

A finite-state morphological analyzer using foma for Klingon, a constructed language

Matthew David Dreselly Thomas
University of Colorado at Boulder
matt@bocosoft.net

Abstract

This paper presents a morphological analyzer for Klingon, constructed language, developed following the official description of the morphology in the Klingon Dictionary (Okrand, 1992).

The open-source tool *foma* for constructing finite-state automata and transducers is used for building the morphological analyzer.

1 Introduction

Klingon was created by Marc Okrand, an American Linguist, for the 1984 movie *Star Trek III*. Klingon is spoken by the fictional Klingons in the *Star Trek* universe, so Okrand designed the Klingon language to sound alien by using grammatical constructs and sound combinations that are rarely used in natural language (Institute, b; Institute, c).

Klingon has a complete grammar that can be spoken and used as a real language. Marc Okrand formalized the Klingon grammar and vocabulary in *The Klingon Dictionary* (Okrand, 1992). The official writing system using the Latin alphabet and is case-sensitive. The alien alphabet, called *pIqaD*, was created for visual effects in the movies by the Astra Image Corporation.

The number of fluent Klingon speakers is not known, but has been estimated between a few dozen (Okrent, 2009) to around 100 (Kelly, 2013). *The Klingon Dictionary* (Okrand, 1992) has sold more than 300,000 copies and has been translated into 4 languages. Klingon has entered into popular culture, and literary works such as Shakespeare's *Hamlet*, *Gilgamesh*, and "A Christmas Carol" have been translated into Klingon. The current lexicon contains almost 3,000 words (Zrajm,) and only grammar and words coined by Marc Okrand are considered canonical by the Klingon speaking community (Institute, a).

2 FST Implementation

The morphological analyzer FST consists of two files; a *lexc* file and a *foma* file. References to *The Klingon Dictionary* and other sources are captured in comments throughout the FST files for reference and as an explanation of implementation decisions.

2.1 *lexc* file

The *lexc* file describes Klingon morphology and contains the basic set of Klingon lexical entries. I use the KLCPI, which are the 502 basic words of first level of KLI's "Klingon Language Certification Program" (Zrajm,). Each lexical entry contains the stem and the english meaning on the upper-side (analysis) of the FST.

Beesley and Karttunen (Beesley and Karttunen, 2002) flag diacritics are used to restrict the stem and affix combinations according to the grammar description.

2.2 *foma* file

The *foma* file creates an FST from the *lexc* and adds morphophonological rewrite rules. The single Klingon morphophonological rule is implemented in the *foma* file as a rewrite rule.

3 Klingon Morphology

Klingon is an agglutinative language that uses Object-Verb-Subject word order. Okrand categorizes all Klingon words into Nouns, Verbs, and Other. The morphological analyzer is designed and tested to work with Nouns and Verbs, however some additional work was done to account for the words in the 'Other' category.

The Noun LEXICON uses the "NOUN-TYPE" feature flag to track the noun category; @U.NOUNTYPE.BODYPART@, @U.NOUNTYPE.CAPABLELANGUAGE@ and @U.NOUNTYPE.OTHER@.

3.1 Nouns

Nouns can take five types of suffixes. Each suffix type, if present, must be appended to the noun stem in a specific order and there can be no more than one suffix per type. Nouns are divided into three categories; body-parts, beings capable of language, and general nouns. Some suffixes can only be appended to certain noun categories or their meaning changes when they are appended.

Below are the noun suffix types and the relevant rules for each type.

Type 1 Augmentative/diminutive

Type 2 Number

| Noun Type | Suffix |
|------------------|--------|
| Body-Part | -pu' |
| Language-Capable | -Du' |
| Generic | -mey |

The *-mey* suffix may also be used with body-part and beings-capable-of-language nouns, but the resulting word acquires an additional meaning “*all over the place*” on top of the plural meaning.

| Klingon | English |
|---------|----------------------------|
| jup | friend |
| jupDu' | friends |
| jupmey | friends all over the place |

Table 1: Generic Plural Suffix Usage with Inherently Singular Nouns

Some nouns are inherently singular or inherently plural. For example the Klingon word *peng* means torpedo (singular) but *cha* means torpedos (plural). The singular forms of these types of nouns may also take the *-mey* generic plural suffix, but the resulting word also acquires the “*all over the place*” meaning.

If a noun is inherently singular or inherently plural, it is marked using the Positive flag diacritic @P.SINGULAR.IRREGULAR@ or @P.PLURAL.IRREGULAR@.

Lastly, nouns in Klingon are not required to take any plural suffix in order to be plural. The singular form of a noun may be used where the plural form would normally be used, and this is perfectly grammatically correct. The meaning is disambiguated using the context of the word in the sentence. Therefore *jup* means both “friend” and

| Klingon | English |
|---------|-----------------------------|
| peng | torpedo |
| pengmey | torpedos all over the place |
| cha | torpedos |
| *chamey | <i>Not allowed</i> |

Table 2: Generic Plural Suffix Usage with Language-Capable Nouns

“friends” depending on the context. There are no articles in Klingon.

The Type-2 suffix group uses the NOUNTYPE, SINGULAR and PLURAL flag diacritics to control which suffixes can be applied and what their meaning is.

Type 3 Qualification

Type 4 Possession/specification

Possession suffixes are selected depending on whether the *object* being referred to is language-capable.

| Possessive Suffixes | Generic | Language Capable |
|---------------------|---------|------------------|
| my | -wIj | -wI' |
| your | -IIj | -II' |
| his, her, its | -Daj | -Daj |
| our | -maj | -ma' |
| your (plural) | -raj | -ra' |
| their | -chaj | -chaj |

Language-capable nouns can still take the generic suffix but the meaning becomes derogatory (Okrand, 1992). The noun type is tracked using the NOUNTYPE flag and in this case the tag [DEROGATORY] is appended to the analysis.

Type 5 Syntactic markers

3.2 Verbs

Verb can take both prefixes and suffixes. Prefixes specify both the subject and object of the verb, including whether or not a subject and/or object exists. Verbs can use either the **Indicative mood** or the **Imperative mood**, which can also affect whether some suffixes can be used. There is no infinite form of a verb in Klingon.

Verbs are categorized into **stative** and **dynamic**. Klingon does not use adjectives so stative verbs are used to describe nouns. For example to say “red ship”, one would use the stative verb *Doq* meaning to be red/orange, “*Duj Doq*”.

Verbs conform to nine ordered suffix types. Just like nouns, each suffix type, if present, must be appended to the verb stem in a specific order. There can be no more than one suffix per type.

Below are the verb suffix types and the relevant rules and flag diacritics for each type.

Type 1 Reflexive, Reciprocal

This suffix type can only be used with the set of prefixes that mean "no object". The *-chuq* reciprocal suffix can only be used when the subject is plural.

Type 2 Volition, Necessity

Type 3 Inceptive, Inchoative

Type 4 Causative

Type 5 Undefined Subject, Capability

Type 6 Perfection, Uncertainty

Type 7 Aspect

Type 8 Honorific

Type 9 Syntactic markers

In addition, a "rover" suffix type can be appended immediately after the verb stem or between the nine suffix types, however rover suffixes cannot be appended after a Type-9 suffix. There are no restrictions on the number of rover suffixes that may be used. A rover suffix's meaning applies to the suffix immediately proceeding it. For example the **negating** rover suffix *be'* can be appended immediately after the verb stem and/or after any other suffix type.

| Klingon | English |
|--------------|-----------------------------|
| Dub | to improve |
| Dubbe' | to not improve |
| DubmoH | to cause to improve |
| DubmoHbe' | to not cause to improve |
| Dubbe'moH | to cause to not improve |
| Dubbe'moHbe' | to not cause to not improve |

Table 3: Generic Plural Suffix Usage with Language-Capable Nouns

Rover suffixes can even be appended after other rover suffixes, e.g. *Dubbe'be'* meaning *to not not improve*.

3.3 Other Words

Pronouns

Numbers

Conjunctions

Adverbials

Exclamations

Names

Questions

4 Klingon Morphophonology

Klingon completely lacks morphophonology with one exception. The endearment noun suffix *-oy* is prepended with a glottal stop ' if the noun stem ends with a vowel.

| Noun Stem | With Endearment Suffix |
|---------------|------------------------|
| jup (friend) | jupoy (dear friend) |
| cha (torpedo) | cha'oy (dear torpedo) |

Table 4: Endearment Suffix

5 Evaluation

More than 150 test cases are contained in a separate *foma* file. These test cases follow Okrand's examples in The Klingon Dictionary. The morphology *lexc* was modified until all test cases passed and recall reached 100%.

In order to obtain a quantitative analysis of the morphological analyzer, I compiled a list of words from The Klingon Dictionary and The Klingon Hamlet (096,). Of the 267 words on the list, the morphological analyzer accounted for 264. The remaining 3 words were *DatDaq*, *tIpuq* and *naDevDaq*. Both *DatDaq* and *naDevDaq* are explicitly forbidden by Marc Okrand in The Klingon Dictionary as there are location nouns with the locative suffix, therefore these words are technically not grammatically. The last word *tIpuq* comes from the Sonnet-18 translation in Hamlet and appears to translate to "bud" (as in rose bud), however this word does not appear in any canonical sources of Klingon. Therefore I consider the morphological analyzer to have achieved 100% recall for the target word list.

References

Hamlet, Prince of Denmark: The Restored Klingon Version. Klingon Language Institute.

Beesley, K. R. and Karttunen, L. (2002). *Finite-State Morphology: Xerox Tools and Techniques*. Xerox Corporation.

Institute, T. K. L. Canonical Sources. http://www.kli.org/wiki/Canonical_sources. Accessed: 2015-12-15.

Institute, T. K. L. Development and Use of the Klingon Language. <http://www.kli.org/about-klingon/klingon-history/>. Accessed: 2015-12-15.

Institute, T. K. L. The Sounds of Klingon. <http://www.speakklingon.info/about-klingon/sounds-of-klingon/>. Accessed: 2015-12-15.

Kelly, S. M. (2013). How Klingon Became a Universal Language. <http://mashable.com/2013/05/18/klingon-star-trek/>. Accessed: 2015-12-15.

Okrand, M. (1992). *The Star Trek: The Klingon Dictionary*. Simon and Schuster.

Okrent, A. (2009). *In the Land of Invented Languages: Esperanto Rock Stars, Klingon Poets, Loglan Lovers, and the Mad Dreamers Who Tried to Build A Perfect Language*. Spiegel & Grau.

Zrajm. Klingon Word Database 1st edition. <http://mashable.com/2013/05/18/klingon-star-trek/>.