Аббр и проб ставятся перед аббревиатурой и именем собственным

### Технический момент: кодировка Вокок

Символы Вофала могут быть записаны в собственной кодировке *Вокок/Wokok* (WOfal KOmplex enKoding) или в Юникоде. В последнем случае систему кодирования будем называть *Вун/Wun* (Wofal UNicode), в ней моды и тюнеры отсутствуют, обратная конвертация в иероглифы невозможна, а в нацик – только если чудесным образом известно в какой, и если одно имаго не записывается в том двумя эмбами. Файлы с Вококом пусть имеют расширение .wkk, с Вуном – .wun

В Вококе патроны простираются в диапазоне 0x00-0x4f (из них цифры – 0x00-0x09, буквы – 0x0a-0x2d), моды – 0x80-0xfe (т.е. 0x80 – это и есть мода ноль). Коды записываются в интелловском, а не в моторолловском формате: младший бит числа расположен ближе к младшим адресам памяти компьютера. Разумеется, “c” – не буква, а шестнадцатеричная цифра 12.

Для уменьшения количества машинных операций при обработке текста, вместе в ряды сгруппированы символы, означающие окончание слова (0x4f-0x7e) или влияющие на отображение последующего символа (аббр и проп влияют, но являются частью слова: вообще все символы 0x0a-0x51 – часть слова). Для уменьшения количества операций преобразования чисел из текстового вида в бинарный и обратно, цифрам присвоены равные им кода. Три раскладки клавиатуры выделены голубым, синим и темно-синим.

| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | .a | .b | .c | .d | .e | .f |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0.** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | a | b | c | d | e | f |
| **1.** | g | ɠ | h | ħ | i | ɪ | ɨ | k | ʟ | m | n | o | ɔ | p | r | s |
| **2.** | ş | ṡ | t | u | v | w | ɑ | z | ʒ | ż | ’ | _ | ¯ | hil | | |
| **3.** | | | | ᘓ | ᘐ | ᘑ | | | | | | | | | ↖ | ↗ |
| **4.** | | | | | | | | | | | | | | | | |
| **5.** | abbr | prop | . | , | - | : | ? | ! | / | | | | | | | |
| **6.** | | | | | | | | | | | | | | | | |
| **7.** | | | | | | | | | | | | | | | | |
| **8.** | | | | | | | | | | | | | | | | |
| **9.** | | | | | | | | | | | | | | | | |
| **a.** | | | | | | | | | | | | | | | | |
| **b.** | | | | | | | | | | | | | | | | |
| **c.** | | | | | | | | | | | | | | | | |
| **d.** | | | | | | | | | | | | | | | | |
| **e.** | | | | | | | | | | | | | | | | |
| **f.** | | | | | | | | | | | | | | | | NiG |

---

Долготу последней фонемы слова невозможно обозначить двоеточием после буквы (т.к. в этой позиции оно является пунктационным знаком), а обозначение долготы удвоением буквы мешает заметить одноморфемные слова и бессознательно распознать целое слово как последовательность сгущения и разрящения чернил. Поэтому после буквы ставим подчеркивание

пробела («_»), которое будем называть *дуратором/durator* – чтобы каждый раз не уточнять подчеркивание графемы или пробела. Заметим, в арабском шадда ّ должна конвертироваться раньше огласовки, т.к. первый эмб под ней может быть гласной, а второй – согласной как в сочетаниях وُّ (u_w), يِّ (i_ɥ).

Введем понятие *байтема/byteme* – последовательность байт, представляющая сигил в той или иной кодировке. Состоит из *патрона/patron* – номера сигила; *моды/mode* – номера, указывающего, какому символу нациков соответствует сигил (например, ت и ة имеют байтемы t4 и t19 в Вофале. Разумеется, непрограммисты видят только патроны. Spell-checker-ы, орфографические корректоры подставляют невидимые моды по словарю конкретного языка.

Прямая, а затем обратная конвертация дают первоначальный текст. В хорошо урегулированном языке прямая делается с помощью регулярных выражений, в плохо – является подстановкой словарной транскрипции с заменой нескольких транскрипционных знаков снова на буквы, а именно тех, которые звучат по-разному в однокоренных словах или словоформах: например, нельзя в конце слов в русском, немецком, нидерландском заменять звонкие согласные на глухие; в фарси – смыгчать (палатализировать) «k, g» и т.д. А обратная конвертация в любом случае держится только на модах.

1. Обратная конвертация
    1.1. мода ноль – особая и показывает, что графема должна быть отброшена при обратной конвертации (“t6” из t6s0 конвертируется в «ц», а “s0” отбрасывается)
    1.2. фонеторы при обратной конвертации отбрасываются, поэтому их неуказанные моды равны нулю
    1.3. мима удаляется при обратной конвертации
    1.4. моды, соответствующие графемам разных нациков, должны быть одинаковы, если одинаковы графемы и эмбов, и имаго
    1.5. если файл начинается с моды ноль, то младшие 15 бит следующих двух байт содержат языковой ключ*/language key* – код языка по ISO 639-3, переведенный из 26-ричной системы счисления (например, “jav” для яванского дает 10*26*26+1*26+22=6808); и тогда текст не содержит мод за исключением позиций, где они нужны для разрешения неоднозначности в пределах одного языка (это даже уже, чем в пределах одного нацика); т.о. в каждом языке моды делятся *решающие/solving* и *факультативные/facultative*
2. Некоторые патроны
    2.1. графемой ʁ обозначим фонему [в] (русскую команду лошади остановиться)
    2.2. e – [ɛ]: русское «э», украинское «е»
    2.3. h – [x]; а также [χ] в случае, если в языке есть мфа-шное [ħ], уже обозначенное как «hʔ»: русское “х” [x], чешское и польское “ch” [x] (и устаревшее индонезийское), суахили “kh” [x], арабское خ [χ]
    2.4. ħ – [ɣ]; а также [ʁ] в случае, если в языке есть [ʕ], уже обозначенное как «ħʔ»: суахили “gh” [ɣ], арабское غ [ʁ]
    2.5. ɠ – [ɦ, h]: чешское и суахили “h”, картавое украинское “ґ” [ɦ]; арабское ه [h]
    2.6. g – [g]: общелатинское “g”, русское “ґ”, украинское “ґ ”
    2.7. i – [i]: общелатинское “i”, русское “и”
    2.8. ɪ – [ɨ, ɪ]: турецкое “ı”, русское “ы”, вьетнамское “ư ” [ɨ]; игбо “ị” [ɪ], дари [ɪ]
    2.9. ɥ – [j]: русское “й”, арабское ي (например, «ɥe, ɥa, ɥo, ɥu, ɥi» – это русские “е, я, ё” после твердого, мягкого знаков и гласной, “ю” и украинское “ ї ”: “съел, семья, моя, съёмка, юла” кодируются как «sɥeʟ, sʲemʲɥa, moɥa, sɥomka, ɥuʟa»)

2.10. o – [ɤ, o]: при наличии обоих фонем этой буквой обозначается только первая, а вторая [o] – как oꟹ

2.11. ɔ – [ʌ, ɔ]: при наличии обоих фонем этой буквой обозначается только первая, а вторая [ɔ] – как ɔꟹ

2.12. r – дрожащие русское “р” и арабское ر [r], недрожащие английское [ɹ] и американское [ɻ] и т.д.

2.13. ɭ – общелатинское “l”, русское “л” и т.д.

2.14. z – [z]: русское “з”

2.15. ş – [ʃ]: турецкое ş, сербско-кириллическое и чешское š, сербско-латинянское и русское “ш”, английское sh, немецкое sch, арабское ش

2.16. ʒ – [ʒ]: русское “ж”, сербско-кириллическое и чешское ž, польское ż, английское zh, персидское ژ

2.17. ṡ – [θ]: (должно быть S̰, не стандартизированное Юникодом, ведь различительные возможности неразрывных с буквой нижних видоизменений исчерпаны уже после первого нижнего видоизменения): английское и суахили th, арабское ث , греческое θ

2.18. ż – [ð]: (пока Z̰ не стандартизировано): английское th, суахили dh, арабское ذ , греческое δ

2.19. w – [w, ʍ] общелатинское “w”, беларусское “ў”, арабское و

2.20. апостраф (’) является буквой и обозначает глоттальную смычку, в частности

2.20.1. в него конвертируются арабские хамза ء , алиф-хамза أ и إ , вав-хамза ؤ , алифМаксура-хамза ئ

3. Дифтонги, трифтонги представляем несколькими буквами, только если удается подобрать такую их последовательность, прочтение которой неносителем языка дает носителю возможность узнать дифтонг, в частности

3.1. dz – пуштунское “ځ”, черногорско-кириллическое “з”, черногорско-латинянское “s”, беларусское “дз”

3.2. dʒ – арабское “ج”, английское “j”, тамильское ஜ, сербско-кириллическое “џ” , беларусское “дж”, сербско-латинянское “dž”

3.3. ts – пуштунское “څ” , русское и сербско-кириллическое “ц”, чешское и сербско-латинянское “c”

3.4. tş – [tʃ] персидское “چ” , английское “ch”, немецкое “tsch”, чешское и сербско-латинянское “č”, сербско-кириллическое “ч”

4. Фонеторы

4.1. ꟹ – *лабиатор/labiator*; превращает в лабиализованные

4.1.1. eꟹ, iꟹ – германские, тюркские, угорские “ö” [ø, œ], “ü” [y, ʏ]

4.1.2. iꟹ – игбо “ụ” [ʊ], дари [ʊ]

4.2. ꜞ – *эмфатор/emfator*; в эмфатические с артикуляцией увулярной или фарингальной: d/dꜞ обозначают ض/د , t/tꜞ – ط/ت , z/zꜞ – ظ/ز , s/sꜞ – ص/س ; k/kꜞ [k/q] – ك (ک в персидском, пушту, урду)/ق , к/қ в казахском; h/hꜞ [χ/ мфа-шное ħ, похожее на громкий шепот] – خ/ح , ħ/ħꜞ [ʁ/ʕ] – غ /ع; ᵹ g/ᵹ gꜞ – ng/nG в серер (дрожащие русское [r]/французское [ʁ] обозначаем не r_/rꜞ_, а одной буквой “r”, равно как и недрожащие английское [ɹ]/немецкое [ʁ] не r/rꜞ – тоже одной “r”, ибо в рамках одного языка им не противопоставлены парный по эмфатичности или длительности)

4.3. ꝛ – *pemp/retr*; в ретрофлексные: в пушто d/dꝛ – ډ/د , t/tꝛ – ټ/ت , r/rꝛ – ړ/ر , n/nꝛ – ڼ/ن ; в урду d/dꝛ [d/ɖ ] – ض/د , t/tꝛ [t/ʈ ] – ٹ/ت , r/rꝛ [r/ɽ] – ڑ/ر ; в хауса r/rꝛ

4.4. ≻, ≺ – начальный и конечный *сплитеры/spliter*; участки слова от начального сплитера до начала слова (пробела или мима) и от конечного сплитера до конца слова читаются при определенных условиях: например, в арабском когда перед начальным сплитером стоит алиф-васла (ٱ), оно читается, только если предыдущее слово оканчивается на согласную; во французском когда после конечного сплитера стоит “nt”, оно читается, только если следующее слово начинается с гласной

5. Символ «hil» (hidden letter, скрытая буква), он же «☒» для Вуна

5.1. в тексте не отображается, но при обратной конвертации превращается в некоторый эмб: например, в русском пара из “s/z” и “t” не произносится перед “şʲ”; и если “☒5” – это “s”, “☒6” – “z”, “☒7” – “t”, то слова “расчёска, разносчик, извозчик, насчёт” предстают как «ra☒5☒7şʲoska, razno☒5☒7şʲik, izvo☒6☒7şʲik, na☒5☒7şʲot» (но в русском если между “s/z” и “t” есть пробел как во фразе “поднос чистый”, то пара фонем не исчезает)

## Арабский и арабопишущие языки

Алиф-васла (ٱ) в начале слова читается, когда предыдущее слово оканчивается на согласную, и не читается, когда на гласную, поэтому после него ставится сплитер, т.е. он конвертируется в «a8≻». По той же причине алиф в составе слитного артикля “ٱل” – в «a5≻». А “ـل” артикля может превратиться как в «ʟ», так и в имаго одного из тринадцати других “солнечных” эмбов. В любом случае после такого имаго ставится проп, и весь артикль превращается одну из комбинаций «a5≻ʟ△», «a5≻ş27△», «a5≻ś5△», «a5≻ź15△», «a5≻s14△», «a5≻s14ʔʊ△», «a5≻z7△», «a5≻z7ʔʊ△», «a5≻t25△», «a5≻d17△», «a5≻t25ʔʊ△», «a5≻d17ʔʊ△», «a5≻r7△», «a5≻n19△». Предлоги «посредством» (bı/ بِـ ), «для» (lı/ لِـ ), союз «и» (va/ و ), пишущиеся слитно с артиклем как “بِالـ , لِلـ , و الـ” (второй даже без алифа), конвертируем в «bıʟ△», «lıʟ△», «vaʟ△».

АлифХанджарийя (надстрочный алиф “ٰ”) преобразуем в «a7_», а фатха-алифМаксура-алифХанджарийя “ىَٰ” – в «a6_8». АлифХамза-фатха (أً), вавХамза-фатха (ؤً), алифМаксураХамза-фатха (ئَ) – в «’3a6, ’5a6, ’6a6»; алифХамза-кясра (إٍ) , вавХамза-кясра (ؤٍ ), алифМаксураХамза-кясра (ئِ) – в «’4i7, ’5i7, ’6i7»; алифХамза-дамма ( أُ ), вавХамза-дамма (ؤُ ), алифМаксураХамза-дамма (ئُ) – в «’3u7, ’5u7, ’6u7». АлифМадда (آ) – сокращенное алифХамза-алиф (أا ), но читаемое уже без глоттальной смычки между двумя гласными – в «a9_». Три отмерших падежных окончания, три танвина «-en ً , -ın ٍ , -un ٌ » и суккун ْ не произносятся, поэтому превращаются в «☒». Алиф и алифМаксура не читаются после танвина «-en», т.е. в сочетаниях « اً , ىً », поэтому в их составе конвертируются в «☒».

Лям-алиф (لا) отдельной моды не требует, т.к. в Юникоде, из которого и в который делаем прямую и обратную конвертации, на самом деле закодирован как последовательность двух букв (ل ا), замена которых на одну графему прописана в OpenType-фунциях файла шрифтов .ttf. Такая же ситуация с лям-алифХамза для даммы (لأ), для кясры (لإ) и с лям-алифМадда (لآ): на машинном уровне они – тоже две буквы (ل أ , ل إ , ل آ). Также в .ttf-файле прописан и выбор начертания в зависимости от расположения пробела до, после, с двух сторон буквы, ни с одной – начальное, конечное, изолированное, срединное.

Факультативные моды не указаны в цветных ячейках, но для полноты картины полные байтемы приведены в серых ячейках под ними. Жухло-зеленым отмечены эмбы и моды, которые исчезнут в следующих языках – персидской группы.

| ب | ت | ة | ث | ج | ح | خ | د | ذ | ر | ز | س | ش | ص | ض | ط |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| b | t4 | t19 | ś3 | d9ʐ0 | h2ʔʊ | h4 | d4 | ź3 | r5 | z3 | s3 | ş6 | s4ʔʊ | d5ʔʊ | t5ʔʊ |

| b3 |
|---|

| ظ | ع | غ | ف | ق | ك | ل | ل | ل | ل | ل | ل | ل | ل | ل | ل |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| z4ʔu | ħ3ʔu | ħ2 | f | k5ʔu | k3 | ɭ10 | ş27 | ś5 | ż15 | s14 | z7 | t25 | d17 | r7 | n19 |
| | | | f5 | | | | | | | | | | | | |

| م | ن | ه | و | و | ي | ى | ◌َ | ◌ِ | ◌ُ | آ | ا |
|---|---|---|---|---|---|---|---|---|---|---|---|
| m | n17 | ɠ | w | _5 | ɨ | _9 | a6 | i7 | u7 | a9_ | ☒8 |
| m4 | | ɠ5 | w6 | | ɨ3 | | | | | | |

| ◌ٓ | ◌ٰ | ◌ً | ◌ٍ | ◌ٌ | ◌ْ | ى | ا | ي | ◌ّ | يّ | وّ | ىٰ | ء | أ | إ | ؤ | ئ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a8 | a7 | ☒1 | ☒2 | ☒3 | ☒4 | ☒9 | _3 | _4 | _2 | _6 | _7 | _8 | ʼ2 | ʼ3 | ʼ4 | ʼ5 | ʼ6 |

В иранском, дари фонемы «ɠ» и «m» записываются двумя эмбами (ح, ه; م, ن), последний из четырех только в сочетаниях «mb», «s» – тремя (ث, س, ص), «z» – четырьмя (ذ, ز, ض, ظ). Для обозначения «k» используется неарабская буква, а артикля «aɭ» нет вообще, поэтому моды при «ɭ» факультативны. Также появляются четыре новые согласные «p, tş, ʒ, g». В иранском “غ, ق” будем конвертировать в разные имаго, в «ħ, kʔu»; т.к. этот язык, в котором они не различаются, является упрощенным случаем дари и таджикского, в которых различаются.

Эмб “айн” (ع) в начале слова и между гласными не читается никогда; предписывается в конце слова и перед согласными произносить его как гортанную смычку – но никто не произносит (хотя и несколько удлиняют предшествующий гласный). Поскольку в настоящий момент в иранском, дари, таджикском нет противопоставления фонем по долготе, айн во всех случаях кодируется как «☒» (а не как дуратор «_»). Также не произносится гортанная смычка, обозначаемая хамзой. Начальное и срединное начертание арабской алифМаксура (ى) неопределено, в иранском ее нет вообще, но конечное и изолированное совпадает с иранской «ɨ», поэтому будет говорить, что в иранском пишется алифМаксура (чтобы каждый раз не уточнять «йаАрабское», «йаИранское»).

Уникальные гласные дари [ɪ, ʊ] представим как ı, ıʚ. Соотносимые фонемы иранского [o, ɒ], дари [ɔ, ɒ], таджикского [ɔ, ɵ] кодируем одинаково, и чтобы не применять фонеторы – как [o, ɔ], т.е. как «o, ɔ» (ближе к иранскому). Также иранские [e, æ], дари [ɛ, a], таджикские [e, a] – как [ɛ, a], т.е. как «e, a» (ближе к дари). Остальные две – [i, u], т.е. «i, u» – одинаковы во всех языках. Дифтонги иранские [eɪ, oʊ], дари и таджикские [aɪ, aʊ] кодируем как в последних двух языках, т.е. как «aɪ, aɪʚ». Специфический падеж изафет в иранском маркируется конечной [e], в дари – [a]. Однако хайеХавваз (ه), представляющий не только «ɠ», но и изафет, изредка обозначает [a] и в иранском. Для унификации выберем [a] маркировкой изафета, т.о. хайеХавваз конвертируется в «ɠ» и «a».

Алиф (ا) – не только безгласая подставка (уже не для хамзы, а для огласовок «e, o», т.е. “◌ِ, ◌ُ ”), но и самостоятельная фонема «a» (когда без огласовок), поэтому может конвертироваться как в «☒», так и в «a». АлифМадда (آ) – всегда «ɔ». АлифМаксура (ى) обозначают не только согласную «ɨ», но и «ɪ» в дифтонге «aɪ» (ای), «i» вне дифтонга и в дифтонге в начале слова (тогда алиф перед ней не читается), изредка «ɔ, a» в конце слова. Вав (و) – не только «v», но и «ɪʚ» в дифтонге «aɪʚ» (او), «u» вне дифтонга и в дифтонге в начале слова (опять тогда алиф не читается), изредка «o» в конце слова, и не произносится в диграфе «h☒» (خو), если тот в начале слова. Насыщенным и темно зеленым выделены имаги и моды персидской группы, темным – которые исчезнут в пушту.

| ب | پ | ت | ث | ج | چ | ح | خ | د | ذ | ر | ز | ژ | س | ش | ص | ض | ط |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| b | p | t4 | s11 | d9ʐ0 | t24ʂ0 | ɠ6 | h | d4 | z8 | r5 | z3 | ʐ | s3 | ʂ6 | s4 | z9 | t5 |
| b3 | p2 | | | | | | h4 | | | | | ʐ5 | | | | | |

| ظ | ع | ء | أ | إ | ؤ | ئ | غ | ف | ق | ک | گ | ل | َ | ِ | ُ | آ | ا | ا |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| z4 | ☒10 | ☒12 | ☒13 | ☒14 | ☒15 | ☒16 | ħ2 | f | kʃ5ʊ | k4 | g | ɭ | a6 | e16 | o14 | ɔ | ☒8 | a5 |
| | | | | | | | | f5 | | | g5 | ɭ10 | | | | ɔ2 | | |

| م | ن | ن | ه | ه | ۀ | و | و | و | و | و | ی | ی | ی | ی |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m4 | n | m5 | ɠ5 | a17 | e6ɦ0e0 | v | u6 | o15 | ɪ8 | ☒11 | ɦ4 | a16 | i10 | ɪ7 |
| | n17 | | | | | v7 | | | | | | | | |