

Classical Sindarin Morphology

Thomas Zink and Andreas Engl

University of Konstanz

{*thomas.zink, andreas.engl*} @uni-konstanz.de

Abstract. All of his life, the linguist and author J.R.R. Tolkien invented languages. Most effort went into his elvish languages, Quenya and Sindarin, and he made changes to them until his death in 1973 [1]. His intention was to create realistic languages, especially in Sindarin, with harmonic sounds. Thus, his languages don't follow logic rules but are full of irregularities. Though from a strict perspective Sindarin cannot be considered a complete language [3], enough published material is available to make a descent description of this beautiful language. This project tries to model the morphology of so called Classical Sindarin nouns. It is based on the the works of David Salo [1] and Thorsten Renk [2].

1 A (very) brief history of Sindarin

Tolkien did not only construct languages, he invented whole worlds in which his languages existed and developed over time. Through the ages movements of peoples, wars and feuds have influenced the languages spoken. This also holds for Sindarin. In Tolkien's writings multiple morphological and phonological changes in written and spoken Sindarin can be observed. The roots of Sindarin reach back until the first age. The first people that awoke in Middleearth were the Quendi which is the elvish word for 'people'(in fact, not the Quendi but the Noeg, Dwarfs, were the first people, but this is another story, told in [4]). They created the first spoken language which is known as Primitive Quendian. When the Quendi split into two Tribes, the Eldar and the Avari, their language also changed, leading to Common Eldarin and Avarian. The Eldar were fond of travelling but some of them only reluctantly travelled to the west. They became known as the Teleri and their language was Common Telerin which developed parallel to Quenya. It is Common Telerin of which all Sindarin dialects are derived, however, the numerous events that led to the foundation of these dialects can't be recounted here. Instead figure 1 shows the development of elvish languages.

Classical Sindarin is the language spoken at the time of "The Lord of the Rings", it is the language of the Sindar, the Grey-Elves, those who still live in Middleearth.

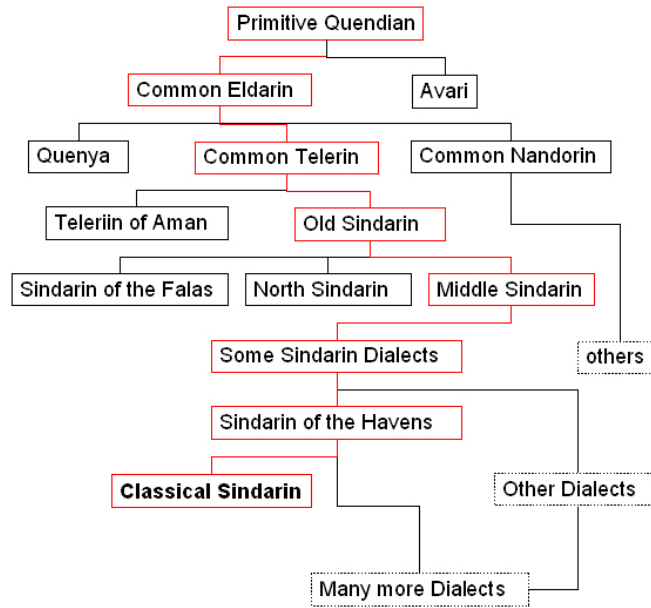


Fig. 1. Development of Elvish Languages

2 Classical Sindarin Phonology

The Sindarin writing system is phonetic with each character denoting a specific sound. The writing system changed drastically between different dialects from simple runes not unlike celtic writing to the beautiful Tengwar modes which are also depicted in the movies. But even Tengwar differs slightly in some dialects. The noldoric mode for instance uses diacritical characters for vowels, the so called tehtars, while the belerian mode uses ordinary tengwar characters. These two modes also differ in the characters used for nasal and glottal sounds. For a nice introduction to Tengwar and it's modes see [5].

The sounds of Classical Sindarin can be described as in table 1 and 2. Sindarin Sounds are pretty much like german sounds except for some sounds like th and dh which more resemble english and others that are trilled like in spanish or russian.

2.1 Vowels

In Sindarin there are short vowels, long vowels, stressed vowels and diphthongs. For correctness we have to add, that in Tengwar no distinction is made between long and stressed vowels. However, Tolkien introduced different characters for long and stressed vowels in his own romanization. Stressed vowels only occur in monosyllables.

	Tincotéma	Parmatéma	Calmatéma	Quessetéma
1	ṽ (tinco)	ṽ (parma)	ṽ (calma)	ṽ (quesse)
2	ṽ (ando)	ṽ (umbar)	ṽ (anga)	ṽ (ungwe)
3	ḥ (púile)	ḥ (formen)	ḥ (harma)	ḥ (hwesta)
4	ḥ (anto)	ḥ (ampa)	ḥ (anca)	ḥ (unque)
5	ṽ (númen)	ṽ (malta)	ṽ (ñoldo)	ṽ (ñwalme)
6	ṽ (óre)	ṽ (vala)	ṽ (anna)	ṽ (wilya)
	ṽ (rómen)	ṽ (arda)	ṽ (lambe)	ṽ (alda)
	ṽ (silme)	ṽ	ṽ (esse)	ṽ
	ṽ (hyarmen)	ṽ (halla)	ṽ (yanta)	ṽ (úre)

Fig. 2. The Tengwar Chart

Long vowels are marked with an acute and stressed vowels are marked with a circumflex. Unfortunately there is no y character with circumflex but with umlaut, which is used instead in table 1. All single character Vowels can be long or stressed but not the diphthongs.

Short	a e i o u y
Long	á é í ó ú ý
Stressed	â ê î ô û ÿ
Diphthong	ae ai au aw ei oe oi ui

Table 1. Classical Sindarin Vowels

According to table 1 the vocals can be defined in xfst as follows (all listings are presented with official 7-bit ascii characters only, thus special characters can not be displayed):

```
define Vshort [ a | e | i | o | u | y ];
define Vlong  [ 'a | 'e | 'i | 'o | 'u | 'y ];
define Vstressed [ %^a | %^e | %^i | %^o | %^u | %^y ];
define Vdiph  [ ae | ai | au | aw | ei | oe | oi | ui ];
define Vowels [ Vshort | Vlong | Vstressed | Vdiph ];
```

2.2 Consonants

Consonants are best grouped by sound. The following table shows all consonantal Sounds.

	labial	interdental	alveolar	palatal	velar	labialized	glottal
unaspirated stops	p b		t d		c g	gw	
nasal	m		n		ng		
trill			rh r				
fricative	f v ph	th dh	s		ch	chw	h
approximant			lh l	j		hw	w

Table 2. Classical Sinarin Consonants

The definitions for xfst thus are:

```
define Cmono [ t | p | c | d | b | g | h | f | v | m | n | r | s | l | w ];
define Cdi   [ gw | ng | rh | lh | ch | chw | hw | th | dh | ph ];
define Consonants [ Cmono | Cdi ];
```

Note that for Consonant and Vowel definitions all digraphs and diphthongs are defined as single characters. Since in Tengwar these sounds are also only a single character, mutations affect the sound itself. This is especially important for some mutation rules as we will see later. Unfortunately the ascii character set cannot represent the numerous sounds of Sindarin adequately.

2.3 Basic Phonology

There are some basic phonological rules which can be easily modeled in xfst.

1. The unvoiced plosives p, t, c never occur following a vowel, but are lenited to b, d, g.
`[t -> d, p -> b, c -> g || Vowels _]`
 There is one exception to this rule: the plural ending -ath (see section Grammar). However, since this rule is not used in this project we set this fact aside.
2. c is always pronounced as k.
`[c -> k]`
3. f at word end is always spelt as v.
`[f -> v || _ .#.]`
4. au at word end is always spelt as aw.
`[au -> aw || _ .#.]`

Rule 2 is dispensable. Tolkien decided to use character c for the sound k so we will stick with it. It is only listed for completeness. When using a lexicon file even the other rules would be pretty useless, assuming the lexicon file is free of typos.

3 Grammar

3.1 Articles

Indefinite Article. There is no indefinite article in Sindarin, thus a noun that has no article is indefinite. Nouns can be singular or plural in an indefinite state.

Definite Article. The singular article is *i*. It strongly binds to the noun so it is concatenated with a *-* leading to *i-*. The singular article triggers the soft mutation (see section Mutations). If it precedes a word beginning with any *i* vowel it is mutated to *ir-*. Note that *_* represents the space character.

```
< ["Art+" .x. {i-}] "@P.NUM.SG@" "@U.MUT.SOFT@" >
```

```
[ {i-} -> {ir_} \ / .#. _ [${i|i|%^i} & ~${Consonants}] ]
```

Genitive Article. The genitive article is *en* and can be translated as *of the*. It is usually only used with a common noun, if the noun is a title (like king) or name the article can be left. The genitive article causes mixed mutation (see section Mutations).

```
< ["Art+" .x. {en_}] "@P.CASE.GEN@" "@U.MUT.MIX@" >
```

In irregular (assumably optional) cases which can not be modeled using rules([2] and [3]), *en* is shortened to *e*. This is not considered in this project.

Plural Article. The plural article is *in* in all cases. It triggers the nasal mutation. Confusingly, if it precedes a word beginning with a consonant it is mutated to *i*. To show the difference to the singular article, the shortened *in* is modeled as *i-* rather than *i-*. This also reflects that the the shortened *in* doesn't bind as heavily to the noun. Strictly this isn't necessary since the number of the noun can also be identified by inspecting the consonant and vowel mutations.

```
< ["Art+" .x. {in_}] "@P.NUM.PL@" "@U.MUT.NAS@" >
```

```
[ "in_" -> "i_" || _ [${Consonants} & ~${Vowels}] ]
```

Other Cases and Prepositions. Special cases arise, if a preposition precedes the articles. Then, depending on preposition stem and ending, article, and mutations, the articles and prepositions are merged to one word. Due to their complexity prepositions are not modeled in this project.

3.2 Plurals

Forming plurals raises a lot of consonant and vowel mutations which are covered in the section Mutations. However, there are other plurals which are covered here in short.

Suffix Plurals. Some words have the base form plural. In these cases the singular is formed by attaching a special suffix. Some of these words are modeled and the required suffix is attached using flag diacritics.

```
< "+Noun":0 "+Sg":{od} "@R.SUFF.OD@">
< "+Noun":0 "+Sg":{og} "@R.SUFF.OG@">
< "+Noun":0 "+Sg":{ig} "@R.SUFF.IG@">
```

-ath Plurals. There is a special suffix *-ath* which can be attached to nearly every word and means something like "all" or the entire set whereas the ordinary plural means some or lot or a subset more than one but less than all. Thus it is also called collective plural. The suffix is altered to *-iath* whenever it follows a monosyllable of the form C - any i vowel - n,r or l. In addition, if the *-ath* suffix is attached all stressed i vowels and diphthongs with an i sound are shortened to i. The *-ath* suffix can not cooccur with any article.

```
< ["+Ath" .x. {ath}] "@R.ART.INDEF@" "@R.NUM.SG@">

[
    ['i|^i|ai] -> i
    \/_ ?* ${ath}|{iath} .#.
]
.o.
[
    [..] -> i
    \/_ .#. Consonants [i|i|^i|ai|ei|oi|ie|io] [n|r|l] _ {ath} .#.
]
```

Other Plurals. There still are a number of other plurals in Common Sindarin which are disregarded in this project for different reasons. One special case for example are defective singulars which are singular words which lost their final or primary syllable. When forming the plural of these words a suffix is attached which consists of the last vowel in the word and the final consonant n. Since it is only possible to identify defective singulars by inspecting Old Sindarin roots they are disregarded.

Another class of plurals are suffixes that mean something like people or horde, double affection plurals, -u and -in plurals.

4 Mutations

There are a large number of mutations in Classical Sindarin. Some are historical and form a cascade of changes from Old Sindarin and even Common Eldarin up to Classical Sindarin. Thus, classification of word stems are done regarding the original Old Sindarin roots. This proved to be quite a problem. In order to model all Classical Sindarin morphological rules, one would have to model all

changes from the base languages up to Classical Sindarin. Unfortunately, this is out of the scope of this project. Additionally, some morphological changes are not quite clear and guesses by authors of numerous works, some of them even contradictial. These rules have been left out or modeled at will by the authors of this project as they seemed to fit. They are marked in the listings with '!!' at the end of a line. Some sounds are completely lost by mutation, this is denoted as the character ' in the rules.

4.1 Consonant Mutations

The consonant mutations describe changes of initial consonants. They basically happen all the time in Sindarin, when a word follows a certain word, a preposition, an article and so on. To complicate things even the articles and prepositions which cause initial consonants to change do mutate often. However, some of these changes are very irregular and dependent on archaic roots and are not modeled in this project. Prepositions and Pronouns are completely ignored.

Consonant mutations are modeled using flag diacritics. In cases where a mutation must occur a flag is set in the lexc source file. If this flag is set the mutation rule is applied.

Soft Mutation. Is one of the most common mutations. Occurs after singular definite article *i*, after most prepositions ending with a vowel and in other cases.

```
[
    d -> dh, b -> v, g -> ',
    t -> d, p -> b, c -> g,
    gl -> 'l, gr -> 'r, gw -> 'w,
    lh -> l, rh -> r,
    h -> ch, s -> h, m -> v,
    hw -> chw !!
    || %@U%.MUT%.SOFT%@ _
]
```

Nasal Mutation. Is another important mutation which occurs after the plural article *in* and some prepositions ending in a consonant.

```
[
    t -> th, p -> ph, c -> ch,
    b -> m,
    lh -> l, rh -> r, !!
    h -> ch,
    hw -> chw !!
    || %@U%.MUT%.NAS%@ _
]
```

```
.o.
[
    d -> n, g -> ng
    || %@U%.MUT%.NAS%@ _ Vowels
]
```

Mixed Mutation. The mixed mutation is caused by the genitive article *en* and some prepositions combined with the article *in*. It is called mixed mutation because it has elements of the nasal and soft mutations.

```
[
    t -> d, p -> b, c -> g,
    lh -> l, rh -> r,
    s -> h,
    h -> ch, hw -> chw !!
    || %@U%.MUT%.MIX%@ _
]
```

Other Consonant Mutations. There still are more consonant mutations in Sindarin, like the Stop, Liquid and Medial Mutations. However, they are caused by prefixes which have not been modeled, so they are disregarded and only mentioned for completeness. For a deeper look into these mutations see [1] and [2].

Table 3 shows Consonant Mutations modeled in this project together with some examples which are included in the lexc source file.

Mutations	Soft	Nasal	Mixed
barad	i-varad	i_meraid	en_barad
cant	i-gant	i_chaint	en_gant
dûr	i-dhûr	i_nuir	en_dûr
draug	i-dhraug	i_droeg	en_draug
flâd	i-flâd	i_flaid	en_flâd
graug	i-'raug	i_groeg	en_graug
hîth	i-chîth	i_chîth	en_chîth
lûg	i-lûg	i_luig	en_lûg
mellon	i-vellon	i_mellyn	en_mellon
naug	i-naug	i_noeg	en_naug
parth	i-barth	i_phairth	en_barth
ras	i-ras	i_rais	en_ras
rhavan	i-ravan	i_revain	en_ravan
sereg	i-hereg	i_serig	en_hereg
taur	i-daur	i_thoer	en_daur

Table 3. Sindarin Consonant Mutations

4.2 Vowel Mutations

In addition to consonant mutations Sindarin also has a lot of vowel mutations, also called affections. The most of them occurred from Old Sindarin roots to Common Sindarin. Vowel affections within Common Sindarin are due to plural changes or caused by articles and prepositions.

Primary Final i Affection. The name originates from the fact that it occurred in Old Sindarin words whenever a syllable preceded the final vowel i. In Common Sindarin it occurs only when forming plurals and only in monosyllable words or final syllables.

```
[
    [a|^a] -> ai, e -> i, %^e -> %^i,
    [o|u] -> y, %^o -> %^y, %^u -> ui,
    au -> oe, ie -> i, io -> y
    || %P%.NUM%.PL%@ ?+ _ [${Consonants} & ~${Vowels}] .#.
]
```

Secondary Affection. Secondary affection occurs on vowels within words that have been altered by primary affection. Practically this only affects multisyllable words when forming plurals.

```
[
    [a|'a|^a] -> e, [o|'o|^o] -> e, ['i|^i] -> i, u -> y, %^u -> u
    || %P%.NUM%.PL%@ ?* _ ?* [ ${ ai | i | y } & ~${ui} ] .#.
]
```

Affection does not take place if the last or only vowel in a word is ui.

Table 4 shows all changes in vowels together with some examples which are also included in the lexc file of the project.

Other Affections. Again there are many more vowel affections in Sindarin which occurred from Old Sindarin to Common Sindarin. These are the Affection, Primary Internal Affection and Double Affection. They are disregarded.

5 xfst Project

The project consists of the following files.

- `sindarin-lex.txt`
The lexicon file.

vowel	becomes		Singular	Plural
	in syllables	nonfinal in final syllables		
a, â	e	ai	las, tâl, adan	lais, tail, edain
e, ê	e	i, î	sereg, têtew, edhel	serig, tiw, edhil
i, î	i	i, î	gîl, gilith	gîl, gilith
o, ô	e	y, ÿ	pôd, onod	pÿd, enyd
u	y	y	tulus	tylys
û	u	ui	dûr	duir
ÿ		ÿ	mÿl	mÿl
ai		ai	taith	taith
ae	ae	ae	gwael, aegas	gwael, aegais
au	au	oe	naug	noeg
ie		i	miniel	minil
io		y	mirion	miryn
ui	ui	ui	iuith	iuith

Table 4. Sindarin Plural Vowel Affections

- `sindarin-rul.txt`
The rule file. It models all morphological rules discussed in this paper.
- `sindarin-def.txt`
Holds all definitions.
- `sindarin.xfst`
The xfst script which contains all definitions, reads the lexicon and rule file and compiles the network.
- `sindarin-stack.xfst`
A working stack file of the compiled project.

Listings of these files are left to prevent redundant information and for spatial reasons. The source files are well documented so we forego detailed explanations here and reference to the source files instead.

All files must be located in the same directory. After starting xfst from that directory the project can be loaded using the command:

```
xfst[0]: source sindarin.xfst
```

If for some reason the source files won't work, the stack of the working project is included as `sindarin-stack.fst`. It should work in any xfst implementation on any platform albeit character encoding. The stack can be loaded with

```
xfst[0]: load stack < sindarin-stack.fst
```

This project only covers Common Sindarin nouns, articles discussed in section Articles, and plurals as discussed in section Plurals.

5.1 Projections

Upper Side Language. The upper side language consists of the base form of the nouns and various tags describing different states the nouns can take. Legal upper side strings are:

- Indef+'noun here'+Noun+Sg
- Indef+'noun here'+Noun+Pl
- Indef+'noun here'+Noun+Ath
- Art+'noun here'+Noun+Sg
- Art+'noun here'+Noun+Pl
- Art+'noun here'+Noun+Gen

Replace 'noun here' with any noun of the wordlist.

Lower Side Language. The lower side language shows the mutated nouns with or without article. That is

- 'noun singular'
- 'noun plural'
- 'noun'ath
- 'noun'iath
- i-'noun singular'
- ir-'noun singular'
- in-'noun plural'
- i-'noun plural'
- en-'noun genitive'

Examples. Some example output the network should give.

```

xfst[1]: down Art+edhel+Noun+Gen
en_edhel
xfst[1]: up en_d^al
Art+t^al+Noun+Gen
xfst[1]: down Art+idhil+Noun+Sg
ir_idhil
xfst[1]: down Art+edhel+Noun+Sg
i-edhel
xfst[1]: up i_threnairn
Art+trenarn+Noun+Pl
xfst[1]: down Art+galadh+Noun+Pl
i_ngelaidh
xfst[1]: down Indef+galadh+Noun+Pl
gelaidh
xfst[1]: up gwan^un
Indef+gwan^un+Noun+Pl

```

```

xfst[1]: down Indef+gwan^un+Noun+Sg
gwan^unig
xfst[1]: down Indef+h^ith+Noun+Ath
hithath
xfst[1]: down Indef+fair+Noun+Ath
firiath

```

And some input the network should reject.

```

xfst[1]: up h^ithath
xfst[1]: up ir_edhel
xfst[1]: up in_gwaith
xfst[1]: down Art+onod+Noun+Ath

```

Known Bugs. There is one known bug which could be reproduced but no solution could be found. It references the mutation of the *-ath* plural suffix to *-iath*. The rule is located on lines 99 - 103 in the `sindarin-rul.txt` source file.

```

[
    [...] -> i
    \/.#. Consonants [i|i|^i|a|e|i|o|i|e|io] [n|r|l] _ {ath} .#.
]

```

Though syntactically correct, the *i* is not inserted. Strangely the rule works when the word symbols *./.* are left out. This, however, leads to wrong plurals as e.g. *idhiliath*. When applied in a clean xfst environment (i.e. only this rule is read as regular expression) the rule matches and the *i* is inserted correctly. Whether this bug is due to conflicting rules of the rulebase or to the limited capabilities of the xfst evaluation version (which always complains that the network is too large) could not be discovered. It is left in the sources and marked as ! BUG.

6 Discussion

This project only covers a small part of Classical Sindarin. Unfortunately, much has to be left undone. Sindarin is a beautiful language with morphological changes and historical background that is amazing for a constructed language. The different modes in writing and different dialects are hard to distinguish for the novice's eye. For that reason the numerous documentations of Sindarin which can be found in the Internet are often contradictal. Even after having studied and compared three different sources of which two are written by linguists some things remain mysteries. Sindarin has never been completely described by Tolkien, all what is known has been derived by analysing the numerous Sindarin writings that exist. Thus, no source is perfect. The authors of this project tried to model rules that seem to be commonly accepted as fact. No claim for completeness or even correctness can be made.

References

1. Salo, D.: A Gateway to Sindarin. The University of Utah Press, Salt Lake City 2004. [1](#), [4.1](#)
2. Renk, T.: Pedin Edhellen, a Sindarin Course. 2004, <http://www.sindarin.de/index.html>. visited 2007. [1](#), [3.1](#), [4.1](#)
3. Fauskanger, H. K.: Of the Tongues of Arda, the invented world of J.R.R. Tolkien. <http://www.uib.no/people/hnohf/sindarin.htm>. visited 2007 [1](#), [3.1](#)
4. Tolkien, J.R.R., Tolkien, Christopher (ed.): The Silmarillion. HarperCollins Publishers 1999. [1](#)
5. Katzer, G.: Tengwar und ihre Verwendung, <http://www.uni-graz.at/katzer/tengwar.html>. visited 2007. [2](#)
6. <http://www.langmaker.com/db/Sindarin>. visited 2007