Kathrin Betzenbichler
Matr.-Nr. 01/796312

Documentation of XLE-Project Sindarin Toy-Grammar

1. Introduction

Sindarin is a fictional language, constructed by the author and linguist J. R. R. Tolkien. It is not based on any existing natural or otherwise existing language and was intended to seem realistic. Besides Sindarin, Tolkien also invented Quenya. Both languages are spoken by Elves, one of the people in his fictional middle-earth universe. He began to develop his Elvish languages in the 1910s and made changes to them until his death in 1973 (vgl. Salo 2004: xiii).

This project is an attempt to model the basic syntactic structure of Sindarin, based on the grammar of David Salo (2004). The Sindarin described in his book is titled as Classical Sindarin, which is focused on the Sindarin spoken during Tolkien's Third Age; the time *The Lord of the Rings* took place.

2. Basic features of Sindarin

Unfortunately, the number of complete Sindarin sentences is very limited. But nevertheless there are some basic features that can be attested of for the language of Sindarin.

2.1 Mutations

A very special characteristic of Sindarin are the so-called mutations. These mutations can be considered as stem allomorphy. Salo (2004) distinguishes between consonant and vowel mutations. The consonant mutations always affect the initial consonant of most Sindarin words, conditioned either by the preceding word (like articles or prepositions), holding a grammatical role or taking the second position of a compound. The vowel mutations on the other hand are only conditioned by phonological environment and resemble vowel harmony. Within the categories consonant and vowel mutations are several subcategories based on different rules a word can be altered on. Because of the rather phonological and morphological nature and the great variety of mutations, there were no rules for them were implemented in this project. Regardless, I implemented several allomorphs of one word in the lexicon, where needed, and used the right variant for the sentences in the testsuite.

2.2 Syntax

"Sindarin is a rather consistently *head-left* language" (Salo 2004: 194). This means that the head of a phrase always precedes its complements. In a VP, the verb precedes its objects, in an NP, nouns are followed by nouns, adjectives or PPs that modify them. In PPs, prepositions are followed by their complements etc.

The basic sentence structure in Sindarin is VS(O). NPs are subjects, performing the action of the verb. The object, also an NP, is the recipient of the action. Verbs can be transitive and intransitive. Intransitives require no object, therefore the sentence structure in an intransitive sentence is VS. Sindarin verbs tend to drop the pronouns functioning as subjects, which is also called null anaphora. Therefore, a well-formed sentence may consist of a verb only. On the other hand, verbless sentences, where the copula is omitted, are also perfectly acceptable, which may lead to a sentence with a noun as a subject only or even a sentence with only an AP (see 3.2 Verbless sentences).

3. Grammar of Sindarin

This section reflects the Sindarin grammar and how it was implemented as a toy-grammar into XLE.

3.1 Sentence structure in the rules section

Sindarin is a language with verb-initial word order and sentences may consist of a VP only, since the pronoun can be dropped. When implementing the optional subject-NP into the rules, I encountered some problems. At first it seemed logical to simply set up the S category with an initial VP, and an optional NP for the subject following. However, this raised some problems in parsing the object-NP inside of a VP. Further reading on how LFG treats the problem of SVO word order, for instance Sadler (1997) and Dalrymple (2001) on the example of Welsh, suggested introducing a category called "I" for the verb, leading to a structure like the following:

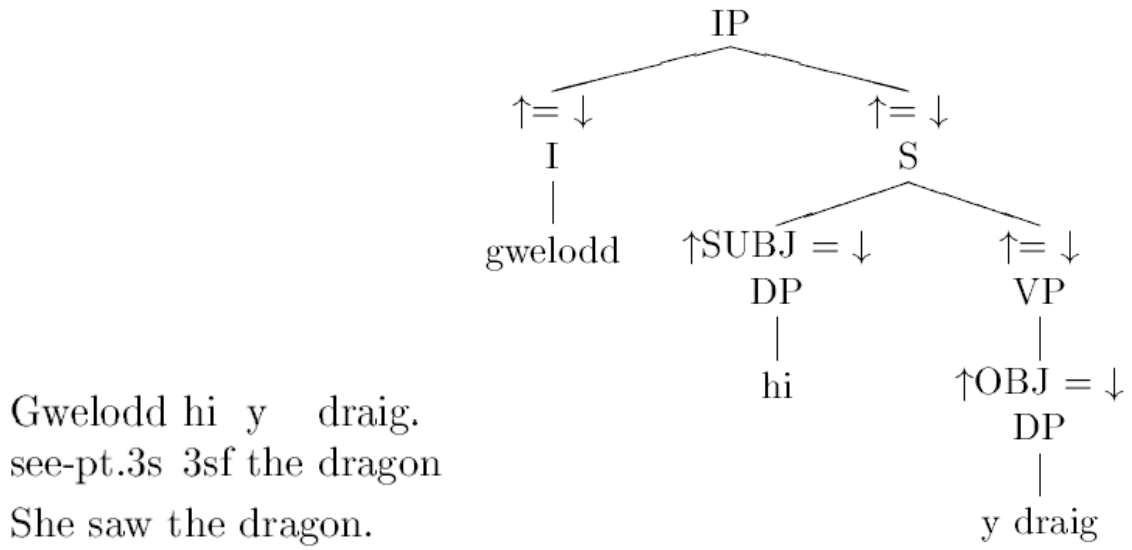

Figure 1: Welsh example of modelling a sentence with VSO word order in LFG (taken from: Sadler (1997: 5f.))

This approach, however, seemed unnecessarily complicated, so I decided to leave the IP out and tried to map everything within the S-category. As it turned out the grammar is working fine by placing the subject-NP within the VP rule, following the verb and preceding the object-NP.

3.2 Verbless sentences

Many sentences in Sindarin can be constructed without an overt verb. This is because the copula „to be“[1] can be omitted, leaving a sentence consisting of just a noun, a complex NP, or an AP with or even without a noun as the head, as in the sentence *mae govannen* “(it is) well met” (vgl. Salo, 2004: 203f.). It is not really specified in Salo (2004), but it seems that in verbless sentences not only the verb, but also the subject is omitted. All sentences described in his grammar are translated as “(it is/there is/these are/…) something”, with “something” seeming in the object-position. Therefore, the rules for verbless sentences can be described as follows: Because verbless and verbal sentences are very different, an early disjunction was implemented for the S rule, with the root-category being S, leading to either verbal or verbless sentences. In verbless sentences an empty-node represents the verb, being an optional transitive variant of “be”. Furthermore, another empty-node stands for the subject, which seems to denote to the third-person singular pronoun “it”. Therefore, the subject predicate is ‘pron’, with subject number being singular and subject person being “3”. The case is nominative. After the empty-nodes either an object-NP follows, or an AP stands alone.

[1] In fact, the only forms of “to be” documented in Sindarin are the gerund *nad* and the imperative *no.*

3.3 Noun phrases

Noun phrases usually consist of a single count, mass or proper noun or a pronoun. Count nouns can take a determiner. Complex noun phrases can be composed of a noun together with another noun or noun phrase (standing in apposition or in genitive NPs), a noun together with an adjective phrase, a noun together with a prepositional phrase or a noun together with a complementizer phrase.

3.3.1 Pronouns

The simplest NP consists of a single pronoun. Many pronouns in Sindarin are unattested or speculative and thus were not implemented. All other nominative, accusative and dative personal pronouns found in Salo (2004) were adopted. The PRON template for pronouns takes variables for the predicate, person, number and the type of the pronoun (PRON-TYPE). It also includes information for the type of noun (NTYPE), being 'pron'. In the lexicon section the variables were assigned the corresponding information, for example the pronoun *im*, having the predicate *im*, first person, singular and being a personal pronoun: PRON * @(PRON im 1 sg pers). Every pronoun in the lexicon also got the correct case assigned by the CASE template.

Gender seems to be only important for the third-person pronoun (vgl. Salo 2004: 107), so I added a gender template (GEND) to assign the correct gender to the corresponding pronouns, even though it makes no difference in the grammar. Besides the neuter, feminine and masculine third person pronouns, there's a special personal pronoun, the general pronoun, which seems to be unmarked for gender and therefore used for any third person singular form in the nominative. The function of this pronoun is quite uncertain, but due to its frequent appearance I decided to include it in the lexicon.

"When the object of a verb is a pronoun rather than a noun phrase, it normally precedes the verb, producing the verb order OVS" (Salo 2001: 204). This was implemented simply as an optional pronoun taking the accusative before the verb in the VP rule.

3.3.2 Nouns (as subjects or objects)

"Sindarin nouns are not marked for case by any overt morpheme" (Salo 2004: 93). This might indicate that the sentence structure in Sindarin is indeed important. Subject-NPs are always at second position, placing object-NPs on third. But since Sindarin is a Prodrop-language the subject-NP can also be left out if it's a pronoun. Now in the case of a sentence like "He/she knows the man", where the pronoun is dropped, how can we tell that the sentence is not "The

man knows", since the verb takes third person singular, which accounts for both? So, in simpler words: How to distinguish between subject- and object-NPs and avoid ambiguity?

"Nouns functioning as accusatives, that is, direct objects of verbs […], are usually given the soft mutation" (Salo 2004: 100). Because I didn't want to map all of the mutations into this project, I decided to just focus on soft mutation on accusatives and give a quick solution for disambiguating. First, I wrote a template MUT for mutations with a variable, allowing to give it different values like "soft". The template was then used in the lexicon with nouns having the soft mutation. After that, I introduced a disjunction for object-NPs inside of the VP rule: Either they can be normal NPs taking the accusative case, or they can be NPs marked as soft mutated by the template. Latter NPs are then preferred by using the OT-Mark NPsoft.

This is a makeshift solution for preventing ambiguity and it only works for nouns which are marked for soft mutation. There are still a lot of nouns which don't undergo mutation and therefore are still ambiguous. Besides, this is not a solution for preventing the not-mutated form to occur in the accusative, when the mutated form should be used. The sentences used in the testsuite, however, should all be correct in this matter.

Nouns functioning as subjects don't undergo mutation. Because it is not described what exactly triggers the soft mutation of object-NPs (is it their syntactic function or the preceding verb or both or something else) I decided to allow a normal NP as the object in verbless sentences.

3.3.3 Determiner

Sindarin has no indefinite article. A noun that has no article is therefore indefinite and can be singular or plural. The singular definite article is *i*. It strongly binds to the noun, so when written, it often is concatenated with a hyphen to the noun. For the sake of simplicity, this was not implemented. Because the singular article triggers mutation, it was often left out in the testsuite sentences.

The definite article for words in plural is *in*. When it precedes a noun with a consonant as the initial sound it is reduced to *i*. The difference to the singular article is that it triggers nasal mutation to the following noun in all cases, so I barely used it in the testsuite.

3.4 Verbal phrases

Verbs in Sindarin can be distinguished in two stem-categories: i-stems and a-stems. This is mostly important for their behaviour regarding mutations. Verbs are inflected for tense (corresponding with mood) and take personal endings to agree with the person and number of

their subject. Additionally, there are two types of verbs in Sindarin: transitives and intransitives, with some verbs being optional transitive. Ditransitives or other transitives are not described. Other forms that Sindarin verbs can take are the impersonal, the gerund and the participle. Besides the indicative mood, verbs can also be imperative.

3.4.1 Basic templates and lexicon-entries for verbs

I set up some basic templates, some of which were predominantly, some were exclusively used for verbs. First of all, the TRANS, INTRANS or OPT-TRANS templates should indicate how many and which arguments the verb is taking. To save typing I introduced templates like PRED (predicate), TENSE, MOOD, ASPECT, PART (for participle), NUM (number) and PERS (person). VPRES and VPAST are the templates for present tense or past tense, invoking the TENSE and MOOD template within, taking the corresponding tense and indicative mood.

In the lexicon, I started each entry with indicating transitivity for the verb, followed by the tense-template and then giving them a number-person-template.

3.4.2 Personal endings

Verbs in Sindarin take different endings, corresponding to the person and number of their subject. No second person endings are found in any published material. The other endings are:

	Singular	Plural
First person	-n	-m
Third person	-	-r

The present tense of a verb is formed by adding the personal suffix to the stem, so the suffix also carries tense information.

Since the endings are carrying information about the subject person and number, I implemented templates for each number-person-combination (1SG, 2SG, 3SG, 1PL, 2PL, 3PL). These are important for both subject-verb-agreement and for the prodrop sentences. Although there are only four different inflectional forms documented, I included templates for all six combinations of number and person. As mentioned, the subject-pronoun of finite sentences can systematically be dropped in Sindarin (since we have different endings for all persons). This is characterized as a prodrop language.

3.4.3 Prodrop

Since the subject position (subject predicate) cannot be empty in XLE without having an empty node, I had to find another solution for defining prodrops. I set up a template called PRODROP, and placed it inside the TRANS and INTRANS templates, so it is always invoked when having a verb (either transitive or intransitive). The PRODROP template denotes as the following: Either we have subject number singular or plural and subject person one, two or three assigned, then it's a prodrop sentence and we have 'pron' as the subject predicate, or it's not a prodrop.

3.4.4 Past tense

Past tense in Sindarin is composed out of the stem and can be formed in at least four different ways: Nasal affixation, reduplication, ablaut and suffixation of *-nt* or *-s*. I only included some sentences in the past tense to the testsuite, since the syntax doesn't change. However, the VPAST template indicates that the verb is in the past tense, the mood is staying indicative.

3.4.5 Participles

Forming the participle of a verb in Sindarin transforms the verb to a more adjectival construction. The same is true for other (existing) languages like English, but in Sindarin the participle behaves even more like an adjective than for instance in English. Thus, it can be placed inside of a noun phrase like a modifying AP, following the head noun. The present participle indicates that the word it modifies is either currently or habitually engaged in an action. Past participles are showing a present state that results from a past action which can (but not must) be still continuing. Unlike in English, the past participle needs no auxiliary verb and doesn't change the sentence structure.

I marked the participle verbs in the lexicon with the template PART, giving them either present or past value. Because they inhibit no longer verb-status, they are not marked for transitivity. However, they still take tenses, so participles in the present were given the VPRES template, participles in the past the VPAST template. In the rules section I introduced a special VP for participles, called VPART. The VPART is restricted to only consist of either a present or a past participle, done via a constraining equation. The VPART was then placed inside of the NP rule.

Additionally, if the word is a past participle it also changes for number, agreeing with the number of the head-noun. So, for past participles I added their number in the lexicon with the number template, giving them either singular or plural as a value. In the rules I used a template for number agreement (NUM-AGR) to check for agreeing number.

3.5 Adjective phrases

Adjective phrases consist simply of an adjective or an adjective combined with modifiers. This could be an adverb or a prepositional phrase used adverbially. They can either precede or follow the adjective. This was implemented in the rules section using a disjunction for the AP rule, making the A either preceding the PP or Adverb (also a disjunction) which modifies it, or following.

Adjectives follow the noun they modify. They agree with the noun in number. This was implemented via the NUM-AGR template. Adjectives in the lexicon are either singular or plural. In the rules section the rules for NPs say that an adjective following a noun must agree in number by invoking the NUM-AGR template for the AP. Adjectives often undergo soft mutation, but sometimes they don't. Following Salo (2004) there is no obvious pattern for this behaviour, so I only used examples in the testsuite found in his grammar.

3.4 Prepositional phrases

Prepositional phrases consist of prepositions as the head, followed by a noun phrase, which is usually undergoing mutation (which again, was not implemented but used the right way in the testsuite). Prepositions in Sindarin show the relationship of one object or event to another (e.g. local or temporal) (vgl. Salo 2004: 139). Since prepositions take an NP as their complement, prepositions can also be followed by a pronoun. In that case it's either a pronoun taking the accusative, like in *maethar yrch ad mín* "Orcs fight against us", or a pronoun in the dative case, for example a construction like *glîr ammen* "he/she sings for us". In the latter case the preposition merges with the pronoun. Apparently only the preposition *an* "for, to" can form the dative with pronouns. This is implemented as the following: In the PP rule, PPs either have a preposition (P) as the head, followed by an NP in the accusative **or** they take an empty node as the missing/merged preposition *an*, followed by an NP as an oblique dative (OBL-DAT), which has to be a pronoun (constraining equation) and takes the dative case (defining equation). The OBL-DAT was then added as an option to the transitive and intransitive template for verbs and the dative pronouns were marked for their case in the lexicon.

4. Known bugs

XLE wouldn't accept *dûr* for some reason (either this is an error in the terminal or in the lfg-file), although it accepted other words like *câb* and *rhîw* with a circumflex. When typing "dûr"

for parsing, it gave me an error because of unknown words. After I changed “dûr” to “dur”, it worked fine.

5. Discussion

Given the limited amount of time and workforce, not all phenomena of the Sindarin language could be implemented. The goal was to give an overview of the basic syntactic structures, how they are realized and how they interact with each other in Sindarin, especially compared to English or German. Additionally, given the fact that Sindarin is a fictional language, documentations of this language are often very open or diffuse and sometimes even contradictory. Tolkien himself unfortunately never gave a formal description of his Sindarin language and therefore everything what is known derived from analysing existing sentences or fragments. This project relies particularly on David Salo’s grammar of Sindarin, which is very detailed and elaborate. Based on his work, the fundamental features of the Sindarin language could be modelled using the XLE tool.

References

Butt, Miriam & Tracy Holloway King & María-Eugenia Niño & Frédérique Segond. 1999. *A Grammar Writer’s Cookbook*. CSLI Publications.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. Syntax and Semantics, volume 34. New York: Academic Press

Sadler, Louisa. 1997. *Clitics and the Structure-Function Mapping*. LFG97 Conference. CSLI Publications.

Salo, David. 2004. *A Gateway to Sindarin: a grammar of an Elvish language from J.R.R Tolkien’s Lord of the rings*. Salt Lake City: The University of Utah Press.