоциации есть факт сознания, а не реальности, и для сознания важна их прочность, а не "правильность" их возникновения. Зачастую "неправильные" ассоциации помогают запомнить объект и проще, и быстрее, и прочнее. Примеров таких "неправильных" ассоциаций масса в учебниках и руководствах по языку, еще больше их можно создать самому (см. далее раздел 2.3). И полезнее делать это самому, поскольку самостоятельная активность сознания это главный двигатель и вдохновитель процесса запоминания. В отношении структурной иероглифики можно утверждать, что явно сформулированный принцип композиционности иероглифов, принцип комбинаторики компонентов является наилучшим подспорьем для активности такого рода, он "развязывает руки" при такого рода работе сознания, снимает все ограничения, которые заданы, например, необходимостью делить иероглиф на фонетик и детерминатив, именно на этот фонетик и этот детерминатив, давно и прочно заданные традицией. Или выделять именно этот ключ в иероглифе. Здесь свобода ассоциативного творчества гораздо полезнее привычки запоминать "правильно".

2.3. Фиксация графико-семантических ассоциаций.

Следующий этап после фиксации образа есть "оживление" его, задание связей и отношений данного образа и его смыслов. Просто запредельная многозначность иероглифов способна обескуражить кого угодно. Но она может и помочь. Кластер сходных значений запоминается проще отдельного значения. Ищите абстрактные значения для таких кластеров.

暖 — тепло; теплый; греть, подогревать — "теплота и ее передача" (включая, кстати, и переносные значения, например, "ставить угощение"). Или любопытный иероглиф 嘴, он идеален для демонстрации такого абстрактного подхода, ибо по сути это любая "выступающая часть": рот, пасть, клюв, носик, мундштук, бутылочное горлышко — нечто длинное и выступающее.

打 - "универсальный", многозначный глагол типа "дать", "выдать", "сделать" и т.п.:

打下 — сделать (сбить [с ног]; свалить);

打中 — дать по цели (попасть в цель, поразить цель);

打回 — дать назад (возвращаться);

打开 — распахнуть (дверь).

Неслучайность структуры иероглифов и неслучайность набора их значений, отработанная веками употребления, иногда поражает своей "логичностью", иногда обескураживает неожиданностью — и то и другое способствует запоминанию. Изучайте поле значений, сочиняйте для них метки смысла.

Используйте побочные ассоциации: самец и гун – 公, обезьяна (猴) и хоу (侯). Никто не говорит, что китайцы называли гунов "самцами", и обезьяна это "звериный" хоу. Но вот, по аналогии с нашим "царем зверей", лев (狮) уж очень похож на "звериного" учителя (师).

看 — рука (手) над глазом (目), человек смотрит вдаль — классика. Основа примера в широко распространенной мнемотехнике смысловой нагрузки ключей. Подобных ассоциаций вы найдете где угодно и сколько угодно. В данном случае ключ - "глаз", а "рука" вовсе не ключ. А почему не наоборот? Не будем обсуждать "правильность" такого рода ассоциаций — графы не ключи, но что нам сейчас до этого за дело? Они вам помогают? И слава богу, пусть работают. Ходульная ассоциация "смотреть — рука, глаз" облегчает жизнь, и это самое главное, а самое интересное, что со временем и совершенно незаметно для вас при достаточном трудолюбии и многократности повторов она заменится нормальной графико-смысловой ассоциацией "看 — смотреть". Тем более, если

вы серьезно отнесетесь к освоению матрицы графов, то при воспроизведении "看" будет не "рука и глаз", а "手 и 目", а "смотреть" не "приложить руку к глазам", а просто "看" как графический образ смысла "смотреть".

Такой фокус проходит не только с ключами. Иероглиф 掰 — прямо таки ребус: делить (分) между двумя руками (手), — "разломить". Один ключ и один "не ключ". Для способности читать, а потом и для умения писать, важно лишь чтобы в голове намертво засела ассоциация образа "掰" со смыслом "разломить".

保 - человечек с копьем в руке, стоит "охраняет", "защищает" нечто. Это не результат "анализа" значений ключей ("человек", "рот", "дерево"), а просто картинка похожая на детский рисунок, который связывает образ и смысл, создает графико-семантическую ассоциацию: "保 — гл. защищать".

2.4. Иероглифические словари и поиск.

Парадокс иероглифических словарей заключается в том, что они созданы для человека уже хорошо знакомого с иероглификой. Первое (и единственное!), что видит человек в незнакомом иероглифе, это его графическую структуру. Он ничего еще не знает о произношении этого иероглифа. Поэтому фонетическая организация словарей для новичка бессмысленна. Словарь должен быть "беззвучным". Единственное на данный момент разумное решение — поиск иероглифов по ключам. Но недостатки ключевой системы тоже давно ни для кого не секрет [1, стр. III-IV]. Русская графическая система оказала заметное влияние на кодировку иероглифов по четырем углам, но при этом сама она исчезла из бумажных словарей, а в электронных так и не появилась. Представить себе начинающего студента владеющего кодировкой иероглифов по четырем углам — это верх необоснованного оптимизма. Человек, решившийся на самостоятельное изучение иероглифики, оказался брошенным один на один с "китайской грамотой".

В чем заключается устойчивость ключевой системы расположения иеро-

глифов в словаре? В чем заключается результативность русской графической системы? Рискну утверждать, что дело в том, что обе эти системы опираются на "иероглифы, принадлежащие письменности (больше нежели звуку)" на то, "из чего они состоят, в каком виде они являются глазу" [1, стр. VI]. Обе эти системы базируются на использовании естественных законов восприятия сложных графических образов. Каждая со своей стороны в прямом смысле слова. Ключевая с начала, русская графическая с конца. При этом ключевая система прямо декларирует наличие неких элементов иероглифов, которые определяют их состав, называя их иногда "ключами", иногда "радикалами". Русская графическая оперирует понятием "черта". Но при внимательном рассмотрении указателей русской графической системы [2] легко заметить, что каждая черта является частью некоего элемента, который на самом деле и относит иероглиф к той или иной группе. Например, группа 8 целиком состоит из иероглифов с элементом 口 в нижнем правом углу, а группа 9 с элементом 日. Группа 34 легко интерпретируется вся как содержащая элемент 力. Группа 38 содержит иероглифы с двумя элементами: 木 и 小, — и легко разделяется на две подгруппы. В группе 45 лишь один иероглиф (及) и его производные выбиваются из общего ряда иероглифов с элементом 父, и т.д. В общем случае, каждая группа содержит 2 или 3 элемента. И это те же самые элементы, что используются классической ключевой системой.

Здесь стоит обратить внимание на неслучайность распределения графов маркера в описании иероглифа (см. далее раздел 3.4.2). Первый граф маркера это всегда первый граф общей последовательности графов, т.е. это всегда верхний левый элемент иероглифа. Второй граф маркера может быть либо вторым, либо последним графом полной последовательности, третий граф почти всегда последний граф, отражающий нижний правый (последний) элемент иероглифа. С графической точки зрения использование таких маркеров представляет собой некий синтез ключевой системы и русской графической. Точнее, и традиционная ключевая, и русская графическая системы отражают, каждая со своей сторо-

ны, общие закономерности восприятия сложных графических образов (см. раздел 2.1).

Возможность свободно комбинировать компоненты для поиска иероглифа в словаре избавляет от необходимости "правильно" выделять ключ иероглифа для его поиска, нет никакой необходимости разбирать и запоминать исключения и нелогичности выделения того или иного ключа в том или ином иероглифе. Свободная комбинаторика компонентов позволяет выбрать из таблицы описания иероглифов все иероглифы, содержащие тот или иной компонент, выбрать из таблицы все иероглифы, содержащие любую мыслимую комбинацию компонентов, независимо от того, являются ли эти комбинации "правильными" или нет с точки зрения традиции. Реализацию такого рода комбинаторики легко переложить на плечи компьютера и избавить пользователя от рутины по поиску результата. Это значительно облегчает жизнь особенно начинающего студента, ибо не требует от него никаких дополнительных навыков, кроме привычки видеть эти компоненты в составе иероглифа, которая легко и просто вырабатывается практикой и не требует длительного сопровождения процесса учителем, постоянно "направляющего" и "подправляющего" студента. Для этого не нужно изощренной алгоритмики как, например, для реализации рукописного ввода, и вполне достаточно возможностей любой самой простой системы управления базами данных. Технически достаточно просто дополнить такого рода поиском любой ныне существующий электронный словарь. При этом возможности поиска иероглифов в словаре упрощаются, а скорость поиска увеличивается кратно.

Естественность восприятия иероглифов как набора компонентов и простота освоения позволяет начинать изучение иероглифики с самого начала обучения. Мало того, структурная иероглифика позволяет не только самостоятельно искать иероглифы в словаре и, таким образом, понимать иероглифические тексты, но и самостоятельно писать иероглифы без общепринятого бесконечного прописывания их на бумаге. Причем писать быстро и понятно для того, кому текст предназначается, писать с первого раза правильно. Реалии ны-

нешней жизни таковы, что рукопись, как бы хорошо она не была сделана, требует ввода в компьютер (или на любое другое цифровое устройство). Серьезные тексты пишутся и передаются с помощью компьютера, а не записок на бумаге. Значит, вопрос сводится к тому, что на основе принципов структурной иероглифики нужно разработать действенный (простой и эффективный) метода ввода. И это было сделано в виде прототипов системы ввода, а затем и некоторых полезных приложений.

3. Методы ввода.

Существует 3 основных (по распространенности) группы методов ввода иероглифов на компьютере.

3.1. Рукописный ввод.

Рукописный ввод полностью эмулирует запись иероглифа на бумаге. Старая и мудрая технология. Иероглиф однозначно определяется последовательностью, размером и взаимным расположением стандартных по форме черт. Машина с помощью специальных и достаточно сложных алгоритмов "распознает" введенный иероглиф и предлагает пользователю вариант (или варианты) для вставки в соответствующую позицию в ожидающей ввода программе (так называемый *фокус ввода*). Говорить о скорости рукописного ввода не приходится, ибо количество черт в иероглифе может достигать нескольких десятков, причем для корректного распознавания изображенного пользователем иероглифа требуется точность воспроизведения по трем параметрам: по последовательности ввода черт, их взаимному расположению и размеру. Конечно, в развитых системах варианты предлагаются уже тогда, когда пользователь еще не закончил ввод, развитые системы распознавания рукописного ввода предполагают алгоритмику адаптации к особенностям ввода конкретного пользователя, но, в целом, ими пользоваться можно только от неизбежности. 人 — очень просто. 犬 — просто. 狗 — это уже заметно сложнее. 猴 — попросите любого новичка написать по-китайски простое слово hóu (обезьяна). А что сказать о иероглифе 藏?..

Понятно, что скорость такого рукописного ввода иероглифов принципиально не много выше скорости письма иероглифов на бумаге. Старая и мудрая технология, но медленная и трудоемкая. Преимущество у рукописного ввода только одно: не требуется нового навыка для тех, кто уже прошел курс обучения

письму на бумаге.

3.2. Фонетический ввод.

Во всех видах фонетического ввода иероглиф вводится по его нормативному звучанию (например, на пекинском диалекте, на кантонском диалекте и т.д.), которое записывается либо латинскими буквами (пиньинь), либо кириллицей (палладий), либо буквами китайского фонетического алфавита (чжуинь). Существуют и другие способы записи произношения иероглифов, но они распространены слабо и принципиально не отличаются от вышеприведенных. Наиболее распространен фонетический ввод по пиньиню. К числу главных недостатков фонетического ввода следует отнести следующие три.

а) Между иероглифом и его звучанием нет взаимно однозначного соответствия. Отдельный иероглиф может произноситься по-разному (иметь до 8-ми вариантов чтения), отдельному слогу может соответствовать несколько десятков иероглифов (~25 тыс. иероглифов основных блоков UNICODE последней версии v.10 "озвучиваются" с помощью всего 1314 слогов китайского языка, в среднем каждому слогу соответствует 19,3 иероглифа; только 57 слогов записываются одним иероглифом, а слогу "и" 4-го тона (yì) соответствует 449 иероглифов). Следовательно, существует проблема выбора искомого иероглифа из всех предложенных системой в ответ на ввод произношения. Понятно, что вышеуказанный выбор из возможных вариантов, снижает скорость ввода, требуя от пользователя дополнительного внимания и дополнительных манипуляций.

б) После завершения идентификации иероглифа, даже если эта идентификация произведена однозначно и никакого выбора уже не требуется, методы фонетического ввода предполагают обязательную манипуляцию (нажатие клавиши на клавиатуре или кнопки мыши, или прикосновения к сенсорному экрану) для фиксации иероглифа в фокусе ввода. Другими словами, в процессе ввода после идентификации каждого иероглифа машина обязательно

"спрашивает" пользователя, согласен ли он с выбором, даже тогда, когда этот выбор однозначен и не предполагает никаких иных вариантов.

в) Фонетический ввод невозможен для иероглифов, произношение которых забылось пользователем или неизвестно ему вообще, что бывает часто, например, при поиске неизвестных иероглифов в словаре.

3.3. Структурные методы.

Устранить недостатки фонетического ввода можно попытаться с помощью структурных методов ввода (англ. *shape-based*). Эти методы, поскольку они основаны на анализе графической структуры иероглифа, не связаны с его произношением, следовательно, позволяют сразу устранить основной недостаток фонетических способов, т.е. позволяют вводить иероглифы с неизвестным произношением.

В первую очередь стоит отметить методы ввода иероглифов *по ключам*, которые полностью эмулируют поиск иероглифа в бумажном словаре. Иероглифы во многих бумажных словарях индексируются по ключам, ключи в этом индексе располагаются в порядке увеличения количества черт. Поскольку ключи используются для индексирования иероглифов, они должны представлять собой отдельные, легко идентифицируемые, часто встречающиеся блоки в составе иероглифов. Чаще всего ключи расположены "в начале" иероглифа: слева или сверху (как следствие того, что изначально иероглифы писались сверху вниз). Часть ключей представляет собой простые черты, часть представляет собой самостоятельные иероглифы, часть ключей имеет варианты, иногда заметно отличающиеся по начертанию от основных ключей. В общем случае, варианты отличаются от ключей тем, что они в данном начертании не могут выступать в качестве отдельного иероглифа, а встречаются только в составе других иероглифов, сохраняя свое значение ключа, т.е. иероглифы, в которые входят эти варианты, в индексе находятся в той же группе иероглифов, что и иероглифы с основным ключом в своем составе. Понятно, что набор ключей

привязан к конкретному словарю, и в реальности количество ключей различается от словаря к словарю по воле составителя, и может достигать вместе с общепринятыми их вариантами почти трехсот штук. Классическим считается набор ключей Канси, существующий с 18-го века и составленный из 214 ключей. Нужно иметь в виду, что ключи традиционно имеют смысловую компоненту: каждый из них имеет "имя" и отождествляется с некой группой явлений, действий, предметов. Эта смысловая компонента ключей широко используется в мнемотехнике для запоминания значений конкретных иероглифов, содержащих конкретный ключ. Таким образом, изучение ключей ныне является обязательным элементом изучения иероглифики вообще.

Способы ввода по ключам используют или виртуальную клавиатуру на экране устройства, или даже рукописный ввод. В структуре иероглифа выделяется единственный компонент — его ключ, оставшаяся часть иероглифа характеризуется одним совокупным параметром — количеством черт иероглифа за исключением черт самого ключа. Таким образом, отличные от ключа компоненты иероглифа не имеют значения: ключ есть основной идентификатор иероглифа, остальные не важны. Совпадают или нет они (эти остатки) с другими ключами, их комбинацией или с их частью — не имеет никакого значения. Пользователю необходимо знать ключ для каждого иероглифа и помнить (или просчитать в уме) количество черт в остатке. Количество черт иероглифа, исключая количество черт самого ключа, является вторым обязательным параметром поиска (см., например, патент [14]). На описываемой в этом патенте виртуальной клавиатуре для ввода по ключам, они расположены по возрастанию количества черт. Пользователь изначально видит только группы ключей от 1 до 10-ти черт (последняя группа включает и все остальные ключи, имеющие более 10-ти черт), и должен указать искомую группу, после чего появляется список ключей с указанным количеством черт, из которого тоже нужно сделать выбор. Пользователь выбирает количество дополнительных черт в оставшейся части иероглифа (количество черт иероглифа, исключая черты ключа). После этого вы-

водится список иероглифов, содержащих данный ключ и указанное количество черт в остатке. Пользователь должен выбрать из этого списка искомый иероглиф. Способ приемлем для поиска иероглифов в словаре, но как способ ввода текста не пригоден ввиду большого количества необходимых манипуляций, большого количества иероглифов-кандидатов в каждой группе "ключ-черты" и, таким образом, низкой скорости такого ввода.

Существуют также способы, использующие идею комбинации ключей (см., например, [15]). Этот патент описывает способ и устройство для идентификации символов в идеографическом алфавите, которые позволяют пользователю графически описать символ с помощью набора его компонентов. Способ использует комбинации из 82-х ключей. Источник не указывает, сколько иероглифов можно ввести с помощью таких комбинаций. Пользователь непосредственно на экране с помощью операций перемещения (drag-n-drop) компонентов располагает их в нужной последовательности. Матрица (или "канва"), где собирается иероглиф, разделена на 9 секторов, и машина "сама" учитывает взаимное расположение компонентов, поскольку "знает" в какой сектор оператор поместил конкретный компонент. Следовательно, этот способ для идентификации иероглифа кроме компонентов учитывает также и их взаимное расположение. Но описать все иероглифы с помощью такого небольшого набора компонентов практически невозможно, поэтому степень неопределенности при вводе остается высокой. Действительно, в описании этого патента иероглиф 若 идентифицируется с помощью трех компонентов. Но точно также, с помощью тех же компонентов идентифицируются еще 4 иероглифа. При данном наборе компонентов-ключей, иероглиф 苦 не отличим для машины от иероглифа 若, например. Следовательно, идентификация неоднозначна и для ввода иероглифа требуется дополнительное внимание пользователя и дополнительная манипуляция по выбору нужного иероглифа из пяти предложенных. К достоинствам этого способа можно отнести использование ключей, как естественного и привычного для китайской иероглифики разделения иероглифов на компоненты.

Следующую группу структурных методов ввода для определенности будем называть способами *структурного кодирования*. Эти способы базируются на идее кодирования графической структуры иероглифа с помощью букв латинского алфавита. Суть их заключается в том, что вводятся некие правила декомпозиции иероглифа на "стандартные", наперед заданные компоненты. Каждый из этих компонентов ставится в соответствие определенной букве латинского алфавита и располагается на соответствующей клавише стандартной латинской клавиатуры. Пользователь нажимает последовательно клавиши, т.е. вводит некую последовательность букв, отражающую последовательность компонентов иероглифа, по этому полученному буквенному коду машина идентифицирует иероглифы и выдает список иероглифов-кандидатов, связанных с введенной последовательностью компонентов. Пользователь должен тем или иным способом указать, какой из иероглифов-кандидатов нужно вставить в текст. Из методов структурного кодирования самые известные это уби (wubi) и цанцзе (Cangjie).

Способ *уби* (см. [16]) использует в разных реализациях от 204 до 227 компонентов, почти половина из них графически совпадает с классическими ключами Канси или их вариантами, часть представлена отдельными элементами ключей, другая часть представлена отдельными чертами (всего около 90 "не ключевых" компонентов, т.е. таких, которые графически не похожи на ключи). Все эти компоненты "привязаны" к 25-ти клавишам стандартной латинской клавиатуры, поэтому каждой клавише соответствует от 3-х до 14-ти компонентов, и, соответственно, от 3-х до 14-ти компонентов имеют одинаковый буквенный код. Логика расположения компонентов на клавиатуре, их привязка к конкретным клавишам, имеет сложную структуру и связана не с удобством ввода, а с попыткой облегчить запоминание этой сложной структуры компонентов. Используемые компоненты не равнозначны. Существует 5 "базовых черт", в соответствии с которыми остальные компоненты делятся на 5 групп по первой черте каждого компонента, 25 "базовых иероглифов", которые, при этом, могут

выступать также и в качестве компонентов (и тогда они вводятся по иным правилам, отличным от правил ввода базовых иероглифов), и обычные компоненты. Правила декомпозиции достаточно сложны и зависят от типа расположения компонентов в иероглифе, таких типов расположения четыре. Декомпозиция некоторых иероглифов требует учитывать взаимное расположение компонентов в виде так называемого "кода отличия". Код отличия - комбинация "номера компонента", который определяется принадлежностью компонента к одной из пяти групп базовых черт и типов взаимного расположения компонентов в иероглифе: слева-справа, сверху-снизу и смешанный тип. Поскольку базовых черт 5, кодов отличия всего 15, следовательно, 15 клавиш, кроме ввода кодов компонентов, выполняют еще и функцию ввода кодов отличия, причем этими кодами являются те же латинские буквы, которые кодируют компоненты. В некоторых ситуациях при декомпозиции важным становится еще один параметр — "последняя черта" компонента. Чтобы минимизировать использование кодов отличия, разработчики ввели для 4-х разных типов иероглифов разные правила определения этой "последней черты" иероглифа.

При вводе компонентов в основном повторяется порядок написания черт в иероглифе, но существуют некоторые исключения. Иероглифы неравнозначны по способу ввода, т.е. для разного типа иероглифов существуют разные алгоритмы ввода. Например, для ввода базовых или "заглавных" иероглифов нужно несколько раз нажать соответствующую клавишу (от двух до четырех раз). Для части иероглифов необходимо просто последовательно нажать клавиши, соответствующие компонентам искомого иероглифа. Для некоторых иероглифов необходимо ввести еще и код взаимного расположения указанных компонентов, упомянутый выше код отличия. Для некоторых иероглифов требуется ввести две буквы, а потом дополнить ввод до четырех символов клавишей L.

Все эти алгоритмические ухищрения требуются для того, чтобы для каждого иероглифа получить уникальный буквенный код, позволяющий идентифицировать его, например:

士 — fghg; 十 — fgh; 人 — w; 员 — km; 呗 — kmy; 手 — rtgh; 金 — qqqq ...

и т. д.

Машина "работает" именно с этими кодами, выбирая нужные иероглифы по мере ввода букв (анализ возможностей способа основан на таблице соответствия буквенных кодов иероглифам, опубликованной в сети [17]). При использовании такого кода, ввод иероглифа требует до четырех манипуляций (нажатий на клавиши), и только 636 иероглифов из всех возможных могут быть однозначно идентифицированы после нажатия одной или двух клавиш. В таблицу иероглифов разработчики сочли возможным добавить и многосложные слова, состоящие из двух иероглифов и целые фразы (до 9-ти иероглифов). При этом код не обеспечивает однозначной идентификации иероглифа или слова, и из всего набора 16,75 % приведенных кодов соответствуют двум и более (в реальности до 45-ти) иероглифам или словам, а 44% иероглифов (или слов), однозначно не идентифицируются вообще, т.е. имеют перекрывающиеся с другими иероглифами коды, и после ввода этих кодов требуется дополнительный выбор пользователя.

В способе *цанце* (см. [18]) используется 24 основных и 87 вторичных "знаков" ("signs"), итого 111 знаков. Из них 44 знака графически отличаются от ключей Канси. Основные знаки совпадают с ключами, а из вторичных некоторые графически совпадают с отдельными ключами, некоторые с их вариантами, некоторые с чертами, а небольшая часть — просто графические компоненты иероглифа, не представленные в предыдущих группах. Из 24 основных знаков лишь 13 совпадают с базовыми иероглифами *уби*. Среди вторичных знаков также есть совпадающие с компонентами *уби*, но есть и отличные. Принцип этого способа заключается, как и в способе *уби*, в "привязке" знаков к кодам латинских клавиш. Для указания знаков используется 24 клавиши. На одну клавишу в *цанце* приходится от 2-х до 8-ми знаков. Даже совпадающие с *уби* знаки, естественно, привязаны к другим клавишам, поскольку принципы компоновки знаков на клавиатуре в этих способах разнятся. Правила декомпозиции способа *цанце* ка-

жуются проще, чем уби, но и в этом способе приходится учитывать разные типы иероглифов: однокомпонентные, двух- и трехкомпонентные (single-units, two-units and three-units characters), — декомпозиция которых отличается друг от друга. За счет того, что количество используемых знаков почти в 2 раза меньше, чем в способе уби, и клавиш для кодировки используется на одну меньше, длина полученного буквенного кода для графически сложных иероглифов становится больше. Для ускорения ввода (уменьшения количества манипуляций) разработчики способа цанцзе добавили правила исключения некоторых знаков для уменьшения количества символов в полученном буквенном коде некоторых иероглифов. Несмотря на это, максимальное количество символов при вводе отдельного иероглифа способом цанцзе все же достигает 5-ти, и всего 238 иероглифов однозначно идентифицируются после нажатия одной или двух клавиш. При этом 22,2 % иероглифов однозначно не идентифицируются. Для таких не однозначно идентифицируемых иероглифов, пользователю предлагается на выбор от 2 до 8-ми вариантов. Это, конечно, лучше, чем в способе уби, но все же взаимно однозначного соответствия между кодом и иероглифом нет (таблицы соответствия буквенных кодов и иероглифов см. [19]).

Для обоих способов характерны одни и те же недостатки: сложность правил декомпозиции, тяжеловесность и изощренность организации буквенных кодов для иероглифов, соответственно, и сложность запоминания необходимой последовательности нажатия клавиш для ввода иероглифов. Отсутствие взаимно-однозначного соответствия кодов и иероглифов требует дополнительного внимания и действий пользователя. Ограниченность поля привязки кодов (25 и 24 клавиши) приводит к тому, что число необходимых манипуляций увеличивается, скорость ввода падает. На современных устройствах с небольшим сенсорным экраном привычные преимущества использования клавиатуры по скорости ввода теряются, ибо сам ввод на этих устройствах, как правило, предполагает использование лишь одной руки, тогда как вторая удерживает устройство. При этом недостатки распределения компонентов на небольшом количестве клавиш

сохраняются. Это распределение неравномерно, оно совершенно не ориентировано на удобство ввода этих компонентов ни по их последовательности, ни по частоте использования, что приводит к излишним перемещениям указателя мыши или пальца на сенсорном экране, соответственно, к снижению скорости ввода.

Вообще правила декомпозиции способов структурного кодирования имеют два аспекта. Первый, это аспект разработчика, для которого правила декомпозиции нужны для того, чтобы получить минимальный приемлемый код для каждого иероглифа. Как описано выше, этот аспект не очень хорошо работает ввиду отсутствия взаимно однозначного соответствия кодов и иероглифов в способах и уби, и цанцзе, несмотря на их сложность, на обилие групп и типов, на многочисленные исключения. Вторым аспектом, это аспект пользователя, который должен знать эту сложную систему правил, соглашений и исключений для того, чтобы в точности повторить действия разработчика и прийти к тому же результату. Понятно, что в конечном итоге пользователь, независимо от правил декомпозиции, должен просто запомнить последовательность нажатия клавиш для ввода того или иного иероглифа. Во время ввода текста ему уже некогда заниматься декомпозицией. Но другого способа, кроме попыток так или иначе провести декомпозицию, получить информацию о нужной последовательности нажатия клавиш у пользователя нет, получить помощь от машины он не может. В самом процессе ввода пользователь даже не видит уже введенных им компонентов, а видит лишь набор букв, отражающий последовательность нажатых им клавиш, т.е. ввод происходит вслепую. В обоих способах структурного кодирования вся информация по декомпозиции и сопоставлению компонентов буквенным кодам скрыта от машины, вся она сосредоточена в голове разработчика. Как уже отмечалось, машина знает только соответствие буквенных кодов иероглифам, полная информация о реальном графическом составе компонентов иероглифов после такого кодирования потеряна безвозвратно и не может быть восстановлена. Поэтому-то от машины невозможно ожидать помощи в ситуации-

ях неопределенности: или пользователь ввел правильную последовательность кодов и получил на выходе искомый иероглиф, или он ошибся и получил совершенно неожиданный результат. Поэтому и возможности этих способов в поиске иероглифов по аналогии с известными, по наличию подобных блоков компонентов, просто по выборке всех иероглифов, имеющих некий конкретный компонент в начале или конце иероглифа, крайне ограничены. Искусственность выделения компонентов, отсутствие наглядности как при декомпозиции, так и при привязке компонентов к ограниченному количеству клавиш, затрудняют для пользователя освоение этих способов.

3.4. Композиционный ввод.

Композиционный ввод призван обеспечить простой, наглядный и эффективный ввод наиболее часто употребляемых в современном китайском языке иероглифов. Наиболее часто употребляемыми иероглифами будем считать иероглифы, которые употребляются в современном китайском языке с частотой большей, чем один раз на миллион иероглифов текста. Частота употребления иероглифов рассчитывалась на основе ланкастерского корпуса иероглифов, составленного Тони МакЭнери (Т. McEneary) и Ричардом Сяо (R. Xiao) в 2006-ом году [11]. Под эффективностью ввода будем понимать величину, обратную количеству манипуляций пользователя, необходимых для однозначной идентификации иероглифа машиной. Чем больше требуется манипуляций пользователя, тем менее эффективным будем считать способ ввода. Понятно, что эффективность ввода прямо влияет на теоретический предел скорости ввода, поэтому при прочих равных условиях достижимая скорость ввода будет выше у того способа, количество манипуляций которого меньше.

Структурная иероглифика предоставляет нам первичный материал для машинной обработки: формальное описание структуры иероглифов, — причем полное, без изъятий и пропусков, формализованное до машинных кодов, при этом оставляет для пользователя наглядное, графическое представление этих

50

кодов в виде графов. Пользователь оперирует графическими образами, машина обрабатывает последовательность кодов, описывающих иероглифы.

Идея композиционного ввода заключается в том, что мы имеем возможность на основе таблицы графов (приложение 1) выбирать нужные нам иероглифы из таблицы иероглифов (приложение 3). К примеру, мы можем попросить машину показать нам иероглиф "состоящий из" или "содержащий в своем описании" единственный граф Λ , мы получим уникальный иероглиф Λ . Если мы попросим машину вывести иероглиф, содержащий в своем описании 2 графа Λ , следующих друг за другом, мы получим единственный уникальный иероглиф $\Lambda\Lambda$. Добавив в запрос еще один граф Λ , мы получим иероглиф $\Lambda\Lambda\Lambda$. Поскольку мы имеем описание полной последовательности компонентов каждого иероглифа, мы можем реализовать любой мыслимый запрос к таблице иероглифов. по сути, мы имеем неограниченные возможности выборки иероглифов, и дело заключается лишь в том, как мы организуем алгоритмы выборки, как оптимизируем их для поиска любого иероглифа с минимальными затратами. Тогда задача пользователя заключается в том, чтобы последовательно вводить нужные ему графы. В соответствии с установленным режимом выборки из таблицы иероглифов машина выдаст ему тот или иной результат.

3.4.1. Матрица графов.

Для ввода графов метод композиционного ввода в нынешней его реализации (см. [20]) использует специальным образом организованную виртуальную клавиатуру, хотя теоретически ничто не мешает реализовать ввод с помощью отдельной аппаратной клавиатуры. Удобство и скорость ввода напрямую зависит от конфигурации такой клавиатуры. Не сложно реализовать клавиатуру как для десятипальцевого слепого письма, так и для ввода на портативных устройствах одной рукой. Существующие прототипы предусматривают обе эти возможности. Матрица графов реализована в них в двух форм-факторах: 20 x 10 и 10 x 10 (рис. 1, 2 и 3).

На матрице 20 x 10 графы расположены в соответствии с уменьшением частоты их использования от центра к периферии. Поскольку, часто используемые графы сконцентрированы в центре матрицы, при вводе требуются минимальные перемещения позиционирующего устройства (указателя мыши, стилуса или пальца на сенсорном экране). Простые подсчеты показывают, что графы расположенные в 10-ти центральных колонках матрицы 20 x 10 составляют 90% графов маркеров (приведенных к частоте встречаемости иероглифов), а графы 10-ти крайних колонок (по 5 с каждой стороны) составляют оставшиеся 10 % в маркерах иероглифов. Из соображений минимизации занимаемого пространства, эти графы можно спрятать до поры до времени и вызывать лишь по мере необходимости отдельной клавишей, подобно тому, как всем привычная клавиша *Shift* на клавиатуре вызывает прописные буквы. Расположение графов на основной матрице 10 x 10 повторяет центральную часть полномерной матрицы 20 x 10, а вторая часть этой уменьшенной матрицы повторяет расположение 10-ти боковых колонок полномерной матрицы, сведенных к центру. Тогда на второй части матрицы принцип распределения графов по частоте использования сохраняется, сохраняется относительное расположение графов и, главное, при переходе с полномерного устройства на портативное и наоборот сохраняется навык, приобретенный пользователем на другом типе матрицы. Запомнить такого ужасающего на первый взгляд вида матрицу на самом деле достаточно легко, тем более, что существует тренажер (см. [20]), с помощью которого неподготовленный пользователь осваивает расположение графов на матрице за 8 — 10 занятий по 15 — 20 минут.

齿	玄	卯	高	音	耳	生	匕	斤	父	方	冫	示	豕	甘	士	匚	酉	鱼	豸
矛	毛	气	廴	几	冂	冫	卅	米	山	口	厂	丁	羊	*	距	彡	彳	黑	瓜
骨	臣	斗	马	舌	工	冂	丁	厶	心	大	二	彡	冂	广	自	首	𠂔	幺	𠂔
缶	风	身	长	青	儿	禾	丩	丨	手	土	丩	一	王	竹	隹	欠	虫	尢	黄
龙	牙	无	石	金	止	目	讠	人	一	口	勺	月	冫	小	又	干	歹	采	白
飞	谷	皿	矢	斗	贝	八	扌	日	亻	乙	白	丶	又	彳	弓	而	非	革	鬼
鬲	豆	支	彡	户	田	己	戈	木	辶	刀	水	女	灬	至	用	𠂔	雨	亡	辰
髟	韦	疒	面	口	见	子	疋	十	宀	卜	力	寸	立	火	氏	父	川	鸟	香
耒	辛	齐	角	页	玉	里	耂	文	门	夕	レ	夂	尸	走	衣	皮	入	麻	瓦
肉	羽	舟	比	行	卩	爻	艮	𠂔	井	牛	车	巾	犬	冫	穴	聿	片	虎	舛

Рис.1. Полномерная матрица графов виртуальной клавиатуры.

耳	生	匕	斤	父	方	冫	示	豕	甘
冂	冫	卅	米	山	口	厂	丁	羊	*
工	冂	丁	厶	心	大	二	彡	冂	广
儿	禾	丩	丨	手	土	丩	一	王	竹
止	目	讠	人	一	口	勺	月	冫	小
贝	八	扌	日	亻	乙	白	丶	又	彳
田	己	戈	木	辶	刀	水	女	灬	至
见	子	疋	十	宀	卜	力	寸	立	火
玉	里	耂	文	门	夕	レ	夂	尸	走
卩	爻	艮	𠂔	井	牛	车	巾	犬	冫

Рис. 2 Компактная матрица графов виртуальной клавиатуры. Часть 1.

齿	玄	卯	高	音	士	匚	酉	鱼	豸
矛	毛	气	廴	几	距	彡	亅	黑	瓜
骨	臣	斗	马	舌	自	首	𠂔	幺	𠂔
缶	风	身	长	青	隹	欠	虫	尢	黄
龙	牙	无	石	金	乂	干	歹	采	白
飞	谷	皿	矢	斗	弓	而	非	革	鬼
鬲	豆	支	夕	户	用	𠂔	雨	亡	辰
髟	韦	疒	面	口	氏	父	川	鸟	香
耒	辛	齐	角	页	衣	皮	入	麻	瓦
肉	羽	舟	比	行	穴	聿	片	虎	舛

Рис. 3 Компактная матрица графов виртуальной клавиатуры. Часть 2.

Матрица графов является основой системы композиционного ввода и она присутствует во всех прототипах необходимых инструментов. В данный момент эти инструменты реализованы на Java и H2 Database Engine, поэтому может быть запущены на любом устройстве, где установлена Java-машина. Тестирование инструментов проводилось на Windows XP, Windows 7, Windows 8, Windows 10, на Linux (CentOS-6.5) и на Mac OS X. Никаких лицензионных ограничений на загрузку и использование программных продуктов нет. Предлагаемые на сайте разработчика инструменты включают программу для запоминания матрицы графов, программу для тренировки "узнавания" и облегчения запоминания иероглифов как набора графов, и программу для ввода иероглифов [21]. Краткое, но достаточно внятное описание инструментов можно посмотреть там же [22]. Понятно, что способ композиционного ввода предполагает его использование в любой операционной системе в качестве метода ввода, и, когда это рано или поздно будет реализовано, тогда и реализация всех вышеупомянутых инструментов изменится. Сейчас важно обратить внимание лишь на принципиальные возможности реализации всех преимуществ композиционного ввода, которые предоставляет нам структурная иероглифика и на алгоритмы, позволяющие реализовать эти преимущества.

3.4.2. Алгоритмы и режимы выборки.

Взаимно-однозначное соответствие иероглифа и его описания с помощью последовательности графов обеспечивает однозначность идентификации любого иероглифа из данного множества. Последовательно вводя графы того или иного иероглифа, пользователь в результате запроса к данной таблице получит в конечном итоге единственный иероглиф, соответствующий данной последовательности графов. При этом, как мы помним, получить уникальный иероглиф можно с гораздо меньшими затратами. Для этого был разработан отдельный алгоритм идентификации иероглифов по введенным графам. По мере ввода графов этот алгоритм извлекает иероглифы из таблицы описания иероглифов, рассматривая введенные графы в следующем порядке:

- 1) один введенный граф алгоритм считает первым графом общей последовательности графов;
- 2) два введенных графа рассматриваются последовательно как:
 - а) первый и последний графы общей последовательности графов или
 - б) первый и второй графы общей последовательности графов;
- 3) три введенных графа рассматриваются как:
 - а) первый, второй и последний графы общей последовательности графов или
 - б) первый, второй и третий графы общей последовательности графов;
- 4) четыре введенных графа рассматриваются как первый, второй, последний и предпоследний графы общей последовательности графов.

Перебор вариантов именно в такой последовательности позволяет однозначно идентифицировать все иероглифы из рассматриваемого данным способом множества. Алгоритм построен таким образом, что он возвращает список иероглифов, в котором в первой позиции списка находится иероглиф строго соответствующий введенным графам, а далее располагаются иероглифы, подобные первому по составу графов, но для однозначности идентификации

требующие ввода дополнительных графов. Таким образом, будем считать однозначно идентифицированным по данному набору графов иероглиф, возвращаемый в списке иероглифов в первой позиции. В результате анализа работы этого алгоритма было обнаружено, что для однозначной идентификации почти всех иероглифов, состоящих из более чем трех графов, за исключением 42-х шт., достаточно учитывать лишь первый, второй и последний графы полной их последовательности. Эти последовательности для каждого такого иероглифа будем называть *fsl-последовательностями* (от англ. "first, second, last"), соответственно, алгоритм вычленения этих последовательностей будем условно называть *fsl-алгоритмом*. Кроме всего прочего, *fsl* ("первый, второй, последний") является удачной мнемоникой для пользователя, особенно на начальном этапе обучения, позволяющей ввести незнакомый иероглиф, еще не зная его маркера (маркеры будут подробно описаны ниже в контексте режимов выборки).

Значение *fsl-последовательностей* при идентификации иероглифов, как мы теперь знаем, вряд ли случайно. Развитие иероглифики происходило веками, и опыт многих поколений по запоминанию и использованию иероглифов привел к тому, что общая структура иероглифов как сложных графических образов постепенно, подспудно и неосознанно, приводилась в соответствие с естественными законами восприятия и распознавания сложных графических образов: целостный охват образа, разделение его на компоненты, переход снаружи внутрь. С точки зрения естественности восприятия, выделение первого и последнего компонента графического образа представляется совершенно неизбежным. Это подтверждается как раз тем, что есть иероглифы (это 40 иероглифов из упомянутых выше 42-х иероглифов), которые не поддаются идентификации по трем графам, и идентифицируются только по четырем: первому, второму, предпоследнему и последнему, - как бы охватывая периметр иероглифа и переходя внутрь его. Если же различия в структуре двух иероглифов (по сути, двух графических образов) сосредоточены в самом их центре, то "добраться" до них через описание с помощью первого, второго и последнего графов не удастся.

Эти два оставшихся из 42-х иероглифов: 暖 и 暖, — каждый состоящий из 5-ти графов (日 一 手 又 и 日 一 手 又), различаются только по среднему (третьему) графу и идентифицируются они только по первому, второму и третьему графам. Также совершенно естественным при распознавании сложных графических образов представляется выделение и фиксация неких "меток" (уже известных нам маркеров), характерных для данных образов, и позволяющих их идентифицировать, не обращая внимания на остальные, незначительные с этой точки зрения детали. Было замечено также, что выборка иероглифов по fsl-алгоритму зачастую "срабатывает" раньше, после ввода только двух графов: первого и второго или первого и последнего, - т.е. и здесь присутствует информационная избыточность, которая может быть устранена. Из этих соображений был проведен дополнительный анализ fsl-последовательностей. На основе "раннего срабатывания" fsl-алгоритма для каждого иероглифа, после обязательной проверки на уникальность, и была выявлена минимальная последовательность графов, которая сохраняет взаимно-однозначное соответствие с соответствующим иероглифом, его маркер. Поскольку, в большинстве случаев fsl-алгоритм срабатывает максимально на трех графах, размер маркеров также не превышает трех графов. Для вышеупомянутых иероглифов, которые fsl-алгоритмом идентифицируются только по четырем графам, также удалось сформировать маркеры из трех графов. Таким образом, для всех иероглифов рассматриваемого множества были сформированы маркеры размером не более трех графов, что позволило организовать ввод иероглифов данного множества с помощью не более чем трех манипуляций. Стоит обратить внимание на то, что маркеры всех иероглифов в первой позиции всегда включают *первый* граф общей последовательности графов. Вторым графом маркера это или *второй* или *последний* граф полной последовательности. Третьим графом маркера это почти всегда *последний* граф полной последовательности, и только у нескольких иероглифов третий граф маркера есть третий граф полной последовательности таковых. Понятно, что эта инвариантность позиции графов маркера отражает естественные закономерности

восприятия графических образов и облегчает запоминание маркеров для каждого иероглифа.

Композиционный ввод при использовании маркеров позволяет однозначно идентифицировать уже после двух манипуляций 72,5% используемых иероглифов (3146 иероглифов), что несравнимо больше, чем уби (651 иероглифов) и цанцзе (238 иероглифов). Максимально необходимое количество манипуляций для идентификации одного иероглифа при настоящем способе ввода равно 3, тогда как в уби — 4, в цанцзе — 5. Также стоит отметить, что описанные способы структурного кодирования после идентификации иероглифа требуют вмешательства пользователя для самого акта ввода, т.е. для переноса иероглифа в фокус ввода. Композиционный ввод в режиме маркеров позволяет машине отслеживать *признак завершения ввода* графов и без участия пользователя автоматически переносить идентифицированный иероглиф в фокус ввода (подробности см. далее). Пользователь, таким образом, "экономит" одну манипуляцию, ему нет нужды указывать машине иероглиф, который должен быть перенесен в фокус ввода. Эта ситуация реализуется при вводе 93,2 % всех вводимых этим способом иероглифов. В оставшихся 6,8 % случаев, когда дальнейший ввод графов в принципе еще возможен (т.е. признак завершения ввода не обнаружен), но требуемый пользователю иероглиф уже идентифицирован машиной, пользователь может отдельной манипуляцией отправить его в фокус ввода.

При использовании композиционного ввода, выбор иероглифа, соответствующего введенным пользователем графам, осуществляется по таблице описания иероглифов в соответствии с предустановленным пользователем режимом выборки. Поскольку в таблице описания иероглифов хранится информация о полной последовательности графов для каждого иероглифа, способ может использовать различные режимы выборки в зависимости от задачи или степени подготовки пользователя. Вообще можно реализовать выборку иероглифов по любой мыслимой комбинации графов: по первому и/или последнему, последовательно один за другим, да хоть через один, — по любой

комбинации. Но опыт показывает, что предпочтительные варианты включают в себя нижеследующие режимы выборки:

Свободный режим — режим выборки всех иероглифов, в которых введенные пользователем графы встречаются в полной последовательности графов в любом порядке. Режим может быть использован для ввода неизвестного пользователю иероглифа, например, с бумажного носителя, для поиска его в словаре или для того, чтобы получить справочную информацию по составу графов или маркера нужного ему иероглифа (см. ниже). Определив несколько графов искомого иероглифа, пользователь вводит их в произвольном порядке, и в возвращаемом списке иероглифов визуально находит нужный.

Последовательный режим — режим выборки иероглифов при котором введенные пользователем графы встречаются в последовательности графов иероглифов по порядку один за другим и не обязательно с начала иероглифа — полезный режим для поиска "по аналогии" иероглифов, имеющих общие блоки графов, которые также могут выступать и в качестве отдельных иероглифов, что, кроме прочего, позволяет отслеживать вхождение простых иероглифов в состав более сложных. Выделение сходных компонентов иероглифов — полезная процедура, она облегчает запоминание иероглифов и широко используется традицией в обучении иероглифике. Теперь это можно делать "на ходу", без специальной предварительной подготовки учебного материала.

Fsl-режим — режим выборки иероглифов, когда введенные пользователем 3 графа система интерпретирует как первый, второй и последний графы искомого иероглифа. Весьма полезный способ для быстрого ввода, если пользователь на начальном этапе обучения еще не помнит маркеры. Этот режим позволяет идентифицировать большинство иероглифов (69 % или 2987 шт.) после ввода не более чем двух графов и почти все (за исключением 40 шт.) после трех. Для ввода оставшихся 40 иероглифов достаточно к трем введенным уже графам добавить предпоследний граф полной последовательности (см. выше описание fsl-алгоритма). А упомянутые выше 2 иероглифа (暖 и 暖) в этом режиме вво-

дятся по первому, второму и третьему своим графам.

По маркерам — режим выборки иероглифов, когда введенные пользователем графы система интерпретирует как графы маркера и возвращает пользователю иероглифы, содержащие введенные им графы в указанном порядке в маркерах таблицы описания иероглифов. Это самый быстрый режим ввода. Как указывалось выше, 80% иероглифов среднестатистического текста на современном китайском языке в этом режиме идентифицируется уже после двух манипуляций.

Ввод по маркерам представляет пользователю еще несколько преимуществ. При вводе по маркерам можно отслеживать уже упомянутый выше признак завершения ввода. Напомню, что во всех способах ввода иероглифов сама процедура ввода разделена на 2 стадии.

1. Процесс идентификации требуемого иероглифа по тем или иным параметрам, введенным пользователем, и представление его пользователю для ввода, а в случае неопределенности, составление списка иероглифов-кандидатов и предоставление его пользователю для выбора необходимого ему иероглифа.

2. Непосредственный ввод искомого иероглифа-кандидата в фокус ввода программы, ожидающей ввода от клавиатуры (он выполняется пользователем "вручную", отдельной манипуляцией).

При вводе по маркерам в ответ на ввод пользователем первого графа, система возвращает список иероглифов, в котором на первом месте стоит иероглиф, маркер которого или состоит из единственного введенного графа, или начинается с введенного графа. Если возвращаемый список содержит единственный иероглиф, значит ввод графов далее не нужен. В противном случае, после ввода второго графа, единственный иероглиф в возвращаемом списке говорит о том же. После ввода третьего графа, если таковой был возможен, продолжать ввод уже бессмысленно, ибо маркеров размером больше трех графов ни для одного иероглифа нет. Размер возвращаемого списка в 1 иероглиф и есть *признак*

завершения ввода, который может быть автоматически обнаружен и использован в алгоритме ввода. После обнаружения признака завершения ввода вторая стадия традиционной процедуры с участием пользователя теряет смысл, без нее можно обойтись, поскольку полученный иероглиф может быть отправлен непосредственно в фокус ввода программы, ожидающей ввода от клавиатуры, автоматически (т.е. в ответ на обнаружение признака завершения ввода). В процессе ввода по маркерам, ввод 93,2 % иероглифов заканчивается по обнаружению признака завершения ввода. Другими словами, в 93,2 % случаев способ композиционного ввода увеличивает скорость ввода за счет устранения одной манипуляции.

Второе преимущество ввода по маркерам заключается в том, что в этом режиме, после ввода очередного графа, с помощью таблицы описания иероглифов очень просто вычислить набор возможных (или "разрешенных") для последующего ввода графов маркера и, таким образом исключить те комбинации графов, которые не встречаются в составе маркеров и в результате возвращают пустой список иероглифов. Для этого достаточно в таблице описания иероглифов выбрать те графы маркера, которые встречаются в следующей позиции маркера после введенного для иероглифов из возвращенного системой на предыдущем этапе списка иероглифов. Тогда, остальные графы, не попавшие в результирующий список, на матрице графов могут быть заблокированы или переведены в неактивный режим. Этот режим блокировки клавиш не только позволяет избавиться от некоторого количества механических ошибок, но также и облегчает освоение способа, предоставляя пользователю визуальную подсказку возможных вариантов непосредственно во время ввода.

И, наконец, третье преимущество заключается в том, что использование композиционного ввода на ранних этапах обучения способствует лучшему запоминанию иероглифов. Если способы структурного кодирования с их сложными правилами декомпозиции могут применяться лишь после того, как иероглиф уже "известен", т.е. легко отличается от других и структура его уже

закреплена в памяти, то начинать писать иероглифы с помощью композиционного ввода можно еще на первых этапах обучения, что как раз способствует пониманию и запоминанию структуры самого иероглифа. Освоение ввода новых иероглифов предельно просто, система ввода предполагает возможность простым нажатием кнопки, "на ходу", не покидая пределы приложения узнать состав любого иероглифа, включенного в систему. Достаточно с помощью удобного режима выборки визуально найти нужный иероглиф и спросить систему о составе его компонентов и маркера.

4. Структурная иероглифика. Проблемы и перспективы.

Структурная иероглифика делает одну простую и очевидную вещь. Она смотрит на иероглиф со стороны его структуры, независимо от его звучания, смысла, независимо от первичного состава его черт, смотрит на него как на композитный образ, собранный как головоломка из нескольких более простых единиц. При этом, не будучи отягощена технологическими ограничениями традиционного способа ввода, поиска и запоминания иероглифов, она вдруг обнаруживает, что иероглиф можно не только "разделить" на компоненты, но и "собрать" его из компонентов можно столь же быстро. Пока непонятно, это "божий дар" или "происки сатаны", но композиционный ввод оказался много эффективнее фонетического, т.е. быстрее, короче по количеству манипуляций, необходимых для ввода отдельного иероглифа. И уже понятно, что использование композиционного ввода несколько смещает акценты методики обучения. И самое главное, что он смещает их в сторону упрощения обучения иероглифики, ускорения обретения навыков чтения и письма, т.е. снижает порог вхождения в область китайского языка, который традиционно кажется непомерно высоким.

"Отстраненность" символа от звука, возможность и необходимость формировать непосредственно графико-семантические ассоциации, минуя привычную для культур с алфавитным письмом связку графико-звуковых и звуко-семантических ассоциаций, позволяет быстрее выработать привычку "немого", беззвучного скорочтения. Конечно, китайские дети, когда учатся читать, сначала проговаривают иероглифы. Но они уже знают, что *bāo* это и "обёртывать", и "обдирать", и учитель им скажет, что в тексте первое будет выглядеть как 包, а второе как 剥. А 和 в тексте может звучать как *hé* и *hè*, и *huó*, и *huò*, и *hú*. И если в тексте нужно перечислить несколько предметов, то нужно написать 和, а не 河, хотя и то, и то звучит одинаково (*hé*). У ребенка просто нет иного выхода, как в тексте обращать внимание на образ и абстрагироваться от звука, который можно, конечно, произнести, но сам по себе звук ничего не даст, от него легко и

отказаться.

Вообще привычка носителей алфавитного письма читать "про себя" исторически достаточно позднее приобретение. Первые тексты писались для их произношения, для их проговаривания вслух. Сегодня нам кажется, что консонантное письмо вообще невозможно ни прочитать, ни понять без огласовки. Отрыв смысла от звука происходил (и происходит) очень медленно. В самом конце 4-го века нашей эры Блаженный Августин описывал способность Амвросия Медиоланского читать "про себя" как редкую и удивительную [23]. К сожалению, мне не известны подобные свидетельства для культур с иероглифической письменностью. Появление в текстах в раннем средневековье пробелов между словами, абзацев, оглавлений, аннотаций по сути произвело "тихую революцию", научив людей в массе "немому чтению" [24, стр. 384]. Смысл стал "отрываться" от звука, передача смысла перестала требовать воспроизведения звука. При этом стоит отметить, что при письме этот отрыв смысла от звука не очевиден, если есть вообще. Во время письма мы, даже если и не проговариваем текст голосом, то "простукиваем" его пальчиками по клавишам, "по буквам". То же простукивание по буквам китайских текстов приводит к необходимости удваивать внимание, описать без ошибок поток звуков, конвертировать их без ошибок в поток графических образов, отдельных от этих звуков. Это привело к появлению странных и очень редких для традиционной иероглифики ошибок, когда в тексте появляется бессмысленный с точки зрения контекста иероглиф, который звучит так же как и правильный, но выглядит иначе. См., например, характерную опечатку, когда в слове tiānyú (天涯) вместо иероглифа 涯 появляется давно устаревший 涯 [13, стр. 68]. Или появление иероглифа 块 в слове fēikuài (飞快) [там же, стр. 37]. А стоит ошибиться при указании тона, как введенный иероглиф, ну, очень похожий на искомый, вообще несет другой смысл и весло (桨) неожиданно превращается в крахмал (浆) [там же, стр. 84].

Кроме акцента на ошибки можно ведь ещё и сознательно "изменять" иероглифы как графические образы, перемещая или вообще удаляя их элементы,

чтобы "сделать" одни иероглиф из другого — весьма частый сюжет детских иероглифических загадок (см. [25]). Например, 抽水泵. Загадки такого типа не решаются с помощью фонетики или анализа значений иероглифов, но легко решаются манипуляциями с образами. Если из помпы откачать (удалить) воду, останется только камень. Почему? 抽水 — откачивать воду; 泵 — помпа, водяной насос. 泵 = 石 (камень) + 水 (вода). Или 大人走了 — большой человек ушёл. Ничего удивительного. Но если иероглиф "человек" ушёл из иероглифа "большой", то остался искомым ответ: иероглиф "один" (大 = 人 + 一). Такого рода манипуляции с образами представляются для китайцев обычным действием. А если к ним добавить иероглифы и их комбинации, используемые пользователями в интернет-пространстве для выражения своих эмоций, отношения и т.п., "то такие «иероглифические» смайлики, включенные в текст, написанный иероглифами же, побуждают читающего все время переключаться между вычлениением из иероглифов их смысла и распознаванием какого-либо образа, переданного ими же" [там же, стр 14]. Значит, это тоже обычные действия и привычка в иероглифе "слышать" не только смысл, но и "видеть" образ этого смысла. Насколько этот привычный "образ смысла" упрощает (или усложняет) его восприятие в тексте?

Конечно, в жизни не все так просто, как в книжках, включая и эту. И трижды прав В.М. Алексеев, когда утверждает, что "иероглиф в китайском иероглифическом языке состоит из трех фазисов, всегда сосуществующих в нераздельности: а) индекс идеограф + b) семантический смысловряд и колорит + с) место среди других" [26, стр. 30]. Только к среднему члену этой формулы необходимо добавить, что "семантический смысловряд и колорит" этот выражены звуковыми средствами китайского языка. И хотя "принцип независимой от звука "глухонемой" письменности всегда считался аксиомой" [там же, стр. 30], иероглиф идет от звука, от слова. Что ни говорите, но "сначала было слово". По меткому замечанию В.М. Алексеева, "на иероглифике никто не говорит", а уж как потом фиксируется это слово в иероглифическом тексте, простите, как полу-

чилось, в меру способностей пишущего, его грамотности, памяти и способности к воспроизведению образов. Отсюда *mómohūhū* и 模模糊糊, и 模模糊糊 и 馍馍糊糊, и всё *mómohūhū*; а *kēdǒu* и 蝌蚪, и 蝌斗 вплоть до 科斗, которое тоже *kēdǒu*. Второй пример, кстати, – отличная демонстрация последовательной утери "никчёмного" детерминатива. Можно долго и успешно бороться за упразднение разнописей иероглифов. Но проблема не в иероглифах, а в специфике передачи устной традиции. Вспомните русские поговорки: "растекаться *мысью* по древу" или "растекаться *мыслью* по древу", "не *мытьём*, так катаньем" или "не *нытьём*, так катаньем". С иероглифами, кажется, проще. В результате работы комиссии по реформе китайской письменности в середине пятидесятых лет прошлого века "из этих 3225 разнописей иероглифов было упразднено 1821 ненужное начертание иероглифов. Примерно тысяча с лишним из них относится к числу 6 – 7 тысяч частоупотребляемых иероглифов." [27, стр. 30] Тем не менее, с написанными уже текстами как-то общаться придется ещё долго.

Нет никаких естественных причин и правил обращения ни с образами, ни со звуками, заставляющими данный образ звучать именно так, и данный звук фиксировать на письме именно таким образом. Фонетика иероглифов изначально произвольна, как произвольна и форма букв, передающих звуки в алфавитных системах письма. И если представителям культур с алфавитным письмом очень трудно избавиться от звукового сопровождения текста, особенно трудно это для носителей флективных языков, то представителям культур с идеографической передачей текстов проще перейти на немое скорочтение, проще "отключить" графико-звуковые ассоциации, когда в них нет необходимости, и использовать непосредственно графико-семантические. "Когда я *пробегаю* газету, то я, наверное, не читаю по буквам и слогам, а именно *пробегаю* длиннейшие слова, в роде: непредусмотрительность, тщательнейший, сверхъестественный и т. д., и делаю это точно так же, между прочим, как и китаец, который *бежит* по строкам, не задерживаясь на иероглифах" [26, стр. 27]. "Это очень верное наблюдение: по буквам или по иероглифам текст читают лишь малограмотные (что, од-

нако, очень часто встречается среди иностранцев, не вполне овладевших иероглификой), но, разумеется, для такого чтения надо предварительно приложить большие усилия при обучении", — так В.М. Алпатов комментирует текст В.М. Алексеева [28, стр. 280]. Возможно, это одна из причин того, что китайские дети быстрее своих европейских, к примеру, сверстников осваивают чтение. Иероглиф уже изначально есть контейнер для передачи смысла. Слово в языках с алфавитной письменностью есть лишь фиксация потока звуков, носителем смысла является устное, произнесенное слово, звук передает смысл.

Не пытаясь выяснить, что было раньше, яйцо или курица, мы все-таки можем полагать, что простота оперирования образами смысла, простота передачи смысла посредством образов сопутствует образному восприятию окружающего мира. И использование фонетического ввода, возврат к графико-звуковым ассоциациям способен если и не разрушить эту гармонию, то нарушить ее основательно. С появлением структурного, не фонетического ввода, достаточно эффективного для того, чтобы использовать его без потери, а на самом деле с заметным увеличением скорости ввода, появляется возможность сформулировать проблему и поставить задачу по выяснению особенностей взаимодействия способов восприятия и методов реализации передачи смысла. Я боюсь говорить о "ценности" для культуры этих разных по сути способов передачи смыслов, об их относительной простоте или сложности, я не знаю, "хорошо" это или "плохо", но на примере структурной иероглифики мы можем увидеть, насколько естественным для человеческого сознания было возникновение такого "странного" (с нашей точки зрения, с точки зрения "алфавитной культуры") метода передачи смысла. Нам кажется, что "зафиксированный" звук способен обеспечить взаимодействие индивидов столь же эффективно как и звуковой сигнал обеспечивает взаимодействие членов стаи во время охоты, например, быстро и надежно. Но после охоты, за пределами стаи, в другом контексте, когда звука уже не слышно, когда, нужно общение с другим "уровнем" бытия (с "небом" или "богом"), или с другим временем (далекими предками или отдален-

ными потомками), оказывается, фиксация образа столь же эффективна как и фиксация звука. Боюсь, как бы даже не более эффективна, если судить по временному отрезку в истории использования именно такого метода фиксации и передачи смыслов. Можно также усмотреть и некое сопутствие (не буду произносить слово "корреляцию", ибо корреляция есть свойства статистики, корреляция считается и выражается числовым языком) и соответствие иероглифики и традиционности как свойства культуры. Уже на первый взгляд, не знаю, верный или не верный, кажется, что использование иероглифов и повышенное внимание к сохранению традиции созданы друг для друга. И опять, использование другого (не лучшего или худшего, просто *иного*) принципа передачи смысла нарушает это соответствие, "надрывает" его. И нынешний китайский подросток, во всю пользующийся фонетическим вводом на своем смартфоне, в ответ на указание ошибочного иероглифа в его тексте, спокойно заявляет: "Какая разница, звучит ведь правильно." Звучит правильно, а пишется не правильно. Ибо при такой степени омофонии как в китайском языке, развести два одинаково звучащих слова (или слога) иначе, чем изобразить их иначе, просто невозможно. И здесь никакой контекст не поможет.

Композиционный ввод формирует другой подход, другой взгляд на иероглиф, иное чем при фонетическом вводе отношение к нему, он не превращает иероглиф в последовательность звуков, он сохраняет целостность графического образа в памяти и обеспечивает чистоту графико-семантической ассоциации. А эта чистота позволяет надеяться, что наряду с "немым чтением" возможно и "немое письмо". Мало того, что такое письмо будет более быстрым, это техника, это мало интересно, но это письмо может иметь совершенно другой результат, не связанный с потоком звуков, не выражающийся звуками живого языка. Одно только упоминание каллиграфии как отдельной области культуры может увести нас слишком далеко в сторону. Но вполне понятные аналогии можно найти и за пределами каллиграфии. Свойства любого числа не имеют никакого отношения к его имени: 12 — "двенадцать", "twelve", "shíèr"; 3 —

"три", "three", "sān". Никогда мы не получим из набора этих звуков числа 9, которое легко получить простым арифметическим действием: $12 - 3 = 9$. Символика формальных систем по сути своей является скорее иероглификой, чем алфавитом, поскольку в формальных системах смысл оторван от звука настолько основательно, что никому не приходит в голову записывать арифметические действия словами: "один прибавить два".

Структурный подход к иероглифике позволил обнаружить некие критические точки в составе иероглифов, которые определяют уникальность каждого иероглифа и, таким образом, обеспечивают скорость процесса распознавания. Это маркеры, это те точки, за которые "цепляется глаз". Невозможность для меня постановки корректной серии экспериментов по выяснению относительной важности этих критических точек привела к неизбежному произволу в выборе маркера. Понятно, первый граф каждого иероглифа доминирует безусловно, но как обстоят дела с остальными? 2051 иероглиф, имеющий более чем два графа в полной последовательности, имеет маркеры из 2-х графов. При этом большая часть из этих иероглифов допускают использование в составе маркера как второго так и последнего графа. Особенно это характерно для относительно простых иероглифов из трех компонентов, и значит, описываемых тремя графами. Сразу, навскидку: иероглиф 以 однозначно идентифицируется по последовательности графов 丩 丶 и точно также по последовательности 丩 人. Какую последовательность лучше использовать в качестве маркера? В данном случае, кажется, последовательность 丩 人 лучше отражает специфику восприятия этого образа, и значит, будет легче запоминаться. Но это не всегда так. Еще 2 примера. Для иероглифа 霞 в качестве маркера срабатывает и 雨口, и 雨又, а для иероглифа 厥 работает и 厂欠, и 厂丩. Особенно последний пример показателен. Для незнакомого с иероглификой человека визуально комфортной кажется последовательность 厂欠, но человеку знакомому с правилами письма иероглифов проще использовать порядок 厂丩, поскольку он уже привык именно в таком порядке писать его. И таких примеров достаточно много. Сторонники рукопис-

ного ввода сразу обратят внимание, что граф $\dot{\cup}$ не может быть первым в описании последовательности компонентов, потому что в большинстве случаев он пишется последним по порядку. Но я-то вижу его первым, ибо расположен он слева. В любом случае, это требует компетентного обсуждения.

Для любой области науки одним из наиболее важных моментов является вопрос о ее эвристической ценности. С прагматической ценностью структурной иероглифики все понятно, она помогает композиционному вводу достигнуть результатов, недоступных иным системам ввода. В композиционном вводе существуют алгоритмы, позволяющие идентифицировать (и, соответственно, ввести на компьютере) любой наперед заданный иероглиф лишь по трем его компонентам. Следовательно, для ввода любого иероглифа достаточно 3-х манипуляций. Для простоты расчета не будем учитывать манипуляции по переводу выбранного иероглифа в поле ввода — иногда при фонетическом вводе их может понадобиться больше 2-х, а при композитном вводе только одна или, в зависимости от установок программы, может не понадобиться совсем. Также не будем учитывать и всякого рода ухищрения предиктивного ввода, которые являются предметом отдельного обсуждения. Но, как мы уже видели, с помощью композиционного ввода 72-х процентов иероглифов требуется вообще только 2 манипуляции. А если учесть реальную частоту встречаемости иероглифов в текстах, эта цифра увеличивается до 80-ти процентов. При фонетическом же вводе для самого краткого слога требуется 2 манипуляции, а в среднем один иероглиф требует 4,2 манипуляции. Композитный ввод, таким образом, ускоряет ввод иероглифов как минимум в 2 раза.

Но гораздо интереснее может оказаться возможность расширить принципы структурной иероглифики на иные системы иероглифического письма. Почти не вызывает сомнений возможность реализовать принципы структурной иероглифики на множестве китайских иероглифов традиционного начертания. Более того, поскольку традиционные иероглифы не проходили через "игольное ушко" реформы по упрощению, принципы их компоновки ближе к изначальным,

мнемотехника не искажена заменой одних элементов другими в процессе реформы, состав их компонентов, на первый взгляд, кажется логичнее и проще. Могут возникнуть новые обстоятельства при попытке применения принципов структурной иероглифики к множеству японских иероглифов, поскольку они тоже изменялись в деталях, но со своей, японской, спецификой.

Решая сугубо практические задачи по разработке систем ввода, по организации словарей, что само по себе ценно и интересно, структурная иероглифики позволяет увидеть нерешенные вопросы и в других областях. Начиная от физиологии и педагогики, например, которые, исходя из специфики восприятия сложных графических образов ребенком, могут понять, в чем специфика обучения иероглифическому письму в отличие от алфавитного. Равно и наоборот: в чем специфика восприятия простых, но "озвученных" образов-букв ребенком, который до начала обучения чтению видит мир "картинок", имеющих имя, и разделение этого имени на звуки для него отдельная и трудная задача. Почему буква в букваре сначала должна иметь имя: "аз", "буки", "веди", "а", "бэ", "вэ", — а лишь потом обретает звук? Почему ребенок должен сначала считать конкретные предметы, а сам счет для него сначала есть лишь последовательность звуков: "один", "два", "три" и т.д., — который лишь потом обретает образ цифр, по сути являющихся графическим образом некоего абстрактного смысла — иероглифами.

Какое влияние специфика обучения иероглифики, её отличие от обучения алфавитному письму оказывает на индивидуальное развитие ребенка? Какое влияние иероглифическое письмо с его спецификой на уровне индивидуального восприятия оказывает на видение мира в целом, на культуру в целом. Насколько далеко заходят различия алфавитной и иероглифической культур?

Структурная иероглифика не только помогает сформулировать эти вопросы, но и позволяет надеяться, что на них когда-нибудь будут найдены и ответы. Думаю, структурная иероглифика будет хорошим подспорьем людям, решившимся отправиться в поиски этих ответов в увлекательное "путешествие

на восток", и особенно тем, кто, решился сделать это, по выражению Василия Павловича Васильева, процитированному во введении, "без учителя, с помощью одного лексикона".

Приложения.

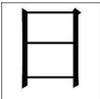
Приложение 1.

Таблица графов.

Для знакомых с любым вариантом иероглифических ключей это описание может показаться избыточным, и можно быстро пробежать его глазами, обращая внимание только на "неключевые" компоненты, а тем, кто с ключами не знаком, будет полезно просмотреть его внимательно. Графы расположены в таблице в порядке убывания частоты встречаемости. Обращаю внимание "традиционалистов", что в первой колонке указан номер по прядку (его можно считать "*частотным рейтингом*"), а не номер ключа в традиционной иероглифике. Третья колонка это действительно только имя графа и ничего больше, поэтому оно достаточно произвольно и не обязано совпадать с названием или значением графически подобного ему ключа. Каждый граф кодирует компонент иероглифа, подобный ему по начертанию, а также компоненты, отмеченные в колонке "варианты". Пустое описание означает, что заметно отличных вариантов начертания компонентов для данного графа нет, естественно, кроме изменения размеров и незначительных изменений пропорций. Все варианты ключей Канси, включенных в набор графов, как правило, отнесены к одному и тому же графу, т.е. варианты таких ключей в составе иероглифов считаются вариантами совпадающих с ними по начертанию графов. Другими словами, если граф совпадает с ключом Канси (или с одним из его вариантов), то все остальные варианты данного ключа кодируются данным графом, и поэтому, как правило, в таблице они сведены в одну строку. Исключение составляют 2 графа: 𠄎 (№ 4), выделенный в отдельный граф вариант ключа 人 (№ 5), и 𠄎 (№ 20), выделенный в отдельный граф вариант ключа 手 (№ 9). Разделение этих двух ключей Канси и использование их вариантов в качестве отдельных графов связано с высокой степенью неопределенности в противном случае. Например, если граф 𠄎

рассматривать как простой вариант графа 手 (как это делает традиция в отношении ключа 手), то для машины иероглифы 右 и 扣 будут неотличимы, ибо последовательность графов обоих иероглифов будет одна и та же: 手 и 口 (первый граф иероглифа 右, как видно из строки 9 таблицы, является вариантом графа 手). То же и в отношении иероглифов 左 и 扛, и в других подобных случаях.

№ п/п	граф	имя	варианты	описание
1.	一	горизонталь		Может быть и слегка наклонной, и заметно наклонной, и вообще кривой, как в последнем примере.
2.	口	рот		Будьте внимательны, чтобы не путать "рот" с "оградой" (№ 50). Они отличаются только размером (единственный случай в иероглифике), но "во рту" ничего не бывает, а в "ограде" можно найти много чего...
3.	乙	второй (циклический знак из 10-ти)		Примем вслед за традицией, что его вариантами следует считать: 乙, 冫, 凵, 凵 and т. п., - не очень наглядно, но "традиционно". Вариантов несколько, будьте внимательны. От простых до изощренных, как в трех последних примерах. Во втором примере сразу 2 варианта "второго".

4.		единица		Она же "человек слева" в традиционном наименовании. Повторяю, это, возможно, важно в мнемонике, но безразлично в комбинаторике графов, как в приведенном примере: в иероглифе "не" трудно нафантазировать себе человека...
5.		человек		Все многообразие (три варианта) "людей" собрано в одной "толпе". Обратите внимание, что "стандартная" откидная вправо без усилия переходит в точку, когда ей это надо, что, собственно, и позволяет объединять их в графе № 10 (с этим каллиграфическим приемом мы не раз еще столкнемся в дальнейшем).
6.		солнце		В примерах два варианта солнца - просто, чтобы глаз привыкал - один предельно вытянут, другой предельно стиснут
7.		охватывать		
8.		белый		

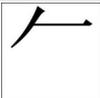
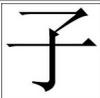
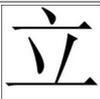
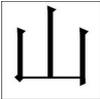
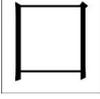
9.	手	рука	我 掰 帮 友	Охватывает все мыслимые варианты классического ключа "рука" за исключением "扌" (№ 20). В комбинации с "копьем" (№ 27) не понятно, кому принадлежит горизонталь. Во втором примере встречается со своим же вариантом, в третьем вытягивает вверх "палец". Ну, и в последнем примере предельно упрощенная "рука", которая у вас постоянно будет путаться с комбинацией "откидная-горизонталь" (№ 12, № 1). Будьте осторожны.
10.	丶	точка	冬 尺	Она же "откидная вправо", т.е. линия идущая слева-сверху вправо-вниз, независимо от ее длины. Три варианта "точек" на примере двух иероглифов в следующей колонке.
11.	土	земля	垆	Отличается от подобного ей "ученого"(№ 102) соотношением горизонтальных черт (у "земли" верхняя короче нижней). В стесненных обстоятельствах слегка "кривит" нижней горизонталью, как, например, рядом с "драконом" (весьма распространенный каллиграфический прием).

12.	丿	откидная	少 夭 囟	Она же "откидная влево". Может "откинуться" чуть-чуть, оставаясь почти вертикалью, может откинуться основательно, как в первом примере, а может почти лечь навзничь, как во втором примере. В третьем варианте редуцирована почти в точку, но всегда сохраняет свое направление сверху-справа влево-вниз.
13.	刀	нож	剪 尔 师 临	В первом примере два "ножа": почти классический, только слегка "стигнут" тяжестью сверху, и над ним сверху "нож сбоку". Во втором иероглифе уже действительно "нож сверху". Далее два варианта "ножа сбоку", но для нас это все "ножи".
14.	木	дерево	枉	Сужается "дерево" по типу "человека" (№ 5), редуцируя "откидную вправо" до "точки" (№ 10).
15.	辶	двигаться	通	Располагается почти всегда с левой стороны иероглифа и подчеркивает его снизу, но иногда "прячется" внутри.
16.	丨	вертикаль		Варьирует только по высоте.
17.	月	луна	那 肯	Кроме стандартного вида есть еще 2 варианта, но узнается всегда.

18.	水	вода	永涵承	У варианта классической "воды" бывает чуть изменен средний крюк. В следующем примере сразу 2 "воды": "вода слева" и часто встречающийся вариант "воды", когда ей тесно. И уникальный случай, когда вода является охватывающим знаком и вмещает в себе 3 горизонтали (пример "пренебрежения" традицией, которая считает, что здесь ключ - "рука").
19.	讠	говорить (сокр.)	言	Полный вариант (встречается значительно реже).
20.	扌	рука (вар.)		Легко узнаваемый знак, вариации только по высоте.
21.	二	два		Кодирует все парные горизонтальные линии.
22.	心	сердце	意蕊性 恭	Слегка сплющивается, располагаясь снизу. Привычный и частый вариант "сердце слева" (пример 3) и достаточно редкий в последнем примере.
23.	丷	рожки	兰业	Располагаются не обязательно сверху, могут быть и ниже, но тогда они "охватывают" какой-нибудь граф. Никогда не встречаются в самом низу.
24.	厶	частный		
25.	女	женщина		
26.	冫	верхушка		

27.	戈	копье	弋	Редуцированный вариант копья, в традиции известен как ключ "охотиться", при декомпозиции часто обозначается как 戈*.
28.	大	большой	因爽拳	Несколько вариантов "большого" вполне узнаваемы и обусловлены только изменениями по высоте или ширине знака.
29.	宀	крыша		
30.	卜	гадание	卩	Оба варианта "гадания" в одном иероглифе.
31.	十	десять		
32.	力	сила		
33.	丅	крюк		
34.	寸	цунь		
35.	纟	нить (сокр.)	纆 紊	Иногда встречается варианты "нити", наверное, проскользнувшие сквозь реформу по упрощению.
36.	疋	нога	是	В некоторых случаях теряет крючок на верхней своей горизонтали.
37.	八	ножки	公 共 兆 亦	Встречаются трех типов, пропускают внутрь себя "посторонние предметы" или способны охватывать другие графы как в двух последних примерах.
38.	又	ладонь	对 祭	Два варианта, второй встречается только в производных от приведенного иероглифа.

39.	卩	бэ́та		Объединяет традиционные "город" и "бугор", ибо по начертанию они не отличимы, а слева или справа он расположен, определяется последовательностью графов в описании иероглифа, посему, разделение на "город" и "бугор" кажется избыточным. Вариантов нет.
40.	目	глаз	且 睪	Часто встречается вариант "глаза" с удлинённой нижней горизонталью. В обоих приведённых примерах требует "особого" описания с помощью дополнительного "пустого" знака - * (см. приложение 2).
41.	王	ван	玨	Прогнозируемый вариант - вполне ожидаемое изменение нижней горизонтали ("два в одном" в примере)
42.	禾	сноп	秝	Вариант образован по тому же типу, что и у "дерева" (№ 14), и "человека" (№ 5) - редукция откидной вправо.
43.	冂	замкнутое пространство	响 冒 澳 鯛	Один из "охватывающих" графов, внутри которого никогда не бывает пусто, за исключением одного иероглифа в комбинации с "оградой" (№ 50). Замечено четыре варианта.
44.	丩	хохол	肖 学	Традиция считает его вариантом "мелкого" (№ 61), но, кажется, проще считать его самостоятельным графом, имеющим два варианта.
45.	己	сам	巳 巳 改	Из трёх вариантов средний "опасен" сходством с вариантом "сидящего человека" (№ 88). См. упоминание в приложении 2.

46.		козырек		В традиционном наборе такого нет, но иногда он сильно упрощает декомпозицию.
47.		сын		Слегка измененная горизонталь образует вариант начертания этого графа
48.		стоять		Располагаясь над "квадратным" (№ 73) и "платком" (№ 84), видимо, для большей устойчивости, опирается не на горизонталь (№ 1), а на "крышку" (№ 94)
49.		гора		
50.		ограда		Сравните с № 2 ("рот"). Если внутри объекта видно нечто, значит это не "рот", а "ограда" (как курьез первый пример - "рот" в "ограде") "Охватывающий" граф. Внутри бывает пусто, но редко (пример 2 интересен комбинацией двух охватывающих графов "замкнутое пространство" и "ограда").
51.		дитё		Стоит обратить внимание на третий пример, в котором этот граф позволяет расположить внутри себя другие графы.
52.		ночь		Иногда путается с "луной" (№ 17), но похожа скорее на "мясо слева" (№ 128)
53.		ворота		
54.		рис		Вариант риса вы уже узнаете и без моей подсказки

55.	ㄥ	крюк вправо	东 乐 甚	В традиционном наборе "ключей" отсутствует. Кроме прямого соответствия картинке кодирует также любой левый нижний угол вплоть до прямого.
56.	厂	обрыв	反	Бывает и с наклонным верхом.
57.	攴	ударять (справа)	攴	Будьте внимательны, имеет тенденцию прикидываться "многоножкой" (№ 63). Включает и редкий ныне вариант - старый ключ "ударять".
58.	文	знак	刘	
59.	丁	тавр	叮 可	Меняет только длину верхней горизонтали.
60.	艸	трава (сверху)		
61.	小	мелкий	尖	Встречается только слегка укороченный вариант.
62.	老	старость (сокр.)	老	Встречается и в полном традиционном начертании как самостоятельный иероглиф как в примере, а также как компонент иероглифов 姥 и 佬.
63.	夂	много ног	处 囟 夜 修	Варианты весьма похожи и узнаваемы. Один из них даже допускает включение внутрь "точки".
64.		две вертикали		В традиции его нет, но очень облегчает жизнь, кодируя парные вертикальные черты.
65.	匕	черпалка		

66.	羊	баран	羔差	Укороченный снизу вариант "барана" теряет "ногу" (или "хвост"?), а иногда просто слегка изгибается.
67.	里	ли		
68.	尸	труп		Не путать с "дверью" (№ 110).
69.	止	останавливать	此	Также вполне предсказуемый вариант.
70.	田	поле		
71.	升	сено		
72.	贝	ракушка		
73.	方	квадратный	万嗷	Иногда теряет верхнюю "точку", располагаясь и самостоятельно, и сливаясь с другими графами.
74.	彳	шаг		
75.	工	работа	功	Предсказуемый вариант.
76.	牛	корова	告牲	Два варианта: "укороченный" и "сжатый"
77.	𦵏	бамбук (сверху)		Традиционный полный "бамбук" кажется, уже и не встречается
78.	火	огонь	燃	Сразу два варианта в одном иероглифе: первый "огонь" просто сжат с боков и легко узнаваем, а второй,.. что ж, привыкайте, что "огонь снизу" тоже "огонь" (очередная жертва в копилку "компактности")
79.	至	достигать		

80.	广	просторное место		
81.	见	видеть		
82.	斤	топор	丘	Опираясь на "горизонталь", меняет свою левую "откидную" на "вертикаль".
83.	示	алтарь	祀	"Алтарь слева" менее похож на свой прототип, чем на вариант "одежды" (№ 117, "одежда слева"). Осторожно!
84.	巾	платок		
85.	冫	холод		похож на вариант воды (№ 18, "вода слева")
86.	西	запад (сверху)	洒	Полный (традиционный) вариант "запада" встречается реже "сокращенного", а в матрице уж очень похож на "вино" (№ 159). Во избежание путаницы в матрицу вынесен сокращенный вариант.
87.	车	телега		
88.	卩	сидящий человек	节范刳	Из трех вариантов второй "опасен" своим сходством с вариантом графа "сам" (№ 45).
89.	良	крепкий	即退	Два стесненных варианта.
90.	犬	собака	独	"Собака слева" встречается значительно чаще своего основного варианта.
91.	生	рождать		

92.		свинья		После комбинации "нож сверху (№ 13)"-"рот (№ 2)" или после "свиной головы (№ 103)" теряет свою верхнюю горизонталь.
93.		рука с палкой		
94.		крышка		"Крыша" (№ 29) без "точки" и без вариантов.
95.	*	пустой граф		Граф-модификатор, так он отображается в описании последовательности графов некоторых иероглифов и в матрице графов — граф, который не имеет собственного образа, но означает некую модификацию последовательности графов или отдельного графа. Смотри приложение 2.
96.		голова		
97.		яшма		
98.		уходить		
99.		ухо		
100.		сладкий		
101.		крестик		
102.		ученый		Китайский "ученый" отличается от "земли" (№ 11) кроме удлиненной первой горизонтали еще и завидной прямоотой, стоит всегда ровно, независимо от размеров или стесненности. Вариантов нет.

103.	𠃉	свиная голова (сокр.)	秉碌缘	Иногда попадает с удлинённым "рылом", прикидываясь "орудием письма" (№ 141), но никогда не компонуется с "вертикалью" (№ 16), "крюком" (№ 33) и "откидной" (№ 12). Иногда, как во втором примере, вытягивает нижнюю горизонталь. Полный вариант встречается нынче редко (пример 3).
104.	金	металл	银淦	Чаще всего употребляется в сокращённом варианте "металл слева". Во втором примере курьезный случай слияния "отца" (№ 131) и "металла".
105.	弓	лук		
106.	隹	бесхвостая птица		
107.	斗	щепка	叫寐藏 鼎	Упрощённый вариант "щепки". Иногда встречается и иной, причём может располагаться и справа (пример 1). Полный традиционный очень редок, при этом слегка варьирует (примеры 2, 3 и 4).
108.	几	столик	微	Уменьшённый столик "сгибает" одну ножку.
109.	青	сине-зелёный		
110.	户	дверь	所	Вариант "старой" двери (остаток традиционного начертания).
111.	用	использовать	甬	Иногда "выпрямляет" откидную.
112.	氏	род	民旅	Верхняя откидная может "маскироваться" другим графом, а горизонталь превратится в откидную.

113.	舌	язык		
114.	自	нос (свой)		
115.	口	окоп	囗	Обычно охватывающий граф сам попал внутрь "ограды" (№ 50).
116.	矢	стрела		
117.	衣	одежда	裒 袁 衤	Неожиданно оказался охватывающим графом. Впускает под шляпку другие графы, а иногда просто теряет ее как во втором примере. Но гораздо чаще встречается сокращенный вариант "одежды" — "одежда слева". Осторожно, этот вариант очень похож на "алтарь слева" (№ 83), добавлена только справа маленькая откидная.
118.	足	ступня (слева)	促	Полный вариант "ступни" достаточно редок.
119.	音	звук		
120.	行	идти	衡	Охватывающий граф. Внутри может располагать много чего. Будьте внимательны при декомпозиции.
121.	穴	пещера	空 罕	Вполне ожидаемая модификация "пещеры" в первом варианте и неожиданная потеря верхней точки во втором...
122.	干	стержень		
123.	石	камень		
124.	而	а, но, да		

125.	长	длинный		
126.	欠	открывать рот		
127.	马	лошадь		
128.	夕	мясо (сле- ва)	肉	Очень похож на "луну" (№ 17) и их часто путают, обратите на это внимание, я специально выбрал шрифт с отличным от "луны" начертанием этого графа. Полная версия встречается значительно реже.
129.	𠂇	когти (свер- ху)	爪	"Когти" в полном варианте встречаются катастрофически редко.
130.	首	макушка		
131.	父	отец		
132.	面	лицо		
133.	𠂇	продвигать- ся		
134.	彡	пух		
135.	角	рог		
136.	川	поток	侃巡𠂇 帶	В первом варианте изгибает последнюю вертикаль. Еще один вариант "потока" (пример 2) и укороченный "вариант варианта" (пример 3). И последняя редукция до состояния трех вертикалей.
137.	皮	кожа (по- кровы)		

138.	高	высокий		
139.	比	сравнивать		
140.	匕	содержать в себе		
141.	聿	орудие письма	尹争妻	Сравните со свиной головой (№ 103).
142.	豸	пресмы- кающееся		
143.	皿	сосуд		
144.	无	без-	僂	Вторая горизонталь обретает крючок слева когда граф повторяется или располагается после "крепкого" (№ 89).
145.	歹	смерть		
146.	非	анти-		
147.	身	тело		
148.	虫	насекомое	聒	Располагаясь над "ракушкой" (№ 72), теряет правую "точку".
149.	支	ветка		
150.	雨	дождь	雪	"Плоский" вариант дождя встречается чаще основного (51:2).
151.	斗	доу		
152.	𦉳	сетка (сокр.)	网	Редкий "полный" вариант сетки. В комбинации с другими графами не замечен.
153.	疒	болезнь		

154.	气	газ		
155.	𠂇	еда (слева)	食	Пребывая в одиночестве, обретает традиционный, "классический" вид.
156.	齐	одинаковый		
157.	入	входить		
158.	卯	четвертый	留卿	Четвертый циклический (из 12). Располагаясь сверху сжимается, как в первом примере. Иногда выступает в качестве охватывающего графа.
159.	酉	вино		
160.	舟	лодка	彤	Был обнаружен только один вариант, когда "лодка" потеряла нижнюю "точку"
161.	片	доска		
162.	牙	зуб		
163.	谷	долина		
164.	采	собирать		
165.	革	сырая кожа (материал)		
166.	风	ветер		
167.	豆	бобы		
168.	亡	ящик		

169.	尢	хромой	𠂔	В этом варианте у "хромого" ноги явно разной длины. Из 18 иероглифов с графом "хромой" такой вариант встречается трижды, когда он "подчеркивает" следующий за ним граф.
170.	臣	слуга	巨	В некоторых случаях утрачивает внутренние вертикальные.
171.	么	крошечный	乡	В варианте "точку" меняет на "откидную". Этот вариант, кстати, встречается в текстах гораздо чаще эталона.
172.	韦	выделанная кожа (материал)		
173.	鸟	хвостатая птица	𪇑	Вариант без внутренней точки. Встречается лишь однажды в качестве самостоятельного иероглифа. В описании требует добавления графа "*" (№ 95).
174.	毛	шерсть		
175.	黑	черный		
176.	辛	горечь		
177.	麻	конопля		
178.	玄	таинственный		
179.	缶	амфора		
180.	鱼	рыба		

181.	羽	перо		
182.	虍	голова тигра		
183.	臼	ступа	輿	Часто выступает в роли "охватывающего" графа, заметно изменяясь только в одном варианте.
184.	龙	дракон		
185.	飞	летать		
186.	鬼	черти		
187.	黄	желтый		
188.	鬲	кувшин		
189.	辰	утро		
190.	骨	кости		
191.	夂	раскоряченные ноги		
192.	髟	волосы	套	традиция утверждает, что это "длинные волосы на голове", но в примере он "волосы" потерял.
193.	香	аромат		
194.	矛	пика		
195.	瓜	дыня		
196.	耒	плуг		

197.	瓦	черепица		
198.	齿	зубы		
199.	肉	след зверя	𠂇	Изменяется только в комбинации с "полем" (№ 70).
200.	舛	ошибка		

Приложение 2.

Коллизии и их разрешения.

Коллизии *первого рода* — неопределенности в последовательности графов. Эти коллизии устраняются добавлением к редкому иероглифу графа-модификатора. Обратите внимание, что добавлен он *в конце* последовательности — это облегчает ввод "исключительных" иероглифов. Для такого рода модификации выбраны, понятно, более редкие иероглифы.

№ п/п	редкий иероглиф	его описание	частый иероглиф	его описание
1.	未	木一*	本	木一
2.	末	木一**	本	木一
3.	杏	木口*	束	木口
4.	鸣	口鸟*	鸣	口鸟
5.	沮	水目*	泪	水目
6.	呗	口贝*	员	口贝
7.	晾	日亼口小*	景	日亼口小
8.	抻	扌丨日*	抽	扌丨日
9.	旧	丨日*	由	丨日
10.	申	丨日**	由	丨日
11.	叭	口八*	只	口八

Коллизии *второго рода* — незначительно отличающиеся иероглифы. Эти

коллизии также устраняются добавкой знака-модификатора к описанию наиболее редкого иероглифа. Такое добавление модификатора часто актуально только для иероглифов состоящих из одного компонента. В иероглифах, состоящих из нескольких компонентов, это добавление модификатора зачастую неактуально и реально не производится, как например, для иероглифа 县, описываемого графами 目 и 厶, эту последовательность графов просто не с чем путать. Иероглифы в строке 3 встречаются с приблизительно равной частотой (~0,15%), поэтому выбор был сделан по сути произвольно, исходя из "логики" картинки (см. строка 4 с еще более редким иероглифом, где используется 2 модификатора).

№ п/п	редкий иероглиф	его описание	частый иероглиф	его описание
1.	日	日*	日	日
2.	目	目*	目	目
3.	己	己*	己	己
4.	巳	己**	己	己
5.	方	方*	方	方
6.	子	子*	子	子
7.	臣	臣*	臣	臣
8.	鸟	鸟*	鸟	鸟
9.	戈	戈*	戈	戈
10.	伐	亻 戈*	伐	亻 戈
11.	沔	水 日*	沔	水 日
12.	裸	衣 田 木*	裸	衣 田 木
13.	晖	日 冫 车*	晖	日 冫 车
14.	邑	口 己 丨*	邑	口 己 丨
15.	赫	土 丿 丨 八*	赫	土 丿 丨 八
16.	宙	宀 丨 日*	宙	宀 丨 日

Таблица иероглифов.

В данном приложении приводится лишь часть таблицы иероглифов, отражающая ее существенные атрибуты. Те поля таблицы, которые, безусловно, важны в работе алгоритмов выборки, предиктивного ввода, блокировки элементов матрицы, но не обязательны для пояснения принципов композиционного ввода опущены. Понятно, что публикация всей таблицы для 4344 иероглифов бессмысленна, здесь приводится только 23 строки из неё в качестве примера. Кому интересен состав всех иероглифов, может обратиться на сайт разработчика [20].

№ п/п	Иероглиф	Последовательность графов	Маркер
1	一	一	一
2	二	二	二
3	三	一二	一二
4	丑	二 一	二一
5	丩	一イ、	一イ、
6	业	ソ一	ソ
7	专	二乙、	二
8	我	手戈	手戈
9	鹅	手戈鸟	手鸟
10	国	口玉	口玉
11	鸟	鸟	鸟
12	乌	鸟*	鸟*
13	世	レ口一	レ口
14	丘	斤一	斤一
15	丙	一人冂	一人
16	东	レ一小	レ小
17	丞	乙水一	乙水
18	丢	丿土厶	丿土
19	员	口贝	口贝
20	呗	口贝*	口贝*
21	喜	十豆口	十豆口
22	暖	日ㄣ一手又	日ㄣ一
23	暧	日ㄣㄣ一手又	日ㄣㄣ

Распределение иероглифов по группам.

Таблица 1. Распределение иероглифов по группам по длине общей последовательности графов (по данным Т. McEnery и R. Xiao [11]. Разница в итоговых суммах происходит от того, что в таблицу иероглифов были добавлены 8 иероглифов, не встречающихся в ланкастерском корпусе).

кол-во графов в полной последовательности	кол-во иероглифов	% в корпусе	накопленные %	кол-во иероглифов в группе	% иероглифов в группе	накопленные % иероглифов в группе
1	145	3,34	3,34	7253559	12,43	12,43
2	963	22,17	25,51	22486596	38,54	50,97
3	1574	36,23	61,74	18256981	31,29	82,25
4	996	22,93	84,67	7683814	13,17	95,42
5	488	11,23	95,90	2076157	3,56	98,98
6	125	2,88	98,78	452481	0,78	99,75
7	40	0,92	99,70	124637	0,21	99,97
8	7	0,16	99,86	16343	0,03	100,00
9	3	0,07	99,93	2159	0,00	100,00
10	1	0,02	99,95	146	0,00	100,00
11	2	0,05	100,00	171	0,00	100,00
итого	4344	100		58353044	100	

Таблица 2. Распределение иероглифов по группам по длине маркеров (по данным Т. McEnery и R. Xiao [11]).

кол-во графов маркера	кол-во иероглифов в группе	% в корпусе	накопленные %	кол-во иероглифов в группе в корпусе	% иероглифов в группе в корпусе	накопленные % иероглифов в группе в корпусе
1	152	3,50	3,50	7317715	12,54	12,54
2	2998	69,01	72,51	41399460	70,95	83,49
3	1194	27,49	100,00	9635869	16,51	100,00
итого	4344	100		58353044	100	

Литература.

1. Васильев В.П., Графическая система китайских иероглифов. Опыт первого китайско-русского словаря. (Репринтное издание, оригинал 1867-го года.) — Санкт-Петербург, 2010.
2. Баранова З.И., Гладков Е.В., Жаворонков В.А., Мудров Б.Г.; Под ред. Мудрова Б.Г., Китайско-русский словарь. — М., Рус. яз., 1988.
3. Сторожук А.Г., Предисловие к переизданию, в кн: Васильев В.П. Графическая система китайских иероглифов. Опыт первого китайско-русского словаря. (Репринтное издание, оригинал 1867-го года.) — Санкт-Петербург, 2010.
4. NJStar Software Company. URL: <http://www.njstar.com/cms>
5. MDBG English to Chinese dictionary. URL: <http://www.mdbg.net/chindict/chindict.php?page=chardict>.
6. Michael Love, личное сообщение, 2015. "Thanks for your email, but we're actually already working on something like this - just refining our component data (in the CHARS tab) first." - 16.09.2015.
7. Ньютон Исаак, Математические начала натуральной философии — Москва, Наука, 1989.
8. Задоев Т.П., Хуан Шуин, Начальный курс китайского языка, Ч.1 — М., Восточная Книга, 2007.
9. The Unicode Standard Version 10.0 - Core Specification, 18.2 Ideographic Description Characters. URL: <http://www.unicode.org/versions/Unicode10.0.0/ch18.pdf>
10. II Core Version 2.2. URL: <http://appsrv.cse.cuhk.edu.hk/~irg/irg/IICore/IICore.htm>
11. McEnergy Tony, Xiao Richard, The Lancaster Corpus of Mandarin Chinese (LCMC), — Lancaster University, 2004. URL: <http://www.lancaster.ac.uk/fass/projects/corpus/LCMC/default.htm> (основной материал находится по адресу URL:

[http://www.lancs.ac.uk/fass/projects/corpus/data/
Chinese_character_frquency_list.zip](http://www.lancs.ac.uk/fass/projects/corpus/data/Chinese_character_frquency_list.zip)).

12. The Unicode Standard, Version 10.0, Kangxi Radicals, Range: 2F00–2FDF.
URL: <http://www.unicode.org/charts/PDF/U2F00.pdf>

13. Китайский язык. 50 текстов для чтения. Начальный и средний уровень.
— М., Восточная книга, 2006.

14. Zhang Jishan, On screen chinese keyboard, Unated States Patent US
6,809,725 B1, 26.10.2004.

15. Lakritz David, Method and apparatus for identifying characters in
ideographic alphabet, Unated States Patent US 5,586,198 A , 17.12.1996.

16. Wubi method: Wikis. URL: http://www.thefullwiki.org/Wubi_method.

17. Wubi86 table. URL: [http://raw.githubusercontent.com/definite/ibus-table-
chinese/master/tables/wubi-jidian/wubi-jidian86.txt](http://raw.githubusercontent.com/definite/ibus-table-chinese/master/tables/wubi-jidian/wubi-jidian86.txt).

18. Overview Of The Cang-jie Method. URL: [http://www.cocoanutstech.com/
cang_jie/www.cjmember.com/the_cj_method.htm](http://www.cocoanutstech.com/cang_jie/www.cjmember.com/the_cj_method.htm)

19. Index: Chinese Cangjie. URL: [https://en.wiktionary.org/wiki/
Index:Chinese_Cangjie](https://en.wiktionary.org/wiki/Index:Chinese_Cangjie).

20. 11-е декабря (эл. альманах), № 5. URL: <http://poutko.ru>

21. 11-е декабря, Загрузки. URL: <http://poutko.ru/China/DownloadsRu.html>

22. 11-е декабря, Краткое описание инструментов. URL: [http://poutko.ru/
China/Archiv/Manual.html](http://poutko.ru/China/Archiv/Manual.html)

23. Августин А., Исповедь. — М.: "Ренессанс", 1991. - Кн.2.III.3.

24. Шкуратов В.А., Историческая психология. Учебное пособие. — М.:
"Смысл", 1997.

25. Рубец, М.В. Визуальное оперирование письменными знаками в китай-
ской культуре: от традиции к кибер-культуре. В сб. Философские проблемы
информационных технологий и киберпространства. – 2018. – Т.15, №2. – С.4–
24.

26. Алексеев В.М., Китайская иероглифическая письменность и её латини-

зация. — Л.: Изд-во Академии наук СССР, 1932 г.

27. Сердюченко Г.П., Китайская письменность и её реформа. — М: Издательство восточной литературы, 1959.

28. В.М. Алпатов, Размышления над книгой В.М. Алексеева "Китайская иероглифическая письменность и ее латинизация", в сб. Общество и государство в Китае: Т. XLIII, ч. 2, —М.: Федеральное государственное бюджетное учреждение науки Институт востоковедения Российской академии наук (ИВ РАН), 2013. — 487 стр.. С. 271-285.

Содержание

1. Структурная иероглифика. Прологомены.....	6
1.1. Элементы иероглифов. Графы.....	6
1.2. Постановка задачи.....	19
1.3. Определения.....	21
1.4. Таблица описания иероглифов.....	23
2. Структурная иероглифика. Область применения.....	25
2.1. Распознавание графических образов.....	25
2.2. Фиксация образа в памяти.....	29
2.3. Фиксация графико-семантических ассоциаций.....	34
2.4. Иероглифические словари и поиск.....	36
3. Методы ввода.....	40
3.2. Фонетический ввод.....	41
3.3. Структурные методы.....	42
3.4. Композиционный ввод.....	50
3.4.1. Матрица графов.....	51
3.4.2. Алгоритмы и режимы выборки.....	55
4. Структурная иероглифика. Проблемы и перспективы.....	63
Приложения.....	73
Таблица графов.....	73
Коллизии и их разрешения.....	93
Таблица иероглифов.....	95
Распределение иероглифов по группам.....	96
Литература.....	97