Uruwi's personal style guide for documents

uruwi

20 April 2018

```
Branch: canon
Version: 0.1
Date: 2017-10-07
```

Contents

1 | Overview of build process

To generate grammars and other documents, XeLaTeX is used. Documents depend on the `common/uruwi.sty` package, which imports dependent packages such as `xcolor` and `tabu`, as well as defining in-house macros such as `\hli` and `\ortho`.

The build process is automated using `make`, which, in addition to invoking XeLaTeX to build the document, generates `dict.tex` files from dictionary files.

1.1 | Lexicon management

Lexicons are stored in `.dict` files, which are plain text files with some formatting info. A typical file will have entries like this:

```
# ramek
: vn?
break, shatter, tear, destroy
n = what was broken was in the way; non-n = what did the breaking
    sought out things to break

# kekoro
: n
most, majority

# malka
: n
quiet, calm, sound

# rajnek
: v
sleep

# ranu
: n
fox

# kretanek
: v
run

# mepek
: v
learn, teach (about)
learn <A> → <A>-\textsf{kejmá mepek}
```

Evidently, each entry is delimited by one or more blank lines. A line starting with an octothorpe gives the entry in the target language. A line beginning with a colon defines the part of speech.

Other lines provide a definition. Some entries require multiple lines; in that case, the subsequent lines will act as usage notes or examples.

The dictionary file rarely includes "grammar words" and numerals, since these are usually defined in the grammar itself.

The dictionary file is converted to a LaTeXfile with the `dict-to-tex.pl6` script. This script also takes a JSON file that specifies styling information and the lexicographic ordering. For instance, the dictionary file for Lek-Tsaro uses this `options.json` file[1]:

```
{
  "styling": "\\textkardinal{%s}",
  "newline": "\\quad ",
  "alphabet": [
    ".",
    "t", "t^y", "k", "x", "s", "s^h",
    "n", "n^g", "n^y", "v", "f", "f^x", "f^h", "m",
    "d", "d^y", "g", "p", "b",
    "h", "y", "c", "r",
    "z", "j", "w", "l",
    "a", "i", "i^o", "i^u", "o", "o^e", "e", "e^o", "u", "u^i", "u^o"
  ],
  "sectionstyle": "\\section*{\\textkardinal{%s}}\n"
}
```

7/dict/options.json

1.2 | Historical tools

Historically:

- Google Docs was used for prototyping language grammars. However, the grammar of Lek-Tsaro was not prototyped in that manner.
- Google Sheets was used for managing lexicons. This was phased out over concerns of using propietary software.

1.3 | Document styling

1.3.1 | Typefaces

The `mathspec` package is used for custom fonts in both text- and mathmode. Gentium is used for the normal font, and VL PGothic for the sans-serif font. The monospace font is not set.

When hacm text is needed, the `uruwi.sty` package is loaded with the `hacm` option, which sets `\kardinal` to the ᒪιμɲcnιl font, modified to include the backslash character (used by Lek-Tsaro). Frequently, superscripts in text are also needed, so `uruwi.sty` is also loaded with the `shortsuper` option, which redefines `\^` to the longer `\textsuperscript`.

[1] I'm showing the Lek-Tsaro options file instead of the one used by Ðraħýl Rase because the latter uses Unicode characters, which don't display quite properly with `listings.sty`.

If the `dhr` option is set, then `uruwi.sty` will also set `\dhrfont` to the зςξ·з̄ʆ̂ (Mîny / Meko) font, which supports *Nasél Tēkel Piva*. A guide to using this font can be found in table 1.1.

1.3.2 | Pages

The first page in a document is the title page. For a conlang, this consists of:

- A horizontal rule.
- A phrase akin to "(language), the language of (place)" in the target language. If said language has its own script, then this should be written in the script, with a transliteration below it in a smaller, sans-serif font.
- The translation of the above, in italics.
- Another horizontal rule identical to one above.
- The name of the author.
- "A complete grammar", first in the native script (if present), then transliterated, then in English.
- The date, at the very bottom.

All of these entries are centred.

For documents other than language grammars, simply include what is relevant to the document.

The title page should be coloured at 25% of one of the predefined colours in `xcolor`.

For more information, consult the title pages of other documents.

The second page contains an optional dedication, followed by metadata. The metadata is set in a monospace font with the following fields:

- `Branch:` This is `canon` for the main branch, and different for experimental branches of the grammar.
- `Version:` A version, updated occasionally.
- `Date:` The date when this version was adopted.

After the metadata is the copyright information.

The second and subsequent pages should be coloured at 15% of the same shade of colour as the title page (so if the title page is at `Thistle!25`, the body pages should be at `Thistle!15`).

1.3.3 | Colours

The following guidelines are used for selecting page colours:

- A daughter language should have the same or similar page colour as its parent.
- On the other hand, a language from a completely different family should have a visibly distinct page colour.

[2]No space!

- If possible, select a colour that matches the character of the language the document covers.
- Avoid overly dark or light colours.
- Avoid grey or other drab colours.
- Favour cool colours (but don't actively avoid warm ones).

Any coloured content in the title page should blend with the page colour.

However, the chapter and section styles don't need to be changed; in fact, none of the grammars change their styles.

1.3.4 | Semantic styles

Strings in the target language inside English text should be wrapped inside an `\ortho` (or `\hortho` for hacm text), ⟨ħanaħâle-tuɣrí⟩.

1.3.5 | Example sentences

Example sentences include the (transliterated) sentence, the gloss and the idiomatic translation. The target-language sentence and the gloss are not aligned; rather, each word is coloured in both the first and second lines[3]. The words of the English translation are coloured to roughly match the target-language sentence:

ʃɔclʃinʾle ɒe ɟɪɻʃɪɒɪ ɟcl ɟʰɪ.eʾle.
want-NEAR-NEG PR.NEAR.SG ring=GEN POS magician-NULL
I don't want the rings of any magician.

Underscores are preferred over periods in morphemes that take multiple words to describe:
ɒɪlcᵓhɪ .əᵓ–ɟʰcn ɾlɔɻɔ, ɷɔɻɪn sɪɾe nɥɪɻɪ ɻɪ neʃ.
eat-GENERIC-Q PR.GENERIC-how_many flower, CMP-NEAR fish cat PR.ANAPH_SUB.INT
>
More fish eat flowers than cats.

If the language at hand exhibits complex morphological behaviour, then it is best to include an underlying form below the surface form (if different):

\ll/ɟɒɥɒ ɟɷɒ ɒhɾdo \ʃob nɷ ʾlɷ ɾb/ɾɒɥɒ ʾlh h \obʃ
ʾlɷ–>h:ɒɥɒ ɟɷɒ ɒhɾdo >ʃob nɷ ʾlɷ ɾbɷ–>nᵠ:ɒɥɒ ʾlh h >obʃ
speaker-0-Ξ walk again SUB-where see speaker rdkbe$_5$-1-Ξ existence DUMMY DESC-Q-where
We went back to the place where we saw the roses. (said in second season)

ʃɪɪɸʃ ʃɪ ʾˡɻcʾˡɻcɷen, nebel \ɻɜzen.
ʃɪɸɪʃ ʃɪ ʾˡɻcʾˡɻcɷen, nebel \ɻɜzen.
lose-NEAR.PAST PR.NEAR DEF~book, give_to-ANAPH_SUB.PAST Ryze-ACC
I gave the book to Ryze.

[3]This was inspired by Isoraḳatheð's practices, although my adaptation is nowhere as extensive as theirs.

Table 1.1: Guide to using the ʒꞔξ·ʒ̄ẞ font.

NTP	Rom	Seq	NTP	Rom	Seq	NTP	Rom	Seq
ε	p	p	ƅ	t	t	ɟ	k	k
φ	s	s	ɛ̀	f	f	Ƈ	n	n
ɜ	m	m	c	ḣ	x	ƨ	ħ	H
o	h	h	6	r	r	ժ	ṡ	S
ƽ	l	l	Ⱡ	v	v	ρ	g	g
ω	ṅ	N	ꝓ	d	d	∩	b	b
ƾ	ż	Z	ყ	z	z	४	ġ	G
β	ḋ	D	∂	ṫ	T			
ʓ	â	A	ʂ	ê	E	ɕ	î	I
ʖ	ô	O	˥	û	U	ɥ	ŷ	Y
ƅ̂	ta	ta	6̂	ra	ra	ɛ̂	pa	pa
ƅ̄	te	te	6̄	re	re	ɛ̄	pe	pe
ƅ̆	ti	ti	6̆	ri	ri	ɛ̆	pi	pi
ƅ̂,	to	to	6̂,	ro	ro	ɛ̂,	po	po
ƅ̄,	tu	tu	6̄,	ru	ru	ɛ̄,	pu	pu
ƅ̆,	ty	ty	6̆,	ry	ry	ɛ̆,	py	py
ɛ̂̀	fa	fa	ɛ̄̀	fe	fe	ɛ̆̀	fi	fi
ɛ̂̀,	fo	fo	ɛ̄̀,	fu	fu	ɛ̆̀,	fy	fy
ʓ̆	aj	Aj	ʂ̆	ej	Ej			
ʖ̆	oj	Oj	˥̆	uj	Uj	ɥ̆	yj	Yj
ŏʓ	ja	jA	ŏʂ	je	jE			
ŏʖ	jo	jO	ŏ˥	ju	jU	ŏɥ	jy	jY
ʓ̄,	aw	Aw	ʂ̄,	ew	Ew	ɕ̄,	iw	Iw
ʖ̄,	ow	Ow				ɥ̄,	yw	Yw
ōʓ	wa	wA	ōʂ	we	wE	ōɕ	wi	wI
ōʖ	wo	wO				ōɥ	wy	wY
ʓ̆,	aẏ	AW	ʂ̆,	eẏ	EW	ɕ̆,	iẏ	IW
ʖ̆,	oẏ	OW	˥̆,	uẏ	UW			
ọ̆ʓ	ẏa	WA	ọ̆ʂ	ẏe	WE	ọ̆ɕ	ẏi	WI
ọ̆ʖ	ẏo	WO	ọ̆˥	ẏu	WU			
‡	0	0	†	1	1	↓	2	2
↑	3	3	↟	4	4	↡	5	5
ⱡ	6	6	ς	7	7	ϟ	8	8
⫝̸	9	9	⸸	X	X	ʭ	E	;
‖	.	.	¦	,	,	∤	?	?
·	/	/[2]	ƅ̂ƽ	kêl	~	⌜⌟	“”	[]

2 | Document structure

A developed grammar will have the following parts:

- A brief overview of the external and internal histories of the language (especially with more developed conlangs)
- Phonology (phoneme inventory, phonotactics and allophony) and orthography (including script if present)
- An overview of the syntax (basic word orders). This can be explored in detail in later chapters.
- A chapter for each main part of speech and its morphology.
- A chapter listing the morphological derivations of this language.
- Some peripheral topics.
- Example texts, if any.
- The appendix. This is where any source code belongs. Other information may go here at one's discretion.
- The lexicon, generated by `workfiles/dict-to-tex.pl6`.

2.1 | Phonology

The consonants and vowels (if the language has both) are provided in a table, using IPA. Cells representing impossible articulations are marked `\invalid`. Any supplementary methods of producing phonemes are mentioned with the appropriate phonemes.

The phonotactics of a language is a necessary part of the phonology and should not be omitted.

A well-developed language will have some degree of allophony that needs to be described. The Ḋraħýl Rase grammar put these rules in a table. Another system is the Uruwian Diachronic Notation (UDN), detailed in Chapter 3.

2.2 | Orthography

The orthography may be provided either with the phoneme inventory or in a separate table (in the same chapter as the phonology).

There are no hard-and-fast rules for orthography, especially with the Latin script. However:

- ⟨ö⟩ and ⟨ü⟩ are best used for some variant of /ø/ and /y/.
- ⟨b d f g h k l m n p s t v w z⟩ should match up with some variant of /b d f g h k l m n p s t v w z/.
- ⟨a e i o u⟩ should match up with some variant of /a e i o u/.
- ⟨r⟩ and its accented counterparts should be used for rhotics.

If the languages uses hacm, then follow these rules:

- The basic hacm letters should correspond roughly to their Arka pronunciations.
- ⟨ɳ⟩ should be used for /r/ or /ɾ/. If the language has both, then prefer /r/.
- ⟨ɻ⟩ should be used for /ɹ/.
- Phonemes not found in Arka should be written with letters modified by superscript letters:
  - ⟨nᵠ⟩ is the best fit for /ŋ/.
  - ⟨ɾʸ⟩, ⟨ɑʸ⟩, etc. for palatal consonants.

2.3 | Syntax

At the very least, this chapter should provide the basic word order, the order of descriptors relative to their antecedents and, if appplicable, whether prepositions or postpositions are dominant.

2.4 | Basic parts of speech

For most languages, this will at least have a chapter for nouns and another for verbs. There will usually be another chapter for morphological derivations.

2.5 | Periphery

Semantics, units of measure, calendar, etc.

2.6 | Appendix

Source code listings (if applicable), and dictionary.

2.7 | The featural approach

Recent conlangs have been inspired by Isoraḳatheð's approach to conlanging, which treats conlangs as collections of *features*. Thus, a list of candidate features is collected for each conlang before it is started.

The `\synopsis` macro is used to provide a synopsis of a feature, although it is not used quite consistently.

3 | The Uruwian Diachronic Notation

The Uruwian Diachronic Notation (UDN) is a notation optimised for listing many sound changes.

3.1 | Rules

The basic structure of a rule is:

$$\alpha \to \omega \quad (\lambda \blacklozenge \rho) \quad [\Gamma] \tag{3.1}$$

where

- α is the string to be replaced.
- ω is the resulting string.
- $\lambda \blacklozenge \rho$ is the *environment* – if this part is present, then the change will happen only when λ is found immediately to the left and ρ immediately to the right of the string to be replaced. If there is an $\neg$ before the parentheses, then the change occurs when this is *not* true.
- Γ is the *global condition* – a predicate on the word that is changed. If it is false, then the change does not occur. Defaults to being always true.

Rules are applied from top to bottom.

Table 3.1: Some basic examples.

$\text{ɬ.l} \to \text{ɬ:}$	Replace [ɬ.l] with [ɬ:]
$\text{n} \to \text{ŋ} \quad (\blacklozenge \text{g})$	Replace [n] with [ŋ] before [g]
$\text{s} \to \text{ʃ} \quad [\#\sigma = 2]$	Replace [s] with [ʃ] in words with two syllables

3.2 | Capture variables

Of course, α and ω need not be fixed strings. One or more characters in α can be a *capture variable*, which can be backreferenced in ω, λ or ρ. Capture variables consist of a class name followed by a subscript. Usually, a language will use three class names: C for consonant, V for vowel and Σ for strings with any number of characters. However, there is nothing limiting a language from using others – e. g. R for rod signal. At most one top-level class variable in α can be anonymous.

A class variable may have zero or more qualifiers: $C_1[+fr]$ denotes a fricative. Class variables in ω might also take qualifiers, in which case this should be interpreted as changing only the traits listed in the qualifiers and leaving the rest the same.

Class variables can be used in λ or ρ, but those mentioned there but not in α may not be backreferenced in ω.

Class variables may also receive an explicit list, such as $C_1\{\mathrm{b}, \mathrm{d}, \mathrm{g}\}$. In that case, a backreference to such a variable may have the same number of elements that correspond to the elements of the original list. Explicit lists can contain fixed strings *or* further anonymous class variables: $C_1\{\mathrm{b}, [+ve]\}$ refers to a consonant that is either [b] or a velar consonant.

Table 3.2: Some basic examples.

$C_1[+ob, +v] \to C_1[-v] \quad (C_2[+ob, -v]\blacklozenge)$	Devoice obstruents following unvoiced obstruents
$\mathrm{n} \to \mathrm{ŋ} \quad (\blacklozenge C_1[+ve])$	Replace [n] with [ŋ] before a velar consonant
$V_1[+r, -l] \to V_1[-r]$	Unround short vowels
$C_1\{\mathrm{b}, \mathrm{d}, \mathrm{g}\} \to C_1\{\mathrm{p}, \mathrm{t}, \mathrm{k}\}$	Change [b d g] to [p t k]

3.3 | Word boundaries

Word boundaries are marked by $\square$. This is commonly used to detect word boundaries, but can also be used to replace strings across words.

Take note that any Γ-variables will refer to the first word covered by α. If you want to refer to other words, use an offset: ${}_{-1}\sigma$ refers to the syllables of the previous word; ${}_{1}\sigma$ refers to those of the next.

Table 3.3: List of commonly-used qualifiers.

Short	Long	Meaning	Short	Long	Meaning
$+lb$	$pa = lb$	Labial	$+lv$	$pa = lv$	Labiovelar
$+av$	$pa = av$	Alveolar	$+pa$	$pa = pa$	Postalveolar
$+rt$	$pa = rt$	Retroflex	$+pt$	$pa = pt$	Palatal
$+ve$	$pa = ve$	Velar	$+uv$	$pa = uv$	Uvular
$+ph$	$pa = ph$	Pharyngeal	$+gl$	$pa = gl$	Glottal
$+pl$	$ma = pl$	Plosive	$+fr$	$ma = fr$	Fricative
$+na$	$ma = na$	Nasal	$+ap$	$ma = ap$	Approximant
$+la$	$ma = la$	Lateral approximant	$+lf$	$ma = lf$	Lateral fricative
$+tr$	$ma = tr$	Trill	$+tp$	$ma = tp$	Tap
$+v$		Voiced	$-v$		Unvoiced
$+a$		Aspirated	$-a$		Unaspirated
$+hi$	$vh = hi$	High	$+mh$	$vh = mh$	Mid-high
$+ml$	$vh = ml$	Mid-low	$+lo$	$vh = lo$	Low
$+vf$	$vf = fr$	Front	$+vc$	$vf = ce$	Central
$+vb$	$vf = bk$	Back			
$+r$	$vr = 1$	Rounded	$-r$	$vr = 0$	Unrounded
$+l$	$l = 1$	Long	$-l$	$l = 0$	Short
$+s$	$s = 1$	Stressed	$-s$	$s = 0$	Unstressed

Table 3.4: Some basic examples.

$C_1[+ob, +v] \rightarrow C_1[-v] \quad (\blacklozenge\square)$	Devoice final obstruents
$\text{ə} \rightarrow \text{æn} \quad (\square\blacklozenge\square V_1)$	Correct English articles
$\Sigma_1 \text{ɚ} \rightarrow \text{mɔɹ}\square\Sigma_1 \quad [\#\sigma \geq 2 \wedge \sigma[-1].r \neq \text{i}]$	Replace "X-er" with "more X" in long words that don't end with [i]