

Orthography, Phonology, Phonotactics and Prosody of the Tokcir Language

by Stephen DeGrace, stevedegrace@yahoo.ca, 2002/04/04

1.0 Introduction

This essay is an attempt to summarise in a complete, collected and current form, the conventions of the Tokcir lanaguage (the product of the NGL, or Next Generation Language, Project) regarding orthography (spelling), textual conventions, phonology, and phonotactics (allowed syllable patterns, consonant clusters, that sort of thing) for the benefit of current or new project members seeking a definitive record in this area. This essay is composed as part of a planned series which will lay down the most basic foundation of the language, specifically intended for conlangers, both new to the project and veteran. These essays are designed to be reference material for workers on the project, not as material for language learners necessarily, and therefore shall not take too much time to explain terms, but as we are not professional linguists and do not expect other workers on this project to be, *some* explanation may be provided, and any questions you have may be posted to NGL@yahoogroups.com where we will attempt to address them. Private questions can be posed to the NGL list moderator at NGL-owner@yahoogroups.com

2.0 Orthography and Phonology

The sounds and spelling of the NGL (Tokcir) language will be dealt with together.

2.1 The Alphabet and Textual Conventions

The following are the sounds of the Tokcir language. I will attempt to give the IPA conventions for all the sounds of the language, plus the nearest English equivalent; in giving English equivalents, except where otherwise explicitly stated, you should assume the English dialect is Standard Canadian (CBC) English, which is close to the dialect spoken by American TV announcers. Most letters are used quite similarly to the most similar IPA symbol.

Vowels:

There are five recognised vowels, which are pretty much exactly like the vowels of Italian. I will attempt to give the IPA convention followed by an English approximation.

a	-	/a/	-	a as in "father"
e	-	/e/	-	a as in "fate"; actually closer to ais as in "français" as said in French
i	-	/i/	-	ee as in "meet"
o	-	/o/	-	o as in "hope", but clear, without a little "w" at per usual English
u	-	/u/	-	oo as in "food"

Consonants:

'	-	/ʔ/	-	glottal stop; say "water" in a Cockney accent, the "t" is ʔ
b	-	/b/	-	b as in "bat"
c	-	/tʃ/	-	ch as in "chair"
d	-	/d/	-	d as in "deed"
f	-	/f/	-	f as in "fate"
g	-	/g/	-	g as in "good"
h	-	/h/ or /j/	-	h as in "hood" or ch as in "Loch Ness"
j	-	/dʒ/	-	j as in "joke"
k	-	/k/	-	k as in "keep"
l	-	/l/	-	l as in "look"
m	-	/m/	-	m as in "maple"
n	-	/n/ or /ŋ/	-	n as in "narrow", "ng" when before k or g
p	-	/p/	-	p as in "peep"
r	-	/r/ or /r̥/ or /r̄/	-	r as in "reap", also some non-English allophones, variable
s	-	/s/ or /z/	-	s as in "sleep" or z as in "zebra", variable
t	-	/t/	-	t as in "teach"
v	-	/v/	-	v as in "very"
w	-	/w/	-	w as in "weep"
x	-	/ʃ/	-	sh as in "sheep"
y	-	/j/	-	y as in "yellow"
z	-	/ʒ/	-	ge as in "garage", or better, j as in French word "je"

Diacritics:

Tokcir has two accents, an acute accent (´) and a grave accent (`). Accents are marked over vowels. Accents have one and only one function in Tokcir: to mark the stressed syllable. Any deviation from normal Tokcir stress is automatically marked with an accent. When stress shifts due to the action of an affix or due to a word being coined with irregular stress, the stress is marked with a grave accent. There are also two single-syllable words in the language which receive a grave accent purely by convention - {òl} and {à}. When an irregular stress is being used for prosodic reasons by some writer, the irregular stress is marked with an acute accent.

In addition there is a third and less official diacritic in use. Tokcir has no official mechanism to mark deletion, as the apostrophe is already pressed into service as a letter of the alphabet (the glottal stop). However, deletion in contractions can be marked textually with a circumflex (ˆ) accent over the nucleus vowel of, if possible, the syllable directly preceding, or, if necessary, the syllable directly following, the deletion.

Textual Conventions:

Firstly, the common conventions for using Tokcir words in English text (English is the *de facto* official working language of the NGL project) and English words in Tokcir text should be

explained. When a section of Tokcir text is quoted in English text, it is common practise to set the Tokcir text in curly braces { } to mark it. When an English word is used in tokcir text, not as a deliberately intended borrowing but as a stand-in for an as-yet-to-be-invented Tokcir word, we use what are called "lazy marks", just a pair of colons ::, to set it off. So, for example, :this: is in lazy marks.

Secondly, there are three common textual shortcuts for common Tokcir words. Their use is not mandatory, but their use is normal in texts even of the highest formality. They are:

q	-	stands for {ku}	-	meaning: "the"
&	-	stands for {et}	-	meaning: "and"
	-	stands for {'ior}	-	meaning: "or"

These are not used at the beginnings of sentences. At the beginnings of sentences, we use {Ku}, {Et} and {'Ior}.

The first letters of the first words of sentences and of proper nouns (places, like {Kanada}, or people, like {Stephen}) are capitalised. The capitals of all letters are as normal. The exception is ' which has no capital. If the rules would require ' to be capitalised, instead we capitalise the letter immediately following. So, for example, a sentence beginning with {'ior} starts {'Ior}.

Sentences begin with capitals and end with periods. Colons and semi-colons work as usual, except that colons have the additional role as "lazy marks". Quotations are as per American English conventions, at least by the convention existing thus far. Questions begin with an inverted question mark (¿) and end with a question mark (?) just like in Spanish. Exclamations begin with an inverted exclamation mark (¡) and end with an exclamation mark (!) just like in Spanish.

2.2 The Names of the Letters of the Alphabet

Below is the list of the letters of the alphabet, in Tokcir alphabetical order, along with their names, written in Tokcir (i.e., pronounced as indicated in section 2.1). The order is starting at the upper-left, going down the column, then up to the top of the next column to the right and continue.

'	a'e	h	ha(eh)*	p	pe
a	a	i	i	r	ir
b	bi	j	ju	s	es(is)**
c	ce	k	ke	t	te
d	di	l	el	u	u
e	e	m	im	v	vi
f	ef	n	en	w	wi
g	gi	o	o	x	ax
				y	ye
				z	iz

q "gif ku"
& "gif et"
| "gif 'ior"

* pronounced /e_j /, used to show the alternate pronunciation of h only

** pronounced /iz/, used to show the alternate pronunciation of s only

3.0 Phonotactics

Phonotactics is the allowed syllable pattern of a language. Tokcir has comparatively simple phonotactics, allowing no consonant clusters. While it allows unlimited combination of its five vowels in pairs, it allows no more than two vowels in a row. Where C represents a single consonant (any consonantal letter of the Tokcir alphabet) and V represents a single vowel (any vowel of the Tokcir alphabet), the following are the allowed syllables of Tokcir. Refer to section 2.1 for what Tokcir considers to be a vowel and what it considers to be a consonant for the purposes of syllable formation.

Allowed Syllables:

V
VC
CV
CVC
CVVC

Pairs of vowels may be strung together into diphongs or the sounds kept fairly separate purely at the speaker's discretion, with no indication being given textually. In addition, it has been reported that in some group members' personal dialects, pairs of vowels *de facto* are given the pronunciation of non-Tokcir-official vowels - for example, {ae} as /æ/. While this is not official practise, it is of some use in developing conventions for borrowing words in to Tokcir which contain non-Tokcir sounds.

Sometimes in adding affixes in the process of derivation, a final syllable of the form CVV will be formed, or an initial syllable of the form VVC will be formed. In either of those cases, the initial or final syllable is made legal by closing it with a glotta stop, '. So VVC becomes 'VVC, and CVV becomes CVV' automatically. The ' is dropped if possible if further affixing is done which would make the ' unnecessary. In addition, in any case where it is considered desirable to keep two vowels separate in order to make the meaning of a compound distinct, the vowels can be separated with a glottal stop. In general, if we need to insert a consonant to make things legal, we always use ' unless specifically commanded to do otherwise by the grammar of the affix in question, and we always put it at the end of the illegal vowel cluster.

Similarly, sometimes the process of affixation may produce a cluster CCV or VCC. In these cases, the illegal CC cluster is broken up by inserting an e between the two consonants. So CC

becomes CeC.

Please refer to Essay 5, section 3.0 for a discussion of compounding. This is relevant because Tokcir does not permit double vowels or double consonants. If compounding causes two of the same letter to abut, rather than producing a double consonant or a long vowel, the effect is to collapse the pair into a single consonant or single vowel.

3.0.1 Tokcir Word Building Blocks

The above syllable pattern rules are generated by an older phonotactics proposal which rather than setting down allowed syllables, sets down allowed "building blocks" instead. While the above is a more common "linguistic" way of understanding syllable pattern, it derives from the system outlined in this section and is subordinate to it. The following are the allowed word building blocks of Tokcir:

CV

VC

CVC

VCV

V (only allowed as stand alone in grammar not vocab)

To be permissible, a word must be able to be analysed as one or some combination of these pieces. If upon compounding or affixation a word is produced which violates this rule, the derived word must be modified to bring it into compliance with legal Tokcir phonotactics. See Essay 5, which discusses word formation, for a more detailed discussion.

3.1 Borrowing Non-Tokcir Words into Tokcir

Now that both phonology and phonotactics have been dealt with the problem of borrowing non-Tokcir words into Tokcir can be addressed. Any time a foreign word is borrowed into Tokcir "strongly", that is, not merely quoted but used as a nativised or semi-nativised word, it must be Tokcirised, that is, given a spelling to make its syllable pattern and sound pronounceable to native Tokcir speakers, who are assumed unless educated in other languages to be bound to the conventions of Tokcir, just as native Anglophones cannot readily pronounce foreign words without Anglicising them. It is assumed that regardless of whether words are Tokcirised in spelling, native Tokcir speakers would Tokcirise them in speech. Things like the names of countries and languages have native Tokcir forms which are all borrowings. Some other words may wish to be borrowed, and individuals always have the options to tokcirise their personal names in print.

The basic principles are that consonant clusters have to be broken up, since Tokcir doesn't allow them, and non-Tokcir vowels and consonants have to be changed to something similar to them which are allowed in Tokcir. Here are some of the existing practises.

Initial consonant clusters are typically broken up by inserting a u, as in, for example, {Furànsa}, "France". A grave accent is used to shift stress away from the inserted u and on to the nucleus of the next syllable. Alternatively, a e may be placed before the beginning of the illegal cluster; this is often employed for clusters beginning with s, and was the general practise in nativising the name of the chemical elements, for example, {eflòrin}, "fluorine". Internal consonant clusters are usually all right unless they are three consonants in a row, in which case a u or sometimes an a may be used to break them up at the discretion of the borrower. Final illegal clusters of two consonants are legalised by adding a a to the end of the word, for example, {sinka}, "zinc", or else by dropping the last consonant, as in {Niufunlan}. "Newfoundland".

In borrowing things like place names, native stress is typically preserved and marked with a grave accent where it differs from normal Tokcir stress.

Sounds are approximated as close as possible and rendered with the proper Tokcir letter. English voiced "th" ("this") is rendered d, unvoiced "th" ("thing") is rendered as t. Non-Tokcir vowels either get a similar Tokcir vowel (sometimes chosen more for orthographical similarity than for phonological closeness) or else are rendered with a Tokcir diphong. For example, the sound /æ/ often gets rendered {ae}, for example, {duràeg}, "drag".

All in all, this is an area which offers a lot of flexibility and in which the codification is not really firm.

4.0 Prosody

Prosody refers to the stress, timing and rhythm of the language.

Firstly, all Tokcir words have stress on the first syllable unless explicitly marked with a grave accent over the alternate syllable receiving stress. This is pretty much all you need to know to begin to work in the language. There is, however, a deeper level to the issue of prosody, which Jack Durst addressed in a proposal made in 1998/12/08. As that proposal was basically accepted without much modification, and as Jack is always laudably terse, I quote it whole and verbatim below. Go a Google groups search on "NGL prosody" for more information.

Justification: As I was working on creating a tape of speech in NGL, I realized that I was using English prosody. Since prosody is a very important part of spoken language, and NGL isn't intended to be an entirely written language, I decided I ought to define a standard.

Existing limitations: Though there are very few existing rules of prosody in NGL, any system which exists must be consistant with the existing rules of orthography, punctuation, and reduction. A great deal of my system has thus been back-derived to make the orthography and reduction rules make sense.

ACCENT AND STRESS:

First of all is the matter of accent. Since NGL orthography has accent marks, we know that there must be some system of accent, and that accent falls on the first syllable of the word except in the case of there being a grave (à) accent mark. From the punctuation rules, we know that another type of accent (marked with the acute (á)) can fall on the accented syllable of an emphasized word or on a normally non-accented syllable of an irregularly pronounced word. A search of Dejanews reveals several comments that accent in NGL is considered to be synonymous with "stress", which implies that it must be a stress accent. Therefore, there must be at least two levels of stress in NGL; however, since the reduction rules do not include syncope based on stress except for VC`- forms, there cannot be such a thing as a weak stress in NGL.

Basic Rule:

In NGL, primary stress falls on the first syllable or the syllable that bears the accent mark; all other syllables bear secondary stress, there is no weak stress. NGL is a syllable-timed language. Sentence stress falls on the focus or on the most emphasized word.

Details:

There are two levels of stress in NGL, primary and secondary, there is no weak stress in NGL. Primary stress is determined lexically. Some morphemes, such as {feàr, itàn, apìru, surì} carry a lexically irregular stress; other morphemes, mainly prefixes, such as {in`-, fon`-, al`-} carry a lexical de-stressing; while a third set of morphemes, often loan words, carry a phonologically shifted stress {estàn, esrìt}. In the absence of any lexical shifting, the stress on a word is on the first syllable. In derived words, stress is determined by the following rules of precedence:

1. If the stress of any morpheme is lexically irregular, the accent belongs to that morpheme.
 - a. If there are two such morphemes in a word, the accent falls on the *first* such place.
2. else, if the first syllable of the word is a de-stressed prefix, the accent falls on the stressed syllable of the next morpheme.
3. Else, unless the first morpheme is phonologically irregularly stressed, the primary stress falls on the first syllable of each word.

There are two irregularities. (1) If a normally accented word follows a one-syllable word with an accent mark, there is no primary stress on that word. This applies to both grave and acute accents. (2) Grammar words except {òl, à} and the (PVS/VXT) verb helpers do not carry primary stress.

A system with no weak stress also simplifies the matter of timing, since

languages without weak stresses are universally syllable-timed. A syllable timed language will take the same amount of time to pronounce a given number of syllables, in contrast to a phrase timed language which will take the same amount of time to pronounce each clause. This does not in any way change the reduction rules, as they were based on the assumption of a syllable-timed language and reduce primarily at the syllabic level anyway, though the certainty of this type of timing may make them more efficient. As a side benefit, syllable-timed languages tend to be less subject to sound changes over time (Compare French, which is phrase-timed, and Spanish, which is syllable timed and see which retains more of the original Latin.) and so more stable in the long-term.

Sentence stress (Implied by the use of the accute) falls regularly on the focus (new information) of simple statements, direct commands, and {i no V} questions. In N-ne and ¿ke? questions, it falls on the question word, in exclamations, it falls on the first word. In answers to questions, it normally falls on the word which answers the question. In any of these cases, the sentence stress can be moved to any word for emphasis, when this is done, the word is optionally marked with an acute (á) accent or a pair of *emphasis marks*. Each clause in a sentence has exactly one sentence-level stress, which falls on the accented syllable of the word stressed and manifests itself as a stronger primary stress on that syllable, and weaker secondary stress on the syllable following it.

INTONATION AND PAUSES:

Secondary is the matter of intonation. As intonation is the richer verbal equivalent of punctuation, it has been designed primarily to work with the accepted punctuation system.

Details:

NGL has three pitches, high, mid, and low. Mid and low correspond roughly to the lower two pitches to English or Spanish; the high tone takes the range comprising both higher pitches in English. A three pitch range of this type provides greater clarity while having an ease-of-learning benefit for speakers of three and four pitched languages. There is one-syllable per pitch except in one word sentences.

These pitches are arranged in specific sequences to signal various sentence events and attitudes. Sentence events:

b/d/a* event punctuation tone/pause pattern

b/d period [.] - low pitch, pause/end of utterance

a new sentence [A] - [pause], mid pitch

b/a comma/semicolon [,/;] - mid pitch, short pause, mid pitch

d focus [stress w/o à] - high pitch

d emphasis [*word*/à] - irregular sentence stress, rise one pitch

2b/d question mark [?] - mid pitch, high pitch, end of utterance
2b/d/a rhetorical question[?]- mid pitch, high pitch, pause, mid pitch
2b/d exclamation point [!] - low pitch, high pitch, pause
b/a inverted mark [¿/¡] - [rise one pitch, short pause,] high pitch
* pitch occurs on the syllable(s) before/during/after event

When no event is happening, the tone is determined by emotion. To express happiness/approval use the high pitch for no event, to express unhappiness/disapproval use the low pitch for no event, to express surprise rise from low to high through the non-event parts of the sentence, for uncertainty do the opposite. To speak matter-of-factly, use the mid tone. Clauses of less than 2 syllables have only a final intonation. Rise one pitch does not apply if you're already in the high pitch. Distanced/formal questions and polite commands are terminated with the sentence pattern.