FST Morphological Analyser and Generator for Mapudüngun

Andrés Chandía
Universitat Pompeu Fabra
Translation and Language Sciences Faculty
E-mail: andres@chandia.net / andres.chandia@upf.edu

Abstract Following the Mapuche grammar by Smeets, this article describes the main morphophonological aspects of Mapudüngun, explaining what triggers them and the contexts where they arise. We present a computational approach producing a finite state morphological analyser (and generator) capable of classifying and appropriately tagging all the components (roots and suffixes) that interact in a Mapuche word form. The bulk of the article focuses on presenting details about the morphology of Mapudüngun verb and its formalisation using FOMA. A system evaluation process and its results are also present in this article.

Keywords Mapudüngun · Mapuche · Morphology · Finite State Transducer · Analyser · Generator · FOMA

1 Introduction

This article explains the morphophonological aspects of Mapudüngun which have to be taken into account when developing a rule based morphological analyser, which is our purpose.

The implementation we have chosen is by means of Finite State Transducers (FST). The language that feed these machines (the FST) is made out of complex regular expressions which need to encode from the language of interest, Mapudüngun in this case, the way the different elements (roots and suffixes) interact, the conditions they have to fulfill in doing so, and the changes this very same interaction produces among the elements.

So, this work tells how we have translated Mapudüngun’s morphophonological behaviour into regular expressions, which have to be as accurate as possible in order to obtain optimal results by means of an FST analyser\(^1\).

Along section 2 and 3 we present the Mapuche language, its typology and morphology are the central topics, which include the suffixes of this language and how the verbs are formed, from the stem to the final form, along with some exceptions and particularities.

Section 4, p. 14, is centred in the computational technology we use and the specific tool to achieve our goal. We begin by explaining what computational morphology implies, and how it can be handled by Finite State Transducers (FST). FOMA is the FST compiling program we use to generate our tools, so we do a review of it. And finally, we refer to the first steps in the incorporation of the Mapudüngun elements into the computational flow of work.

Section 5, p. 21 presents the embodiment of the processes and phenomena explained in the section describing Mapudüngun into the code the compiler is capable of interpret and process. The techniques applied in order to encode the different parts of Mapudüngun are also described, and the rules that manage their interaction and changes derived from it; and how the different aspects of Mapudüngun morphology are treated from the computational point of view. We will explain in detail the stem typology, and the strategies to manage them; the interaction of suffixes after the stem, verb paradigms and verb nominalization. The mobility of some suffixes and the special behaviour of some verb roots are also presented in this section.

In section 6, p. 41, we do account of some Mapudüngun realizations that come from other sources and dialects, different from Smeets’ work which is the base of our development. We explain how and why we have incorporated them into our system.

\(^1\) The system user interface is is available on http://www.chandia.net/dungupeyem and the code is on http://www.chandia.net/dungupeyem/repositorio
A brief count on the FST analyser comes in section 7, p. 46, where we display data on the amount of lexicon, suffixes, and rules, besides the compilation values.

Section 8, p. 47 brings up the subject of assessment. We explain how our machine has been evaluated, the parameters taken into account. We introduce some other machines to compare to, and also another similar system; all for the sake of an accurate comparison and subsequent evaluation of the outcome the system produces. In this section there is also a comprehensive analysis of the forms that were not recognized by the system, and the reasons for that.

The final section 9, p. 54 (before the conclusions, p. 46), is just a brief account of the web interfaces we have developed to access our tools. We simply show the elements found on these interfaces and how to operate them.

2 Mapudüngun, the Mapuche language

Along this section we present Mapudüngun, its location, typology and a basic description of its conformation, which includes the phonemes it presents and their graphic representation. Morphology comes next, where we present some information about Mapudüngun suffixes and how the verbs are formed. Finally, we introduce the stems formation and its particularities. Mainly, we present the morphophonological aspect of some specific phenomena of Mapudüngun. A complete description of the language is found in the book we base our analyser upon: “A Grammar of Mapuche” by Ineke Smeets [21].

Mapudüngun is an isolated language2 spoken actively by approximately 144,000 people in Chile [Zúñiga 2006] [24], as well as by some 8,400 people in Argentina [Instituto Nacional de Estadísticas y Censos 2005], virtually all of whom are bilingual in Spanish [Sadowsky, S. 2013: 87-96] [18].

The word Mapudüngun is a compound of two nominal roots, mapu meaning ‘soil, land, earth, ground, country’; diengu meaning ‘language, matter, subject, tongue (as in mother tongue)’. Mapudüngun is usually translated as ‘the language of the land, the speaking of earth’ or ‘lengua/habla de la tierra’ in Spanish.

2.1 Mapudüngun: polysynthetic and agglutinative

Polysynthesis means that there are many elements or morphemes in (verb) forms, which is typical of the Native American languages, Mapudüngun among them.

In agglutinative languages a series of concepts are distributed in several morphemes [Zúñiga 2006: 199] [24]. Agglutination is when morphemes are inside words, not altering their own form and being identifiable in different contexts, as in Basque, Turkish, Quechua and Mapudüngun. The original meaning of the stem is modified by the affixes attached to it.

In Mapudüngun, verbs may contain many morphemes, e.g., di-tu-l-me-tu-a-fi-ñ ‘I will reach it, I will find it’. This word has eight significant elements.

“The fact that the language is polysynthetic means that it is rich, in terms of the ability to create new words ” [Zúñiga 2006: 202] [24].

In Mapudüngun, nominal forms are simple, while verbal ones are extremely complex, presenting a good number of derivative and inflectional morphemes. They can realize as univalent, only one actant as subject; bivalent or monosynthetic verbs, two actants, subject and object; and trivalent or bi-transitive verbs, three actants, subject, primary object and secondary object with the semantic roles of agent (A), human receiver (R), and inanimate patient or theme (T), respectively.

“In verbal phrases there are morphemes that behave as verbal derivatives (verbalizer, causative, transitivizer, benefactive/malefactive, modal, locative and directional, manner; and affixes that are part of non-finite verb forms), also obligatory verbal inflectional suffixes (time, mode, person and number) and facultative (negation, aspect, passive, reflexive/reciprocal/medial and mediative)” [Fernández-Garay & Malvestitti 2002: 36-37] [7].

2.2 The Mapuche alphabet

Smeets states that 19 consonants [table 1] and 6 vowels [table 2] form the Mapuche phonemic system.

| Table I Consonants [Smeets, I. 2008: 23] [21] |
|-----------------------------------------------|
| Labial | Interdental | Palatal | Retroflex | Velar |
| Plosives | p | t | ch | tr | k |
| Fricatives | f | s | sh | | |
| Glides | w | y | r | q |
| Nasals | m | n | ñ | ng |
| Laterals | l | ll | | |

Smeets also adds loaned sounds from Spanish b, d, g (as in Spanish “bote”, “duende” and “guerra” respectively) and the voiceless fricative x (as in Spanish ”jefe”). She does not include the interdental series present in some Mapudüngun variants, usually represented as ’l’, ’n’, ’t’; because the dialect she studied did not present it, and her data, “in agreement with Croese’s findings, do not call for a distinction between

---

2 The relationship between Mapudüngun and others aboriginal American languages has not yet been established.
3 Mapudüngun morphology

In this section, and along this article, we mainly refer to the verb morphology because it is the more complex part of Mapudüngun, and virtually all morphophonological changes are found inside the verb form. Other categories of words are mentioned because they occur as verbal stems together with a verbalizing suffix. In other cases they are used to bringing up a special case, or because a specific suffix also interacts with nouns, adjectives, adverbs, etc. in a nonverbal form. This does not mean that our work only covers the Mapuche verb; but for making this article not too extensive we do not expand the topic to all parts of speech (categories).

3.1 Verb suffixes

We begin by exposing suffixes because they occur in almost all verb stems, the only stems where a suffix do not occur are those formed by a single verbal root. But in Mapudüngun, adjective, adverb, noun and other roots need a suffix to become verbal stems.

To simplify, we call verb stem to any form, simple or complex, to which suffixes are bonded in order to form a complete verb predication. A simple stem is made of one root only, a complex stem may imply two roots, a root with some suffixes, or a combination of them all. It may be argued that these are lemmas instead, but as we say, to keep it simple, we call all these forms, verb stems. More details are found in sections 3.1.8 Verb derivational nominalization, p. 8; 3.2 Verb stems, p. 9 and 5.2.1 Stems codification, p. 29.

After the stem, in a Mapuche verb form, suffixes “occur in a more or less fixed position relative to one another” [Smeets, I. 2008: 17] [21]. But also there are quite a few incidental factors that shape the Mapuche complex verb form.

Verb suffixes are located on one of the thirty-six slots assigned to the verb form on the basis of their relative position and function. Slot 1 occupies word final position and slot 36 is next to the root. The order of these slots determine the morphotactics of the verb forms. Some slots host a few mutually excluding affixes, some of them may present variation in their form and some others may be zero markers. Some suffixes may exclude others from different slots for grammatical or semantic reasons.

Even though it is not rare to find up to seven or eight suffixes following the root (see E1), verbs usually contain between four and six suffixes in spontaneous speech.

In the following lines we try to graphically represent three different Mapuche verb forms. S represents the stem. Every dot represents a slot; the leftmost dot is slot 36, the rightmost dot is slot 1. X is a suffix occurrence in a slot. Ø is also a suffix occurrence but with a null morpheme, which is a morpheme that has no phonemic or graphic realization.

| Front | Central | Back |
|-------|---------|------|
| High  | i       | ü    |
|       | u       |      |
| Mid   | e       |      |
|       | o       |      |
| Low   | a       |      |

Table 2 Vowels [Smeets, I. 2008: 25] [21]

The interdental” and the alveolar series l, n, r. ”A tentative conclusion might be that the distinction is dying out” [Smeets, I. 2008: 31] [21].

Minimal intransitive verb 2nd person plural
S . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X X X . . .

Minimal transitive verb 2nd → 1st persons plural
S . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . X Ø X . . .

Representation of example E1
S . . . X . . . X . . . X X X . . . . . . . . . X . X Ø X X

(E 1) Verb with 10 suffixes [Smeets, I. 2008: 443 (76)] [21]
nü-nie-hma-r-pu-tu-e-y-iñ-mu
‘they continued to take it away from us’
Root: nü- → TV.nü_tomar

Suffixes:
1. -nie- Progressive persistent (+PRPS.nie32)
2. -hma- Indirect object (+IO.hma26)
3. -r- Interruptive (+ITR.r18)
4. -pu- Locative (+LOC.pu17)
5. -tu- Iterative/restorative (+RE.tu16)
6. -e- Internal direct object (+IDO.e6)
7. -y- Indicative (+IND.y4)
8. -Ø- First person (+1.Ø3)
9. -iñ- Plural (+PL.iñ2)
10. -mu Dative subject(+DS3A.mew1)

In example E1 the root and suffixes are displayed as items to better identify them, but the analyser output is visualized linearly, as follows:
- TV.nü_tomar+PRPS.nie32+IO.hma26+ITR.r18+LOC.pu17+RE.tu16+IDO.e6+IND.y4+1.Ø3+PL.iñ2+DS3A.mew1

Analysis tags express, starting from the left, the abbreviated name of the part of speech (PoS) or suffix. PoS are introduced by a - (minus) sign, suffixes, by a + (plus) sign. -TV is ‘transitive verb’, -IV is ‘intransitive verb’, -Ø is ‘noun’, etc.

Concerning suffixes, +PRPS is ‘progressive persistent’, +IDO is ‘internal direct object’, +PL is ‘plural’, etc. A complete list of tags meaning is found in annex 11.1 Tags meaning, p. 55.
After the abbreviated name of the PoS or suffix, separated by a dot, it is the root or suffix standard form. Roots are followed by their meaning in Spanish with an underscore _ as separator. _nu_ tomar in E1. For the already mentioned suffixes, the forms are _n_e_, _e_ and _i_n_ respectively.

The number at the end of each tag indicates the slot (the position in the verb chain) where the verb suffix is located.

For instance:

- TV._nu_romar: "the transitive verb root _nu_ which means 'tomar' in Spanish ('take')"
- PRPS._nie32: "the progressive persistent suffix, which form is _nie_, is located in slot 32".
- IDO._e6: "the internal direct object suffix, which form is _e_, is located in slot 6, etc.

These items will be the result from the analyser we have built, and similar structure and information will be shown in all the examples.

Slots 1 to 15 hold inflectional suffixes in fixed positions. Slots 16 to 27 hold derivational suffixes, some of which are mobile. Slots 28 to 36 hold derivational suffixes in fixed positions, except for the rather mobile suffix _-w_, which usually fits in slot 31 and marks reflexivity/reciprocity. Mobile suffixes are assigned to their most usual position. "A difference in order of the suffixes does not always result in a semantic difference" [Smeets, I. 2008: 177] (see sections 3.1.5, p 5 and 5.3.4, p. 38).

3.1.1 Verbalizers (slot 36)

Nouns, adjectives, adverbs and numerals (roots) "can be changed into verbs by means of suffixation" [Smeets, I. 2008: 304] [21]. "There are six verbalizing suffixes. They immediately follow the root and fill slot 36" [Smeets, I. 2008: 121] [21].

1. Suffix _-Ø_- indicates the verbalization of a noun, adjective, numeral and a number of adverb roots.
2. Suffix _-_ verbalizes noun, adverb, numeral roots and the interrogative pronoun _nute_- 'how much'.
3. Suffix _-nge_- can verbalize noun, adjective, numeral roots and the interrogative element _chum_- 'how'. A verb formed with _-nge_- is intransitive.
4. Suffix _-_ verbalizes adjectival roots.
5. Suffix _-_ verbalizes noun roots.
6. Suffix _-_ verbalizes noun roots.

( E 2 )

mapu-che 'person of the land (Mapuche person)'
-NN.mapu_tierra-NN.che_persona

mapu-che-nge 'I am a person of the land (a Mapuche)'
-NN.mapu_tierra-NN.che_persona
+VRB.nge36-IV+IND1SG.n3

3.1.2 Stem formative (slot 36)

Reduplication is another resource in Mapudungun, and reduplicated roots are also used to form verbs, but for doing so, they are obligatorily followed by a verbalizing suffix, even when it is a reduplicated verb root, in this case Smeets calls these suffixes "stem formative in reduplicated roots (SFR)" and they are also assigned in slot 36. There are four stem formatives:

1. Suffix _-Ø_- occurs when the reduplicated root is an onomatopoeia or a verb. The resulting verb is intransitive.
2. Suffix _-nge_- is added to reduplicated verb roots, the resulting verb is intransitive.
3. Suffix _-tu_- is added to reduplicated noun or verb roots. The resulting verb of a reduplicated verb root has the same valence as the single form.
4. Suffix _-ye_- is added to reduplicated verb roots, the resulting verb is transitive.

( E 3 ) [Smeets, I. 2008: 481 (33)] [21]

\[ hiwa-hiwa-tu-fu-n 'I always did my best' \]
-IV.niwa_esforzar-RVBR+SFR.tu36+IPD.fu8+IND1SG.n3

3.1.3 Derivational suffixes (slots 16 to 35)

From slot 16 to 27 the suffixes mostly act as semantic modifiers. From slot 28 to 35, they have an aspectual or valency function. Suffixes most commonly used are:

- Causatives _-l_- (e.r. 3 11.2.6) and _-_ (e.r. 11.2.6), slot 34. These suffixes make the event denoted by the stem to be actually applied or happen, in this sense, they operate as transitivizers also.
- Factitive _-ka_- (e.r. 11.2.6) and transitivizer _-tu_- (e.r. 11.2.6), slot 33. It indicates that the agent causes the event denoted by the verb to take place, often it also adds intensive value.
- Reflexive/reciprocal _-w_- (e.r. 11.2.6), slot 31. It indicates reflexivity when combined with a singular subject. The reflexive morpheme -(u)w- indicates reflexivity or reciprocity when it combines with a dual or plural subject.
- Stative _-le_- (e.r. 11.2.6), slot 28. It denotes a state which may or may not involve agentivity on the part of the subject. With a few verbs, it may denote either an ongoing event or the resulting state. It may be used to indicate a quality or characteristic that is not permanent or intrinsic.

\[ 3 \] e.r. stands for "examples references". In the referred section there is a list of the examples that contain the suffix being mentioned. The entire list is under annex 11.2.6-Examples by suffixes, p. 58.
Beneficiary -el- (e.r. 11.2.6), slot 27. It makes the (animate) patient become the beneficiary of the event.

Passive -nge- (e.r. 11.2.6), slot 23. It indicates that a participant, a 3rd person with the role of agent, is not found in the sentence described by the sentence, but outside the speech act.

1st person agent -w- (e.r. 11.2.6), slot 23. It indicates a non-declared participant to be determined by the context. Which is a 1st person non-singular, the agent, and implicitly includes the listener who is the patient.

Thither -me- (e.r. 11.2.6), slot 20. It indicates that the denoted situation involves motion away from the speaker or another orientation point, with a connotation of temporariness.

Persistence -we- (e.r. 11.2.6), slot 19. It indicates a situation which persists after a previous event has taken place.

Hither -pa- (e.r. 11.2.6), slot 17. It indicates that the denoted situation either involves a movement towards the speaker or takes place at a location near the speaker. It may indicate a development towards the present.

Locative -pu- (e.r. 11.2.6), slot 17. It indicates that the event takes place away from the speaker. It does not imply motion and indicates a permanent situation.

3.1.4 Inflectional suffixes (slots 5 to 15)

Among these suffixes are those that indicate aspect, tense, negation and truth value:

Pluperfect -wye-, slot 15. Indicates that the event takes place before the past or future orientation moment (see following example and E5).

( E 4 ) [Smeets, I. 2008: 69 (62)]
tripa-wye-ye 'he had left'
-IV.tripa_salir+PLPF.wye15+IND.y4+3.Ø3

Constant feature -ke- (e.r. 11.2.6), slot 14. Indicates a constant or characteristic feature of the subject.

Proximity -pe- (e.r. 11.2.6), slot 13. It seems to indicate an event or a feature in the recent past, a strong probability and doubt.

Reportative -rke-, slot 12. It indicates that the situation has not been directly witnessed; the speaker has been informed by others, has heard rumours, or has deduced it (see following example).

( E 5 ) [Smeets, I. 2008: 254 (1)]
futa-nge-wye-rke-ye 'she had been married, they say'
-NN.futa_marido+VRB.nge36-IV+PLPF.wye15+REP.rke12+IND.y4+3.Ø3

Affirmative -lle-, slot 11. It adds emphasis (E59).

Non-realized situation -a- (e.r. 11.2.6), slot 9. It denotes a non-actual fact. The situation will take place after the orientation moment.

Impeditive -fu- (e.r. 11.2.6), slot 8. It denotes that the event does not conclude as expected or that it can not be completed.

Pluperfect -mu- (e.r. 11.2.6), slot 7. It indicates that an event is realized before an orientation moment in the past. It occurs in complementary distribution with the pluperfect -wye-, slot 15.

Constant feature -ye-, slot 5. As suffix -ke-, slot 14, it also denotes a characteristic or constant feature, and they appear in complementary distribution (E162).

3.1.5 Suffix mobility

Smeets identify suffixes from slots 28 to 36 as fixed suffixes, and from slots 16 to 27 as mobile, later ones appear in non-common positions respect to other suffixes. A detailed list of the mobile suffixes with their usual position (slot) follows:

Repetitive/Restorative -tu- (e.r. 11.2.6), slot 16. It indicates that a situation is repeated or restored.

Hither -pa- (e.r. 11.2.6), slot 17. (Explained in 3.1.3).

Persistence -we- (e.r. 11.2.6), slot 19. (Explained in 3.1.3).

Thither -me- (e.r. 11.2.6), slot 20. (Explained in 3.1.3).

Immediate -fem-, slot 21. It denotes immediate action (see following example).

( E 6 ) [Smeets, I. 2008: 271 (20)]
nye-nge-fem-ûy 'it was brought immediately'
-IV.ye_traer+PASS.nge23+IMM.fem21+IND.y4+3.Ø3

Sudden -rume-, slot 21. (E162). It denotes sudden action.

Play -kantu-, slot 22 (E200). It denotes an action performed in jest, for fun, not in earnest, or just to pretend to be doing.

Simulative -fuluw-, slot 22. It indicates simulation, not real intention to do something (see following example).

( E 7 ) [Smeets, I. 2008: 265 (9)]
illku-le-fuluw-ûn 'I pretended to be angry'
-IV.illku_enojar+ST.1e28+SIM.fuluw22+IND1SG.n3

Passive -nge- (e.r. 11.2.6), slot 23. (Explained in 3.1.3).
Pluralizer -ye-, slot 24 (E91). It is especially used with intransitive verbs which take a 3rd person subject. With a 1st or 2nd person plural subject, it indicates a numerous subject. With transitive verbs, it indicates that numerous patients of the event.

Force -fal-, slot 25 (E155). It indicates either that there is a necessity or obligation for the subject to perform the action, or that the subject orders someone else to perform the action.

Beneficiary -el- (e.r. 11.2.6), slot 27. (Explained in 3.1.3).

Stative -le- (e.r. 11.2.6), slot 28. (Explained in 3.1.3).

Reflexive/Reciprocal -w- (e.r. 11.2.6), slot 31. (Explained in 3.1.3).

Transitivizer -tu- (e.r. 11.2.6), slot 33. It may be added to intransitive and transitive verbs, and it adds an object. With intransitive verbs, the form has one object. With transitive verbs, the form has two objects.

Mobility does not imply a semantic change, and as more suffixes a verb presents less displacement occurs. See the following examples:

( E 8 ) [Smeets, l. 2008: 270 (19)] [21]
ngilla-1-me-mu-y-iñ 'you went to buy for us'
-TV.ngilla_comprar+BEN.el27+TH.me20+2A.mu23
+IND.y4+1.Ø3+PL.iñ2

( E 9 ) [Smeets, l. 2008: 263 (11)] [21]
i-me-we-ke-lá-y 'he no longer always eats there'
-TV.i_comer+TH.me20+PS.we19+CF.ke14+NEG.1a10
+IND.y4+1.Ø3+PL.iñ2

( E 10 ) [Smeets, l. 2008: 421 (62)] [21]
pütu-yekó-me-tu-y-ng-ün 'they drank all the time'
-TV.püto_beber+ITR.yekó18+TH.me20+RE.tu16
+IND.y4+3.ng3+PL.ün2

In the examples above, the thither suffix -me- presents three different positions respect to the other suffixes. In E8 it is close to slot 27, displaced beyond slot 23. In E9 it occurs in its usual position, just before the suffix -we- , slot 19. Finally, in E10, it appears between suffixes of slots 18 and 16, to the right of its usual position.

3.1.6 Verb paradigms

In the previous section we have skipped suffixes from slots 10 and 6, they take part in the transitive verb paradigm, we include them here. Suffixes of slot 23 are also included in this paradigm together with those of mood, person, number and dative subject of slots 4, 3, 2 and 1 respectively.

Negation, positioned in slot 10, may actually be part of transitive and intransitive forms. There are three negation morphemes, one per mood, reason to show them in the verb paradigms.

The simplest verb form is intransitive, less suffixes than in transitive forms are mandatory: mood (slot 4), person (slot 3) and number (slot 2). See examples below:

( E 11 )
küpa-y-m-i 'you (sg) came'
-IV.küpa_venir+IND.y4+2.m3+SG.12

( E 12 )
küpa-la-y-m-u 'you two did not come'
-IV.küpa_venir+NEG.1a10+IND.y4+2.m3+DL.u2

( E 13 )
küpa-no-l-m-ün 'if you (pl) do not come'
-IV.küpa_venir+NEG.no10+CNĐ.14+2.m3+PL.ün2

The imperative mood have forms for 1st person singular; 2nd person singular, dual and plural; and for 3rd person (undefined number). Indicative forms of 1st person dual and plural may be used adhortatively. Negation suffix for imperative is -ki- (slot 10), which always co-occur with the conditional marker -l- (slot 4). Negation of the adhortative forms, which are indicative, is accomplished by the -ki-l- combination of imperative negation and conditional mood mark when the intention is imperative (adhortative). See examples below (the complete conjugation of the intransitive verb küpa- 'to come' is in annex 11.3, table 13):

( E 14 )
küpa-m-u 'come, you both!'
-IV.küpa_venir+IMP.Ø4+2.m3+DL.u2

( E 15 )
küpa-ki-l-chi 'I better not come'
-IV.küpa_venir+NEG.ki10+CNĐ.14+IMP1SG.chi3

( E 16 )
küpa-y-u ind: 'we both came' imp: 'let we both come'
-IV.küpa_venir+IND.y4+1.Ø3+DL.u2

( E 17 )
küpa-ki-l-y-u 'let we both not come'
-IV.küpa_venir+NEG.ki10+CNĐ.14+1.y3+DL.u2

The transitive paradigm demands more suffixes to reflect the relations between agent, patient and object. Interacting also the suffixes -w- 1st person agent, and -mu- 2nd person agent, slot 23; -e- internal direct object and -f- external direct object, slot 6; and -Ø- dative subject, 1st or 2nd person

---

4 Even though the form is the same, we have labelled it +CNĐ 'conditional' and +CNĐ 'conditional marker in imperative forms, to better distinguish them'.
agent, and -mew- ~ -ew- dative subject, 3rd person agent (no number), slot 1. A complete explanation of the transitive paradigm is in chapter 26 "Slots" of Smeets 2008 [21]. See examples below (the complete conjugation of the transitive verb pi- 'to say (to tell)' is in annex 11.4, table 14 and the negative imperative forms in annex 11.5, table 15):

(E 18)  
pi-e-y-u 'I told you'  
-TV.pi+IDO.e6+IND.y4+1.Ø3+DL.u2+DS12A.Ø1

(E 19)  
pi-mu-l-i 'if you (non-sg) tell me'  
-TV.pi+2A.mu23+CND.l4+1.i3+SG.Ø2

(E 20)  
pi-fi-m-u 'tell him, you both'  
-TV.pi+EDO.fi6+IMP.Ø4+2.m3+DL.u2

(E 21)  
pi-ki-fi-l-y-iñ 'let us (pl) not tell him'  
-TV.pi+NEG.ki10+EDO.fi6+CNI.l4+1.y3+PL.iñ2

3.1.7 Verb inflectional nominalization

In a Mapuche sentence, subordinates are derived from verbs, nominalized by inflectional nominalizers. These suffixes share position in slot 4 with mood markers, therefore, a verb form is either finite or nominalized. Finite forms take mood, person and number. Nominalized forms can not take those suffixes, taking instead one of the inflectional nominalizers.

Besides as subordinates of verbs, nominalized verbs may also act "as subject, direct object, instrumental object or complement noun phrase, indicating an event as such, a participant, an instrument, time, place, reason, purpose or background event" [Smeets, I. 2008: 188] [21]; as noun modifiers, and as predicates in nominal sentences.

"Some nominalized forms can be used as a finite verb form. The subject of a subordinate is usually indicated by a possessive pronoun, which immediately precedes the subordinate. However, when a subordinate is used as a temporal or causal clause, or as a finite verb form, the subject is indicated by a personal pronoun" [Smeets, I. 2008: 189] [21]. There are seven inflectional nominalizers:

Agentive verbal noun -t-
This suffix may denote an event as such; an instrument or location, and the patient or agent of an event.

(E 22)  
tüfa ñi pi-e-t  
-AP.tüfa_este -SP.ñi_mi_su  
-TV.pi_decir+IDO.e6+AVN.Ø36+CONT.Ø2

Subjective verbal noun -lu
This suffix denotes the subject of an event. It may also be used as an active participle, and form a temporal or causal clause.

(E 27)  
pichi che kim-nu-lu  
-AP.pichi_pequeño -NN.che_persona  
-TV.kim_saber+NEG.no10+SVN.Ø36+CONT.Ø2

Instrumental verbal noun -m
This suffix may indicate an instrument, a location, or an event as such. In combination with -a- non-realized action (slot 9), it may indicate purpose. With -ye- constant feature (slot 5), it forms a temporal clause.

(E 24)  
închi family lleg-mu-m  
-NN.-PP.nînchî_nosotros  
-AP.ta_el-SP.yîn_nuestro-s  
-IV.lleg_crecer+PLPF.mu7+IVN.Ø4+2.m3+DL.u2

Objective verbal noun -el
This suffix expresses a passive participle, indicating the patient of the event. It can also be used to indicate an event as such; and rarely it is also used as an instrumental or locative.

(E 25)  
kuyfi pichi-ka-el  
-AP.kuyfi_antes  
-AP.pichi_pequeño+VRB.Ø36+CONT.ka16+OVN.Ø36+CONT.Ø2

Instrumental verbal noun -a-
This suffix may indicate an instrument, a location, or an event as such. It can also be used as an active participle, and form a temporal or causal clause.

(E 23)  
fûta-nge-wma-rke  
-NN.fûta_marido+VRB.nge36-IV+CSVN.wma4+REP.rke

Objective verbal noun -Ø
This suffix expresses an event as such, without time mark. It can convert the form into an adjective denoting an attribute or quality of the modified noun. It can also form a noun denoting a person or thing involved in the event referred to by the verb. It is usually translated as an infinitive: küdaw 'the work', küdaw-ün 'to work'.

(E 26)  
pütrem-tu-n kûme-la-y  
-NN.pütrem_tabaco+VRB.tu36+PVN.n4

Subjective verbal noun -lu
This suffix denotes the subject of an event. It may also be used as an active participle, and form a temporal or causal clause.

(E 27)  
pichi che kim-nu-lu  
-AP.pichi_pequeño -NN.che_persona  
-TV.kim_saber+NEG.no10+SVN.Ø36+CONT.Ø2

'he has been married, some say'
Transitive verbal noun -fiel
This suffix may be used as an infinitive, passive participle, locative or instrumental.

(E 28 ) [Smeets, I. 2008: 237 (16)] [21]
ínche múle-y mi pe-a- fiel
-PP.ínche_yo -IV.müle_estar+IND.y4+3.Ø3
-SP. mí_tuyo -TV.pe_ver+NRLD.a9+TVN. fiel14
 'I have to see you (sg)’ lit: ‘I am in your will be seen’

3.1.8 Verb derivational nominalization

Some non-verbal suffixes can turn a verb into an adjective or a noun; the stem may be formed by a unique root, a verbal compound, a verbalized root, a verbalized compound or a reduplicated root; or even by a complex stem, a root followed by some suffixes, mainly from slots 35, 34 or 33.

-fal +ADJDO indicates that the event denoted by the verb can actually be done.

(E 29 ) [Smeets, I. 2008: 312 (12)] [21]
pepi-l-fal 'feasible, practicable'
-IV.pepi_poder-hacer+CA.l34+ADJDO.fal

-fe +NOMAG denotes a characteristic agent.

(E 30 ) [Smeets, I. 2008: 311 (1)] [21]
kofke-tu-fe 'bread eater'
-NN.kofke_pan+VRB.tu36+NOMAG.fe

-nten +ADJQE indicates that the event denoted by the verb may be realized quickly and/or easily.

(E 31 ) [Smeets, I. 2008: 312 (14)] [21]
afü-nten 'it gets quickly cooked'
-IV.afü_cocinar+ADJQE.nten

-we +NOMPI denotes a characteristic place or instrument.

(E 32 ) [Smeets, I. 2008: 312 (9)] [21]
püra-püra-we 'stairs'
-IV.püra_subir-RVBR+SFR.Ø36+IV+NOMPI.we

3.1.9 Non-verbal suffixes

In the previous section there were already presented four suffixes that can turn verbs into adjectives or nouns; these resulting forms may, in turn, be complex verb stems, i.e., a verb converted into an adjective or a noun may be used as a verb stem, it may be “re-verbalized”.

There are other suffixes that act upon non-verbal forms; the final form, i.e., "non-verb + suffix", in its turn can also be a complex verb stem (see 3.2 Verb stems, p. 9). Some of these suffixes change the class (category) of the form they are attached to, and some others do not.

Class-changing suffixes (CC)

-chi +ADJ changes a noun or nominalized verb into an adjective.

(E 33 ) [Smeets, I. 2008: 114 (25)] [21]
lef- chi che 'runner’ lit: ‘running person’
-IV.lef_correr+SVN.Ø4+ADJ.chi
-NN.che_persona

-tu +ADV changes a noun or nominalized verb into an adverb.

(E 34 ) [Smeets, I. 2008: 114 (b)] [21]
amu-n-tu 'going’, 'on my way there’
-IV.amu_ir+PVN.n4+ADV.tu

Non class-changing suffixes (NCC)

-ke +DISTR is affixed to adjectives, adverbs and numerals. It indicates a whole consisting of several component parts, each of which has the feature expressed by the form it accompanies.

(E 35 ) [Smeets, I. 2008: 112 (17)] [21]
küla-ke 'a threesome'
-NU.küla_tres+DISTR.ke

-em +EX is affixed to a noun of which indicates that is dead or no longer in function or existence.

(E 36 ) [Smeets, I. 2008: 110 (6)] [21]
ley-tüfa ñi küdaw- yem 'this was my former job’
-SP.ñi_mü mi su-SP.küdaw_trabajo+EX.em

-ntu +GR it refers to a group as a whole or a place which is characterized by the presence of many items referred to by the noun.

(E 37 )
küra-ntu 'scree'
-NN.küra_piedra+GR.ntu

-rke +REP indicates that the situation or thing expressed by the form it accompanies has not been witnessed by the speaker himself. The speaker has been informed by others, he has heard rumours or he has deduced a conclusion. It may express surprise after the sudden realization of something.

(E 38 ) [Smeets, I. 2008: 110 (8)] [21]
trewa-rke! 'a dog’!, 'what a big dog’!; 'it must have been a dog’ (at wondering about who ate the meat that disappeared’
-NN.trewa_perro+REP.rke

-we +TEMP indicates a period subsequent to an orientation moment.
FST Morphological Analyser and Generator for Mapudungun

( E 39 ) kechu-we antü ‘in five days’
-NU.kechu_cinco+TEMP.we -NN.antü_sol_día
-wen +REL refers to the people relation indicated by the noun it accompanies.

( E 40 ) kompañ-wen iñchiu ‘we are partners’ lit: ‘we both are partners of one another’
-NN.kompañ_compañero+REL.wen -PP.iñchiu_nosotros-dos

3.1.10 Instrumental object suffix -mew

This suffix may never be part of a complex stem but it may be added to nominalized verbs (E 43), nouns and pronouns. It indicates instrument, place, time, cause and is used in comparative and partitive constructions. It may also refer to the circumstances under which an event takes place. See next examples:

( E 41 ) [Smeets, I. 2008: 62 (5)]
anel-tu-fi-ñ kiñe kuchillo-mew ‘I threatened him with a knife’
-TV.anel_amenazar+TR.tu33+EDO.fi6+IND1SG.n3
-NU.kiñe_uno -NN.kuchillu_cuchillo+INST.mew

( E 42 ) [Smeets, I. 2008: 62 (9)]
uma-pu-n ta-ñi peñi-mew ‘I stayed at my brother’s’
-IV.uma_pernoctar+LOC.pu17+IND1SG.n3
-AP.ta_el-SP.ñi_mi_su
-NN.peñi_hermano+INST.mew

( E 43 ) [Smeets, I. 2008: 62 (7)]
are-tu-n-mew monge-li-y ‘he lives on borrowing’
-IV.are_prestar+TR.tu33+PVN.n4+INST.mew
-IV.monge_vivir+ST.1e28+IND.y4+3.Ø3

3.2 Verb stems

We have classified different types of stems depending on the way they are composed. In this respect we do not strictly follow Smeets. Verbs stems are completed by the verbalizing suffix (+VRB) when there is no verb root present, or a stem formative (+SFR) when there is a reduplicated root.

The following list shows the different types of stems from the simplest to the most complex ones:

Simple stems
- Verb root

( E 44 ) [Smeets, I. 2008: 64 (29)]
amu-y-ng-un ‘they (pl) went’
-IV.amu_ir+IND.y4+3.ng3+PL.un2
- Verb compound (verb root + verb root)

( E 45 ) [Smeets, I. 2008: 420 (54)]
ammu-mayna-tu-e-n-ew ‘he made me stumble’
-IV.amu_ir-TV.mayna_atar-los-pies-CR.TV
+TR.tu33+IDO.e6+IND1SG.n3+DS3A.e1w

- Verb compound (verb root + non-verb root / non-verb root + verb root)

( E 46 ) [Smeets, I. 2008: 401 (38)]
ad-kintu-a-’l ‘to have a look’
-NN.ad_forma-TV.kintu_mirar+NRLD.a9+OVN.e14

- Non-verb root +VRB (verbalizer suffix)

( E 47 ) [Smeets, I. 2008: 68 (56)]
küla antü-nge-y ‘it was three days ago’
-NU.küla_tres -NN.antü_sol_día+VRB.nge36-IV+IND.y4+3.Ø3

- Reduplicated root +SFR (stem formative suffix)

( E 48 ) [Smeets, I. 2008: 112 (22)]
aku-aku-nge-y ‘continually arrive (e.g. letters)’
-IV aku_llegar-RVBR+SFR.nge36-IV+IND.y4+3.Ø3

- Non-verb compound (non-verb + non-verb) +VRB

( E 49 ) [Smeets, I. 2008: 123 (11)]
trewa-ad-nge-y ‘he has dog face’
-NN.trewa_perro-NN.ad_cara+VRB.nge36-IV
+IND.y4+3.Ø3

Complex single root stems
- Numeral + non class-changing suffix +VRB

( E 50 ) [Smeets, I. 2008: 400 (30)]
kiñe-ke-l-y ‘he gave one to each of them’
-NU.kiñe_uno+DISTR.ke+VRB.136
+EDO.f16+IND.y4+3.Ø3

- Adjective + non class-changing suffix or inflectional nominalizer +VRB

( E 51 ) [Smeets, I. 2008: 473 (40)]
pichi-n-tu-fe-nge-y ‘he is playful’
-AJ.pichi_pequeño+PVN.n4+VRB.tu36
+CF.ke14+IND.y4+3.Ø3

- Question +VRB + inflectional nominalizer +VRB

( E 52 ) [Smeets, I. 2008: 243 (52)]
chum-ngi-n-tu-y-m-i-’ ‘what do you (sg) think (about it)?’
-QC.chum_cómo+VRB.nge36-IV+PVN.n4+VRB.tu36
+IND.y4+2.m3+SG.i2

- Adjective + inflectional nominalizer +VRB + derivational nominalizer +VRB

( E 53 ) [Smeets, I. 2008: 375 (25)]
awka-n-tu-ma-ýe ‘he is playful’
-AJ.awka_salvaje+PVN.n4+VRB.tu36+OMAG.fe
+VRB.nge36-IV+IND.y4+3.Ø3
Adverb + non class-changing suffix or inflectional nominalizer + optional class-changing suffix +VRB

(E 54) [Smeets, I. 2008: 383 (18)] [21]
\[áu-nu-tu\-y-ng\-ün \(\text{they were more}\)\]
- AV.áu_much+PVN.n4+VRB.tu36
+IND.f4+3.ng3+PL.trn2

Noun + non class-changing suffix + inflectional nominalizer + class-changing suffix +VRB

(E 55) [Smeets, I. 2008: 411 (53)] [21]
tukuyu-ke-chi\-le-wü\-y \('it looks like (long) fabric')
- NN.tukuyu_tela+DISTR.ke+SVN.04+ADJ.chi
+VRB.Ø36+ST.1e28+REF.w31+IND.f4+3.Ø3

Noun + optional transitivizer or facitve + optional reflexive + optional non-realized + class-changing suffix or non-class-changing suffix or inflectional nominalizer +VRB

(E 56) [Smeets, I. 2008: 90 (34)] [21]
\[áš-ka-w\-ün-nger\-y \('he is capricious')\]
- NN.ad_costumbre+FAC.kas3+REF.w31+PVN.n4
+VRB.Ø36-IV+IND.f4+3.Ø3

Verb + optional causative + optional transitivizer or facitve + optional reflexive + optional static + optional hither + optional non-realized + inflectional or derivational nominalizer +VRB

(E 57) [Smeets, I. 2008: 225 (243)] [21]
\[llüka-nten-ngen\-wma \('I was someone who easily gets afraid')\]
- IV.1lluka_temer+ADJQE.nen+VRB.nten36-IV
+CSVN.wma4

Reduplicated verb root + causative (verb +CA + verb +CA)

(E 58) [Smeets, I. 2008: 412 (68)] [21]
ap-üm-ap-üm-ye-ngen\-y \('we have gradually been finished off')
- IV.af_acabar+CA.n34+RVRB+SVF.ye36-IV
+PASS.ngen23+IND.f4+3.Ø3

In the last example, the form corresponds to two roots and two suffixes, but it is actually a reduplicated stem of one root with a suffix attached, reason to list it as a single root complex stem.

Complex compound stems are the most complex ones, they are not listed here but on the section Complex compound stems, p. 34, where we expose them together with the encoding expressions and rules that manage them, these stems are formed by two roots and at least one suffix apart from the verbalizer.

3.3 Special verbs

Some roots (verbs and non-verbs), due to semantic or grammatical reasons, must co-occur with certain suffixes when forming complete verb forms. There are some exceptions and/or conditions needed for these roots to behave this way. Smeets writes about the conditions, we have found the exceptions.

3.3.1 Question roots

Interrogative roots may be verbalized, but not all forms take the same verbalizers (see D25).

chem- 'what, which' may be verbalized by suffixes -Ø- and -ye-, see following examples:

(E 59) [Smeets, I. 2008: 434 (86)] [21]
\[chem-\text{le-a-i-e \('whatever they would do')}\]
- QC.chem_qué_cuál+VRB.Ø36+AFF.11e11+RNL.aa9
+CND.14+3.e3

(E 60) [Smeets, I. 2008: 128 (39)] [21]
\[chem-\text{ye-\text{w-üy-\text{m-u \('how are you both related?')}\]
- QC.chem_qué_cuál+VRB.ye36+REF.w31
+IND.f4+2.m3+DL.u2

\text{chuchi-~tuchi-} 'which' is verbalized by the null suffix -Ø-, see following example:

(E 61) [Smeets, I. 2008: 405 (7)] [21]
\[\text{chuchi-\text{kün-\text{a-f \('how should they carry on')}\]
- QC.chuchi_cuál+VRB.Ø36+PFPS.künu32
+RNL.aa9+OVN.e14

\text{chum-} 'how' is verbalized by -Ø- or -nger-, see following examples:

(E 62) [Smeets, I. 2008: 416 (15)] [21]
\[chum-la-e-n-\text{e-f \('he did not do anything to me')}\]
- QC.chum_cómo+VRB.Ø36+NEG.1a1f+IDO.e6
+IND1SG.n3+DS3A.e1

(E 63) [Smeets, I. 2008: 225 (246)] [21]
\[chum-\text{neg-\text{wma \('how it was')}\]
- QC.chum_cómo+VRB.nten36-IV+CSVN.wma4

\text{tunte-~chunte-} may be verbalized by suffixes -Ø-, -1- and -ntu-, see following examples:

(E 64) [Smeets, I. 2008: 114 (6)] [21]
\[\text{tunte}-\text{n-tu \('for how long')\]
- QT.tunte_cuánto+VRB.Ø36+PVN.n4+ADV.tu

\text{ye-\text{w-üy-m-u \('how are you both related')\}

\text{chuchi-~tuchi-} 'which' is verbalized by the null suffix -Ø-, see following example:

(E 61) [Smeets, I. 2008: 405 (7)] [21]
\[\text{chuchi-\text{kün-\text{a-f \('how should they carry on')}\]
- QC.chuchi_cuál+VRB.Ø36+PFPS.künu32
+RNL.aa9+OVN.e14

\text{chum-} 'how' is verbalized by -Ø- or -nger-, see following examples:

(E 62) [Smeets, I. 2008: 416 (15)] [21]
\[chum-la-e-n-\text{e-f \('he did not do anything to me')}\]
- QC.chum_cómo+VRB.Ø36+NEG.1a1f+IDO.e6
+IND1SG.n3+DS3A.e1

(E 63) [Smeets, I. 2008: 225 (246)] [21]
\[chum-\text{neg-\text{wma \('how it was')\}
- QC.chum_cómo+VRB.nten36-IV+CSVN.wma4

\text{tunte-~chunte-} may be verbalized by suffixes -Ø-, -1- and -ntu-, see following examples:

(E 64) [Smeets, I. 2008: 114 (6)] [21]
\[\text{tunte}-\text{n-tu \('for how long')\]
- QT.tunte_cuánto+VRB.Ø36+PVN.n4+ADV.tu

\text{ye-\text{w-üy-m-u \('how are you both related')\}

\text{chuchi-~tuchi-} 'which' is verbalized by the null suffix -Ø-, see following example:

(E 61) [Smeets, I. 2008: 405 (7)] [21]
\[\text{chuchi-\text{kün-\text{a-f \('how should they carry on')\]
- QC.chuchi_cuál+VRB.Ø36+PFPS.künu32
+RNL.aa9+OVN.e14

\text{chum-} 'how' is verbalized by -Ø- or -nger-, see following examples:

(E 62) [Smeets, I. 2008: 416 (15)] [21]
\[chum-la-e-n-\text{e-f \('he did not do anything to me')}\]
- QC.chum_cómo+VRB.Ø36+NEG.1a1f+IDO.e6
+IND1SG.n3+DS3A.e1

(E 63) [Smeets, I. 2008: 225 (246)] [21]
\[chum-\text{neg-\text{wma \('how it was')\}
- QC.chum_cómo+VRB.nten36-IV+CSVN.wma4

\text{tunte-~chunte-} may be verbalized by suffixes -Ø-, -1- and -ntu-, see following examples:

(E 64) [Smeets, I. 2008: 114 (6)] [21]
\[\text{tunte}-\text{n-tu \('for how long')\]
- QT.tunte_cuánto+VRB.Ø36+PVN.n4+ADV.tu

\text{ye-\text{w-üy-m-u \('how are you both related')\}
3.3.2 Deictic verbs

"Deictic verbs are derived from the roots fa- 'to become like this' and fe- 'to become like that'. These roots do not occur without a derivational suffix. A verb which is derived from the root fa- denotes a situation which is contextually determined. A verb which is derived from the root fe- denotes an instance which is situationally determined" [Smeets, I. 2008: 321] [21].

Smeets says that deictic verbs do not occur without a derivational suffix, but there is a case in which fe- directly takes an inflectional suffix without any derivational one:

(E 67) [Smeets, I. 2008: 462 (61)] [21]  
kom fe-le-y 'they all are that way'  
-IV.fe_ser-eso+IND.y4+3.Ø3  
+PT.llemay_seguro_cierto
For the verb kopüd- 'to lie down on one’s belly' there are no examples but the meaning it takes with +PFPS.künu32, +PRPS.nie32 and +ST.1e28. We have found examples in other texts, but some of them show a different behaviour to what Smeets explains, i.e., they do not present the “obligatory” suffixes.

(E 75 ) [Febrés, A.]*

kopu- 'to face down lying on the floor or head down, or half bent the body.'
-IV.kopüd_yacer-boca-abajo+PVN.n4

For verbs kudu- 'to lie down' and külu- 'to lean on one’s elbow', Smeets herself presents examples contradicting its obligatory co-occurrence with the treated suffixes. Other texts also contradict her (Augusta, F. [3], Febrés [3], Zúñiga [24], Mösbach [14]).

(E 77 ) [Smeets, I. 2008: 349 (17)] [21]
kudu-pu-a-el 'to go to bed'
-IV.kudu_yacer+LOC.pu17+NRLD.a9+OVN.e14

(E 78 ) [Smeets, I. 2008: 244 (4)] [21]
kudu-nu-l-m-i 'if you do not go to bed'
-IV.kudu_yacer+NEG.no16+CN.D.14+2.m3+SG.i2

(E 79 ) [Smeets, I. 2008: 526 (lür)] [21]
külü-a-y antü 'the Sun will lay down'
-IV.külü_apoyar+NRLD.a9+IND.y4+3.Ø3
-NN.antü_sol

For likosh- Smeets gives the meaning it takes with two of the suffixes but no example. The example we show comes from Augusta, F. [3].

likosh-küle- 'to squat, to crouch'
likosh-künu-w- 'to squat down, to crouch down'
[Smeets, I. 2008: 528] [21]

(E 80 ) [Augusta, F. ] [3]
likök-küle- 'to be snuggled'
-IV.likosh_acurrucar+ST.1e28+PVN.n4

For payla- there are also no examples in Smeets but the definition in combination with two of the three suffixes. Examples coming from other sources contradict Smeets’ observations.

payla-le- 'to be lying on one’s back'
payla-künu-w- 'to lie down on one’s back'
[Smeets, I. 2008: 543] [21]

For potri- 'to lean', Smeets gives two examples and the meaning acquired with one of the three suffixes, and in a compound. We did not find examples of potri- in other texts. We believe that potri- and potrü- 'to bow forward' are two variants of the same verb, even though Smeets defines them differently. The interchangeability between i and e is not rare. She provides two different translations for an example with potrü-, the second one in alignment with the same example using potri- instead. Compare:

(E 83 ) [Smeets, I. 2008: 549 (potri-)] [21]
potri-tripa-n ti wangku-mu 'I toppled out of the chair'
-IV.potri_inclinar-IV.tripa_salir-CR.IV
+IND1SG.n3
-AP.ti_el -NN.wangku_silla+INST.mew

(E 84 )
potrü-tripa-n ti wangku-mu
1. 'I fell backward from the chair' [Smeets, I. 2008: 62 (11)]
-IV.potrü_inclinar-IV.tripa_salir-CR.IV
+IND1SG.n3
-AP.ti_el -NN.wangku_silla+INST.mew

potri-le- 'to be leaning (over)'
potri-tripa- 'to topple'
[Smeets, I. 2008: 543] [21]

(E 85 ) [Smeets, I. 2008: 296 (20)] [21]
potri-künu-w-ün 'I bent forward'
-IV.potrü_inclinar+PFPS.künu32
+REF.w31+IND1SG.n3

We did find examples of potrü- in other texts, and as in previous cases, they show no concordance with Smeets statements:

(E 86 ) [Augusta, F. ] [3]
potrü-w-ün 'to buck'
-IV.potrü_inclinar+PS.we19+PVN.n4

* All the examples that come from Augusta, F., Febrés, A. and Valdivia, L. has been consulted on-line on the CORLEXIM site [3].

10 This example seems to support Smeets’ statements, if the phonological changes are that static -le- drops its vocalic element and the indicative 1s person adds an epenthetic schwa in presence of the previous consonant. Normally, -le- keeps the vocal and the following suffix, beginning in consonant, remains the same.
For the verb *rekül-* 'to lean' there are some examples in other texts contradicting Smeets.

(E 87) [Augusta, F.] [3]

**rekül-tu-n** 'to cuddle up, to lie down'
-IV.rectul_apoyar+TR.tu33+PVN.n4

**rekül-tu-we** 'back (of something)'
-IV.rectul_apoyar+TR.tu33+NOMPI.we

For the verb *iūnif-* ~ *ünif-* 'to lie down on the floor' we did not find examples other than the one given by Smeets, see example E73, p. 11.

For the verb *wira-* 'to sit down with spread legs', Smeets presents no examples but definitions in combination with two of the three suffixes:

- **wira-kūnu-** 'to adopt a position with the legs apart'
  - wira-le- 'to sit with the legs apart'
  [Smeets, I. 2008: 576] [21]

Examples from other texts:

(E 89) [Augusta, F.] [3]

**wira-le-n** 'to be with legs open'
-IV.wira_sentar-con-las-piernas-abiertas
+ST.le28+PVN.n4

(E 90) [Mösbach, E. 1936] [14]

**wira-l-küle-chi** 'the open ones'
-IV.wira_sentar-con-las-piernas-abiertas
+CA.134+ST.le28+SVN.Ø4+ADJ.chi

(E 91) [Mösbach, E. 1936] [14]

**wira-l-kūnu-ye-ngko-fu-y** 'they remain open'
-IV.wira_sentar-con-las-piernas-abiertas
+CA.134+PFPS.kūnu32+PLR.ye24+PASS.nge23
+CF.ke14+IPD.fu8+IND.y4+3.Ø3

3.3.4 Verbs that need a directional

There is another group of verbs that require a directional to be expressed if they are not part of a compound or take a transitivizer or causative suffix. Directional suffixes are *-me*-thither (slot 20), *-pa*- hitter (slot 17) and *-pu*- locative (slot 17). The affected verbs are:

- **antūi-** 'to spend a day'
- **fül-** 'to come near'

\[\text{antūi-} 'to spend a month'
\]

\[\text{fül-} 'to approach'
\]

\[\text{nge-} 'to have been' (it only requires *-me*-thither or *-pa*-hitter)
\]

\[\text{pīl-} 'to come near'
\]

\[\text{ru-} 'to pass, to go through' (it only requires *-me*-thither or *-pa*-hitter)
\]

\[\text{tripantu-} 'to spend a year'
\]

And the following compounds:

- **kim-kon-** '(know-enter-) to find out, to understand' requires *-pa*- hitter
- **kim-pūra-** '(know-go_up-) to realize' requires *-me*-thither or *-pa*-hitter
- **trem-tripa-** '(grow_up-go_out-) to become conscious of while growing up' requires *-pa*-hitter
  [Smeets, I. 2008: 325, 326] [21]

For **antūi-** 'to spend a day' there are some examples refuting Smeets’ observations, i.e., showing the occurrence of the verb without the directional or nor the transitivizer or causative:

- Augusta, F. [3] **antū-n, antū-le-iñ, antū-y, antū-ñma-le-n, antū-ñma-n**
- Febrés, A. [3] **antū-n, antū-ku-n**
- Mösbach, E. [14] **antū-y**
- Smeets, I. [21] **antū-le-y, antū-a-y, antū-y, antū-le-chi**
- Zūñiga, F. [24] **antū-y**
- Valdivia, L. [3] **antū-n, antū-n-ku-n**

For **fül-** 'to come near', it happens as with **antūi-**:

- Augusta, F. [3] **fül-küle-n, fül-ma-n, fül-ün**
- Mösbach, E. [14] **fül-a-n, fül-el, fül-la-e-y-ew, fül-küle-le-n, fül-ma-ngu-fu-lu, fül-e-y**
- Smeets, I. [21] **fül-küle-n**

For **küyen-** 'to spend a month', Smeets gives no examples, and we have also found contradicting ones from other authors:

- Augusta, F. [3] **küyen’-ün**
- Febrés, A. [3] **küyen-*, küyen-a-y**

For **llekü-** 'to approach', Smeets gives one example, and other authors have contrary examples.

(E 92) [Smeets, I. 2008: 503 (elu-)] [21]

**llekü-pu-el** 'to come near'
-IV.llekü_acercar+LOC.pu17+OVN.el4

It looks like Pascual Koña uses *wira-* with the sense of "two things that spread apart", not only the legs. Mösbach collected Koña’s memoires in the book cited in reference [14].
For *nge- ‘to have been’, Smeets remarks that an exception is when the negation marker *-la-* co-occurs. But we have found that the exception also applies with the negation marker *-no-*. The first two following forms found in Smeets are also realized without directional suffixes:

Exceptions: *nge-,* nge- *

Negation *-la-*: *nge-la-* y, *nge-la-n*, *nge-we-la-* y, *nge-ke-la-* fu-y, *nge-we-tu-la-* y

Negation *-no-*: *nge-nu-* n, *nge-ke-nu-* lu, *nge-nu-n-* mu

For *pülle-* ‘to come near’, there are contrary examples even in Smeets:

Augusta, F. [3] *pülle-le-n*, *pülle-küma-n*, *pülle-lu*, *pülle-ye-leu*, *pülle-le-chi*, *pülle-le-nu-chi*, *pülle-nie-gel-chi*, *pülle-ke-ñma-w-kiule-y*, *pülle-ñma-w-kiule-y-u*

Mösbach, E. [14] *pülle-ñma-w-ke-chi*

Smeets, I. [21] *püllle-ye*, *püllle-le-

For *ru-* ‘to pass, to go through’ the occurrence are as Smeets describes them, except for two examples found at Febrés, A. [3] (*ru-n*, *ru-a-n*), which we have no way yet to confirm as right or wrong.

( E 93 ) [Augusta, F.] [3]

**ru-l-pa-n** *nutram-pe-lu* ‘interpreter, translator’

-IV. _ru_pasar+CA.134+HH.pa17

-NN. _nutram_conversacion+PX.pe13+SVN.lu4

For *tripantu-* ‘to spend a year’ there are also examples from other authors not supporting Smeets’ findings:

Augusta, F. [3] *tripantu-le-n*, *tripantu-n*, *tripantu-y*, *tripantu-chi*

Febrés, A. [3] *tripantu-n*, *tripantu-a-n*, *tripantu-y*

Mösbach, E. [14] *tripantu-a-m*, *tripantu-n*, *tripantu-el*

Valdivia, L. [3] *tripantu-n*

Finally, for compounds *kim-kon-’* to find out, to understand’, *kim-püra-* ‘to realize’ and *trem-tripa-* ‘to become conscious of while growing up’, we have found no other examples than Smeets’, who also gives a contradictory example: E94.

( E 94 ) [Smeets, I. 2008: 447 (26)] [21]

**kim-kon-y**-i/i* ‘we had become aware’

-IV. _kim_saber-IV.kon_entrar-CR.IV

+IND.y4+1.Ø3+PL.in2

( E 95 ) [Smeets, I. 2008: 447 (24)] [21]

**kim-kon-pa-n’* ‘I understood, I realized’

-IV. _kim_saber-IV.kon_entrar-CR.IV+HH.pa17

+IND1SG.n3

( E 96 ) [Smeets, I. 2008: 381 (1)] [21]

**kim-püra-me-n’* ‘I became aware, I came to appreciate’

-IV. _kim_saber-IV.püra_subir-CR.IV+TH.me20

+IND1SG.n3

**kim-püra-me-pa-n’* ‘I realized’

-IV. _kim_saber-IV.püra_subir-CR.IV+TH.me20

+HH.pa17+IND1SG.n3

( E 97 ) [Smeets, I. 2008: 446 (11)] [21]

**kim-püra-me-pa-n’* ‘I have come to know him’

-IV. _kim_saber-IV.püra_subir-CR.IV+TH.me20

+HH.pa17+EDO.fi6+IND1SG.n3

( E 98 ) [Smeets, I. 2008: 262 (10)] [21]

**trem-tripa-pa-y’* ‘they grew up knowing (about)’

-IV. _trem_crecer-IV.tripa_salir-CR.IV

+IND.y4+3.Ø3

3.4 Morphophonology

As we have mentioned in the introduction, interaction among roots and suffixes creates different contexts in which the form of these elements may be affected. Most common changes correspond to epenthesis and elision, but there are also cases of phoneme alternation, some are obligatory and others, optional. We present all of these changes in section 5.1 Phonological changes into spelling, p. 21, where, at the same time, we explain how these variations have been encoded to be processed by the FST analyser.

4 The computational approach

In this section, after a basic introduction to computational morphology and Finite State Transducers (FST), we explain how morphophonologic phenomena of Mapudüngun have been encoded in order to process Mapuche words through FST analysis, and obtain a proper identification of the parts (roots and suffixes) forming these words.

We do a quick review on FOMA implementations. FOMA is the FST compiling program we use to generate the analyser (see section 4.4). And finally, in section 4.5, we enter into Mapudüngun encoding, starting by the alphabet and finishing by the lexicon: roots and suffixes.

4.1 Computational morphology

Computational morphology is the branch of computational linguistics concerned with word structure. Two kinds of processing are of interest. One is morphological analysis, by which a surface word form is analysed into a lexical representation, consisting of the word’s component morphemes or grammatical features. The other is morphological generation, by which a lexical representation is converted to a

---

12 In this section we follow Gasser [2011: 55] [8].
surface word form. Consider the Mapuche verb lelien 'you looked at me'. A basic morphological analysis would simply segment the word into the morphemes that make it up, as seen below:

(E 100)

lelien → le-li-e-n

A more abstract lexical representation output would indicate the lexical and grammatical significance of the morphemes. The word lelien could be represented at the lexical level as shown right below:

(E 101)

le-li-e-n 'you looked at me / look at me!'

\[
\text{-TV,li} \text{.mi} \text{ar}^{13}+\text{IDO}^{14}.e6+\text{IND1SG.n3}+\text{DS12A.Ø1}
\]

This chain of tags represents the root of the verb leli-meaning 'look at', the internal direct object -e-, the portmanteau suffix indicative 1st person singular -n, and the null dative subject for 2nd or 1st agent persons.

Three types of information are required to perform morphological analysis: a lexicon, the morphotactics and the (alternation) rules.

A lexicon is composed of roots or stems, which combine with grammatical morphemes to yield surface word forms.

Morphotactics refer to constraints on the order and class of the morphemes that make up a word within a particular category. For example, the morphotactics of Mapudüngun verbs specify the following minimal sequence of morphemes: mood (indicative, conditional or imperative, slot 4), subject (slot 3) and number (slot 2).

Alteration rules are responsible for the variation of forms that morphemes take in the presence of other morphemes. For example, the portmanteau suffix of 1st person indicative mood takes two forms, one before vowels, another before consonants, where an epenthetic -u appears.

Together, knowledge of alternation rules, morphotactics, and the forms of roots or stems in the lexicon represent the morphology of a given language.

Morphological analysis may be efficiently handled by finite state transducers (FST). An FST is a network of states and transitions between them, and the analysis of a word is a path through this network. Each of the transitions along

13 The analyser lexicon is collected with the Spanish translation, that’s why all the analyses presented in this article carry the root meaning in Spanish.
14 Tags meaning are found in annex 11.1 Tags meaning.
15 We follow Smeets’ descriptions throughout this work, but it is important to know that in this issue there is discrepancy among authors of Mapudüngun descriptive grammars. Basically, what Smeets [2008] [21] identifies as “agent-patient paradigm” is what Zuthiga [2006] [23] calls “verbal inversion”, and Salas [2006] [19] “person focalization”.
16 A transducer is a device or machine that converts energy from one form into another, e.g., a microphone is a transducer that converts the vibrations captured from the air into analogous electrical impulses. An FST converts a chain of symbols into another chain of symbols.
17 In this section we follow the explanation given by Ríos [2015: 18-21] [17].
18 miaw-ün ‘I wandered’, miaw-üy-m-i ‘you wandered’, miaw-a-n ‘I will wander’, miaw-a-y-m-i ‘you will wander’.

4.1.1 Finite state method

A finite state transducer (FST) is a piece of software that operates as an enhanced finite state machine (FSM) which in its turn is capable of representing and operating over finite state networks (FSNs).

4.1.2 Finite state transducers (FSTs)

There is an important distinction between FSMs that are one-sided, and FSTs that have an upper and a lower side, or more generally, an input and an output level. Since an FST has two sides, it can not only decide if a given word is part of its regular language, but it will also return the corresponding output to the given input [Beesley & Karttunen 2003: 11][1].

An FST accordingly implements a relation between two regular languages: an upper side and a lower side regular languages, and it literally “transduces” strings from one language into the other. In a non-deterministic FST it may produce more than one possible outputs for a given string.

See figure 1 for an example of an FST that contains the relation of two of the following four word forms with the Mapudüngun root miaw- ‘to wander’, and their respective morphological analysis:

\[
\text{miaw-a-n ‘I will wander’, miaw-a-y-m-i ‘you will wander’}
\]
Andrés Chandía, andres@chandia.net

1. Finite state transducer for miaw- with Ind1Sg, present/past or future

(E 102)

miawün → miaw+Ind1Sg3  
miawymi → miaw+Ind4+2p3+Sg2  
miawan → miaw+Nrld9+Ind1Sg3 (in figure 1)  
miawaymi → miaw+Nrld9+Ind4+2p3+Sg2

Note that the transducer contains an empty transition $\epsilon: \epsilon$ which makes the NRLD suffix -a- optional. The transducer in figure 1 may be applied in both directions:

Given $miawüymi$ 'you wandered' as input, applied in "upward" direction, it produces:

-IV.miaw_merodear+IND.y4+2.m3+SG.i2 as output. This is the procedure for morphological analysis.

Given -IV.miaw_merodear+IND1Sg.n3 as input, applied in "downward" direction, it produces $miawün$ 'I wandered' as output. This is the procedure for generation.

Composition is a "hard to handle" concept in finite state processing. However, here it suffices to affirm that a cascade of rules compiled into finite state transducers may be combined into a single equivalent FST via composition. See figure 2.

2. A cascade of rules compiled into finite state transducers may be combined into a single equivalent FST via composition. This mathematical possibility, shown by Johnson, may be performed in practice using a finite state software [Beesley & Karttunen 2003: 35] [1].

(E 103) $[a \Rightarrow c \_ [r|t]]$;

(E 104) $[y \Rightarrow \{ie\} \_ [r]\{st}\]$

4.1.3 Two levels morphology

The upper language, also called the abstract level, the lower language also called surface level, and the relation they establish as part of an FST are well explained in Beesley & Karttunen [2003] [1]. Related images (figures 3 and 4) are presented here just to illustrate a general idea about these three concepts: upper and lower languages, and the relation among them.

Table 3 shows that relations contain pairs of strings. For analysis, the lower language is used as input, and the upper language is produced as output [figures 3 and 4]:

Table 3 Article/Determiner/Quantifier distinctions.

| Upper: the+Art+Def | Upper: a+Art+Indef |
|--------------------|--------------------|
| Lower: the         | Lower: a           |

Fig. 3 Analysing canto [Beesley & Karttunen 2003: 13] [1]

Fig. 4 Another path in the Spanish morphological analyser [Beesley & Karttunen 2003: 13] [1]

Regular expressions as rules In the development of this system we mainly used two types of operators "restriction and replacement" ($\Rightarrow$, $\Rightarrow$), complemented with composition ($\circ$), context ($|)$), and some others operators.

Restriction is one of the fundamental functions in two-level calculus:

(E 103) $[a \Rightarrow c \_ [r|t]]$;

(E 104) $[y \Rightarrow \{ie\} \_ [r]\{st}\]$. 

\[\text{Rule FTS 1} \quad \text{Intermediate string} \quad \text{Rule FTS 2} \quad \text{Intermediate string} \quad \text{Surface FTS n} \quad \text{Surface string}\]

\[\text{COMPOSITION} \quad \text{COMPOSITION} \quad \text{COMPOSITION} \quad \text{COMPOSITION} \quad \text{COMPOSITION} \quad \text{COMPOSITION} \]

\[\text{Underlying string} \quad \text{Underlying string} \]

\[\text{Table 1 Article/Determiner/Quantifier distinctions.}\]

| Upper: the+Art+Def | Upper: a+Art+Indef |
|--------------------|--------------------|
| Lower: the         | Lower: a           |

\[\text{Fig. 1 Finite state transducer for miaw- with Ind1Sg, present/past or future}\]

\[\text{Fig. 2 A cascade of rules compiled into finite state transducers may be combined into a single equivalent FST via composition. This mathematical possibility, shown by Johnson, may be performed in practice using a finite state software [Beesley & Karttunen 2003: 35] [1].}\]

The number at the end of each suffix tag indicates its assigned slot. For a complete explanation on how to read suffixes tags see the note following E1, p. 3.
E$^{104}$ denotes the relation in which the string "y" is transformed into "ie" (here a condition is introduced, in this case is the context represented by the twin pipes (||) when followed by "r" or "st"; so, "ugly" becomes "uglier" or "ugliest", and "pretty" becomes "prettier" or "prettiest"$^{20}$.

Composition is the concatenation of rules, for example, E$^{104}$ could be decomposed into two concatenated rules, i.e., they together form a "composition" of rules:

(E 105) \[y \rightarrow \{ie\} || _r . o. \]

(E 106) \[y \rightarrow \{ie\} || _r , _{st}\];

Or it could be represented stating two different contexts for the change to occur:

(E 106) \[y \rightarrow \{ie\} || _r , \{st\};\]

In general, regular expressions may be redundant because some of the operators could easily be defined in terms of others, which means that a string and its restrictions may be expressed in different ways by means of regular expressions.

For more detailed descriptions on regular expressions operators, please consult: FOMA’s Regular Expression Reference: [https://code.google.com/archive/p/foma/wikis/RegularExpressionReference.wiki](https://code.google.com/archive/p/foma/wikis/RegularExpressionReference.wiki)

Another excellent reference is Beesley & Karttunen’s "Finite State Morphology" book [1].

4.2 The bases of the analyser

The main script of the analyser is the series of regular expressions (regexps) encoding the Mapudiñgun grammar. This is where the different parts forming the Mapuche language are declared: roots, suffixes, particles, etc., and the rules for them to interact in the way Mapudiñgun accepts it.

We follow Smeets description of Mapudiñgun: "A Grammar of Mapuche" [21] to implement the analyser and generator. We have begun treating only one variant or dialect of this language, which is known as central Mapudiñgun. The analyser, which is basically the same as the generator, has little broader rules that accept mainly spelling variants, that in some cases come from different dialects, but generally correspond to a certain syncretism (and confusion) generated after the different spelling proposals and the influence of Spanish orthography.

The spelling proposal we follow is AMU, "Alfabeto Mapuche Unificado" [Sochil 1986, 1988] [23]. This "grafe-mario" is also known as the "academic proposal".

Having half mind in computational technology and the other half in linguistics, finding a description of a language such as the one Smeets does of Mapudiñgun in her thesis work, leads straight to think in a computational implementation of her grammar. The suffixes organized in slots and the description of interaction rules are the words that reflect what regular expressions can encode. We did further research on other descriptions of Mapudiñgun while implementing the analyser but only to compare descriptions (Other sources of Mapudiñgun grammars we have consulted are Fernández-Garay & Malvestiti [7], Lonkon [12], Ragileo [15], Salas [19] and Zúñiga [24]).

4.3 The analyser

This is a rule based morphological analyser (and generator) applied to the Mapudiñgun language. It was built with finite state transducers using the algorithms developed by Mans Hulden$^{25}$ for his FOMA$^{22}$ project, an open source application to compile finite state transducers.

The rules that were generated for the analyser, as well as the tags used in its outputs, are based on the description and study made by Dr. Ineke Smeets of Mapudiñgun, and published in her book "A Grammar of Mapuche" [Smeets, I. 2008] [Ref].

4.4 FOMA implementations

The path we follow is Annette Ríos work on Quechua, she have developed various tools, the main one being the finite state morphological analyser and generator [Ríos 2015] [17].

Ríos’ development for the analyser and generator was with XFST and other tools released by Xerox [Beesley & Karttunen 2003] [1]. She used FOMA for the spell checker. We decided to use FOMA for everything, mainly because it is open source software, and we wanted to develop a totally free set of linguistic tools for Mapudiñgun.

FOMA is a very well-known software used by many linguists and computer developers for a wide range of tasks, but mainly for language applications. Searching on the Internet we have looked for the FOMA implementations$^{24}$ per year, shown in table 4:

| Year | FOMA implementations |
|------|----------------------|
| 2008 | 14                   |
| 2009 | 15                   |
| 2010 | 16                   |
| 2011 | 17                   |
| 2012 | 18                   |
| 2013 | 19                   |
| 2014 | 20                   |
| 2015 | 21                   |
| 2016 | 22                   |
| 2017 | 23                   |
| 2018 | 24                   |
| 2019 | 25                   |
| 2020 | 26                   |

20 The actual rule to treat this behaviour is much more complex, it is presented in this simplistic way just as an example.

21 [https://en.wikipedia.org/wiki/Mans_Hulden](https://en.wikipedia.org/wiki/Mans_Hulden)

22 [https://fomafst.github.io/](https://fomafst.github.io/)

23 [https://www.degruyter.com/view/product/22765](https://www.degruyter.com/view/product/22765)

24 This information was compiled using Google Scholar only. Some publications mention more than one FOMA implementation, some times at different years; to simplify, we have counted the publications about FOMA implementations, and we have left out those that mention FOMA only as a reference. The results vary as the search is repeated.
2019), Turkish [Yıldız, Avar, & Erçan. 2019], Indonesian [Larasati et al. 2011], Japanese [Sim. 2013], Arabic [Attia, Al-Badrashiny, & Diab. 2014], Kazakh [Kessikbayeva & Cicekli. 2014] and others\(^{25}\). We list here some of the implementations counted in table 4.

| Year | Implementations |
|------|-----------------|
| 2020 | 12              |
| 2019 | 17              |
| 2018 | 16              |
| 2017 | 14              |
| 2016 | 10              |
| 2015 | 11              |
| 2014 | 13              |
| 2013 | 09              |
| 2012 | 12              |
| 2011 | 13              |

4.5 The encoded Mapuche alphabet

We present in this section the alternative representation of some graphemes in the analyser, alternative from those of Smeets. Finally, the rules implying variation in the spelling of words, generated either by elision, epenthesis or other similar phenomena.

As it also deals with orthography, this section brings up the subject of the spelling unifier (see 6.1, p. 41).

The sigma alphabet of the analyser includes every non-empty symbol that appears in the network, either by itself or as a component of a symbol pair [Beesley & Karttunen 2003: 57, 58] \(^1\). For example, the sigma alphabet of the network compiled from [cat "+Noun":0] consists of the symbols a, c, t and +Noun [Beesley & Karttunen 2003: 62] \(^1\). Then, the sigma alphabet is a list of all the single symbols that occur either on the upper or lower side of the arcs [Beesley & Karttunen 2003: 99] \(^1\) (see 4.1.2 Finite state transducers (FSTs), p. 15).

The analyser’s sigma, among all the symbols it comprises, contains the Mapuche alphabet:

\[
\begin{align*}
\text{VOW} & \{ a | e | i | o | u | ü \}; \\
\text{SVW} & \{ w | y | g \}; \\
\text{CON} & \{ ch | d | f | k | l | ll | m | n | ñ | ng | p | r | s | sh | t | tr \}; \\
\end{align*}
\]

\(\{\text{VOW, SVW, CON}\} \in \Sigma\) of the analyser

In order to apply morphotactics (see 4.1 Computational morphology, p. 14) and rules, the alphabet has been divided into three groups:

4.5.1 Vowels

For this group of characters (see table 2, p. 3) Smeets’ proposal was adopted without any variation. They conform the group named VOW.

It is worth to mention though, that "Mapudüngun has six vowel phonemes, /ë e o o/... It should be noted that vowels of Mapudüngun have traditionally been treated as the five vowels of Spanish (/i e o u/), with identical stressed and unstressed allophones, plus a high central unrounded vowel /i/ (commonly known as the ‘sixth vowel’) having a mid central allophone [u] in unstressed position" [Sadowsky, S. 2013] \(^{18}\).

The previous paragraph comes to say that the grapheme ü represents two sounds: /i/ and /u/ which realize in complementary contexts.

\(^{25}\) There are many more FST implementations for different languages, with other available FST compilers; but as we have developed our own with FOMA, we just list those.
4.5.2 Semivowels:

Smeets identifies them as glides (see table 1, p. 2). She includes the \( r \) and counts the glides among the consonants, while in the analyser these are separated from the consonants to form the semi-vowels group, except for the \( r \). With this categorization we can treat certain phonological phenomena that involve these semi-vowels, such as elision and epenthesis (see 5.1, p. 21). We could have called them glides as Smeets does, but following Beesley & Kartunnen [2003] [1], we have called them semi-vowels. Anyway, they do occur as both, semi-vowels and glides\(^26\).

The sound \(/s/\) usually represented by a \( g \), Smeets represents it by a \( q \) to reflect a difference in sound, which is softer than the one she represents using \( g \), for instance, in the loan from Spanish *guayta* 'cookie', so written in Smeets' work. She spells the Mapuche word for 'seven' as *regle*, while we spell it *regle*. This word and all those that Smeets spells with \( q \) have been traditionally spelled with \( g \). This distinction does not affect the meaning of words, then \( g \) spelling has been adopted by us to represent all the \(/s/\) variants.

For the other two graphemes, \( w \) and \( y \), no changes were made. All three are grouped under the SVW denomination.

4.5.3 Consonants:

Contrary to Smeets, who states that there are 19 consonants, and because we have counted out the three glides identified in the previous paragraphs as semi-vowels, there are 16 recognized consonants.

Smeets uses \( d \) to represent the voiceless interdental fricative \(/\theta/\) to make clear the difference with the \( d \) from Spanish loans. We use \( b \) because this distinction does not affect the meaning of words.

For the interdental series \( t', n', l' \), we have eliminated the apostrophe to interpret these letters as the alveolar counterpart. Letters with apostrophe appear mainly in the older texts. In more recent texts they sometimes appear, even though not consistently. Some times the writer introduces an interdental in a word, later it does not, not even for the same word. Due to this misleading usage of the interdental variant and the "dying out" of the distinction [Smeets, I. 2008: 31] [21], the inclusion of interdental has been avoided for morphological generation, while they are accepted as variants in the analysis.

The elimination of apostrophe broadens the possibilities of analysis. The words and roots to be analysed are collected in the lexicon, if it appears just *newen* 'force, strength' in the lexicon, the variant *n'even* would be taken as an unknown root, while by eliminating the apostrophe before analysis, it allows the transducer to analyse any introduced variant, either *n'even, n'even*, *newen* or *newen* as the same root.

Loaned sounds \(/b/\) (bilabial, plosive, voiced), represented by \( b \), and \(/x/\) (velar, fricative, voiceless), represented by \( j \), from Spanish, are not included in our system. They are converted into the corresponding letter of the Mapudüngun alphabet by means of the spelling unifier (see 6.1 below); \( b \) into \( f \) and \( j \) into \( k \), which are the most usual conversions we have found reflected in some dictionaries: *jabón* → *kafom*; *burro* → *furiku*, etc. [Febrés 1882\(^28\)].

Other possible conversions are directly reflected in the lexicon: *vaca* → *waka*\(^29\); *vehículo* → *weikulo*; *voto* → *woto*; *jamón* → *kïmon*; *junio* → *kunio*, etc.

Within the consonants group, CON, note the representation for digraphs (*{ch}, {ll}, {ng}, {sh}, {tr}*) in curly braces. This indicates that these two symbols together form a single representation of a sound. In other words, the concatenation of these two symbols is invariable and univocal.

4.6 Intermediate representation symbols

These symbols are created to treat different morphophonological changes that occur in the language due to the interaction between suffixes, between roots, and among all of them together. It may be said that they are part of the alphabet, at least in an abstract level, because they are used to represent a certain stage in a change process, from which the surface form (the one we write or read) arises. A few of them are listed here as an example, the complete list of these symbols with their function is found in annex 11.2.7 Intermediate language symbols, p. 60:

(E 107)
"\( @G \)" is used to treat ephenthesis of glottal stops represented by \( g \) in the spelling, see 6.4.2, E116, E117 and R1, p. 22.

(E 108) "\( @D \)" is used to treat schwa insertion represented by \( ù \) in the spelling, see E110, E111 and E112, p. 21.

(E 109) "\( @K \)" is used to treat radical alternation in some intransitive verbs which change their last consonant from \( g \) to \( k \) if they are in contact with the causative suffix -\( ûm \)-, which transitivizes them, see D19 and D19, p. 28.

\(^{26}\) Glides immediately precede the vowel, semi-vowels immediately follow the vowel, both are less sonorous than the vowel.

\(^{27}\) Not in all dialects this sound is voiceless, prove of that is in the early transcriptions made by Jesuit monks, all of them used \( d \), which represents a voiced sound.

28 Consulted at http://corlexim.cl <2019-07-11>. More examples are found in Febrés dictionary [1882] [6].

29 Today it is well-known that in Spanish there was never a difference in pronunciation of words spelled with \( b \) or \( v \) (see http://lema.rae.es/dpd/srv/search?id=d45ahCOieD6TjHKns8 for more information), otherwise, probably the inclusion of vaca into Mapuche lexicon should have been as *faka*. 
4.7 Roots encoding

In the FST, this section headline is "Read in roots", because the system reads the files containing the roots lexicon and incorporate them into the analyser. Lists are separated by part of speech (grammatical categories).

(D1)
define AVROOT @re"roots/avroot.lex";
define NROOT @re"roots/nroot.lex";

The expression above introduces the lists of roots for adverb and noun categories, so any listed noun may be found throughout "NROOT", which encodes nouns as shown below:

(D2) Sample of nouns lexicon:

|"-NN"{.aylen_brasa}]: ["@G"{aylen}]
|"-NN"{.aywin_sombra}]: ["@G"{aywin}]
|"-NN"{.chachay_papa-affectuoso}]: [{chacha} {tatay}]
|"-NN"{.chadi_sal}]: {chadi}
|"-NN"{.chaf_cascara_piel-de-frutas}]: {chaf}
|"-NN"{.chafid_bagazo}]: {chafid}

In D2 "-NN" stands for (simple) noun or nominal (root). "@G" is an intermediate language tag to treat glottal stop insertion, see 6.4.2, E116, E117 and R1, p. 22.

In the same way D1 shows noun and adverb roots encoding, there is a file per part of speech and other forms (verbs, adjectives, particles, interjections, etc.).

We have divided the lexicon in two major groups: the roots that may be verbalized, and the forms that can not be verbalized. Among the roots, all except for the verb roots may be independent (without suffixes) words.

Roots (verbalizable)
1. File ajroot.lex: adjectives
2. File avroot.lex: adverbs
3. File ivroot.lex: intransitive verb roots
4. File names.lex: proper nouns
5. File nroot.lex: nouns
6. File nuroot.lex: numerals
7. File onroot.lex: onomatopoeia
8. File gqroot.lex: question forms
9. File tigroot.lex: transitive verb roots
10. File itj.lex: interjections
11. File negs.lex: negation particles
12. File numbers.lex: numbers
13. File partls.lex: particles
14. File perspr.lex: personal pronouns
15. File posspr.lex: possessive pronouns
16. File prep.lex: prepositions
17. File puncts.lex: punctuation marks

Tags meaning is found in annex 11.1. There is a list of tags assigned to every part of speech and suffix with the name (in Spanish) identifying them, on: http://www.chandia.net/glosas-del-dungupeyem

4.8 Suffixes encoding

As every suffix is assigned to a slot, which in turn is encoded in a file, the addition of such information to the main script is carried out by calling these files under the script section "Read in slots". Each file contains the fillers of the corresponding slot. For example:

(D3) define NEG @re"slots/1-15-Inflectional-Suffixes/slot-10.aff";

D3 is the definition of an expression named "NEG", which in turn, is composed by the regular expressions stored in the text file "slot-10.aff", for which its complete location route is indicated: "slots/1-15-Flexive-Suffixes/" [See figure 5].

![Fig. 5 Directory tree structure for the location of a suffix file.](image)

This file, deployed in D4 below, contains negation suffixes filling slot 10:

(D4) ["+NEG".la10]: {la}
| ["+NEG"{.ki10}]: "@NK"
| ["+NEG"{.no10}]: [{no}|{nu}]
| ["+NEG"{.kino10}]: [{kino}|{kino}];

30 This regexp is written in only one line, but for the sake of better understanding each of its parts is exposed in separated lines.
D4 expresses four suffixes filling slot 10. They are separated by the pipe (vertical bar |) symbol. Every suffix section indicates its upper or abstract level (right of the colon) and lower or surface level (left of the colon). Every abstract form is returned in the analysis process after a cleaning task leaves only the category tag, the suffix form and the slot number: "+NEG.no10". These same forms, before being cleaned up, are used to apply morphotactics in the abstract level at analysis and generation processes. At the surface level we find the forms given by analysis and/or the tags that trigger a process of replacement, like in the case of "@NK" (in D4 above) which further down the script has a rule replacing it for certain form in a defined context.

The complete list of files per slot containing the suffixes is given in annex 11.2, p. 56.

5 The Mapudünung FST analyser

In this section we expose and explain how the different aspects of Mapudünung morphology are treated in order to implement our FST analyser and generator. The order of presentation is not necessarily the same as in the encoding script. There are some encoding already presented in previous sections: the Mapuche alphabet [section 4.5, p. 18]; the spelling unifier [section 6.1, p. 41]; the lexicon’s inclusion [section 4.7, p. 20], and the suffixes’ inclusion [section 4.8, p. 20].

We begin by explaining phonological changes (5.1), including special cases (see 5.1.2, 5.1.3, 5.1.4, p. 26). Then we move on to the stems typology and the strategies to manage them. The interaction of suffixes after the stem comes next, this section introduces the verb paradigms and verb nominalization. It also brings up the subject of the mobility of some suffixes and how to deal with it.

Some verb roots present a special behaviour, we treat them at the end of this section.

5.1 Phonological changes into spelling

The occurrence of roots and suffixes in a verb form generates certain phonological changes when interacting with their neighbours suffixes or roots. In Mapudünung that interaction may be between suffixes, between the root and the consecutive suffix, between roots inside a compound stem, or between all the previous ones inside complex stems.

These changes are encoded in the "lower" or "surface" side of the language (see sections 4.1.2 and 4.1.3, p. 15), therefore, they are under the section "Lower rules" in the FST script, and they affect the word form.

There are frequent phonological changes in Mapudünung, which are reflected in spelling, except for the epenthesis of the voiced velar fricative /ɣ/ represented by g. Which is optional between the sequences ii, uu, ûû, and obligatory between the sequences ae, ea, ee, ai, ia, where the first vowel of the sequence is the last of a suffix, and the second vowel of the sequence is the first of a consecutive suffix.

The epenthesis of schwa (sometimes i, e, u) is obligatory between a form (root or suffix) ending in consonant or semi-vowel and the following form beginning in consonant or semi-vowel.

( E 110 ) [Smeets, I. 2008: 353 (43)] [21] kim-uw-küle-y-iñ 'we (pl) know each other'
-IV.kim_saber+REF.w31+ST.le28+IND.y4+1.Ø3+PL.iñ2

( E 111 ) [Smeets, I. 2008: 68 (56)] [21] puw-ûn 'to arrive'
-IV.puw_llegar+PVN.n4

( E 112 ) [Smeets, I. 2008: 109 (1)] [21] mamüll-entu 'grove' –NN.mamüll_árbol+GR.ntu

"A schwa is optionally inserted between a consonant and the suffix sequence -l-e +CND.14.3.e3 and between a consonant and the suffix sequence -y-iñ +IND.y4+1.Ø3+PL.iñ2 and -û+y +IND.y4+1.Ø3+DL.u2" [Smeets, I. 2008: 51]. Sometimes it is i instead of schwa.

( E 113 ) [Smeets, I. 2008: 52] [21] kim-l-e → kim-ûl-e 'it he knows'
-IV.kim_saber+CND.14.3.e3

( E 114 )
lef-y-u → lef-ûy-u 'we both ran'
-IV.lef_correr+IND.y4+1.Ø3+DL.u2

( E 115 ) [Smeets, I. 2008: 52] [21] lef-ûy-iñ → lef-ûy-û 'we (pl) ran'
-IV.lef_correr+IND.y4+1.Ø3+PL.iñ2

Lexical forms must be collected in an intermediate form with the appropriate tags to later apply the rules transforming them into the final surface forms, e.g., the portmanteau suffix for indicative, 1st person, singular may occur as -ûn-, -ûñ-, -n- or -ñ-. It is encoded as:

(D 5 ) ["+IND1SG{.n3} ] : ["@0@" [n|ñ] ];

The intermediate form is encoding two variants: the tag "@0@" followed by n or ñ.

The rules to generate the four forms mentioned above are:

(D 6 ) ["@0@" -> u || [CON|SVW].#. ] ]
.o. ["@0@" -> @ || VOW _ ];

In Smeets’ texts, this /ɣ/ is never transcribed, but in texts from other authors is sometimes present: kellu-ke-tge-n 'you helped me'
-NN.kellu_ayuda+VRB.Ø36+CF.ke14+IDO.e6+IND1SG.n3 +DS1A.Ø1 [Mösbach, E. [14]].
These are two rules concatenated by the composition symbol .o. (see figure 2). They are applicable not only to this case, but wherever the tag "@U" is found. The first rule indicates that the tag "@U" is replaced by i in the context (twin pipes || indicate the context) where a consonant or semi-vowel precedes it, or when it is at the beginning of the word (.#. "word boundary character"), the tag position being marked by the underscore "_.". The second rule indicates that the tag is deleted if preceded by a vowel. In combination with the "intermediate representation", the FST compiles the four possible forms for this suffix.

The suffixes affected by these rules are the following ones:

- More implicated object (s29) -I- ["@U"1]
- Satisfaction (s25) -ñmu- ["@U"{ñmu}] 32
- Interruptive (s18) -r- ["@U"r]
- Reportative (s12 & NCC) -ke- ["@U"{rke}]
- Conditional (s4) -l- ["@U"l]
- Plain verbal noun (s4) -n- ["@U"n]
- Indicative 1st singular (s3) -n- ["@U"n]
- Adjectiviser quick & easy (NOM) -nten- ["@U"nten]

Some suffixes may trigger an -i- or an -u- (the first two), or only an -u- (the last two):

- Pluperfect (s15) -wy- ["@U"{wy}]
- Complete subjective verbal noun (s4) -wma-
- Reflexive/reciprocal (s31) -w- ["@U"w]
- 1st person agent (s23) -u- ["@U"u]

Tags are created and assigned arbitrarily. For the rules to work it is necessary to place the corresponding tag in the appropriate position. As the rule is applied to the surface level, the tag is placed at that same level, which is encoded to the right of the colon, e.g.:

(D 7) ["+PLPF". {ýwye15}] : ["@U"{wy}];

The encoding above incorporates the pluperfect marker into the system. To the left of the colon is the upper or abstract level, the analysis representation. To the right, it is the lower level, the surface representation, where the tag is added preceding the suffix as the initial character. Once the tag is appropriately replaced, there is another rule that eliminates the unused tags.

Epenthesis of a glottal stop is optional between the ending vowel of a root and the initial vowel of a following root in compounds.

(E 116) [Smeets, I. 2008: 52] [21]
dewma-iyal-la-y → dewma-giyal-la-y "he did not prepare food"

32 Some suffixes have multiple forms, as the case presented in D5 and D6; or satisfaction -ñmu- that it may also be -ñmo-. For simplicity, we present here only the most common form of the suffix.

To apply the rule described above, we have marked, in the lower level, all the roots beginning in vowel, placing a "@G" before the root; for example:

(E 117)
["-AV"{aymûn_bastante}]: ["@G"{aymûn}]
["-IV"{chew_estornudar}]: ["@G"{chew}]
["-IV"{uma_dormir}]: ["@G"{uma}]
["-NN"{antû_sol}]: ["@G"{antû}]
["-TV"{ingka_defender}]: ["@G"{ingka}]
["-TV"{ungûm_esperar}]: ["@G"{ungûm}]

The rules encoding this change are the next ones:

(R 1) **Glottal stop insertion (between roots)**

define RuGlot ["@G" (->) g || VOW _] .o. ["@G" -> Ø];

The first rule, to the left of the concatenation (.o.) symbol, encodes the optionality by enclosing the direction operator between parenthesis. The change applies only when the tag is preceded by a vowel, otherwise, the second rule transforms it into 0 (zero = null character), i.e., it is deleted.

A glottal stop is also optionally added between suffixes, but only in the cases where the sequence ii is generated. Instead of adding a tag to the suffixes, this change is treated with a more general rule:

(R 2) **Glottal stop insertion (between suffixes)**

define RuleEligGem [ii] (->) [igi];

Applying this rule we can analyse words like the following one:

(E 118) [Smeets, I. 2008: 51] [21]
leli-l-i-iñ → leli-l-i-gin "if we look"
-IV.leli_mirar+CND.l4+1.i3+PL.iñ2

Two equal consonants geminate in careful speech, and became a single sound in colloquial speech. And this fact is sometimes transcribed into written text.

(E 119) [Smeets, I. 2008: 51] [21]
kon-nu-i-i → konu-i-i "if I do not enter"

TV.kon_entra+NEG.nolØ+CND.l4+1.i3+SG.Ø2

33 iyal is a lexicalized form for “food” that may be analysed as -TV.i_comer+HRLD.a9+ØVI.e14 ‘what will be eaten’.

34 This is a wider rule that treats elision, epenthesis and gemination of other phonemes, see R3.
To treat gemination we use the same type of rule shown in R2. The following rule not only encodes the gemination of n, but also of m and e:\[35\]:

(R 3) **Gemination simplification**

define RuleEliGem \{nn\} (\rightarrow) n,
\{mm\} (\rightarrow) m, \{ee\} (\rightarrow) e;

(E 120) [Smeets, I. 2008: 437 (17)]
fillem-mew \rightarrow \text{fillem}_+\text{toda-clipse-de-cosas}+\text{INST}.mew

The non-realized affix -a- (s9) separates itself from a preceding a- inserting -y in between. In this case we have an intermediate representation of the suffix (D8), and a couple of rules treating the corresponding tag "@Y" in context (R4):

(D 8) ["+NRLD": \{a9\}]: ["@Y"; a];

(R 4) **y epenthesis**

define RuTrEPENTHy ["@Y" (\rightarrow) y | [a|.#. ] _];

This concatenated rule says that "@Y" becomes y when preceded by an a or at the beginning of the word, and it becomes a null character when preceded by any character except a.

(E 122) [Smeets, I. 2008: 63 (19)]
tripa-a-n \rightarrow tripa-y-a-n 'I will leave'
-IV.tripa_salir+NRLD.a9+IND1SG.n3+DS12A.Ø1

Pronouns engü 'they (dl)' and engün 'they (pl)' are optionally realized as yengu and yengün respectively, either isolated or forming part of a compound where the previous element ends in vowel. Here also applies definition D8.

(E 123) [Smeets, I. 2008: 95 (ii)]
tufa-yengü 'these two'
-DP.tufa_este-PP.engu_ellos-dos

The sequence ae is optionally simplified as a.

(E 124) [Smeets, I. 2008: 52]
leli-l-a-e-y-u \rightarrow leli-l-a-y-u 'I shall not look at you (sg)'
-IV.leli_mirar+NEG.1a10+IND.e6+IND.y4+1.Ø3+DL.u2+DS12A.Ø1

5.1.1 Interaction between suffixes of slots 10 to 4

Most common suffixes in a verb form are those located between slots 10 and 4. There are multiple morphophonological changes depending on the suffixes occurring. The sequence a-e mentioned above is only one of them.

All the suffixes triggering morphophonological changes that need special rules and belonging to this series of slots are encoded as follows:

(D 9) Slot 10: Negation\[37\]
["+NEG": \{.la10\}]: \{la\} | ["+NEG": \{.ki10\}]: [k"@NK"];

(D 10) Slot 9: Non-realized situation
["+NRLD": \{a9\}]: ["@Y"; a];

(D 11) Slot 8: Impeditive
["+IPD": \{.fu8\}]: [f"@IP"];

(D 12) Slot 6: Internal and external direct objects
["+EDO": \{.fi6\}]: ["@ED"];| ["+IDO": \{.e6\}]: ["@ID"];

(D 13) Slot 4: inflectional nominalizers
["+OVN": \{.el4\}]: ["@EL"];

(D 14) Slot 3PTMT: Portmanteau morphs
["+IND1SG": \{.n3\}]: ["@0"[n|ñ]];

The following set of rules deals with the suffixes shown above:

(R 5) **Negation for imperative forms** (ki \rightarrow k)
define RuNegKi ["@NK" (\rightarrow) [i|e] | [e]"@ID"];

35 In the case of e the phonological term is not gemination but lengthening.
36 Term negation (X). Any single symbol except X. Equivalent to [\! - X] [Hulden, M. in https://code.google.com/archive/p/foma/wikis/RegularExpressionReference.wiki]

37 Some slots of this series encode more suffixes than the ones displayed here, which are the ones having relevance for the rules generated and explained in this section.
From slot 10, there are two negation suffixes that can play a role in this set of interconnected rules: one is the negation for indicative -la-, the final e has to be taken into account when interacting with +IDO suffix -e-, slot 6. The other negation suffix is -ki- for imperatives. It may drop its final i, or replace it by e when followed by e, which is also the form of the +IDO suffix encoded as "@ID" in the intermediate representation. This is what previous rule R5 manages. "When the negative marker -ki- (slot 10) is followed by e, the sequence ie is optionally replaced by ee or contracted to e" [Smeets, I. 2008: 52 (8.1.4.3)] [21].

(E 128) [Smeets, I. 2008: 52 (8.1.4.3)] [21] sngu-we-ki-e-i-l ~ sngu-we-ke-e-i-l ~ sngu-we-ke-i-l 'don't speak to me anymore'
-NN. dngu_palabra+VRB.036+PS. we19 +NEG. ki10+IDO. e6+CNI.14.1.13+SG. Ø2+DS12A. Ø1

Another suffix that has the same conditions of interaction as the negation suffix -la-, with +IDO suffix, is +NRLD suffix -a-, slot 9. This suffix may also be realized as -ya- (see D8 and R4). We will recall -la- NEG and -a- +NRLD further down in R10.

The occurrence of the suffixes sequence +IPD -fu- and +IDO -fi- may yield -fui- or -fwi- in Smeets' texts, but also -fui- in some other texts.

(E 129) [Smeets, I. 2008: 39 (c)] [21] angkad-fu-fi-n → angkad-fu-fi-n 'the one I had taken on the back of my horse'
-TV. angkd_llevar-en-ancas+IPD. fu8+EDO. fi6 +IND1SG.n3

(E 130) [Mösbach, E. 1936: 16] [14] kim-la-fu-fi-i-n → kim-la-fu-fi-i-n 'I didn't know (about) that'
-TV. kim_saber+NEG. lai1+IPD. fu8+EDO. fi6 +IND1SG.n3

(R 6) Impeditive + EDO

Define RuTrIPDEDO

[["@IP" -> [u"@1"|w"@2"] || _ "@ED"] .o.
["@ED" -> [fi|i] || "@1" _ ] .o.
["@ED" -> i || "@2" _ ]);

R6 encodes the changes exemplified in E129 and E130. RuTrIPDEDO is composed of three concatenated rules. First sub-rule states that "@IP" is either transformed into "u@1" or "w@1" when followed by "@ED", which is expressed in the realization context (the underscore marks the position of the treated element): || _ "@ED". It must be taken into account that rules are applied sequentially, so, when the transformation of "@IP" is carried out, this very same tag, which indicates the context for the subsequent change of "@ED", is lost (it has been transformed in something else). This is why when changing "@IP", new context marks are given for the subsequent "@ED" change. These new tags ("@1" and "@2") are only used as context marks to continue processing the forms. In a later step, context tags are cleared out.

Second sub-rule of RuTrIPDEDO (concatenated by .o.), states that "@ED" is either transformed into fi or i when preceded by "@1".

Third sub-rule states that "@ED" is replaced only by i when preceded by "@2". "@1" and "@2" were established as contextual marks by the previous rule. Rules are applied sequentially.

Whenever +IPD (impeditive) is followed by +EDO (external direct object), the context is given, so the rule is applied accepting three combinations for analysis.

The process just described is illustrated in figure 6 below. It is important to be aware that it shows the generation direction because it is easier to explain and understand. Also note that the FSTs are reversible, therefore rules may be applied backwards, i.e., in the analysis direction. The resulting analysis for any of the three possible spellings -fui-, -fwi-, -fui- will yield the analysis "+IPD. fu8+EDO. fi6".

Fig. 6 Concatenated rules for generation process: simplified view of rule RuTrIPDEDO that processes the interaction between suffixes impertive -fu- and external direct object -fi-.

When +IPD -fu- occurs followed by +IDO -e- they yield the form -fue-, but they may optionally yield the contracted form -fe-.

(E 131) [Smeets, I. 2008: 52 (8.1.6)] [21] ellka-ì-ke-rke-fu-ì-y-ew → ellka-ì-ke-rke-fu-ì-y-ew 'she used to hide it, they say'
-TV. ellka_ocultar+CA.134+CF. ke14+REP. rke12 +IPD. fu8+IDO. e6+IND. y4+3. Ø3+DS3A. ew1
+IPD and +OWN may co-occur in a sequence yielding -fu-el or -el. Smeets gives no examples to this respect, but she writes "The suffix -fu- may occur in indicative and conditional forms and in subordinates except those marked with the plain verbal noun suffix -n +PVN (s4) or the completive subjective verbal noun suffix -uma +CSVN (s4)" [Smeets, I. 2008: 231] [21]. Then, the sequence +IPD +OWN is feasible, and we do have found examples where -el contracts with a previous form ending in e, and examples of -fu-el from other authors:

(E 132 ) [Smeets, I. 2008: 245 (10) [21]
nie-el → nie' to have had'
-TV.nie_tener+OWN.e14

(E 133 ) [Smeets, I. 2008: 249 (7) [21]
kúdaw-pe-el → kúdaw-pel ‘the own job’
-IV.kúdaw_trabajar+PX.pe13+OWN.e14

(E 134 ) [Smeets, I. 2008: 411 (53) [21]
fende-ke-el-chi → fende-kel-chi ‘a sold thing’
-TV.fende_vender+CF.ke14+OWN.e14+ADJ.chi

(E 135 ) [Zúñiga, F. 2006: 144 (54) [21]
ng-we-ke-no-fu-el → ng-we-ke-no-fel ‘to be no more'
-IV.ngwe_serestar+PS.we19+CF.ke14+NEG.no10+IPD.fu8+OWN.e14

(R 7 ) Impeditive + Internal Direct Object or Objective Verbal Noun
define RuTrIPDIDO0VN
[["@IP" -> [u"@3"] |"@3"] | _ ["@ID" |"@EL"]
.o. ["@ID" -> e, "@EL" -> el | | "@3" _ ];

R7 encodes the changes exemplified in E131 and E135. RuTrIPDIDO0VN is composed of two concatenated rules. The first rule states that "+IPD" is either transformed into "u@3" or "@3" (the last one being the elision of u) when followed by "@ID" or "@EL". Once that conversion is done, if the sequence is completed by "@ID", this one is transformed into e, yielding two possible intermediate representations, fu@3e or f@3e. If the sequence is completed by "@EL", this one is transformed into el, also yielding two possible intermediate representations, fu@3el or f@3el. Then "@3" is wiped out giving fue or fe for "+IPD+IDO", and fuel or fel for "+IPD+OWN".

R8 converts "+IP" into u in any other context, yielding fu, as shown in E136:

(R 8 ) Impeditive
define RuTrIPD ["@IP" -> u];

(E 136 ) [Smeets, I. 2008: 63 (17) [21]
kutran-fu'n ‘to have been ill’
-NN.kutran_enfermedad+VRB.Ø36+IPD.fu8+IND1SG.n3

R9 converts "@ED" into fi when is preceded by anything but "@IP", as shown in E137:

(R 9 ) External direct object
define RuTrEDO ["@ED" -> {fi} | | \"@IP" _ ];

(E 137 ) [Smeets, I. 2008: 65 (31) [21]
allkū-tu-nie-fi-n ‘I am listening to him’
-TV.allku_oir+TR.tu33+PRPS.nie32+EDO.fi6+IND1SG.n3

(R 10 ) Internal direct object
define RuTrIDO ["@ID" -> e | | [CON|SVW] _ ]
.o. [\"@ID\" -> e | | a _ @"OE" |\"OE\"]
.o. [\"@ID\" (->) e | | [VOW|"OE"] _ ];

R10 contemplates other possible contexts of realization for +IDO. It becomes e when preceded by any consonant or semi-vowel.

(E 138 ) [Smeets, I. 2008: 87 (21) [21]
kim-e-y-u ‘I recognized you’ lit: ‘I knew you’
-TV.kim_saber+IDO.e6+IND.y4+1.Ø3+DL.u2+DS12A.Ø1
+IDO also becomes e when is preceded by a and is followed by the intermediate form "@U"n, which corresponds to the portmanteau suffix for indicative, 1st person singular; which in turn, it either ends the verb form (because it implies the presence of the null suffix -@ +DS12A following it), or it is followed by -ew +DS3A.

(E 139 ) [Smeets, I. 2008: 157 (17) [21]
pe-e-n ‘you saw me’
-TV.pe_ver+IDO.e6+IND1SG.n3+DS12A.Ø1

(E 140 ) [Smeets, I. 2008: 94 (63) [21]
pe-me-e-ew ‘there he saw me’
-TV.pe_ver+TH.me20+IDO.e6+IND1SG.n3+DS3A.ew1

The last context of realization for +IDO says that "@ID" is optionally transformed into e, which means that it may be elided, when preceded by a vowel or by the tag for negation in imperatives "@NK" (see R5 and examples E128 and E141).

(E 141 ) [Smeets, I. 2008: 94 (63) [21]
ina-ni-a-Ø-lu-mu ‘they have been followed’
-AV.ina_detrás+VRB.Ø36+PRPS.nie32+NRLD.a9+IDO.e6+SVN.lu4+DS3A.mew1

The last rule of the set treating suffixes between slots 10 and 4, deals with the objective verbal noun suffix +OWN-el.

(R 11 ) Objective Verbal Noun
define RuOWN
["@EL" -> [l|{e1}] | | [a|e|"@ID"] _ ]
.o. ["@EL" -> {e1}];
R11 specifies that "@EL", which is how +OVN is encoded, may be converted into r or el when preceded by a, e or "@ID" (e.g. 11.2.6). And in any other context it will be converted into el, see examples below:

(E 142) [Smeets, L. 2008: 114 (26)] [21]
lang-üm-êl chi ufisha 'killed sheep'
-IV.La_morir+CA.m34+OVN.el4+ADJ.chi
-NN.ufisha_oveja

(E 143) [Smeets, L. 2008: 189 (45)] [21]
pi-el-mew 'of what is said'
-TV.pi_decir+OVN.el4+INST.mew

(E 144) [Smeets, L. 2008: 189 (46)] [21]
entu-el 'what is taken (out)'
-TV.entu_sacar+OVN.el4

5.1.2 Special case: suffix -nge-

Corresponding to the verbalizer (see 3.1.1) or stem formative (see 3.1.2) located in slot 36, or to the passive suffix, slot 23, the form -nge- may alternate with -ngi- when followed by the indicative suffix -y-, and the verb form corresponds to the 3rd person non-specified for number. All of which is encoded by the next rule:

(R 12) Alternative nge form

define RuNGE ['@EY' (->) i || _ "@O"y] .o. ['@EY' -> e];

For this rule to work, all mentioned suffixes were encoded as follows:

(D 15)
"[+VRB"{.nge36}"-IV"] : [{ng}"@EY"];
"[+SFR"{.nge36}"-IV"] : [{ng}"@EY"];
"[+PASS"{.nge23}] : [{ng}"@EY"];

An example of each case is rendered below:

(E 145) [Smeets, L. 2008: 456 (3)] [21]
wentru-ngi-y 'they were men'
-NN.wentru_hombre+VRB.nge36-IV+IND.y4+3.Ø3

(E 146) [Smeets, L. 2008: 305 (2)] [21]
weyel-weyel-ngi-y 'he always swims'
-IV.weyel_nadar+RVBR+SFR.nge36-IV+IND.y4+3.Ø3

(E 147) [Smeets, L. 2008: 445 (3)] [21]
elu-ngi-mapu 'he was given land'
-TV.elu_dar+PASS.nge23+IND.y4+3.Ø3
-NN.mapu_tierra

5.1.3 Special case: verb i- 'to eat'

The verb i- may be realized as i-, iy- or yi- depending on the context. To apply the rules (R13) that regulate this verb form in the different contexts, the verb i- has been encoded as follows:

(D 16) ["-TV".i_comer]: "@i";

First sub-rule of R13 avoids i- to be recognized and analyzed as the final i of any word by deleting it.

(R 13) Forms of verb i- 'to eat'

define Verbi
["@i" -> @ || _ #.] .o.
["@i" -> "@G"{iy} || _ [a|e|"@Y"|"@ID"|"@EL"]] .o.
["@i" -> [yi] || _ k [i|ü|"@NK"], "@i" _ , _ "@i"] .o.
["@i" (->) [yi] || _ w] .o.
["@i" -> "@G"{i}];

Second sub-rule converts "@i" into the intermediate form "@G"{iy} which set the verb ready to be part of a compound where "@G" will optionally become g if preceded by a vowel (see E116, E117, p. 22; R1 and R2, p. 22). This change is carried out when "@i" is followed by a, e or the intermediate forms "@Y", "@ID", "@EL".

(E 148)
dewma-iy-a-l-mew → dewma-gly-a-l-mew 'while preparing food'
-TV.dewma_hacer+TV.i_comer-CR.TV+NRLD.a9
+OVN.el4+INST.mew

(E 149) [Smeets, L. 2008: 204 (125)] [21]
i-el → iy-el 'what had been eaten'
-TV.i_comer+OVN.el4

Third sub-rule in R13 converts "@i" into iy in different contexts:

1) when "@i" is followed by ki, kü or the intermediate form "@NK" (see R5 and E128, p. 23):

(E 150) [Smeets, L. 2008: 445 (3)] [21]
i-ki-fi-l-nge → yi-ki-fi-l-nge 'you need not eat it'
-TV.i_comer+NEG.ki10+EDO.fi6+CNI.l4
+IMP2SG.nge3

2) "@i" is converted into iy when preceded by itself and followed by itself, this way the rule is managing reduplication of the verb root i- 'to eat':

(E 151) [Smeets, L. 2008: 307 (8)] [21]
i-i-kânu-fi-n → yi-i-kânu-fi-'l 'I ate it quickly'
-TV.i_comer+RVBR+SFR.Ø3-IV+PFPS.kânu32
+EDO.f16+IND1SG.n3
In the fourth sub-rule, "@i" is optionally converted into yi when followed by w:

(E 152 ) [Smeets, I. 2008: 263 (11) ]  
*we-me-ke-la-y* 'he no longer goes there to eat (as he used to)'
- TV. i_comer+PS.wel19+TH.me20+CF.ke14+NEG.la10+IND.y4+3.Ø3

(E 153 ) [Smeets, I. 2008: 260 (4)]  
*we-la-n*  →  *yi-we-la-n* 'eat no more'
- TV. i_comer+PS.wel19+NEG.la10+IND1SG.n3

Finally, fifth sub-rule states that "@i" is converted in the intermediate representation "@G"i in any other context. So, if it forms part of a compound being the second member, it can up bring a g before itself when the previous element of the compound ends in vowel (see E148).

(E 154 ) [Smeets, I. 2008: 309 (4)]  
*hüra-fi-l* 'i ate reluctantly'
- TV. i_comer+AIML.puda+EDO.fi6+IND1SG.n3

(E 155 ) [Smeets, I. 2008: 43 (f)]  
*fal-aun* 'i must eat'
- TV. i_comer+FORCE.fal25+IND1SG.n3

5.1.4 Special case: verb entu- 'to take out'

The verb entu- may be realized as *ntu-, entu- or nentu- depending on the context. To apply the rules (R14) that regulate this verb form in the different contexts, the verb entu- has been encoded as follows:

(D 17 ) ["-TV*.entu_sacar_quitar]: ["@VE"{ntu}];

(R 14 ) **Forms of verb entu- 'to take out'**

define Verbentu

["@VE"  →  [["@G"e]{ne}]  ||  [.#.|a|i]  _  ] .o.
["@VE"  →  [ne]{g}  ||  u  ] .o.
["@VE"  →  e  ||  [d|f|m]  _  ] .o.
["@VE"  →  {ne}  ||  [e|u]  _  ] .o.
["@VE"  →  e  ];

There are 5 different contexts shaping the form of this verb. First sub-rule of R14 indicates that "@VE" may be converted into (the intermediate form) "@G"e or ne at word beginning, or after a or i, in both last cases the tag "@G" is optionally converted into g (see R1 and E116, p. 22):

(E 156 ) [Smeets, I. 2008: 448 (32)]  
*entu-fi-y-i'n* 'we took him out'
- TV. entu_sacar+EDO.fi6+IND.y4+1.Ø3+PL.in2

(E 157 ) [Smeets, I. 2008: 318 (8)]  
*inentu-y* 'they fixed a date'
- TV. entu_sacar-NN.an2tu_dia+IND.y4+3.Ø3

(E 158 ) [Smeets, I. 2008: 407 (24)]  
*tayma-entu-ngay-pa-y* 'they were taken out there'
- TV. tayma_eliminar−TV. entu_sacar−CR.TV
+PASS.nge23+HH.pa17+IND.y4+3.Ø3

(E 159 ) [Smeets, I. 2008: 315]  
*witra-entu-n* 'i pulled out'
- TV. witra_levantar−TV. entu_sacar−CR.TV
+IND1SG.n3

(E 160 ) [Smeets, I. 2008: 409 (40)]  
*dulli-entu-a-y-i'n* 'we will choose him'
- TV. dulli_elegir−TV. entu_sacar−CR.TV
+NRLD.a9+IND.y4+1.Ø3+PL.in

(E 161 ) [Smeets, I. 2008: 553 (rapi-)]  
*rapi-entu-y* 'he threw up'
- TV. rapi_vomitar−TV. entu_sacar−CR.TV
+IND.y4+3.Ø3

Second sub-rule of R14 converts "@VE" into ne after i or eliminates it, which means that the verb may be realized as *ntu- when the tag is eliminated, or nentu- when the tag is converted:

(E 162 ) [Smeets, I. 2008: 318 (8)]  
*wemü-entu-ngay-ye-m* 'they were suddenly expelled without realizing (it)'
- TV. wemü_perseguir−TV. entu_sacar−CR.TV
+PASS.nge23+SUD.rume21+CF.ye5+IVN.m4

(E 163 ) [Smeets, I. 2008: 556 (rufü-)]  
*rufü-entu-me-ki-y* 'he is busy serving out there'
- TV. rufü_servir−comida−TV. entu_sacar−CR.TV
+TH.me20+CF.ke14+IND.y4+3.Ø3

Third sub-rule converts "@VE" into e after d, f or m yielding the entu- form of the verb:

(E 164 ) [Smeets, I. 2008: 405 (7)]  
*ad-entu-a/ 'how to settle'*
- NN.ad_forma−TV. entu_sacar+NRLD.a9+OVN.el4

(E 165 ) [Smeets, I. 2008: 201 (102)]  
*ütrüf-entu-fi-n* 'i have thrown it away'
- TV. ütrüf_tirar−TV. entu_sacar−CR.TV
+EDO.fi6+IND1SG.n3

(E 166 ) [Smeets, I. 2008: 486 (16)]  
*kim-entu-a-n* 'i shall declare'
- TV. kim_saber−TV. entu_sacar−CR.TV
+NRLD.a9+IND1SG.n3

Fourth sub-rule converts "@VE" into ne when preceded by e or u yielding the nentu- form of the verb:
5.1.5 Special case: radical consonant alternation before causative -üm-

Mapuche verb roots which have an intransitive meaning may be transitivized by adding causatives suffixes -el-, -ül-, -üm-, slot 34, the factitive -ka- or transitivizer -tu- suffixes, slot 33. Few roots undergo a change through this process, actually, Smeets says that it is an "unproductive relic phenomenon" [Smeets 2008: 53] [21]. She gives the following exhaustive list:

af- 'to come to an end' → ap-üm- 'to finish'
lef- 'to run' → lep-üm- 'to make run (animals)'
traf- 'to fit in/on' → trap-üm- 'to cause to fit in/on'
legg- 'to come up (plants)' → llek-üm- 'to plant' (tr.), but lleg-üm- 'to make come up'
nag- 'to go down' → nak-üm- 'to carry down', but nag-üm- 'to take down'
là- 'to die' → lang-üm- 'to kill'

We have also found other cases:

trof- 'to explode, crack' (itr.) → trop-üm- 'to crack' (tr.)
nel- 'to get loose' → nel(k)-üm- 'to let loose, to set free'
liü- 'to burn' (itr.) → liüp-üm- 'to burn' (tr.), 'to set fire'

As the goal of the FST is analysis, the system was set for the maximum analysis possible. So, instead of introducing both forms (intransitive and transitive), only the intransitive verb form was introduced in the lexicon, together with the creation of a rule to handle the radical change.

D 18 ) Encoding of forms with radical change:

The forms on the list have a tag on the right side. "@FP" when the root has to end in $f$ for the intransitive meaning and in $p$ for the transitive sense. "@NG" appears when the root does not change anything regarding intransitiveness, and add $ng$ when transitive. "@GK", transitive ending in $g$, transitive ending either in $g$ or $k$. The later are the ones with a "but" on the "list of forms with radical change" (18).

D 19 ) Causative -üm- encoding:

R 15 ) Radical consonant alternation before -üm-

define RuCAlt01

define RuCAlt02

The above set of rules is similar to the one defined by RuTriPDED0 (R6 p. 24), in the sense that it follows the same logic. Basically, when any of the tags "@NG", "@FP" or "@GK" enters in contact with "@ÜC", the transitivizing option is activated imposing a new context tag "@4", "@5", "@6" to allow the subsequent change of "@ÜC" into $ü$. After these two steps, context tags ("@4", etc.) are wiped out.

Rule RuCAlt02 operates on the intransitive change, i.e., it either eliminates the tag or transforms it into the intransitive form. The following analyses show that processes described above are successfully carried out:
le-üm-kantu-nge-y 'they made it run' (they made a mare run for exercise)
-IV. lef_correr+CA. üm34+PLAY.kantu22
+PASS. nge23+IND.y4+3.Ø3

traf-me-n 'I went to meet' (somebody)
-IV. traf_encajar+TH. me20+IND1SG. n3

traf-üm-a-fi-n 'I will gather' (it)
-IV. traf_encajar+CA. üm34+NRLD. a9
+EDO. fi6+IND1SG. n3

lleg-mu-m 'where it had grown up'
-IV. lleg_crecer+PLPF. mu7+IVN. m4

lleg-üm-fi-y-u 'we brought him down'
-IV. lleg_crecer+CA. üm34+EDO. fi6+IND1SG. n3
+1.Ø3+DL. u2

na-üm-fi-y-n 'it went down / the going down'
-IV. nag_bajar+CA. üm34+EDO. fi6+IND1SG. n3
+NEG. no10+NRLD. a9+OVN. el4

5.2 Morphotactics: constructing the verb form

As it was explained in Morphotactics, p. 17, morphotactics is the set of constraints that regulates the co-occurrence of morphemes. Once the lexicon and suffixes that interact in the verb form are declared (see 4.7 Roots encoding, p. 20 and 4.8 Suffixes encoding, p. 20), it is necessary to regulate their interaction.

We have introduced the Mapuche verb form in section 3, p. 3. In a concise way, the verb is a stem followed by a series of suffixes that complete the verb form.

5.2.1 Stems codification

Section 3.2 Verb stems, p. 9 exposes different stem configurations. Most simple stem type is formed by a single verbal root. Verb suffixes may be added immediately to this type of stem. See example E 44, p. 9.

Other type of stem that accepts verb suffixes immediately is the simple (implying no suffixes) compound where one of the members is a verbal root, the other member may be another verbal root, an adjectival, adverbial, nominal, numeral or a question root. See examples E 45, E 46, p. 9 and following table.

| Stem | Suffixes |
|------|----------|
| Verbal root | +Suffixes |
| Verbal root + Non-verbal root | +Suffixes |

Table 5 Simple stem forms

(D 20) Simple stems encoding

define CMPVBVAL [CAjVbVSTEM|CAvVbSTEM|CNnVbSTEM|CQtVbSTEM|CVbAjSTEM|CVbAvSTEM|CVbNnSTEM|CVbVbSTEM];
CAjVbVSTEM Complex adjective+verb compound stem  
CAvVbSTEM Complex adverb+verb compound stem  
CnVbSTEM Complex noun+verb compound stem  
Cq1VbSTEM Complex question+verb compound stem  
CVbAjSTEM Complex verb+adjective compound stem  
CVbAVSTEM Complex verb+adverb compound stem  
CVbnSTEM Complex verb+noun compound stem  
CVbVSTEM Complex verb+verb compound stem

Definition D20 above, defined as CMPVBVAL, digests, among others in the FST script, the possible Mapuche simple stems. CMPVBVAL stands for "verb compounds with their corresponding valence" (see 5.2.1 below).

**Compounds encoding.** In D21, formCNnVbROOT encodes the form of a "noun + verb" compound. The whole form is enclosed in brackets and the tag "-NVCR" is attached to it. Then, sub-rule CnVbSTEM applies neutralization of tags (see 5.2.1 below) and verb valence (see 5.2.1 below) to the compound.

(D21 ) Noun + verb compound  
define formCNnVbROOT [[NROOT [TVROOT|TVROOT]]"-NVCR"];  
define CnVbSTEM [RuIVCNnVb .o. RuTVCNnVb .o. RuChnVb01 .o. [neutCNnVb .o. formCNnVbROOT]];

Neutralization of tags. Compound stems, complex stems (see Complex single root stems, p. 33) and complex compound stems (see Complex compound stems, p. 34), are qualified as "complex" because they incorporate suffixes into the stem, and have their own rules of interaction. For this reason, PoS and suffixes tags are converted into different tags while applying the inner compound rules (R16). We call this process "neutralization" because it makes general rules not affect stems. This change is reverted before the analysis output, so the user does not have to interpret a wider set of tags. Neutralization is applied first, and then compound rules are applied to the resulting form, therefore the rules are generated taking into account the converted tags (see CnVbSTEM in D21). R16 is an example of how neutralization is applied:

(R16 ) Neutralization of PoS and suffixes tags (sample)  
define NeutNn ["-nn0" <- ";NN"];  
define NeutIV ["-iv0" <- ";IV"];  
define NeutTv ["-tv0" <- ";TV"];  
define NeutAdjdo ["+adjdo0" <- "+ADJDO"];  
define NeutCa ["+ca0" <- "+CA"];  
define NeutDistr ["+distr0" <- "+DISTR"];  
define Neutih ["+ih0" <- "+HH"];  
define NeutNomag ["+nomag0" <- "+NOMAG"];  
define NeutPvm ["+pvm0" <- "+PVN"];  
define NeutRef ["+ref0" <- "+REF"];  
define NeutTh ["+th0" <- "+TH"];  
define NeutTr ["+tr0" <- "+TR"];  

Valence in compounds. When one of the roots in a compound is a verb and the other is not, the resulting compound gets the valence from the verb root. When both members of a compound are verb roots, the valence is derived from the second. This needs to be encoded because transitive verbs take suffixes that intransitive ones do not.

(R17) Valence in verbal compounds (sample)  
define RuIVCNnVb ["-CR.IV" <- "-NVCR" || 
"-nn0" $["-iv0"]$ ];  
define RuTVCNnVb ["-CR.TV" <- "-NVCR" || 
"-nn0" $["-tv0"]$ ];  
define RuIVCVbVb ["-CR.IV" <- "-VCR" || 
["-tv0"] $["-iv0"]$ ];  
define RuTVVCbVb ["-CR.TV" <- "-VCR" || 
["-tv0"] $["-iv0"]$ ];

R17 has two examples of valence application, one for "noun + verb" compounds and another for "verb + verb" compounds. Each of them have a rule for transitive and another for intransitive valences. The tag "-NVCR", that was added to the compound in rule formCNnVbROOT (D21), is transformed into "-CR.IV" when preceded by the sequence of neutralized tags "-nn0" ? "-iv0" (?) indicates zero or more elements in between). This process establishes the intransitive valence for this compound. Transitive process is analogous.

When the compound is made up of two adjectives, two nouns or two verbs, we need to process the compound in a way to not accept equal roots, in which case it would not be a compound but a reduplicated root.

(D22 ) Verb + verb compound  
define formCVbVbROOT [[%< [TVROOT|IVROOT] %# < [IVROOT|TVROOT] %#] "-VCR"];  
define neutCVbVb [NeutIV .o. NeutTv];  
define preCVbVbROOT [._eq(formCVbVbROOT, %<, %#)];  
define CVbVbSTEM [RuIVCVbVb .o. RuTVVCbVb .o. [neutCVbVb .o. formCVbVbROOT]];  

As in the case of the "noun + verb" compound (D21), in D22 the first rule formCVbVbROOT defines the elements and their order in the compound, but it also adds some marks to the roots. Both roots are marked with < on the left and # on the right: %< ROOT %. The % (percentage) escapes the symbols to read them literally. Then, the whole form is enclosed in brackets and the tag "-VCR" is attached to it. The next rule neutCVbVb, defines the tags to be neutralized (see Neutralization of tags, p. 30).

Rule preCVbVbROOT filters from the output side of formCVbVbROOT all those strings where some sub-strings

\[39\] This notation is equivalent to ?* "-iv0" ?*: "-iv0" surrounded by none or any amount of elements to the right and to the left.
occurring between the delimiters < and # are different. This rule\[^{40}\] is meant to treat reduplicated roots, but we have modified it a little, so we can apply it to the compounds in order to not analyse reduplication as composition. Actually, what we do is subtract from the form (defined by \(\text{formCVbVbROOT}\)) the result of the calculus made at \(\text{preCVbVbROOT}\), obtaining only those forms where both members are different.

Finally, \(\text{CVbVbSTEM}\) holds the result of applying neutralization, valence definition (see [Valence in compounds](#)), and the subtraction explained in the previous paragraph.

All type of stems will be later collected under the rule \(\text{VERBSTEM}\) (D28, p. 34), where \(\text{CNPBVAL}\) (see table 5 and D20) is one of them.

Stems formed with a verbalizer suffix. One more degree of complexity is given by the necessity of some single roots or compounds of adding a verbalizing suffix in slot 36 (see section 3.1.1 Verbalizers (slot 36), p. 4) to be used as verb stems. Single roots that need this kind of suffix are adjectives, adverbs, nouns, numerals, onomatopoeia, proper nouns and question forms. Reduplicated roots of any category also need these suffixes, which are called "stem formative" in this case (see section 3.1.2 Stem formative (slot 36), p. 4). Compounds where none of the two roots forming them is a verb, also need a verbalizer in slot 36.

Table 6 summarizes what have been explained in the previous paragraph.

### Table 6 Simple stem forms: 1° complexity

| Stem | Verbalizers Slot 36 | Suffixes |
|------|---------------------|----------|
| Non-verbal root | +VRB | +Suffixes |
| Non-verbal compound | +VRB | +Suffixes |
| Reduplicated root | +SFR | +Suffixes |

\(\text{(D 23 ) Verbalizers (slot 36)}\)

\[
\begin{align*}
&["+\text{VRB}\{.036\}] : \emptyset \quad \text{["+\text{VRB}\{.nge36\}"-IV"] : \{\text{n}@\text{EY}\}} \\
&["+\text{VRB}\{.tu36\}] : \{\text{tu}\} \\
&["+\text{VRB}\{.ntu36\}] : ["@N"\{\text{tu}\}] \\
&["+\text{VRB}\{.136\}] : 1 \\
&["+\text{VRB}\{.ye36\}] : \{\text{ye}\};
\end{align*}
\]

\(\text{(D 24 ) Single non-verbal roots + verbalizer}\)

define \(\text{SPNVBROOT}\) [AJROOT|AVROOT|NROOT|NUROOT|PROPN|QROOT] SVRB;

\(\text{SPNVBROOT}\) states that any of the single roots it collects must be followed by a verbalizer (collected under \(\text{SVRB}\)) in order to occur with verbal suffixes. R18 exposes the rules that regulate the suffixation of verbalizers by category (see section 3.1.1, p 4):

\(\text{(R 18 ) Non-verb roots forming verb stems}\)

define RuAj ["-AJ"|"-CAJ"] => _ ?\^\{.036|{.136}|{.nge36}|{.ntu36}\};

\(\text{Rule RuAj in R18 allows adjectives, compounds made of two adjectives and complex adjective stems (see \text{reftp:59 Complex single root stems}, p. 33) to be completed as verbal stems by suffixes -Ø-, -l-, -nge- or -ntu-, slot 36 (see section 3.1.1 Verbalizers (slot 36), p. 4).}\)

\(\text{Rule RuAv allows adverbs and complex adverb stems (see 5.2.1, p. 33) to be completed as verbal stems by suffixes -Ø-, -l-, -nge- or -ntu-, slot 36.}\)

\(\text{Rule RuNn allows nouns, proper nouns, nominal compounds and complex noun stems (5.2.1, p. 33) to be completed as verbal stems by suffixes -Ø-, -l-, -nge- or -ntu-, slot 36.}\)
Rule RuNű allows numerals to form verbal stems with suffixes -Ø-, -1- or -nge-, slot 36.

Rules RuQc and RuQt regulate verbalizing suffixes for question roots, there are only four question roots and they have diverse behaviour, so they have been encoded distinctively, as shown in D25:

(R 20) Reduplicated roots stem formation

Reduplicated root stems. As shown in table 6 (p 31), reduplication, even verbal one, needs what Smeets calls a stem formative (slot 36) to further attach verb suffixes. We explain the case of nominal reduplication encoding, which is analogous to the other two types, verbal and onomatopoetic.

(R 19) Nominal root reduplication

First rule in R19 neutralizes the nominal tag (see 5.2.1, p. 30). Second rule marks the reduplicated element and adds a tag to the entire structure. In InsNRoot, [...] (Epsilon modifier)4 produces a rule that inserts one instance of \(<\text{NROOT}\) in between > and "\(\text{~RNRN}\)", which is the right side of the form defined in the previous rule NROOTx2. Rule APPLYYN combines and applies previous configurations. Finally, REDNNROOT cleans < and > from the grammatical representation, filter the form out of the previous rule, and clean any tag from the lexical side, to end up in a clean analysis (see examples E32, p. 8; E48, p. 9; E146, p. 26 and E151, p. 26):

(R 20) Reduplicated roots stem formation

define REDSTEMS [REDONROOT|REDVBROOT| REDNNROOT] SSFR;

define RuRdOnSt ["~RNRN" => _ ?* {.Ø36}];
define RuRdVbSt ["~RVRK" => _ ?* {Ø36}];
define RuRdNnSt ["~RNNR" => _ ?* {Ø36}];
define RDROOTSTEM [RuRdOnSt .o. RuRdVbSt .o. RuRdNnSt .o. REDSTEMS];

R20 assigns the appropriate +SFR to each type of reduplicated root to convert them into verbal stems. Reduplicated noun, onomatopoeia and verb roots forming stems are collected in RDROOTSTEM, which in turn will be part of the VERBSTEM definition (see D28, p 34).

Non-verbal compounds. One type is made up by two adjectives, which is recognized as an adjectival compound; another types are "adjective + noun", or two nouns, both recognized as nominal compounds. Another compound, not registered by Smeets, but present in other authors’ texts, is "numeral + noun", also recognized as nominal compound.

As single non-verbal roots, these compounds may be verbalized by a suffix of slot 36. The same suffixes that verbalize single adjectives, verbalize also adjective compounds. The same suffixes that verbalize single noun roots, verbalize nominal compounds. These are collected in their own rule: CPNVBROOT in D27. Then, CPNVBSTEM applies verbalization restrictions:

(R 19) Nominal root reduplication

define NROOTNT [NeutNn .o. NROOT];
define NROOTx2 [%< NROOTNT %<~RNNR>];
define InsNRoot [... -> %< NROOTNT > | %> $[..]~RNNR];
define APPLYYN [NROOTx2 .o. InsNRoot];
define REDNNROOT [%< %<%> .o. _eq(APPLYYN, %<%>) .o. %<%>~RNNR -> Ø];

First rule in R19 neutralizes the nominal tag (see 5.2.1, p. 30). Second rule marks the reduplicated element and adds a tag to the entire structure. In InsNRoot, [...] (Epsilon modifier48) produces a rule that inserts one instance of \(<\text{NROOT}\) in between > and "\(\text{~RNRN}\)", which is the right side of the form defined in the previous rule NROOTx2. Rule APPLYYN combines and applies previous configurations. Finally, REDNNROOT cleans < and > from the grammatical representation, filter the form out of the previous rule, and clean any tag from the lexical side, to end up in a clean analysis (see examples E32, p. 8; E48, p. 9; E146, p. 26 and E151, p. 26):

(R 20) Reduplicated roots stem formation

define REDSTEMS [REDONROOT|REDVBROOT| REDNNROOT] SSFR;

define RuRdOnSt ["~RNRN" => _ ?* {.Ø36}];
define RuRdVbSt ["~RVRK" => _ ?* {Ø36}];
define RuRdNnSt ["~RNNR" => _ ?* {Ø36}];
define RDROOTSTEM [RuRdOnSt .o. RuRdVbSt .o. RuRdNnSt .o. REDSTEMS];

R20 assigns the appropriate +SFR to each type of reduplicated root to convert them into verbal stems. Reduplicated noun, onomatopoeia and verb roots forming stems are collected in RDROOTSTEM, which in turn will be part of the VERBSTEM definition (see D28, p 34).

―

48 Epsilon modifier [...] The LHS of a rule may be wrapped in the epsilon modifier, in which case any epsilons on the LHS get a special interpretation, where only one empty string is assumed to exist between each symbol in the input string. For example, the rule: [...] -> x will produce a transducer that maps the input string a unambiguously to xxx. Also, [...] will simply produce a rule that inserts one instance of the RHS whenever the context is matched: [...] -> x will map aaa to xaxaxax.

Hulden, M. in https://code.google.com/archive/p/foma/wikis/RegularExpressionReference.wiki
Complex single root stems. As we have explained before (Neutralization of tags, p. 30), stems made up by a single root, a compound or a reduplicated root that incorporates at least one suffix (rarely more than three) into the structure are considered "complex stems".

Complex single root stems (one root plus one or more suffixes forming a verb stem, see 3.2, p. 9) that we encode are adjectival, adverbial, nominal, numeral, questions and nominalized verbs (for the later see sections 3.1.7 Verb inflectional nominalization, p. 7; and 3.1.8 Verb derivational nominalization, p. 8).

We explain here the complex nominalized verb stem. The other ones follow the same procedure with the appropriate rules for their category; they were listed in item 3.2, p. 9 as "Complex single root stems".

(R 21) Complex nominalized verb stem: 1st step

define formCXVBROOT [[IVROOT|TVROOT] (CA) (TRFAC) (REF) (ST) (JH) (NRLD) [FLECNOM|NMZ] SVRB];

In R21 we have a composition of 9 rules, some of them including two or three sub-rules. As in the treatment of compounds, we first define the form and order of elements in the stem. In this case formCXVBROOT states that the stem begins with a transitive or intransitive verb root. Then there is a series of suffixes that are optional, which means that they can co-occur (rarely more than three) in any combination, respecting the order. These suffixes are causative -l- or -m-, slot 34; factitive -ka- or transitivizer -tu-, slot 33; reflexive/reciprocal -w-, slot 31; stative -le-, slot 28; hither -pa-, slot 17; and non-realized situation -a-, slot 9. Then come the obligatory nominalizers, those may be inflected (see 3.1.7, p. 7) or derivational (see 3.1.8, p. 8). A verbalizer, slot 36, completes the stem.

(R 22) Complex nominalized verb stem: 2nd step

define neutCXVb [NeutAdjdo .o. NeutAdjqe .o. NeutCa .o. NeutFac .o. NeutHh .o. NeutNrld .o. NeutNomag .o. NeutPvm .o. NeutRef .o. NeutSt .o. NeutSvn .o. NeutTr .o. NeutTv .o. NeutTv]; define CXVBROOT [neutCXVb .o. formCXVBROOT];

define RuCxV01 [-$$["-iv0"|"-tv0"]**"+[OVN]"|"+IVN"|"+TVN"|"+AVN"|.l34]|"+CSVN"|"+NOMPT"|"+NOM"] ];

Tag neutralization of all the members in the stem, and the application to the stem form comes in the second and third rules. Rule RuCxV01 specifies which of the nominalizing suffixes do not form part of this stem. Those are not neutralized because while forbidding them, they need no further interaction rules.

(R 23) Complex nominalized verb stem: 3rd step

define RuCxV02 ["+pvm0"|"+nomag0"|"+adjqe0"]

We explain here the complex nominalized verb stem. The other ones follow the same procedure with the appropriate rules for their category; they were listed in item 3.2, p. 9 as "Complex single root stems".

(R 24) Complex nominalized verb stem: 4th step

define CXVBSTEM [RuCxB01 .o. RuCxB02 .o. RuCxB03 .o. RuCxB04 .o. RuCxB05 .o. RuCCXBSt .o. RuPr50 .o. CXVBROOT];

Final rule CXVBSTEM compiles all together producing the final possible forms for this type of stem. All complex single root stems are also collected in VERBSTEM (see D28).

Complex reduplicated root stem. As it was explained with example E58, p. 10, this stem is listed as a single root complex stem because it is "one" root and "one" stem that reduplicate, i.e., the whole stem reduplicates. This form was not encoded as a compound nor as a single root, but in the section that deals with reduplicated roots. The difference between the rule for this case and the one presented in 19 for nominal reduplication, is that the root is encoded together with the suffix, and that the identifying tag suits with the category of the stem, see below:

(R 25) Verbal root reduplication

define VBROOTTNT [[IVROOT|TVROOT] (CA)]; define VBROOTx2 [%< [IVROOT|TVROOT] (CA) %>="-RVBR"]; define RuVbCa [-$$["-IV"|"-TV"]**"+[OVN]"|"+AVN"|.l34]|"+CSVN"|"+NOMPT"|"+NOM"] ];

define RuCxB01 [-$$["-iv0"|"-tv0"]**"+[OVN]"|"+IVN"|"+TVN"|"+AVN"|.l34]|"+CSVN"|"+NOMPT"|"+NOM"] ];

The differences we have mentioned in the paragraph above are found in the line starting with "define VBROOTx2", where there is an optional suffix CA (causative), and the corresponding tag for the reduplicated verb root "-RVBR". The causative suffix -śim- is the only one found in a reduplicated stem, at least in Smeets' texts.
Complex compound stems. Basically, this type of stem is formed in the same way as the "complex single root stem" (p. 5), but implicating two roots. Complex compound stems (see p. 9) that we encode are:

- adjective (+ suffixes)\(^{42}\) + noun (+ suffixes): see E192;
- adjective (+ suffixes) + verb + nominalizer, see E193;
- adverb (+ suffixes) + verb, see E188;
- noun (+ suffixes) + noun (+ suffixes), see E194;
- noun (+ suffixes) + verb, see E189;
- verb (+ suffixes) + noun, see E191;
- verb (+ suffixes) + verb, see E190;

All complex compound stems are collected together with simple compounds in CMPVBVAL, see D20. And as for the previous types of stems, the later ones are also collected in rule VERBSTEM (D28), which is summarized in table 7:

( D 28 ) Verb stems

**Table 7 Mapudüngun stems**

| Stem | Verbalizers | Suffixes |
|------|-------------|----------|
| verbal root | +SVR |  |
| verbal root + verbal root | +SVR |  |
| verbal root + non-verbal root | +SVR |  |
| non-verbal root + verbal root | +SVR |  |
| non-verbal compound | +SVR |  |
| reduplicated root | +SVR |  |
| root + suffixes | +SVR |  |
| root + suffixes + root | +SVR |  |
| root + suffixes + root + suffixes | +SVR |  |

Different conformations of stems where identified in section 3.2 Verb stems, p. 9; in this point, we expose the rules that regulate the interaction among the elements introduced above, roots (section 4.7) and suffixes (section 4.8), which take part of the different types of stems.

Complex compound stems "Adverb + optional causative + optional transitivizer + verb"

Rule: \(^{43}\) AVROOT (CA) (TR) [IVROOT2|TVROOT2];

\(^{42}\) Parenthesis express optionality.

\(^{43}\) Rules are presented here in a simple way, just to show the elements involved, but actually, rules are much more complex in the system because they have to deal with the generation of the compounds, the addition of tags to carry out the processes, and the elimination of these tags once used. See this example of one of the simplest rules in the FST script, which does not have to deal with the addition of a verbalizer.
+SVN.lu4

"Verb + optional causative + optional transitivizer or factitive + optional reflexive + optional hither + noun"
Rule: [TVROOT|IVROOT] (CA) (TRFAC) (REF) (HH) NROOT2;
(E 191 ) [Smeets, I. 2008: 456 (8)]

kim-el-tu-che-ke-fu-y 'he used to teach people'
TV.kim_saber+CA.l34+TR.tu33-NN.che_persona +CF.ke14+IPD.fu8+IND.y4+3.Ø3

"Adjective + transitivizer or factitive + noun + optional derivational nominalizer + VRB"
Rule: AJROOT (TRFAC) NROOT2 (NMZ) SVRB;
(E 192 ) [Smeets, I. 2008: 90 (36)]

wisa-ka-sungu-n, ta eymi 'what a dirty talker you [are]!'
AJ.weda_malo+FAC.ka33-NN.düngu_palabra +VRB.Ø36+PVN.n4 -AP.ta_el -PP.eymi_tu

"Adjective + optional transitivizer or factitive + verb + derivational nominalizer + VRB"
Rule: AJROOT (TRFAC) [IVROOT2|TVROOT2] NMZ SVRB;
(E 193 )
küme-ka-puru-fe-nge-y 'he is (always) a good dancer'
AJ.küme_bueno+FAC.ka33-IV.puru_bailar 
+NOMAG.fe+VRB.nge36-IV+IND.y4+3.Ø3

"Noun + optional transitivizer or factitive or non class-change suffixes + noun + optional non class-change suffixes or derivational nominalizers +VRB"
Rule: NROOT (TRFAC|NCC) NROOT2 (NCC|NMZ) SVRB;
(E 194 ) [Smeets, I. 2008: 459 (36)]
ta-yiñ pu peñi-wen-lamngen-wen-nge-n 'we are all related as brothers and sisters' lit: ‘this is our brothers relation sisters relation’
AP.ta_este-SP.yiñ_nuestro COLL.pu 
-NN.peñi_hermano+REL.wen-NN.lamngen_hermana 
+REL.wen+VRB.nge36-IV+PVN.n4

5.3 Morphotactics of verb suffixes

In section "3.1 Verb suffixes", p. 3, we have explained that suffixes belonging to the same slot are mutually exclusive. There are about eighty verbal suffixes spread in thirty-six slots. Some suffixes exclude others for grammatical or semantic reasons, for example, once a verb has taken an inflectional nominalizer, slot 4, it can not take suffixes of mood (slot 4), person (slot 3) and number (slot 2).

To start treating suffixes co-occurrence, we first established the suffix sequence with all the possible variants generated by suffix mobility (see 3.1.5, p. 5 and 5.3.4, p. 38), see next rule:

(D 29 ) Verb suffixes define VERBSUFFIX [(REF) (EXPOO) (PASS) (REF) (TR) (CA) (REF) (TRFAC) (FORCE) (BEN) (FORCE) (PRSPFPFS) (REF) (HH) (CIRCINT) (PLAYSIM) (MIO) (STPR) (BEN) (OS) (IMMSUD) (PLR) (IO) (PASS) (FORCESAT) (PLR) (FORCE) (TH) (PASS1A2A) (PLAYSIM) (IMMSUD) (TH) (PS) (ITR) (HHLOC) (TH) (PS) (REF) (RE) (RECONT) (PLPF15) (CF14) (PX) (REP) (RE) (AFF) (NEG) (NRLD) (IPD) (PLPF07) (EIDO) (CF05) [[[(MOOD) [PERSON|PTMT] (NUMBER)] (DS)] [[FLECNOM|NMZ] (DS) (NCC) (CC) (INST)]]];

The names or tags appearing in D29 encode the suffixes assigned to each slot (see 4.8 Suffixes encoding, p. 20).

The first thing that may call the attention is repetition of some tags in different positions, e.g. PASS, REF, FORCE, IMMSUD, etc. This is to deal with suffix mobility (see 3.1.5, p. 5 and 5.3.4, p. 38).

Also note that almost all suffixes are marked as optional, they are between parenthesis, except for PERSON, PTMT, FLECNOM and NMZ. The Mapuche verb is either finite (PERSON and PTMT) or nominalized (FLECNOM or NMZ). These are the obligatory suffixes for those forms.

Methodology. To encode suffixes occurrence in the Mapuche verb form, we started incorporating the minimal verb form, i.e., an intransitive verb root plus suffixes expressing mood, person and number (see annex 11.3 "Conjugation of the intransitive verb kürpa-‘to come’, p. 63), which are obligatory in a finite verb form. So we continued adding the transitive verb related suffixes. So, we first established a set of rules dealing with the minimal forms for both, intransitive and transitive verbs (see annex 11.4 "Conjugation of the transitive verb pi-‘to say (to tell)’", p. 64).

5.3.1 Verb paradigms

In D29 above, the last two lines reflect the two forms a verb may take. Penultimate line corresponds to finite forms; in slot 4 is mood, in slot 3 is person or the portmanteau morphs.

44 Portmanteau morphs which include a subject marker are assigned subject position (slot 3) [Smeets 2008: 152] [21]. We have also seen that assigning portmanteau morphs in this position allows the conditional marker, obligatory in negative imperative forms, appears in its natural position, slot 4 for mood.
(see slot-ØPTMT.aff in annex 11.2.1, p. 49); in slot 2 is number, and dative subject (used in transitive forms) is in slot 1.

It was also necessary to incorporate suffixes assigned to slots 23 and 6, as they complete the transitive verb paradigm (see section 3.1.6 Verb paradigms, p. 6), and negation suffixes in slot 10, even though they are not strictly obligatory and occur in transitive and intransitive forms, they complement with mood suffixes and have a particular incidence in the case of imperative negative forms, (see annex 11.5 "Negative imperative forms of the transitive verb pi-’to say (to tell)’, p. 65).

Table 8 Intransitive and transitive suffixes per slot

| Slot | 23 | 10 | 6 | 4 | 3 | 2 | 1 |
|------|----|----|---|---|---|---|---|
| Tr   | neg. | mood | pers. | ptmt | num. | - |
| Itr  | neg. | obj. | mood | pers. | ptmt | num. | dative subj. |

Table 8 shows suffixes per slot implied in transitive and intransitive Mapuche verbs. Not all suffixes in table 8 co-occur in a transitive form, for instance, agent markers (slot 23) do not co-occur with direct objects (slot 6) or dative subjects (slot 1).

To regulate the verbal paradigms, thirty-three rules were necessary, some of them containing sub-rules, and some including the interaction with inflectional nominalizers, slot 4. No reference to mood, person or number may be made when a verb takes one of the nominalizers, but nominalized verbs may include agents (E195, E196) or objects with the corresponding dative subject (E195, E198):

(E 195) Smeets, I. 2008: 269 (11) [21] müttrüm-uw-‘lu ‘his calling to’
-TV. müttrüm_llamar+1A.w23+SVN.lu4

(E 196) Smeets, I. 2008: 269 (14) [21] fey-pi-mu-a-fiel ‘what you will tell me’
-TV. feypi_decir+2A.mu23+NRLD.a9+TVN.fiel4

(E 197) Smeets, I. 2008: 394 (38) [21] chem-pi-e-t-ew ‘what they where told by’
-QC. chem_qué-TV. pi_decir-CR.TV
+IDO.e6+AVN.t4+DS3A.ew1

(E 198) Smeets, I. 2008: 485 (5) [21] pe-fi-lu ìchče ‘at my seeing her’

-PP.îńche_yo
-TP.pe_ver+EDO.fi6+SVN.lu4

(R 26) Dependency rule 1

define RuDp01 ["+DS3A" | "+DS12A"] => "+IDO" ? _;

R26 is what we call a "dependency" rule, it says that for suffixes +DS3A and +DS12A to occur it must previously occur the suffix +IDO, i.e., +DS3A and +DS12A depend on +IDO occurrence.

(R 27) Prohibition rule 10

define RuPr10 [~"+CMD" ?* ["+1"{.Ø3}] | ["+3"{.Ø3}{{.ng3}}]]];

R27 is a prohibition rule. The combination of symbols ~+$46 can be read as "it can not be the case that", and the rest of this regexp is read as "the conditional is followed by a 1st person suffix in its null form -Ø-, or the 3rd person suffix in its forms null or -ng-.

(R 28) Obligation rule 9

define RuOb09 [["+NEG"{{.ki10}}{{.kino10}}] => _ ?* "+CNI"];

R28 is an obligation rule which regulates the obligatory occurrence of the conditional marker when there is a negation in the imperative form (see e.r. 11.2.6 and annex 11.5 Negative imperative forms of the transitive verb pi-’to say (to tell)’, p. 65).

5.3.2 Nominalized verbs

Last line of D29 (p. 35), reflects the form of a nominalized verb, either by inflectional (see section 3.1.7 Verb inflectional nominalization, p. 7) or derivational (see section 3.1.8 Verb derivational nominalization, p. 8) nominalizers. In both cases a nominalized verb may be followed by a dative subject (see E197), a non class-changing suffix (see NCC. aff in annex 11.2.4, p. 58), a class-changing suffix (see CC. aff in annex 11.2.4, p. 58), or the instrumental suffix (see section 3.1.10 Instrumental object suffix -mew, p. 9).

To regulate verb nominalization twelve rules were added. Note that these rules regulate co-occurrence among the suffixes of "[FLECNOM|NMZ] (DS) (NCC) (CC) (INST)", and some times with suffixes from other slots; but in general, there are other rules to deal with co-occurrence of these suffixes, or the ones belonging to the transitive and intransitive paradigms, and the derivational ones.

45 "It is remarkable that the subject-object paradigm is completed with suffixes which occupy a position in between derivational suffixes, away from the inflectional block at the end of a verb form. The suffixes -mwe- +2A and -nwe- +1A share their position, slot 23, with the passive marker -nge- [Smeets, I. 2008: 161] [21].

46 -X calculates the complement of X, i.e., finds all the elements in the group that are not part of X, or that are not X. $X$ denotes the language that contains a sub-string drawn from the language X [Hulden, M. in https://code.google.com/archive/p/foma/wikis/RegularExpressionReference.wiki].
(R 29) Nominalization prohibition for completive subjective verbal noun

Define RuPr19 [-$["+CSVN" ?* ["+DS3A" | "+DS12A"] | "+INST" | "+ADJ"]];

R29 forbids dative subject suffixes (slot 1), instrumental, or adjectivizer (class-changing suffix), to appear when the verb has been nominalized by the "completive subjective verbal noun" (slot 4).

(R 30) Obligation for agentive verbal noun

Define RuOb12 ["+AVN" => "+IDO" $[_] ["+DS3A" .ew1]];  

R30 forces the "agentive verbal noun" (slot 4) to occur together with the "internal direct object" (slot 6) and the "dative subject for 3rd person agent" (slot 1) in its form -ew (see E22, p. 7 and E197, p. 36).

(R 31) Only plain verbal noun may be adverbialized

Define RuDp05 ["+ADV" => "+PVN" ?* _];

R31 states that the class-changing suffix -tu may only adverbialize a verb nominalized by the "plain verbal noun" -n- (see E34, p. 8). In other words, the adverbializer depends on the "plain verbal noun" to occur with a verb.

Examples of inflectionally nominalized verbs may be found through e.r. 11.2.6, e.r. 11.2.6, e.r. 11.2.6, e.r. 11.2.6, e.r. 11.2.6, e.r. 11.2.6 and e.r. 11.2.6.

Examples of derivationally nominalized verbs are: E29, E30, E31, E32, E87, and the following ones:

(E 199) [Smeets, I. 2008: 314 (Ø)] [21]

anù-m-ka ‘planting’
-IV.anù_sentar+CA.m34+FAC.ka33+NOM.Ø

(E 200) [Smeets, I. 2008: 314 (Ø)] [21]

-ül-kantu ‘song’
-NN.ûl_canto+VRB.Ø36+PLAY.kantu22+NOM.Ø

(E 201) [Smeets, I. 2008: 314 (Ø)] [21]

yall-tuku ‘illegitimate child’
-NN.yall_hijo-de-un-hombre
-IV.tuku_poner+CR.TV+NOM.Ø

(E 202) [Smeets, I. 2008: 314 (Ø)] [21]

ru-pa ‘time’
-IV.ru_pasar+HH.pa17+NOM.Ø

(E 203) [Smeets, I. 2008: 312 (8)] [21]

angkù-m-tu-we ‘poison’, ‘device to dry things’
-IV.angkù_secar+CA.m34+TR.tu33+NOMPI.we

5.3.3 Occurrence of suffixes between slots 5 and 35

There are thirty more rules to regulate the occurrence of suffixes that are not obligatory in the minimal transitive or intransitive forms. Most of the rules come from descriptions of the suffixes made by Smeets, for example the rule for the reflexive reciprocal -w-. "The suffix -w- does not combine with a suffix in slot 23, 6 or 1. The reflexive morpheme -w- may occur with intransitive verbs, i.e., with verbs which do not take a suffix in slot 6" [Smeets, I. 2008: 291] [21]; R32 reflects the previous description:

(R 32) Reflexive do not occur in transitive forms

Define RuPr48 [-$["+REF" ?* ["+REF" | "+PASS" | "+1A" | "+2A" | "+IDO"]];

In the description of (non) combinations of +REF, Smeets also mentions suffixes of slot 1. These are not collected in R32 because there is a previous dependency rule (R26, p. 36) stating that dative subjects (slot 1) need the +IDO suffix (slot 6) to occur, as this one is forbidden to occur with +REF the condition does not fulfill for the +DS (slot 1) to occur.

(R 33) More involved object obligatory contexts

Define RuOb16 ["+MIO" => "+CIRC" ?* _, _ ?* ["+PASS" | "+EDO" | "+TVN"]];

The rule presented in R33 derives from what we have found in Smeets’ examples, there are no explicit combination rules for the more involved object suffix labelled +MIO. It is important to rule this suffix due to its form -1- after vowel, -ul- after consonant or semi-vowel, sometimes -el- after r. These forms coincide with other suffixes forms like stative’s or causative’s ones, which are proximate in their occurrence position, therefore, they may be erroneously identified.

There are fourteen examples given by Smeets, where +MIO (slot 29) is present. In six of them is preceded by +CIRC (slot 30), circular (erratic) movement suffix -iaw-. There are another six where it co-occurs with +EDO (slot 6), external direct object suffix -fi-. One where it co-occurs with +PASS (slot 23), passive -nge-. And one more where it co-occurs with +TVN (slot 4), transitive verbal noun -fiel-.

+MIO "indicates a more direct, intense or complete involvement of the patient in the event" [Smeets, I. 2008: 287] [21]; +CIRC "denotes an ongoing event which involves movement in no particular direction" [Smeets, I. 2008: 288] [21].

It is not clear for us the semantic or grammatical relation between +MIO and +CIRC, but when these two suffixes co-occur, nor +PASS nor +EDO occur. On the other hand, the other three suffixes in the rule have a grammatical relation with +MIO. For +EDO, the external direct object, +MIO gives a further degree of prominence to the object, see following example:

(E 204) [Smeets, I. 2008: 288 (6)] [21] kayla-tu-kùnu-I-Hiñ ‘I lied to him’
The same happens with the objects denoted in a passive +PASS construction (E205), or in a transitive +TVN clause (E206). In the case of a verb nominalized by the transitive verbal noun suffix, +CIRC and +TVN may co-occur (E207), rule R33 does not prevent it.

\[(E\ 205)\ \{Smeets,\ l.\ 2008: \page{397}\ (62)\} [21] \]
yiñ ngünen-ka-i-ngewe-no-a-m ‘we are no longer deceived’

\[-SP.yin.nuestro\]

\[-NN.ngünen_engaño+VRB.Ø36+OSC.kasa33+MIO.129+PASS.ngé23+PS.wi19+NEG.no10+NRLD.a9+IVN.m4\]

\[(E\ 206)\ \{Smeets,\ l.\ 2008: \page{288}\ (5)\} [21] \]
eymi mi wirar-ül-meke-ke-fiel-mew ifiche ‘you are always shouting at me’

\[-PP.eymi_tu-SP.mi_tu_tuyo\]

\[-IV.wirar_gritar+MIO.129+PR.meke28+CF.ke14+TVN.fiel4+INST.mew\]

\[(E\ 207)\ \{Smeets,\ l.\ 2008: \page{398}\ (5)\} [21] \]
äl küdaw-kiaw-ul-öl-fiel pu ülmen ‘he worked around for the rich people’

\[-SP.ni_mi_su\]

\[-IV.kudaw_trabajar+CIRC.iaw30+MIO.129+COLL pu -NN.ülmen_adinerado\]

\[(R\ 34)\ \text{Circular movement context restrictions}\]

\text{define RuPr41: } [-$("+CIRC"|"+INT"|"+ST"|"+PR") ?* 

\["+ST"|"+PR")]

5.3.4 Treating suffix mobility

As it was explained in section 3.1.5, p. 5, some suffixes may occur in different positions, this is call "mobility". To deal with it, we have declared the slot containing the suffix in all the positions it may appear along the suffixes chain (see D29, p. 35). When a slot holds more than one suffix and only one of them is mobile, we have created a new file containing only the mobile suffix. For example, slot 17, encoded in file slot-17.aff, holds hither -pa- and locative -pu- and only +HH is mobile. We have created the file slot-17M.aff that holds the hither only, and is declared as HH in the script. In D29, HH is found once alone and once together with LOC, as HHLOC, these are the two positions where it may occur; the same stands for the other mobile suffixes.

In R34 above, also in R32, p. 37, there are suffixes repeated on both sides of the expression, left and right of the ?* symbols. These are prohibition rules, interpreted as "it can not be the case that", as we have explained for R27, p. 36. So, "it can not be the case that +ST co-occurs with +ST, or +PR with +PR", in R34. And "it can not be the case that +REF co-occurs with +REF", in R32, p. 37. All of this is to avoid that the same suffix may be recognized twice in a verb form, due to being declared in two different positions. These rules are also used to exclude the co-occurrence of these suffixes with different ones for grammatical or semantic reasons, e.g., R34 states that "it can not be the case that +CIRC co-occurs with +ST, one expresses the opposite idea of the other: circular movement / stative; or +INT (intensifier) with +PR (progressive)."

5.3.5 Over-generation

One of the problems derived from encoding homograph suffixes, or suffixes that do not phonetically realize (null suffixes $\emptyset$) is that first ones may be mistakenly recognized, and the seconds could be virtually recognized at any position. To solve these issues we have generated some rules that do not have to do with the Mapudüngun morphotactics, but help in preventing wrong recognition. These rules have derived from the observation of analysing results (see Ambiguity, p. 48).

\[(R\ 35)\ \text{Forbidden null morpheme sequence occurrence}\]

\text{define RuPr46: } [-$("+NOM" ?* 

\["+[OVN].Ø4] 

\["+SVN].Ø2] 

\["+PR" | 

\["+AVN" | 

\["+SVN" | 

\["+ADV"]];

Rule R35 forbids the nominalizer which has a null form -Ø- to be followed by other suffixes that also have a null realization: +OVN, +SVN, +IMP, +1, +3, +SG, +DS12A

Complete forms. There is another set of eight rules that acts upon the entire verb forms when the stems are verbalized roots, complex stems, compounds or complex compounds. Some suffixes in the stem condition the entire form.

\[(R\ 36)\ \text{nominalization of verbalized noun}\]

\text{define RuPr51: } [-$("+[OVN]" 

\["+AVN" | 

\["+SVN" 

\["+[ADV]]];

The prohibition rule R36 states that a noun or proper noun once verbalized may be nominalized by any of the inflectional nominalizers but not adverbialized. Note that in this rule coexist neutralized tags belonging to the stem with non-neutralized ones that belong to suffix chain of the verb.

5.3.6 Special roots

In this section we explain how roots of Deictic verbs, p. 11, and Defective verbs, p. 11, are encoded. Also, some other special cases are explained:

\text{List of special roots}\n\["+AV"|"+SG01"\{ful_cerca\}: \{ful\]
In the third possible context, these roots must co-occur with hither -pa-, locative -pu- (slot 17), or thither -me- (slot 20):

( E 209 ) [Smeets, I. 2008: 419 (46)]

\textit{pülle-pu-el} 'going to a near place'

-AV.pülle_cerca+VRB.Ø36+LOC.pu17+OVN.el14

(R 38) Realization contexts for 2nd type of defective verbs

define RuVSC02 ['"0SC02" => \_ ?* ["-CR.IV"] ["-CR.TV"], \_ ?* ["+ST"] ["+PRPS"] ["+PFPS"] ["+LOC"] \{.no10\} ["+NRLD"]] .o.

\$ ['"0SC02" ?* ["-CR.IV"] ["-CR.TV"] ?* ["+ST"] ["+PFPS"]];

Rots tagged @SC02 have different contexts of realization too, but they also have forbidden contexts, this is why R38 is a concatenation of two rules. The first rule treats @SC02 roots in compounds, when this roots are before tags -CR.IV, -CR.TV is because they form part of a compound, -CR stands for "compound root":

( E 210 ) [Smeets, I. 2008: 522 (külu-)]

\textit{külu-ru-pa-n antü} 'when the sun is going down'

-IV.külu_apoyar-IV.ru_pasar-\textit{CR.IV, IV}+HH.pal7+PVN.n4 -NN.antü_sol

The second context, obligatorily places @SC02 roots together with stative -le- (slot 28), progressive persistent -nie- or perfect persistent -kunu- (slot 32), locative -pu- (slot 17), negation for conditional -no- (slot 10) or non-realized situation -a- (slot 9):

( E 211 ) [Smeets, I. 2008: 281] [21]

\textit{rekül-kunu-w-øy} 'thenn, he leaned over'

-IV.rekül_apoyar+PFPS.kunu32+REF.w31+IND.y4+3.Ø3

The concatenated prohibition rule states that @SC02 roots can not be part of a compound and form a verb taking the suffixes stative -le- (slot 28), progressive persistent -nie- or perfect persistent -kunu- (slot 32).

(R 39) Realization contexts for 1st type of compounded defective verbs

define RuVSC03 ['"0SC03" ?* "0SC05"] => \_ ?* ["+TH"] ["+HH"];

R39 obliges compounds made with @SC03 and @SC05 roots to occur together with markers for thither -me- (slot 20) or hither -pa- (slot 17):
(E 212 ) [Smeets, I. 2008: 262 (9) ] [21]  
\textit{kim-pùra-me-pa} ': there I realized'  
- TV. kìm squared - IV. pura squared - subir - CR. IV  
+ TH. me20 + HH. pa17 + IND1SG.n3

(E 213 ) [Smeets, I. 2008: 381 (1) ] [21]  
\textit{kim-pùra-me} ': I came to appreciate'  
- TV. kìm squared - IV. pura squared - subir - CR. IV + TH. me20  
+ IND1SG.n3

(R 40 ) Realization contexts for 3rd type of defective verbs  
define RuVSC04  
\$\left[ \text{"@SC06" p* ["+ST" | "+PR" | "+PRPS" | "+PFPS" ]} \right]$

Prohibition rule R40 forbids roots tagged @SC06 to take suffixes stative -le- or progressive -meke- (slot 28), progressive persistent -nie- or perfect persistent -kùnu- (slot 32).

(R 41 ) Realization contexts for 4th type of defective verbs  
define RuVSC05 \$\left[ \text{"@SC07" p* ["+CIRC" | "+ST" | "+PR" | "+PRPS" | "+PFPS" ]} \right]$

Prohibition rule R41 forbids roots tagged @SC07 to take suffixes of circular movement -iaw- (slot 30), stative -le- or progressive -meke- (slot 28), progressive persistent -nie- or perfect persistent -kùnu- (slot 32).

(R 42 ) Realization contexts for verb \textit{nge}- 'to be'  
define RuVSC08 \$\left[ \text{"@SC08" \Rightarrow p? ["+CA" | "+TR" ] , p? ["+HH" | "+TH" ] , p? ["+NEG" ] . o} \right]$

\$\left[ \text{"@SC08" p? ["+CIRC" | "+PR" | "+PRPS" | "+PFPS" ]} \right]$

R42 is another rule made by composition, the first sub-rule defines the contexts where the verb root \textit{nge}- 'to be / to have' must happen. The first context for this root, tagged @SC08, demands it to be followed either by +CA causative (slot 34) -l- or -m-, or by +TR transitivizer (slot 33) -tu-:

(E 214 ) [Smeets, I. 2008: 126 (28) ] [21]  
\textit{nge-me-fi-hi} ': I have taken them'  
- TV. nge squared - me squared + CA. 134 + TH. me20 + EDO. fi6 + IND1SG.n3

The second possible context for \textit{nge}- makes it occur with hither -pa- (slot 17) or thither -me- (slot 20):

(E 215 ) [Smeets, I. 2008: 231 (1) ] [21]  
\textit{nge-me-fu} ': I was there'  
- TV. nge squared - me squared + IPD. fu5 + IND1SG.n3

47 In this example both suffixes appear, but only one is obligatory.

48 The obligatory suffixes of this and next rule co-occur in this verb, which is another possible context.

(E 216 ) [Smeets, I. 2008: 534 (mungel) ] [21]  
\textit{nge-\textit{pa}} -: \textit{yaw-\textit{ki}} - y\textit{m-i} 'You were hanging around here'  
- TV. nge squared - ser squared + HH. pa17 + CIRC. iaw squared + CF. ke14  
+ IND. y4 + 2.m + SG. i2

The last possible context for \textit{nge}- says that it must be followed by a negation marker, slot 10. This rule may be read as an exception to the previous rules: \textit{nge}- must be followed by -l-, -m-, -tu-, -pa- or -me-, except when it takes a negation suffix, making previous suffixes optional:

(E 217 ) [Smeets, I. 2008: 407 (17) ] [21]  
\textit{nge-ke-la-fu-y} ': they were not'  
- TV. nge squared - ser + CF. ke14 + NEG. la10 + IPD. fu8  
+ IND. y4 + 3.Ø

(E 218 ) [Smeets, I. 2008: 194 (64) ] [21]  
\textit{nge-nu} ': there was not'  
- TV. nge squared + NEG. no10 + PVN.n4

Finally, the prohibition concatenated rule for \textit{nge}- 'to be' prevents its co-occurrence with +HH, +TH and the negation for indicative -la-:

(R 43 ) Realization contexts for verb \textit{pepi} 'to be able to'  
define RuVSC07 \$\left[ \text{"@SC09" \Rightarrow p? ["+CA" {,ul34} ] | +FAC ]} \right]$

The root \textit{pepi} 'to be able to' in a single root stem must always be followed by -l- causative form or by factitive -ka-:

(E 219 ) [Smeets, I. 2008: 402 (45) ] [21]  
\textit{pepi-ka-w\textit{u-n}} ': the setting of preparations'  
- TV. pepi - poder - hacer + FAC. ka33 + REF. w31 + PVN.n4

(E 220 ) [Smeets, I. 2008: 545 \textit{pepi} ] [21]  
\textit{pepi-\textit{fa-l-fa-y}} ': it can not be done'  
- TV. pepi - poder - hacer + CA. 134 + FORCE. fa125  
+ NEG. la10 + IND. y4 + 3.Ø

(R 44 ) Realization contexts for verb \textit{ru}- 'to go through'  
define RuVSC07 \$\left[ \text{"@SC10" => p? ["+CA" | "+CF" ]} \right]$

\$\left[ \text{"@SC10" \Rightarrow p? ["+CA" {,u134} ] | +FAC ]} \right]$

Root \textit{ru}- 'to go through' does not occur without direction markers -me- (thither slot 20) or -pa- (hither slot 17), even in complex stems or compounds.

(E 221 ) [Smeets, I. 2008: 247 (3) ] [21]  
\textit{amu-ru-me-y} maw\textit{u-n} - mew 'he went through the rain'  
- TV. amu- ir - IV. ru - pasar - CR. IV + TH. me20  
+ IND. y4 + 3.Ø

\textit{NN. maw\textit{u-n} lluvia} + INST. mew

(E 222 ) [Smeets, I. 2008: 515 \textit{kata}- ] [21]  
\textit{kata-ru-\textit{me-y}} 'it pierced through'  
- TV. kata - perforar - IV. ku - pasar - CR. IV  
+ CA. 134 + TH. me20 + IND. y4 + 3.Ø
Deictic roots. See section 3.3.2, p. 11. Root fa- is tagged @FA and root fe- is tagged @FE, this allows us to apply rules that verbs formed from these roots require.

(R 45) Realization contexts for deictic verbs

Define RuVSC09 ["@FA" => _ ?* ["+CA"|"+ST"]], o. ["@FE" => _ ?* ["+CA"|"+ST"], _ $["+IND"] "+v3"];

R45 states that verbs containing fa- ‘to be like this’ or fe- ‘to be like that’ must also contain causative suffixes -l-, -m-, or stative suffix -y-. However, verbs derived from fe- do not obligatory fit this rule when they end in -y which corresponds to indicative mood, 3rd person (see examples E26 to E72, p. 11).

Verbs with causative and factitive. Finally, there is a number of verbs, which roots are tagged @CF, that "do not take the causative suffix -l- +CA (slot 34) without simultaneously taking the factitive morpheme -ka- +FAC (slot 33)" [Smeets, I. 2008: 301] [21]:

(R 46) Realization contexts for deictic verbs occurring with causative

Define RuVSC10 ["@CF" => _ ?* ["+CA"|"+CA0"].134]] => _ ?* ["+FAC"|"+fac0"]];

R46 do not force roots marked @CF to be followed by causative suffix, instead, it states that the sequence "@CF-l-" must co-occur with factitive -ka-. All of this may be read as "if a @CF root is followed by +CA it must also follow +FAC. This rule also covers complex stems an complex compounds by means of "neutralized" tags (see Neutralization of tags, p. 30).

6 Beyond "A grammar of Mapuche"

"A grammar of Mapuche" [Smeets, I. 2008] [21], our development base, describes the central Mapuche dialect. We have added into the analyser words that are not in Smeets’ work. Compounds that she have not found throughout her study but some other authors mention. Also some minor dialectal variations.

6.1 The spelling unifier

There is a significant variation in Mapudüngun spelling, mainly due to the existence of different spelling proposals, together with the strong influence of Spanish orthography. Some texts may present a mixture of these orthographic proposals along with Castilianized orthography. This is something to sort out before analysing a text, either with rule based analysers or statistical ones, because the divergence in input means more rules for the first ones and poor results for the second ones.

The task of the unifier is to replace characters; from a series of possible inputs, a single output is returned. This process is called “unification”. Initially, the idea was to change different graphemic proposals for the Mapuche language into one single spelling. But the strong influence of Spanish orthography in written Mapudüngun was noticed along the way. Therefore, the final implementation reflects mainly this fact, including anyway a couple of rules related to some of the graphemic proposals. This process is embedded in the analyser.

Mapudüngun vowels should not have accentuation marks, if there is any, it is transformed into its non accented version:

\[\hat{a} \rightarrow a, \hat{e} \rightarrow e, \text{ etc.}\]

The morphological analysis is performed on words in lowercase letters. The unifier section makes the main FST
interpret every uppercase letter as a lowercase one, which does not mean it gives a lowercase output, as it is demonstrated below:

Kasinta → -PN.Kasinta_Jacinta

Other changes are:

- b → f
- ca → ka, co → ko, cu → ku
- ce → se, ci → si
- gue → ge, gui → gi
- hua → wa, hue → we, hui → wi, huo → wo, huu → wu, when h is not preceded by c
- ha → a, he → e, hi → i, ho → o, hu → u when h is not preceded by c
- j → k, qu → k
- v → u when v is between consonants or semi-vowels, v → f
- otherwise
- q → g when q is not followed by u
- tx → tr, x → tr
- z → d

From all these changes, only the last three and the first part of the fourth backwards (v into u) do not have to do with Spanish but with some graphemic proposal for Mapudüngun.

The Mapuche alphabet we encode has twenty-five graphemes, five of which are digraphs. There are six vowels, three semi-vowels and seventeen consonants:

a, ch, d, e, f, g, i, k, l, ll, m, n, ng, ñ, o, p, r, s, sh, t, tr, u, ü, w, y

There are additional graphemes that we accept for analysis, which belong to other graphemic proposals for Mapudüngun or to Spanish:

b, c, h, j, l', n', q, t', tx, x, v, z

6.2 Lexicon

We have augmented our lexicon mostly from Augusta’s dictionaries [3]. But not only new words have been introduced, also many variants to the already collected words from Smeets, do not forget that Mapudüngun spelling is not fixed yet. But what most variants generates are the differences in pronunciation of some sounds, for instance, final -r is usually interchanged by final -Ñ; ù in any positions is commonly interchanged with u, i and sometimes e, or the other way around; tr with ch, d with s, etc., see table 9 for some examples:

| PoS | Smeets | Variant | Meaning |
|-----|--------|---------|---------|
| -AJ | arken | arkeñi | evaporated |
| -AJ | koll-ñollí | koll-ñollí | brown, reddish brown, beige |
| -AJ | - | liuke | clean, clear, pure (water) |
| -AV | kitù-ñití | kitu | alone, self, own |
| -AV | - | kashill | near |
| -AV | - | kiti | even, also |
| -IV | - | yawa | make noise |
| -IV | - | witra | get up |
| -IV | chekod-likoñ | likoñ | to squat, to sit down on one’s heels |
| -NN | achowall | achow-achowall | chicken |
| -NN | chafo | chafo | cough, catarrh, cold |
| -NN | - | dagahsa-dawilla | river shrimp |
| -TV | ingka | inka | to defend |
| -TV | - | kedín-keñiñ | to shear |
| -TV | ùtríñ | ùtríñ | to throw |

6.3 Williche verb forms

"In Williche[50], the southernmost dialect of Mapudungun, transitive verbs expressing the 1 → 2 relationship (with a total number of participants greater than two) is indicated by the combination of -e- and a 2nd person subject marker in slot 3 (Mosbach 1962: 80, and Augusta 1903: 84–86 cited by Salas 1979a: 307), e.g. pe-e-y-m-i ‘I saw you (sg)’, pe-e-y-m-u ‘I saw you (dl)’, pe-e-y-m-in ‘I saw you (pl)’ “[Smeets, I. 2008: 160] [21]. This very same relations are included in a single form of central Mapudungun when participants are more than two +1A.w23+IND.y4+1.Ø3+PL.1ñ2, which may be disambiguated by means of personal or possessive pronouns. But Smeets also gives as example pe-e-y-m-i ‘I saw you (sg)’ which are actually two participants, central Mapudungun and Williche also differ in this form. Note that Williche forms are ended by the null morpheme of dative subject +DS12A demanded by the +IDO marker -e-, see the following examples to compare:

( E 231 ) central Mapudungun 1s → 2s [Smeets, I. 2008: 157 (20)][21]
pe-e-y-u ‘I see you (sg)’
-TV.pe_ver+IDO.e6+IND.y4+1.Ø3+DL.1ñ2+DS12A.Ø1

50 -NW.willi_sur-NW.che_persona ‘Southern people’
6.4.1 Different indicative form

"The mark for indicative mood is -i-. It appears as a vowel if the root ends in a consonant, as a semi-vowel -y- if the root ends in a vowel other than i-, and it does not appear if the root ends in i-.\textsuperscript{51} [Zúñiga, F. 2006: 105] [24]. Instead of the "root ending", it should be said "the preceding sound", because this one may actually belong to the root, but also to a previous suffix. We have found that it realizes as -i- after semi-vowel too (E237); and is a null suffix when found either preceded or followed by i (E239). In Smeets work, the indicative is either -y- after vowel or semi-vowel, or -üy- after consonant; there is a -ı- variant for the latter.

To treat the variant presented by Zúñiga, we have added an @IZ tag to the already existent encoding of +IND suffix, D30. And added a new rule to deal with contexts of conversion of the intermediate representation @IZ (R51):

\[(D 30) \ ["+IND".y4]: [["@01" |"@0"].y] |"@IZ"]

\[(R 51) \ Zúñiga's indicative form\]
define RuIndZu ["@0IZ" -> 0 | | i , , i ] .o. ["@0IZ" -> i ];

D30 encodes indicative mood suffix, we have introduced the @IZ tag to treat Zúñiga’s variant. Preceding this new tag is the suffix as Smeets presents it (E238). @0 is for "CON-üy-", "VOW|SVW-ý-"; @UI is for "CON-ý-".

R51 defines, by context, the form @IZ should take, transforming it always into i, except when occurs either before (E239) or after (E240) i, in which case it is transformed into 0, acting as a null suffix. Otherwise, the tag remains until the end of the process, when it is cleared out (E236, E237 and E238).

( E 236 ) [Zúñiga, F. 2006: 105 (Cuadro III-3a / 2a)] [24]
kon-l-m-i 'you enter'
-IV.kon_entrar+IND.y4+2.m3+SG.12

( E 237 ) [Zúñiga, F. 2006: 283 (pewma)] [24]
chum-yaw-i-m-i 'what are you doing around?'
-QC.chum_cómo+VRB.Ø36+CIRC.iaw30
+IND.y4+2.m3+SG.12

( E 238 ) [Zúñiga, F. 2006: 105 (Cuadro III-3b / 2a)] [24]
tripa-y-m-i 'you leave'
-IV.tripa_salar+IND.y4+2.m3+SG.12

( E 239 ) [Zúñiga, F. 2006: 227 (84)] [24]
amu-a-h- 'let us go'
-IV.amu_ir+NRLD.a9+IND.Ø4+1.Ø3+PL.in2

( E 240 ) [Zúñiga, F. 2006: 105 (Cuadro III-3c / 2a)] [24]
pl-m-i 'you say'
-IV.pldecir+IND.Ø4+2.m3+SG.12

\textsuperscript{51} Translation is ours, the original is in Spanish
6.4.2 Glottal stop before +IDO

When the internal direct object -e- is preceded by a-, there is an optional epenthesis of a glottal stop in between, reflected as -ñ- in spelling. As there is already a rule that treat glottal stop epenthesis in compounds where the second root starts in vowel. We just added the appropriate tag @G to the +IDO suffix, D31 (see E116, E117, E118, R1 and R2, p. 22):

( D 31 ) Encoding of the +IDO suffix
["+IDO"{.e6}]: ["@G""@ID"]

(E 241) [Zúñiga, F. 2006: 274 (56)] [24]
kullī-a-gē-ñ-"u-"i will pay you both'
-TV.kullī.pagār+NRLD.a9+IDO.e6
+IND.y4+1.Ø3+DL.u2+DS12A.Ø1

(E 242) [Zúñiga, F. 2006: 279 (109)] [24]
elu-tua-ge-ñ-"give it back to me"
-TV.elu_dar+RE.tu16+NRLD.a9+IDO.e6
+IND1SG.n3+DS12A.Ø1

(E 243) [Zúñiga, F. 2006: 130 (note 10)] [24]
elu-la-ge-ñ-"you did not give me"
-TV.elu_dar+NEG.la10+IDO.e6+IND1SG.n3+DS12A.Ø1

6.4.3 Nominal compounds

We have only explained the Mapuche verb throughout this article. We have not introduced the nominal forms as such, as one of the possible verbal stems. Nominal forms are much more simpler than the verbal ones. We do not want to make this article too extensive adding encoding details that are well explained with the verb form, it suffices to say that there is nominal compounding in Mapudüngun.

Zúñiga defines püle ‘by, towards’ as a post-position [Zúñiga, F. 2006: 195] [24]. Smeets defines it as a post-position too, ‘side’; but also as a noun [Smeets, I. 2008: 69 (10.4) [24], which is how we have incorporated püle ‘side, direction’ into our lexicon.

Even classifying püle as a post-position, in many occasions this word forms compounds in Zúñiga’s texts, which are not recognized by the Smeets derived rules of the analyser.

(R 52 ) Nominal compound (püle)
define formCXNN
["%<DEMPR|IVROOT|INTPR]%<%<NROOT%>#-"CNN"];
define preCXNN
[≡(formCXNN, %<, %<)];
define RuleCXNN1
["-NN" ⇒ ["@TYA","@FA","@T"] ?* _ ];
define CXNN
[CLEANu .o. RuleCXNN1 .o.
formCXNN - RuleCXNN1 .o. CLEANVFD];

The mechanism for compounding was already explained at Compounds encoding, p. 30, so here we add that in R$52$ the form of these nominal compounds is defined as having a first member that may be a demonstrative pronoun, an intransitive verb or an interrogation pronoun. RuleCXNN1 specifies which forms, out of these categories, are actually accepted to form the compound with the noun as a second member. The specific forms has been tagged to filter them out, @TYA, @FA and @T respectively; actually, only one member of each category is tagged. In the examples we show compounds that are not recognized following Smeets, compounds that take püle as a second member (In Smeets, püle may be the second noun in a nominal compound):

(E 244) [Zúñiga, F. 2006: 275 (65)] [24]
fa-püle ‘around here’
-TV.fa_ser-esto-NN.püle_lugar

(E 245) [Zúñiga, F. 2006: 191 (Cuadro III-17)] [24]
kañ$^{53}$-püle ‘somewhere else’
-AJ.ka_otro-NN.püle_lugar$^{54}$

(E 246) [Zúñiga, F. 2006: 273 (49)] [24]
tie-püle ‘over there’
-IP.tüye_aquel-de-allá-NN.püle_lugar

(E 247) [Zúñiga, F. 2006: 182 (95.d)] [24]
tuchi-püle ‘wherever’
-IP.chuchi_qué_cuál-NN.püle_lugar

(E 248) [Zúñiga, F. 2006: 275 (67)] [24]
üye-püle ‘over there’
-AV.üye_allí_allá-NN.püle_lugar

Augusta’s nominal compound. Another nominal compound integrated into the analyser was found in Augusta, F. [3]. It has a numeral as first member and a numeral as second member:

(E 249) [Augusta, F. (epuange)] [24]
epu-ange ‘two faces’
-NU.epu_dos-NN.ange_cara

$^{53}$ The verb root tagged @FA is also identified as a deictic verb with the same tag (see Deictic roots, p. 41).

$^{54}$ This is an epenthetic ñ.

$^{54}$ There is a colloquial expression in Chile, ‘salta pal lao’ which means something like ‘I don’t believe you’, ‘are you kidding’, ‘you better not...’, depending on the situation. Pevillafu says that this expression is equivalent to kaipüle in Mapudungun.

$^{55}$ Two faces, (be of) two faces, a designation of a two faces father regulator, either because they represent two sexes, or because with this expression they allude to the benign and serene heaven and to the unfavourable heaven; or to the severity and benignity that the supreme being can demonstrate to men. In addition, the idea is applied to both God and Mayorwekufü ‘major daemon’. V. Augusta (1910, p. 227) [Augusta, F. epuange] [3]
6.4.4 Instrumental and ad-position mew

Smeets classifies -mew as an instrumental suffix that follows nouns, deverbal nouns and pronouns (see Smeets, I. 2008: 61 - 67, "10.1 The instrumental -mew ~ -mu" [21]).

Zúñiga defines mew as an ad-position that is realized separated from the noun or deverbal noun that follows, but together with pronouns or adverbs (see Zúñiga, F. 2006: 194 - 197, "4.1 Las adposiciones y los sustantivos relacionales" [24]).

In order to recognize mew as an independent form, but still as the instrumental suffix, it was declared as such in the non-verbal section of the script:

-NU.kiñe_uno-NN.püle_lugar

Lonkon calls these elements, "modal prefixes". She identifies four of them: kalli, kim-, küpa- and pepi-, giving them the same values as Smeets. But Lonkon considers them prefixes, therefore, attached to the verb they are moulding, except for the permissive/enabling kalli [Lonkon, E. 2011: 249] [12].

Zúñiga identifies two of these modal elements: kim- and pepi-. He says that they form part of "complex verb stems", which means that both verb roots, the modal and the moulded one, form a compound. However, he says, they can also be expressed as separated elements, then he calls them "pre-verbal particles" [Zúñiga, F. 2006: 136] [24].

Zúñiga also displays a list of verb roots that form complex verb stems, and, in fact, he devotes a paragraph to explain that they may be formed by radical concatenation, and by nominal incorporation. In the list is found the verb root wiño- "re-", "return, come back". In the list are also kalli- and küpa-, but not shinge- [Zúñiga, F. 2006: 179] [24].

Zúñiga explains that these forms are frequently found as pre-verbal particles, i.e., separated from the main verb, which is reflected in spelling. He treats them as radical concatenation, though.

Salas calls them "modals", and he identifies kim- 'know', küpa- 'wish' and pepi- 'be able'. He says that they function as prefixes of single and complex stems [Salas, A. 2006: 192] [19].

Augusta defines wiño- as a suffix equivalent to the prefix morpheme "re-", which mainly expresses the idea of redoing the action indicated by the following verb. We have realized that wiño also expresses the idea of retrospective and/or backwards action [Augusta, F. wiño] [3].

Among the examples given by Augusta, there are some that show wiño separated from the main verb, and others, forming a compound with the moulded verb.

Many Mapudüngun native speakers perceive wiño separated from the verb it is moulding, and they reflect it as Zúñiga shows in its examples.

An affix does not realize isolated in Mapudüngun, it must be attached to a verb or another root (adjective, noun, etc.). It can neither work as a root accepting suffixes to be attached. wiño, on the other hand, also works as a verb root that may be inflected by attaching suffixes to it. Therefore, wiño meets the auxiliary definition given by Smeets.

wiño ~ wiño either forms a compound, originates an inflected verb, or is a separated element. If some native speakers perceive it as a separate element, it may indicate that it is a modal element, thus, following Smeets, this verb would fulfil the "auxiliary" function as she defines it. Or the "pre-verbal particle" function, as Zúñiga calls it.

Anyway, not all Mapudüngun native speakers spell wiño ~ wiño separated from the main verb. Many of them use it in the verb as a compound, i.e., as a "modal prefix", as Lonkon calls it.
We have added wüño as auxiliary into the lexicon list through the following entry:

(D 32) "-XV"{.wüño_re_-volver-a}]: [{wüño}|{wiño}]

(E 255) [Zúñiga, F. 2006: 148 (63)]

wüño-witra-me-tu-a-fiel 'to go there to recover them'
-IV.wüño_volver-IV.wütra_levantar-CR.IV
+TH.me20+RE.tu16+NRLD.a9+TVN.fiel4

(E 256) [Augusta, F. contestar][3]

wüño fey-pin 'to answer'
-XV.wüño_re-_volver-a -TV.feypi_decir+PVN.n4

(E 257) [Lonkon, E. 2017]
wiño wütra-m-püra-m-nge-tu-a-fu-y 'it revitalized'
-IV.wütra_levantar+CA.m34-IV.püra_subir-CR.IV
+CA.m34+PASS.nge23+RE.tu16+NRLD.a9+IPD.fu8+IND.y4+3.Ø3

6.6 Proposing -ña as adverbializer

Smeets lists -ña as an unproductive suffix [Smeets, I. 2008: 116] [21], giving a series of examples with this suffix:

1. fücha-ña 'very long' (fücha 'long')
2. we-ña 'very new' (we 'new')
3. wesha-ña 'very bad' (wesha 'bad')
4. rume-ña 'extremely' (rumé 'very')
5. welu-ña 'wrong, reversely' (welú 'but, wrong, reversely')
6. alí-ña 'for a long time' (alí 'much'), cf. alí-ña-mew 'much later, a long time after that'
7. fentre-ña ~ fentre-yma 'for a long time' (fentre 'much')
8. epu-ña 'with the two of us' (epu 'two')
9. ka-ruka-ña 'neighbour' (ka-ruka 'neighbour')

Except for the examples 5, 8 and 9, it is quite evident the part of the meaning contributed by -ña, 'very'. In other texts we have found:

1. weda-ña 'evil, too bad' (weda 'bad') [Zúñiga, F. 2006: 270 (15)] [24]
2. pichi-ña 'just now, recently' (pichi 'little') [Zúñiga, F. 2006: 271 (30)] [24]
3. fücha-ña 'very big' (fücha 'big, old') [Augusta, F. (füchana)] [3]
4. lekú-ña 'very close' (lekú 'close, near') [Augusta, F. (lekú-ña)] [3]

We have not gone through an exhaustive research of this suffix, that is why it is just a proposal. We think it is quite clear that -ña adds the 'very' part of the meaning, but we have observed that it is applied only to adjectives and certain adverbs with this sense, and only as part of the nominal compounding. There is only one case, mentioned by Smeets, where it seems to be part of a verbal stem, but there are no examples of it; so in the cases of E258 it could be either the indirect object marker, slot 26, or the experience suffix, slot 35, which share the form -ña:

(E 258) [Smeets, I. 2008: 574 (wesha)]

wesha-ña-w-e (Vi) to break down, to fall apart, to become a bad person;
wesha-ña-w-küle-e (Vi) to be broken/in pieces, to feel awful;

(D 33) "+ADV"{.añma}]: ["@Ø"{ma}];

(R 53) Prevent adverbializer to appear as verbal suffix define RuPr51 ~$[$["+ADV"{.añma}]]];

(R 54) Apply adverbializer only to adjectives and adverbs define NvsOpRu01 ["[+ADV"{.añma}] => ["-AJ"|"-AV"] ?* _ ];

D33 encodes -ña as adverbializer. R53 prevents it to appear along the verb sequence of suffixes. R54 restricts it to co-occur only with adjectives and adverbs. The other forms, nouns, numerals, that occur with this suffix are collected as lexicalized forms in the lexicon:

(D 34) ["-AV".epuña_con-nosotros-dos]:["@G"{epuña}]

7 Analyser dimensions

This section exposes data referent to the amount of each type of element interacting in the system: lexicon, suffixes, rules, states, etc.

There is a flow chart (deployed in figures 9, 10 and 11) showing the interconnection of all the analyser elements along the process in annex 11.6, p. 66.

Roots (verbalizable lexicon): 2,096

Adjectives: 128
Adverbs: 24
Intransitive verbs: 257
Proper nouns: 68
Nouns: 1,325
Numerals: 14
Onomatopoeia: 12
Questions: 4
Transitive verbs: 264

Non verbalizable lexicon: 266

Adverbs: 88
Anaphoric pronouns: 5
8 Evaluating the analyser

8.1 Corpora in use

8.1.1 Gold standard

We have collected a corpus made of sentences coming from "A grammar of Mapuche" [Smeets, I. 2008] [21], which is our "Gold standard" corpus. Words in the corpus were analysed and disambiguated by Smeets.

The Gold Standard corpus includes all the sentences from chapters 10 to 18, and 21 [Smeets, I. 2008: 61-116, 121-128] [21]. These chapters deal with nouns, adjectives, adverbs, numerals, demonstratives and anaphoric pronouns, personal pronouns, possessive pronouns, interrogative pronouns, suffixation and verbalization. The corpus also contains all the seventeen texts of "Part VIII - Texts" [Smeets, I. 2008: 369-487] [21], which is obviously the most abundant source of Mapuche writings in the Gold standard. Texts titles are:

- Text 1. Demons
- Text 2. Work
- Text 3. Youth
- Text 4. Missionary
- Text 5. The war
- Text 6. An old man
- Text 7. Olden times
- Text 8. Conversation about demons
- Text 9. Conversation about youth
- Text 10. Conversation about work on big farms
- Text 11. Conversation about land disappropriation
- Text 12. Our reservation
- Text 13. My father
- Text 14. Brick
- Text 15. Song 1
- Text 16. Song 2
- Text 17. Song 3

8.1.2 Control corpus

Out of the Gold standard we have extracted a control corpus of 240 sentences containing a total of 1,671 words, which correspond to 650 forms.

ñi trewa, ñi ñarki ka ñi kawell

In example E259 there are seven words, but 5 forms which are ka, kawell, ñarki, ñi and trewa. The 3 ñi words count as 1 form.

The control corpus is used to check results obtained from the analyser. A correct analysis for every word of this corpus must appear in the output.
8.1.3 Comparison corpus

This corpus is collected from Zúñiga’s texts [Zúñiga, F. 2006: 270 (15)] [24]. It is made of 170 sentences that contain 1,256 words which correspond to 511 forms (see previous section 8.1.2 and example E259). The texts are extracted from “Mapudüngun. El habla Mapuche”, chapter V. Textos en mapudüngun [Zúñiga, F. 2006: 266 - 288] [24]. Text titles are:

1. Feychi ngürü aflngünengelu ’That crafty fox’
2. Mawün ‘Rain’
3. Pewma ‘A dream’
4. Ngillañ mawün ‘Asking for rain’
5. We tripantu ’New year’
6. Pausa_Historia ’Pause_History’
7. Abuela_Voz ’Granny_Voice’

8.2 Establishing the ambiguity parameter

Ambiguity. In any language some words present ambiguity. These word forms, known as homographs, correspond to different meanings, which are disambiguated by context; for example, the English word drop means ‘a little amount of liquid’ in “a drop of coffee stained my letter”. And it means ‘to fall’ or ‘let fall’ in “do not drop papers on the floor”. The only way to know which meaning drop is referring to, i.e., the only way to “disambiguate” it, is by putting it into context.

In Mapudüngun there are quite few ambiguous words (homophones which once written become homographs) that appear very often in texts, like ka, fey and most of the words ending in n which is the shared form between the “1st person, singular” suffix and the “plain verbal noun” suffix, examples of them are below (see 5.3.5 Over-generation, p. 38):

\[
\begin{align*}
ka & \rightarrow - AJ.ka_otro \\
ka & \rightarrow - CJ.ka_y \\
ka & \rightarrow - PT.ga_cierto_mencionado \\
fey & \rightarrow - AV.fey_entonces \\
fey & \rightarrow - DP.fey_que_aquel_ese \\
fey & \rightarrow - IV.fey_ser-e+IND.y4+3.Ø3 \\
fey & \rightarrow - PP.fey_el-ella_elllos \\
küdawün & \rightarrow - IV.küdaw_trabajar+IND1SG.n3 \\
küdawün & \rightarrow - IV.küdaw_trabajar+PVN.n4
\end{align*}
\]

Mapudüngun ambiguity calculus. If we add up all the analysis of three previous words: 9, and divide it by the number of words we are taking into account: 3, we obtain the "average Mapudüngun ambiguity" (ama) of these words: 3. We use this result as a reference parameter when comparing analysis results.

As we have explained before, the development of the analyser arises from Smeets’ description of Mapudüngun morphology [Smeets, I. 2008] [21]. We have first developed an analyser that strictly fits with Smeets work, we have generated an FST from it, we call it "SmeetsAnalyser". Then, we have expanded the analyser in lexicon and rules that fit another variants of Mapudüngun, provisionally, we call it Düngupeyem⁶³.

To calculate the average Mapudüngun ambiguity, and establish it as a reference parameter, we have analysed the control corpus with the "SmeetsAnalyser".

Control corpus analysed with "SmeetsAnalyser"
- word-forms = 650
- produced analyses = 2,232
- unknown words = 2

\[
\frac{2,232}{650} - 2 = 3.59 \text{ ama}
\]

Following this calculus we obtain 3.59 ama, this means that every word-form has an average of 3.59 possible analyses, this is our reference parameter.

Increased ambiguity calculus. The incorporations we have made in the system, those mentioned in section 6, [p. 41], increase the ambiguity in analysis results because they imply more ways of analysis for every form. Increasing the lexicon also produces this effect.

Now we calculate the "average increased ambiguity" (aia) to be able to compare the results and verify if the system is still reliable.

Control corpus analysed with Düngupeyem
- word-forms = 650
- produced analyses = 2,477
- unknown words = 2

\[
\frac{2,477}{650} - 2 = 3.82 \text{ aia}
\]

As expected, ambiguity rises, but not too much, less than 0.24 points per word-form, which indicates that the reliability of the analyser is quite good.

8.2.1 Comparison corpus results

To check if the additions (see 6, p. 41) we have made to the FST really worth it, we analyse the comparison corpus.

⁶³ düngu ‘word, speech, language’; pe ‘proximity suffix, it indicates physic and or temporal proximity with the action expressed by the verb; ye ‘constant feature suffix’; m ‘it indicates instrument or location’. All together it is something like ‘instrument always used to do things with language’. We intended to say ‘language tool’. That is one of the reasons to be a provisional name. ‘Tool’ in Mapudüngun is küdaw-ka-we, so another possibility is to call our analyser Düngu-ka-we. Another reason to be a provisional name is that we have found no native Mapuche yet, who validates the name.
Comparison corpus analysed with "SmeetsAnalyser"
- word-forms = 511
- produced analyses = 1,368
- unknown words = 120

$$\frac{1,368}{(511 - 120)} = 3.49 \text{ aa}$$

Comparison corpus analysed with Düngupeyem
- word-forms = 511
- produced analyses = 1,828
- unknown words = 10

$$\frac{1,828}{(511 - 10)} = 3.64 \text{ aa}$$

At first glance, these results confirm a good performance of Düngupeyem, the difference in the average ambiguity (aa) even diminishes from 0.24 to 0.15. But there is a factor that we did not considerate in the calculus with the control corpus, it would have made no difference in results. There were too few unrecognized words (only 2), and the same amount for both analysers.

Analysing the comparison corpus, we obtain 120 unrecognized words from "SmeetsAnalyser", and only 10 from Düngupeyem. To measure the impact of this factor in the ambiguity index, we have used a "root guesser" to count the possible analyses the unknown words generate. Then, we add the resulting possible analyses to the known analysis and recalculate the ambiguity index.

120 unrecognized words from "SmeetsAnalyser"
- words producing no possible analyses = 49
- words producing possible analyses = 71
- total number of possible analyses = 636

$$\frac{1,368 + 636}{(511 - 49)} = 4.33 \text{ aa}$$

10 unrecognized words from Düngupeyem
- words producing no possible analyses = 3
- words producing possible analyses = 7
- total number of possible analyses = 56

$$\frac{1,828 + 56}{(511 - 3)} = 3.70 \text{ aa}$$

These calculi confirm that the analyser is reliable, even more after adding the considerations for other dialects of Mapudungun and more lexicon. The average analyses (aa) have raised, but too little as to consider it a problem, only 0.06 between the analyses that does not take into account the possible analyses for unknown roots (3.64), and the one that does take them into account (3.70).

8.3 Comparing against other analysers

For this purpose we have used the Gold standard corpus (see 8.1.1, p. 47) to train two different analysers, and we have analysed the comparison corpus (see 8.1.3, p. 48) with these tools and our analyser to compare results.

8.3.1 Trainable systems to compare against

A trainable system receives a disambiguated corpus to learn from, i.e., it stores in its memory (or database) the correct forms that have been introduced in the training process. Then, using different algorithms, it compares an input text to the stored data and tags input accordingly. The trainable systems we have compared our analyser against are RFTagger [Schmid & Laws 2008] and Morfette [Chrupala et al. 2008].

RFTagger. This is a tool for the annotation of text with fine-grained part of speech tags. It is a Hidden-Markov-Model tagging method that is particularly suitable for PoS tag sets with numerous detailed tags [Schmid & Laws 2008: 1] [20]. An HMM part of speech tagger calculates the most likely sequence of PoS tags for a given word sequence. The difference with our analyser is that RFTagger identifies a complete form and tags it accordingly, while our morphological analyser breaks down the input form and gives a tag for each part that constructs the full form.

Morfette. This is a tool for supervised learning of inflectional morphology. Given a corpus of sentences annotated with lemmas and morphological labels, and optionally a lexicon, Morfette learns how to morphologically analyse new sentences. This is a data-driven modular probabilistic system that learns to perform morphological tagging from morphologically annotated corpora. The system is composed of two learning modules that use a maximum entropy classifier to predict morphological tags. The third module dynamically combines the predictions of the Maximum-Entropy models and generates a probability distribution over the sequences of tag-lemma pairs [Chrupala et al. 2008: 1] [4].

Training RFTagger and Morfette. The text of the Gold standard corpus has to be tokenized, one token per line. For each token, a tag is entered after a tab space. For verbs,
the tag shows 38 slots for attributes. 36 verbal slots and two additional ones for nominal, temporal, instrumental or nominalizing suffixes; each unavailable attribute\(^67\), which fits into a slot, is indicated by a 0 (zero), as shown in the following example:

(E 260) *amukefuy IV.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.+CF@ke14.0.0.0.0.0.+IPD@fu8.0.0.0.0.+IND@yy4.+3@Ø3.0.0.0.

For non-verbal forms, the tag holds four positions for attributes, the first one occupied by the category of the form and the last three by possible nominal suffixes, as shown in the following example:

(E 261) *peñimu ‘at my brother’s’
peñimu NN.0.0.+INST@mew

Morfette requires a bit more information. There is an additional column between the form and the tag to enter the lemma, where the root was added instead. Unlike RFTagger, attributes of the tag are not preceded by dots, compare E260 with the following example:

(E 262) *amukefuy IV ‘he used to go’
amukefuy amu IV.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.+IPD@fu8.0.0.0.0.+IND@yy4.+3@Ø3.0.0.0.0.

Both machines were trained in their standard way, that is, with their default settings, except that in the case of Morfette we chose the vector representations of words (see E260, E261, E262). The vector representation for tags uses a 0 (zero) for each possible attribute, where the realized ones replace them in the appropriate position. It is the only configuration that RFTagger accepts.

8.3.2 Results from the three analysers

Results from the two comparison tools are not ambiguous, they output only one tag per form, which gives us a binary value, either right or wrong tag assignment. They always set a tag for every form, so there are no unknown forms neither. We plan to add a disambiguator to the system so we can analyse words in context and, instead of outputting all possible analyses per word, only output the most suitable one.

The training corpus, the "Gold standard", is made up of 1,220 sentences, which in turn are made up of 9,209 tokens of which 1,998 are punctuation marks; when subtracting the latter, the total of unambiguously annotated word-forms is 7,211. This material is actually insufficient\(^68\) for an accurate training that can obtain acceptable results. Note that this is one of the main reasons for developing a morphological analyser based on rules, as there are not enough Mapuche annotated corpora available to carry out the data-driven type of approach with computational tools.

### Table 10 Analysing results for the comparison corpus (see 8.1.3)

| Data type | RFTagger | Morfette | Düngupeyem |
|-----------|----------|----------|------------|
| Tokens    | 1,471 - 100% | 1,471 - 100% | 1,471 - 100% |
| Right     | 723 - 49.1% | 786 - 53.4% | 1,460 - 99.2% |
| Wrong     | 749 - 50.8% | 685 - 46.5% | 0 |
| Unknown   | -       | -       | 11 - 0.75% |
| Analyses  | 1,471 | 1,471 | 4,533 |

**RFTagger results.** This tool can not be blamed for this low performance (nor Morfette either). The insufficient amount of training material is the cause of the poor results. RFTagger was not developed for agglutinative languages explicitly, or the complexities of Mapudungun. Also keep in mind that agglutinative languages have potentially an infinite number of words, as English, for example, has potentially an infinite number of phrases. Therefore, it is more difficult for the system to learn from such a variety of forms how to tag accurately.

**Morfette results.** This tool can handle the analysis process a little better than RFTagger. More words are correctly analysed, and less words are incorrectly analysed.

Throughout the results validation we have noticed that Morfette breaks down the forms to apply its inferred assumptions on analysed suffixes. But both machines fail though, in assuming almost all the words starting by a capital letter as a proper noun.

**Düngupeyem results.** To make the results comparable, we have generated a mathematical formula that approach the two types of data and give us a way to explain the numbers. The average ambiguity (aa) of the comparison corpus (see 8.2.1, p. 49) is 3.64 aa. On subtracting the number of analyses corresponding to that percentage, of Düngupeyem’s correct analyses (1,460) we relate this amount to those of

\(^67\) Grammatical categories are collected with attributes, which are the aspects that detail the category, for example, the word "plastic" has the category Noun; a fine-grained tag for it would be NSM, where N stands for "noun", S for "singular" and M for "masculine", which are the attributes.

\(^68\) In data driven tools the more training data receive the tool the better it performs its tasks. Text applied tools use to be trained with thousands of sentences. For the tools we have used (RFTagger and Morfette) we believe that about 15 thousand sentences would give good quality results; in Smeets’ work there are about 1,800 morphologically tagged sentences.
Morfette and RFTagger right analyses. Then, we recalculate the percentage obtaining a closer degree of correspondence among the three machines results:

\[
\begin{align*}
3.64 & \times 100 = 53.29 \text{ (ambiguous words)} \\
1.460 & = 53.29 = 1.406.71 \\
1.471 & = 100\% \text{ (analysed words)} \\
1.406.71 & = 95.62\% \text{ (correctly analysed words)}
\end{align*}
\]

After these results, it is again clear the system’s reliability, and the analysers’ high degree of accuracy. We see that regardless counting not recognized words as if they were wrong analyses, failed analyses reach only 0.75%, while success overcomes 95%. The other two tools are around 50% in right and wrong analyses.

8.4 Comparing against a Quechua analyser

Our work follows the path that Annette Ríos has drawn for Quechua. This step is not an exception. Although there are notable differences, Düngeyem’s evaluation is largely based on the same process for Quechua FST tools carried out by Ríos [Ríos, A. 2015: 36-39] [17].

The Quechua system is made up of 5 cascading transducers [Ríos, A. 2015: 22] [17], and some of these FSTs use a disambiguator for specific parts of the forms. For example, to unravel whether a root is nominal or verbal; or on another transducer to reveal if a suffix is of one or another type, etc. So the final Quechua form has been disambiguated throughout the analysis process.

Training material also differs. Quechua system used about 3,000 disambiguated sentences. Mapuche training material (see 8.2, p. 48 and 8.3.1, p. 49) consists of less than the half of those sentences, 1,220 to be exact. In both cases, however, this is too little material to train data-driven systems well enough.

Comparison corpus are similar. Ours contains 170 sentences with 2,207 tokens (1,475 words + 732 punctuation marks) (see 8.1.3, p. 48), Quechua has 322 sentences containing 2,142 tokens [Ríos, A. 2015: 36] [17].

Although the procedure was also performed differently, the final counting (the addition of results) may serve as comparison data. In the Quechua case, both texts of the comparison corpus were tagged separately, to later compare them with each other. In the Mapuche case, by contrast, texts were tagged at once.

8.4.1 Quechua results

Quechua results are transferred and presented in the style used in this article, different from their original presentation, for comparison reasons70.

Table 11 Analysis results. RFTagger / Morfette / Quechua system

| Data type | RFTagger | Morfette | Quechua |
|-----------|----------|----------|---------|
| Tokens    | 2,142 - 100% | 2,142 - 100% | 2,091 - 97.6% |
| Right     | 1,459 - 68.1% | 1,505 - 70.2% | 2,041 - 95.2% |
| Wrong     | 683 - 31.8% | 637 - 29.7% | 50 - 2.33% |

Table 12 Results RFTagger / Morfette - Quechua / Mapudüngun

| Data type | RFTagger | Morfette |
|-----------|----------|----------|
| Quechua   | 68.11% | 31.88% | 70.26% | 29.73% |
| Mapudüngun| 49.12% | 50.88% | 53.43% | 46.57% |
| Difference| 18.99% | 19% | 16.83% | 16.84% |
| Average difference | 18.99% | 16.83% |

RFTagger results. Table 11 shows that increasing the amount of training data (1,220 to 3,000 sentences), results improve consequently for both, RFTagger and Morfette.

RFTagger worked well with 68% of the Quechua forms, which is 19% more than with Mapuche ones. The percentage of wrongly analysed forms is also better, being 19% lower. Thus, performance of RFTagger with Quechua was 19% better than with Mapudüngun (see table 12). However, be aware of the differences between the two evaluation processes (see 8.4, p. 51).

Morfette results. This tool also performs better for Quechua (compare results in table 12). As with Mapuche texts, Morfette performs better than RFTagger in the Quechua texts tagging.

Morfette results with Quechua texts were 16.8% better than with Mapuche texts (see table 12). It produced 16.8% more of correct analysis, and the same percentage less of incorrect analyses. The difference is smaller, as it is in the Mapuche texts analysis.

Recall that the percentage of correctly analysed words of our analyser was 95.62% (see Düngeyem results, p. 50).

70 Data are taken from table 2.13: Evaluation: Disambiguated Texts [Ríos, A. 2015: 38] [17].
table 11 shows, in the case of the Quechua system, 95.2% of correctness. Despite the differences between both systems (see 8.4, p. 51) results are similar. So, if Ríos considers her machine performs successfully based on her analyses results, we may go along with the same consideration for our system.

8.5 Not analysed!

In analyses of the control corpus (8.2, p. 48), made up of Smeets’ sentences [Smeets, I. 2008] [21], there were two unknown words. We analyse these cases below.

8.5.1 Smeets’ not analysed words.

eluwün-antü ‘funeral day’ [Smeets, I. 2008: 402 (74)] [21]. Presented this way, this word is a nominal compound. But we have not collected eluwün ‘funeral’ as a noun because it is actually a nominalized verb:

(E 263 ) [Smeets, I. 2008: 404 (7)][21]
el-uw-ün ‘funeral’, lit: ‘the leaving behind/going’
-TV.el_dejar-atrás_partir+REF.w31+PVN.n4

antü ‘day’
-NN.antü_sol_día_tiempo

This is another type of compound that we have not encoded in our system. We think it is not worth to include it because, up to this moment, it has only appear two or three times. Adding variants of whatever form increases the ambiguity, which negatively affects the performance of the analyser.

mapuche-domo ‘Mapuche woman’ [Smeets, I. 2008: 399 (183)] [21]. This is another nominal compound that Smeets presents as made up two nouns. But these are actually three noun roots:

(E 265 ) [Smeets, I. 2008: 117][21]
mapu-che ‘land person’
-NN.mapu_tierra-NN.che_persona

(E 266 ) [Smeets, I. 2008: 76 (20)][21]
domo ‘woman’
-NN.domo_mujer

A compound of three roots is not contemplated in the system either, for the same reasons of the previous compound. But, a solution to successfully analyse these type of compounds may be to cascade another FST capable of analysing three roots compounds after the main FST is unable to do it, this way we do not increase ambiguity in the main analyser.

8.5.2 Zúñiga’s not analysed words.

In the analyses of the comparison corpus (8.2.1, p. 49), made up of Zúñiga’s sentences [Zúñiga, F. 2006: 266 - 288] [24], there were ten unknown words (puru ‘to dance’ is repeated, so, actually nine). In this section we try to find out why these words could not successfully go through the analyser.

are-tu ‘borrowed’ [Zúñiga, F. 2006: 280 (122)] [24]. In Zúñiga and Augusta, F. [3], this root is an adjective. From the perspective of Spanish or English it should be the participle form of the verb. Smeets and Augusta collect it as verb, are-‘to lend’, and when followed by the suffix -tu-, are-tu- ‘to borrow’.

"An -n form occurs as an adjective denoting an attribute or quality of the modified noun" [Smeets, I. 2008: 190][21], among other functions of the plain verbal noun. This lead us to think that aretu may be analysed as are-tu-Ø, where the plain verbal noun is elided, or realized as a null suffix.

A solution, then, to correctly analyse this form, would be to add the null form of the plain verbal noun to the list of suffixes. The problem is that this action would enormously increase the ambiguity, because every nonverbal root allowed to take the verbalizer Ø, would be analysed as root and root+VRB.Ø3+PVN.Ø4.

ina-lef-nepe-n ‘I startled wake up’ [Zúñiga, F. 2006: 283 (Pausa_Historia)] [24]. From the meaning Zúñiga gives to this word, we think that it is a short form for:

(E 267 ) ina-lef-ün_nepe-n lit: ‘I woke up on the run’
-AV.ina_a-través-IV.lef_correr-CR.IV
+PASS.nge23+PX.pe13+PVN.n4
-IV.nepe_despertar+IND1SG.n3

ngen-ko ‘god of the waters’ [Zúñiga, F. 2006: 285 (we tri-pantü)] [24]. It is the same case as eluwün-antü ‘funeral day’ (see eluwün-antü ‘funeral day’, p. 52). In Smeets is found as in the following example:

(E 268 ) [Smeets, I. 2008: 138 (41)][21]
oge-n ko lit: ‘owner/master of the water’
-TV.oge_tener+PVN.n4
-NN.ko_agua

puru ‘to dance’ [Zúñiga, F. 2006: 287 (Pausa_Historia)] [24]. We think it is the same case as are-tu, even though, this is a single root, and is not collected as ‘danced’, not by Zúñiga nor by Augusta (see are-tu ‘borrowed’, p. 52). So, it is probably:

(E 269 ) puru-Ø ‘to dance’
-IV.puru_bailar+PVN.Ø4
The first reason for this analysis not being produced by our machine is the three roots stem, as we have explained in mapuche-domo 'Mapuche woman' (p. 52). Then, we are guessing the stem to be composed as uf-kün-tukur-.

Smeets collects ūfi- ‘to become tight, to tighten’ [Smeets, I. 2008: 567] [21]. Augusta, F. [3], ū-fûn ‘tighten the straws with (when roofing)’, uf-tükû-n ‘tighten with tools’. But also ūfi-n ‘tighten with something to tie’. So ūf- and ūfi- are, very likely, variants of each other.

In Smeets’ dictionary are kim-tuku- ‘to have known/un- derstood for some time’ and kim-tu- ‘to remember’ [Smeets, I. 2008: 559] [21]. Febrés’ dictionary presents kûn-tûku-l-en ‘make someone else to remember’; kûn-tûku-n ‘to remember’; kûn-tûku-pe-m ‘the memory’ Febrés, A. [3]. It is quite probable that kim- and kûn- are also variants of each other. tuku- and tûku- are undoubtedly variants of each other, and it can not be tû-kû- because there is no attested -ku-suffix.

wima-kûtû-ye-nge-y ‘she was whipped’ [Zúñiga, F. 2006: 281 (123)] [24]. A more literal translation would be ‘she was taken to be whipped all along’ if the parts forming this verb are the ones we suggest: wima- ‘dipstick, thin stick’. -kûtû- it seems to correspond to a suffix that adds the sense of ‘all along’.

Augusta defines it as ‘suffix and post-position: (Variant used in Pangi). From (temporarily). kûtu kûtû ‘from a long time’. Conjunction: and even, until’ Augusta, F. [3].

Valdivia recognizes in it a locative sense, ‘from and to; fa kûtû tiye kûtû ‘from here to there’ Valdivia, L. [3].

Smeets only recognizes it as an adverb, kûtu ~ kûtû ‘even, also’. There is a root, though, that Smeets collects as weñangkû- ‘to get sad’ [Smeets, I. 2008: 573] [21]. Augusta collects it as weñang- ‘to have pain, annoyance, desire’, but also weñang-kû-n ‘get sad’. So, maybe -kûtû- is formed by two suffixes -kû-tu-, being -kû- this suffix of the ‘all along’ sense, and -tu- the transitivizer suffix. This is not totally rare in Mapudüngun, see the case of llemay in the note to E67 (p. 11).

Concluding, the verb is made up of a complex stem, which is formed by a ‘noun root + 1 or 2 suffixes + verb root’, possibly:

\[ \text{(E 274) wima-kû-tu-ye-nge-y} \]
\[ \text{-NN.wima_vara+THR.ku35+TR.tu33} \]
\[ \text{-TV.ye_traer_llevar+PASS.nge23+IND.y4+3.Ø3} \]

\[ \text{(E 275) wima-kû-tu-ye-nge-y} \]
\[ \text{-NN.wima_vara+THR.kutu35} \]
\[ \text{-TV.ye_traer_llevar+PASS.nge23+IND.y4+3.Ø3} \]

We have tagged THR the new suffix, from ‘through’, that comes from the idea of ‘all along’. An we have assigned it to
slot 35, which seems to be a suitable position for this suffix, of course all of this is tentative.

wirea-n-püra-may-a-n 'That raise it up' [Zúñiga, F. 2006: 284 (Ngillañmawün)] [34]. The division we propose for this verb would give a meaning like 'I rose it up in assent' or 'I rose up my assenting X'. This way, the stem is complex and formed by three roots, a not allowed analysis in our system. The analysis would be:

(E 276) wirea-n-püra-may-a-n
-IV.witra_levantar+PVN.n4-IV.püra_subir
-IV.may_asentir+CR.TV+NRLD.a9+IND1SG.n3

9 Public user interface

In this section we briefly present the exploitation interface we have developed for open access to our analyser.

The URL to access it is:
http://www.chandia.net/dungupeyem

Fig. 7 Analyser public web interface.

Numbers in figure 7 mark:
1. Text box to paste or type Mapuche words to be analysed.
2. An e-mail field to add the user address in order to receive comments about unknown words. And the "analyse" button to submit the text.
3. A field that allows to upload a .doc, .docx, .odt or .txt file to be analysed.
4. A text field to input analyse glosses in order to generate Mapuche words.
5. A link to a contact form in case the user needs some feedback from us.

Fig. 8 Analysis results on screen.

Figure 8 shows the analysis results for the word amu-ke-fu-y 'he used to go', where all tags are in blue and show the tag name when hovered.

When the user uploads a file to be analysed, the system gives a link back to download a text file containing all the analyses.

The user can review information about the tags on a complementary page called "Glosas del Dungupeyem". There is also a search block available where tags and suffixes may be queried.

Whenever a user submits an unknown word to the system, or uploads a file to analyse, we receive an e-mail with this information to check if there were some issues, and improve the system if necessary.

10 Conclusions

The FST analyser has been developed and it has proved to be a reliable tool, of course it is improvable, and it can extend its use to other tools, something that we have already started to do, but not explained in this article, which is devoted to the analyser implementation, specifically in the Mapuche verb treatment.

Throughout this article, the process and results have been described in detail to explain the quality, scale and precision of the system. We continue working on tools derived from the analyser, which is the basic system for current and future work on automatic processing of Mapudungun.

Acknowledgements We deeply and sincerely thank Iñaki Alegria who have guided us along the confection of this article. He has spent countless hours of revision to improve and present this work in a clear, concise and understandable manner.

References

1. Beesley, K. and Karttunen, L., Finite State Morphology. CSLI Studies in Computational Linguistics. CSLI Publications, Stanford, U. S. A. (2003).
2. Castro, R. and Rios, A., Allin Qillqay! A Free Online Web spell checking Service for Quechua, in Ugaz Burga, J. E., Gonzales Sánchez, S. R., and Torres Guerra, C., editors, Memoria - VI Congreso Internacional de Computación y Telecomunicaciones (COMTEL) 2014, Fondo Editorial de la Universidad Inca Garcilaso de la Vega, Lima, Peru, pages 23–30 (2014).
3. Chandía, A. et al. CORLEXIM. Corpus lexicográfico del mapudungun. <http://corlexim.cl> [April, 2021]
4. Chrupala, G. et al., Learning Morphology with Morfette. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., and Tapia, D., editors, Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008), European Language Resources Association (ELRA), Marrakech, Morocco, (2008).
5. Corporación Nacional de Desarrollo Indígena. Gobierno de Chile, Temuco, Chile (2005).
6. Febré, A., Diccionario Araucano-Español ó sea Calepino Chileno-Hispano Por el P. Andrés Febré de la Compañía de Jesus. Reproducido textualmente de la edición de Lima de 1765, por Juan M. Larsen. Con un Apéndice sobre las lenguas Quichua, Aimará y Pampa. Buenos Aires: Juan A. Alsina (1882).
7. Fernández-Varay, A. and Malvestitti, M., Formas no finitas del Mapudüngun en dos variedades de la Argentina. In IX Congreso de la Sociedad Argentina Lingüística: 1–14. Universidad Nacional de Córdoba, Córdoba, Argentina, (2002).
8. Gasser, M. Computational Morphology and the Teaching of Indigenous Languages. Proceedings of the First Symposium on Teaching Indigenous Languages of Latin America. CLACS & MLCP, Indiana University Bloomington & Association for Teaching and Learning Indigenous Languages of Latin America (ATLILLA), (2011).
9. Guevara, T. Las últimas familias e costumbres araucanas. Imprenta, litografía i encuadernación "Barcelona", Santiago, Chile, (1913).
10. Hulden, M., Fast approximate string matching with finite automata. Proceedings of the 5th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland, 25 November 2011 - 27 November 2011, 51-55, (2011).
11. Jurafsky, D., Definition of Minimum Edit Distance. CS 124: From Languages to Information. Stanford University, California, U. S. A. (2012).
12. Lonkon, E., Morfología y Aspectos del Mapudüngun. Universidad Autónoma Metropolitana. Unidad Iztapalapa. México D. F., México, (2011).
13. Lonkon, E., Políticas públicas de lengua y cultura aplicada al mapudüngun. El pueblo mapuche en el siglo XXI. Propuestas para un nuevo entendimiento entre culturas en Chile. Centro de Estudios Públicos, Santiago, Chile (2017).
14. Mösbach, E., Vida y costumbres de los indígena araucanos en la segunda mitad del siglo XIX. Imprenta Universitaria. Santiago, Chile (1936).
15. Ragileo, A., Gramática del idioma Mapuche. Public domain, (1982).
16. Ríos, A., Spell checking an agglutinative language: Quechua. In: 5th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland, 25 November 2011 - 27 November 2011, 51-55, (2011).
17. Ríos, A., A Basic Language Technology Toolkit for Quechua. Faculty of Arts of the University of Zurich. Zurich, Switzerland, (2015).
18. Sadowsky, S. et al., Illustrations of the IPA: Mapudüngun. Journal of the International Phonetic Association 43(1). 87–96. doi:10.1017/S0025100312000369, (2013).
19. Salas, A., El mapuche o araucano. Centro de Estudios Públicos, 2nd edition. Santiago, Chile, (2006).
20. Schmidt, H. and Laws, F., Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. In Proceedings of the 22nd International Conference on Computational Linguistics, volume 1 of COLING ’08, pages 777–784. Association for Computational Linguistics, Stroudsburg, U. S. A., (2008).
21. Smeets, I., A Grammar of Mapuche. Mouton de Gruyter. Berlin, Germany. New York, U. S. A., (2008).
22. Sochil (Sociedad Chilena de Lingüística). Encuentro para la Unificación del Alfabeto Mapuche. Proposiciones y Acuerdos. Arturo Hernández Sallés, Coordinador del Encuentro. Temuco: Pontificia Universidad Católica de Chile, (1986).
23. Sochil (Sociedad Chilena de Lingüística). Alfabeto Mapuche Unificado. Temuco: Pontificia Universidad Católica de Chile, (1988).
24. Zúñiga, F., Mapudüngun. El habla mapuche. Centro de Estudios Públicos. Santiago, Chile, (2006).

11 Annexes

11.1 Tags meaning

Parts of Speech (and other elements)
- AJ: Adjective
- AV: Adverb
- AP: Anaphoric pronoun
- CJ: Conjunction
- COLL: Collectivizer
- CR.: IV: Intransitive verb compound
- CR.: TV: Transitive verb compound
- DP: Demonstrative pronoun
- FE: Foreign expression
- TJ: Interjection
- TP: Interrogative pronoun
- IV: Intransitive verb
- LOC: Locative
- MW: Noun
- BBR: Digit (number)
- NG: Negation particle
- NU: Numeral
- ON: Onomatopoeia
- PCT: Punctuation mark
- PN: Proper name
- PP: Personal pronoun
- PPF: Possible proper noun
- PR: Preposition
- PT: Particle
- QC: Interrogative chum
- QT: Interrogative tunte
- RNNR: Replicatory noun
- RONR: Replicated onomatopoeia
- RVBR: Replicated verb
- SP: Possessive pronoun
- TV: Transitive verb
- UNK: Unknown root or word
- XV: Auxiliary

Suffixes
+1: 1st person
+1A: 1st person agent
+2: 2nd person
+2A: 2nd person agent
+3: 3rd person
+ADJ: Adjectiviser
+ADJDO: Adjectiviser doable
+ADJQE: Adjectiviser quick and easy
+ADV: Adverbializer
+AFF: Affirmative
+AIML: Aimless/involuntary
+AVN: Agentive verbal noun
+BEX: Benefactive
11.2 Suffixes by slot

11.2.1 Slots 1 to 15: inflectional suffixes

- **slot-01.aff**: Dative Subject
  - Ø: 1st or 2nd person agent "+DS12A"
  - ew ~ -mew ~ -mu: 3rd person agent "+DS3A"

- **slot-02.aff**: Number
  - Ø: Singular (1st & 2nd indicative) "+SG"
  - -i: Singular (2nd indicative & conditional) "+SG"
  - -u: Dual "+DL"
  - -ii: Plural (1st indicative & conditional) "+PL"
  - -iin: Plural (2nd all moods & 3rd indicative) "+PL"

- **slot-03.aff**: Person
  - Ø: 1st non-singular "+1" & 3rd "+3" persons
  - -i: 1st person "+1"
  - -m: 2nd person "+2"
  - -e: 3rd person "+3"
  - -ng: 3rd person non-singular "+3"
  - -y: 1st "+1" & 3rd "+3" persons agent

- **slot-03PTMT.aff**: Portmanteau [mood, person, number]
  - -n: Indicative, 1st person, singular "+IND1SG"
  - -chi: Imperative, 1st person, singular "+IMP1SG"
  - -nge: Imperative, 2nd person, singular "+IMP2SG"
  - -pe: Imperative, 3rd person "+IMP3"

- **slot-04.aff**: Mood
  - Ø: Imperative "+IMP"
  - -l: Conditional "+CND" & "+CNI"
  - -y: Indicative "+IND"

- **slot-04NF.aff**: inflectional nominalizers
  - Ø: Objective "+OVN" & subjective "+SVN" verbal noun "+OVN"
  - -el: Objective verbal noun "+OVN"
  - -fie: Transitive verbal noun "+TVN"
  - -lu: Subjective verbal noun "+SVN"
  - -m: Instrumental verbal noun "+IVN"
11.2.2 Slots 16 to 27: mobile derivational suffixes

- slot-05.aff: Constant feature
  -ye: "+CF"
- slot-06.aff: Internal & External direct objects
  -e: Internal direct object "+IDO"
  -fe: External direct object "+EDO"
- slot-07.aff: Pluperfect
  -mu: "+PLPF"
- slot-08.aff: Impeditive
  -fu: "+IPD"
- slot-09.aff: Non-realized situation
  -a: "+NRLD"
- slot-10.aff: Negation
  -ki: Imperative "+NEG"
  -la: Indicative "+NEG"
  -no: Conditional "+NEG"
- slot-11.aff: Affirmative
  -ille: "+AFF"
- slot-12.aff: Reportative
  -rke: "+REP"
- slot-13.aff: Proximity
  -pe: "+PX"
- slot-14.aff: Constant feature
  -ke: "+CF"
- slot-15.aff: Pluperfect
  -wye: "+PLPF"

11.2.3 Slots 28 to 36: fixed derivational suffixes

- slot-22.aff: Play & simulation
  -faluw: Simulation "+SIM"
- slot-23.aff: Passive, 1st & 2nd persons agent
  -w: 1st person agent "+1A"
  -wu: 2nd person agent "+2A"
  -nge: Passive "+PASS"
- slot-23M.aff: Passive mobile
  -nge: "+PASS"
- slot-24.aff: Pluralizer
  -ye: "+PL"
- slot-25.aff: Force & satisfaction
  -fal: Force "+FORCE"
  -hte: Satisfaction "+SAT"
- slot-25M.aff: Force mobile
  -fau: "+FORCE"
  -sl: Passive "+PASS"
  -slot-26.aff: Indirect object
  -lma: "+IO"
- slot-27.aff: Beneficiary
  -el: "+BEN"
- slot-28.aff: Stative & progressive
  -le: Stative "+ST"
  -meke: Progressive "+PR"
- slot-28M.aff: Progressive mobile
  -le: "+ST"
- slot-29.aff: More involved object
  -l: "+MIO"
- slot-30.aff: Circular movement & intensive
  -iaw: Circular movement "+CIRC"
  -tie: Intensive "+INT"
- slot-31.aff: Reflexive/reciprocal
  -we: "+REF"
- slot-32.aff: Progressive persistent & perfect persistent
  -kim: Perfect persistent "+PFPS"
  -nie: Progressive persistent "+PRPS"
- slot-33.aff: Transitivizer & factitive
  -ka: Factitive "+FAC"
  -tu: Transitivizer "+TR"
- slot-33M.aff: Transitivizer mobile
  -tu: "+TR"
- slot-34.aff: Causatives
  -l: "+CA"
  -mr: "+RA"
- slot-35.aff: Experience & oblique object
  -ma: Experience "+EXP"
  -ye: Oblique object "+OO"
- slot-36S.aff: Stem formative
  -Ø: "+SFR"
  -nge: "+SFR"
  -tu: "+SFR"
  -ye: "+SFR"
11.2.4 Nominal suffixes

- slot-36V. aff: Verbalizers
  -Ø: "+VRB"
  -l: "+VRB"
  -nge: "+VRB"
  -ntu: "+VRB"
  -tu: "+VRB"
  -ye: "+VRB"

- slot-36V. aff: Verbalizers
  -Ø: "+VRB"
  -l: "+VRB"
  -nte: "+ADV"
  -nta: "+ADV"
  -ntu: "+ADV"
  -tu: "+ADV"
  -ye: "+ADV"

- slot-36V. aff: Instrumental
  -mew ~ mur: "+INST"

- slot-36V. aff: Nominalizers
  -Ø: nominalizer "+NOM"
  -fal: Doable "+ADJD"  
  -fe: Agentive "+NOMAG"
  -nte: Quick & easy "+ADJQE"
  -we: Place or instrument "+NOMPT"

11.2.5 Other suffixes

- OS. aff: Aimless/involuntary
  -piada: "+AIML"

11.2.6 Examples by suffixes

Verbalizer -Ø- (slot 36): E25, E26, E55, E59, E61, E62, E64, E69, E128, E136, E141, E170, E192, E200, E205, E209, E225, E226, E233, E237.

Verbalizer -l- (slot 36): E50, E65.

Verbalizer -nge- (slot 36): E2, E5, E23, E47, E49, E52, E53, E56, E57, E63, E145, E193, E194.

Verbalizer -ntu- (slot 36): E66.

Verbalizer -tu- (slot 36): E26, E30, E51, E52, E53, E54, E204.

Verbalizer -ye- (slot 36): E60.

Stem formative -Ø- (slot 36): E32, E151.

Stem formative -nge- (slot 36): E48, E146.

Stem formative -tu- (slot 36): E3.

Stem formative -ye- (slot 36): E58.

Causative -l- (slot 34): E29, E90, E91, E93, E131, E191, E214, E220, E222, E224, E226, E228, E230.

Causative -m- (slot 34): E58, E72, E142, E199, E203, E208, E257.

Factitive -ka- (slot 33): E56, E192, E193, E199, E205, E219, E226, E228, E230.

Trasitivizer -tu- (slot 33): E41, E43, E45, E87, E137, E188, E191, E203, E271, E272, E274.

Perfect persistent -kānu- (slot 32): E61, E73, E85, E91, E151, E189, E204, E211.

Progressive persistent -nie- (slot 32): E1, E121, E132, E137, E141.

Reflexive/reciprocal -w- (slot 31): E55, E56, E60, E85, E110, E211, E219, E263.

Circular movement -iaw- (slot 30): E207, E216, E237.

More involved object -l- (slot 29): E204, E205, E206, E207.

Stative -le- (slot 28): E7, E43, E55, E68, E70, E71, E74, E80, E82, E89, E90, E110, E224, E226, E227.

Beneficiary -el- (slot 27): E8, E207

Indirect object -āma- (slot 26): E1, E230.

Force -fal- (slot 25): E155, E220.

Passive -nge- (slot 23): E6, E58, E72, E91, E147, E158, E162, E167, E172, E185, E189, E205, E257, E267, E271, E274, E275.

2nd person agent -mu- (slot 23): E8, E19, E196, E272.

1st person agent -w- (slot 23): E195, E233.

Simulation -faluw- (slot 22): E7.

Play -kantu- (slot 22): E172, E200.
Immediate -fem- (slot 21): E6.
Sudden -rum- (slot 21): E162.
Thither -me- (slot 20): E8, E9, E10, E74, E96, E97, E98, E125, E140, E152, E163, E173, E212, E213, E214, E215, E221, E222, E255.
Persistence -we- (slot 19): E9, E74, E76, E86, E128, E135, E152, E153, E205.
Interruptive (one) -r- (slot 18): E1, E167.
Interruptive (repeated) -yeku- (slot 18): E10.
Hither -pa- (slot 17): E70, E93, E95, E97, E98, E158, E190, E202, E210, E212, E216, E223, E224, E273.
Locative -pu- (slot 17): E1, E42, E77, E92, E167, E209.
Continuative -ka- (slot 16): E25, E251.
Repetitive/restorative -tu- (slot 16): E1, E10, E70, E226, E242, E252, E255, E257, E270, E271, E272.
Pluperfect -wye- (slot 15): E4, E5.
Constant feature -ke- (slot 14): E9, E51, E72, E91, E131, E134, E135, E152, E163, E189, E191, E206, E216, E217, E228, E254.
Proximity -pe- (slot 13): E93, E133, E267.
Reportive -kke- (slot 12): E5, E131.
Affirmative -lle- (slot 11): E59.
Negation (imperative) -ki- (slot 10): E15, E17, E21, E128, E150, E181.
Negation (indicative) -la- (slot 10): E9, E12, E26, E62, E66, E74, E116, E124, E127, E130, E152, E153, E180, E217, E220, E243, E277.
Negation (conditional) -no- (slot 10): E13, E27, E78, E119, E135, E185, E205, E218.
Non-realized situation -a- (slot 9): E28, E46, E59, E61, E73, E77, E79, E122, E125, E126, E141, E148, E160, E164, E166, E169, E174, E185, E186, E196, E205, E239, E241, E242, E255, E257, E276, E277.
Impeditive -fu- (slot 8): E3, E91, E129, E130, E131, E135, E136, E189, E191, E215, E217, E257.
Pluperfect -mu- (slot 7): E24, E175.
Internal direct object -e- (slot 6): E1, E18, E22, E45, E62, E65, E101, E121, E124, E126, E127, E128, E131, E138, E139, E140, E141, E168, E197,  E231, E232, E234, E235, E241, E242, E243.
External direct object -fi- (slot 6): E20, E21, E41, E50, E73, E98, E129, E130, E137, E150, E151, E154, E156, E165, E174, E176, E178, E181, E198, E204, E208, E214, E230.
Constant feature -ye- (slot 5): E162.
Imperative -Ø- (slot 4): E14, E20.
Conditional -l- (slot 4): E13, E19, E59, E77, E112, E113, E118, E119.
Conditional (in imperatives) -l- (slot 4): E17, E21, E128, E150.
Indicative -y- (slot 4): E1, E4, E5, E6, E8, E9, E10, E11, E12, E16, E18, E26, E28, E43, E44, E47, E48, E49, E50, E51, E52, E53, E54, E55, E56, E58, E60, E65, E66, E67, E68, E69, E79, E91, E94, E99, E110, E114, E115, E116, E121, E124, E131, E138, E145, E146, E147, E152, E156, E157, E158, E160, E161, E163, E167, E168, E169, E170, E171, E172, E178, E180, E186, E189, E191, E193, E211, E216, E217, E220, E221, E222, E224, E225, E228, E229, E231, E232, E233, E234, E235, E236, E237, E238, E241, E254, E257, E270, E271, E272, E273, E274, E275, E277.
Subjective verbal noun -Ø- (slot 4): E33, E55, E72, E90.
Objective verbal noun -el- (slot 4): E25, E46, E61, E77, E92, E125, E132, E133, E134, E135, E142, E143, E144, E148, E149, E164, E185, E209.
Transitive verbal noun -iel- (slot 4): E28, E196, E206, E207, E255.
Subjective verbal noun -ie- (slot 4): E27, E93, E141, E182, E188, E190, E195, E198.
Instrumental verbal noun -m- (slot 4): E24, E162, E175, E205.
Plain verbal noun -n- (slot 4): E26, E34, E43, E51, E52, E53, E54, E56, E64, E69, E75, E76, E80, E81, E86, E87, E88, E89, E111, E179, E183, E187, E192, E194, E210, E218, E219, E223, E256, E263, E267, E268, E276.
Agentive verbal noun -t- (slot 4): E22, E197.

Completeive subjective verbal noun -wma- (slot 4): E23, E57, E63, E71.

1st person non-singular -Ø- (slot 3): E1, E8, E16, E18, E94, E110, E114, E115, E121, E124, E138, E156, E160, E178, E224, E231, E233, E239, E241, E272.

3rd person non-singular -Ø- (slot 3): E4, E5, E6, E9, E26, E28, E43, E47, E48, E49, E50, E51, E53, E55, E56, E58, E65, E66, E67, E68, E69, E79, E91, E99, E116, E131, E145, E146, E147, E152, E157, E158, E161, E163, E167, E168, E169, E170, E171, E172, E180, E186, E189, E191, E193, E211, E217, E220, E221, E222, E225, E228, E229, E257, E270, E271, E273, E274, E275.

Imperative 1st person singular -chi- (slot 3): E15.

1st person -i- (slot 3): E19, E118, E119, E128.

3rd person -e- (slot 3): E59, E113.

Indicative 1st person singular -ne- (slot 3): E2, E3, E7, E41, E42, E45, E62, E70, E72, E73, E74, E82, E83, E84, E85, E95, E96, E97, E98, E101, E102, E112, E126, E127, E129, E130, E136, E137, E139, E140, E151, E153, E154, E155, E159, E165, E166, E173, E174, E176, E177, E184, E204, E208, E212, E213, E214, E215, E226, E227, E230, E242, E243, E251, E252, E253, E267, E276.

3rd person non-singular -ng- (slot 3): E10, E44, E54, E254.

Imperative 2nd person singular -nge- (slot 3): E150, E181.

2nd person -m- (slot 3): E11, E12, E13, E14, E20, E52, E60, E78, E216, E232, E234, E235, E236, E237, E238, E240, E277.

1st persons agent -y- (slot 3): E17, E21.

Singular -Ø- (slot 2): E19, E119, E128.

Singular -i- (slot 2): E11, E52, E78, E216, E232, E236, E237, E238, E240, E277.

Plural -i-Ø- (slot 2): E1, E8, E21, E94, E110, E115, E118, E156, E224, E233, E239.

Dual -u- (slot 2): E12, E14, E16, E17, E18, E20, E60, E112, E114, E121, E124, E138, E178, E231, E234, E241, E272.

Plural -iun- (slot 2): E10, E13, E44, E54, E235, E254.

Dative subject 1st or 2nd persons agent -Ø- (slot 1): E18, E101, E121, E124, E126, E128, E138, E139, E168, E231, E232, E234, E235, E241, E242, E243.

Dative subject 3rd person agent -mew- (slot 1): E1, E22, E45, E62, E65, E127, E131, E140, E141, E197.

nominalizer -Ø-: E199, E200, E201, E202.

Adjective -chi-: E33, E55, E72, E90, E134, E142.

Ex (discontinuative) -em-: E36.

Doable -fal-: E29, E170.

Agentic -fe-: E30, E53, E193.

Distributive -ke-: E35, E50, E55.

instrumental -mew-: E41, E42, E43, E83, E84, E120, E143, E148, E206, E221, E251, E252, E253, E254.

Quick & easy -nten-: E31, E57.

Group -ntu-: E37, E112.

Aimless/involuntary -puda-: E154.

Reportative -rke-: E23, E38.

Adverbializer -tu-: E34, E64.

Place or instrument -we-: E32, E87, E203.

Temporal -we-: E39.

Relative -wen-: E40, E194.

11.2.7 Intermediate language symbols

"@1", "@2", "@3". Provisional contextual marks to treat changes derived from the interaction between implicative -fus- (slot 8) and internal direct object -e-, external direct object -fi- (slot 6) or objective verbal noun -el- (slot 4); see R6, R7 and figure 6.

"@4", "@5", "@6". Provisional contextual marks to treat radical consonant alternation of some intransitive verbs when interacting with causative -im-, which translivizit them. "@4" operates in the insertion of ng; "@5" operates in the interchange between f and p; "@6" operates in the interchange between g and k; see 5.1.5 and R15.
"@∅∅" treats the elision of the final e of progressive persistent -nie- (slot 32) when followed by a; see 5.3.6.

"@∅D" stands for the external direct object -fi- (slot 6) while processing its interaction with implicative -fu- (slot 8) or objective verbal noun -el- (slot 4); see D12, R6, R9, figure 6 and 1st item on this list.

"@∅I" treats the change of e into i of constant feature -ke- (slot 14) when followed by ü or y; see 5.3.6.

"@∅L" stands for objective verbal noun -el- (slot 4) while processing its interaction with implicative -fu- (slot 8), internal direct object -e- or external direct object -fi- (slot 6), depending on that, it could take the form -el- or -l-; see D13, R7, R11, R13, figure 6 and 1st item on this list.

"@∅Y" treats the optional realization as -ngi- or -nge- forms; see 5.1.2, R12, D15, D23.

"@∅P" regulates the alternation of the radical consonant f into p of intransitive verbs when transitivized by causative -üm- (slot 34); see 5.1.5, D18, R15 and 2nd item on this list.

"@∅G" treats the optional epenthesis of a glottal stop between vowels belonging to different morphemes, usually roots in compounds; see 6.4.2, D2, E117, R1, R13, R14, D18, 5.3.6, D31, D34.

"@∅K" regulates the alternation of the radical consonant g into k of intransitive verbs when transitivized by causative -üm- (slot 34); see 5.1.5, D18, R15.

"∅i" is used to regulate different realizations of the verb i- 'to eat', which depends on the morphophonological context where it appears; see 5.1.3, D16, R13.

"∅I" stands for the internal direct object -e- (slot 6) while processing its interaction with negation for imperative -ki- (slot 10), implicative -fu- (slot 8) or objective verbal noun -el- (slot 4); see D12, R5, R7, R10, R11, R13, D31, figure 6 and 1st item on this list.

"∅IP" stands for implicative -fu- (slot 8) while processing its interaction with internal direct object -e-, external direct object -fi- (slot 6) or objective verbal noun -el- (slot 4); see D11, R6, R7, R8, R9, figure 6 and 1st item on this list.

"∅IZ" treats the indicative (slot 4) alternative realization -i-, see 6.4.1, D30, R51.

"∅K" treats the insertion of kāu in stative -le- (slot 27) giving -kile-, mandatory when preceded by semi-vowel or any consonant but m, optional when preceded by m or u.

"∅KG" enables the optional initial k of perfect persistent -kün- (slot 32) when preceded by d, f or m.

"∅KI" treats the form alternation of circular movement -kiaw- (slot 30) after semi-vowel or any consonant but m; -yaw- after m or any vowel.

"∅KU" (capital u with diaereses) treats the insertion of kāuí at the beginning of intensive marker -tie- (slot 30) when preceded by g, i, n, r or u.

"∅L" manages the alternation between -e- and -le- that may be added to causative -l- (slot 34) and benefactive -el- (slot 27). Both may take l, el, lel forms depending on previous phoneme.

"∅M". In Mapuche speech, interjections are important because they mark the intention in discourse. Hm is widely used but its transcription may bare one or more m. We use the next rule to recognize the interjection spite being written with different amounts of m: (R 55) hm interjection recognition define RuM ["M" -> m*];

"∅N" regulates the epenthesis of n at the beginning of a form which is obligatory when the form is preceded by a, e, i, o, and optional when preceded by u, ü or at form boundary; see D23.

"∅NG" regulates the radical addition of ng in some intransitive verbs when transitivized by causative -üm- (slot 34); see 5.1.5, D18, R15.

"∅NK" is used to treat the negation for imperative -ki-, it optionally takes the forms -ke- or -k- when followed by e; see D4, D9, R5, R10, R13.

"∅R" regulates the insertion of n for experience marker -ma- (slot 35) when preceded by vowels.

"∅RF" treats the optional n insertion in some forms ending in vowel as word ending or when followed by consonant.

"∅U" indicates the insertion of a u between a form beginning in w and a previous consonant or semi-vowel; see 5.1.

"∅U" (capital u with diaereses) is used to treat schwa insertion represented by ü in the spelling, this is the most common epenthesis in Mapudüngun; see D5, D6, 5.1, 5.1, D7, D14, R10, R12, D30.
"@ÜC" (capital u with diaereses) is used to treat schwa-insertion in causative suffixes -l- and -m- (slot 34), but also the radical consonant alternation of some intransitive verbs which take this suffix, see 5.1.5, D19, R15.

"@ÜI" (capital u with diaereses) is used to treat the optional alternative form of indicative -iy- (slot 4) after f, l, m; see D30.

"@ÜÑ" (capital u with diaereses) regulates the form of indirect object (slot 26) and adverbializer -ïma-. They realize as -ma- when preceded by f, l, sh. They realize as -ïma- when preceded by e, i, a, u. They realize as -ïma- or -yma- when preceded by a, û. And they realize as -ïûma- when preceded by semi-vowel or ch, d, k, ll, m, n, ñ, ng, p, r, s, t, tr; see D33.

"@VE" treats the form of the verb entu- 'to take out', this is explained in 5.1.4, p. 27, see D17, R14.

"@Y" regulates the insertion of a y between a form beginning in a or e and a previous a, e or boundary; see D8, R4, D10, R13.

"@YY" it also regulates the insertion of a y, but in different contexts. Obligatory after vowel or form boundary, optional after semi-vowel, in both cases the following form begins in e.
11.3 Conjugation of the intransitive verb **küpa-** ‘to come’

The root **küpa-** is returned as **-IV. küpa+venir** by the analyser. To make the conjugation reading simpler, only the Mapuche root is presented in the table. Imperative forms preceded by an * are identical to those of indicative mood, and may be used adhortatively.

**Table 13** Conjugation of the intransitive verb **küpa-** ‘to come’

| Person | Indicative | Conditional | Imperative |
|--------|------------|-------------|------------|
| 1s     | **küpa-n** | **küpa-l-i** | **küpa-chi** |
|        | -IV. küpa+IND1SG.n3 | -IV. küpa+CND.14+1.13+SG.Ø2 | -IV. küpa+IMP1SG.chi3 |
| 1d     | **küpa-y-u** | **küpa-l-i-u** | *küpa-y-u* |
|        | -IV. küpa+IND.y4+1.Ø3+DL.u2 | -IV. küpa+CND.14+1.13+DL.u2 | -IV. küpa+IND.y4+1.Ø3+DL.u2 |
| 1p     | **küpa-y-iñ** | **küpa-l-i-iñ** | *küpa-y-iñ* |
|        | -IV. küpa+IND.y4+1.Ø3+PL.iñ2 | -IV. küpa+CND.14+1.13+PL.iñ2 | -IV. küpa+IND.y4+1.Ø3+PL.iñ2 |
| 2s     | **küpa-y-m-i** | **küpa-l-m-i** | **küpa-nge** |
|        | -IV. küpa+IND.y4+2.m3+SG.i2 | -IV. küpa+CND.14+2.m3+SG.i2 | -IV. küpa+IMP2SG.nge3 |
| 2d     | **küpa-y-m-a** | **küpa-l-m-a** | **küpa-m-u** |
|        | -IV. küpa+IND.y4+2.m3+DL.u2 | -IV. küpa+CND.14+2.m3+DL.u2 | -IV. küpa+IMP.Ø4+2.m3+DL.u2 |
| 2p     | **küpa-y-m-un** | **küpa-l-m-un** | **küpa-m-un** |
|        | -IV. küpa+IND.y4+2.m3+PL.ün2 | -IV. küpa+CND.14+2.m3+PL.ün2 | -IV. küpa+IMP.Ø4+2.m3+PL.ün2 |
| 3      | **küpa-y**   | **küpa-l-e**  | **küpa-pe** |
|        | -IV. küpa+IND.y4+3.Ø3 | -IV. küpa+CND.14+3.e3 | -IV. küpa+IMP3.pe3 |
| 3d     | **küpa-y-ng-u** |             |             |
|        | -IV. küpa+IND.y4+3.ng3+DL.u2 |             |             |
| 3p     | **küpa-y-ng-ün** |             |             |
|        | -IV. küpa+IND.y4+3.ng3+PL.ün2 |             |             |
11.4 Conjugation of the transitive verb *pi* - ‘to say (to tell)’

- **TV.pi_decir.** Only the Mapuche root is shown in the table. Imperative forms with * are identical to the indicative counterpart, those marked with ** may be used adhortatively.

### Table 14 Conjugation of the transitive verb *pi* - ‘to say (to tell)’

| Persons | Indicative | Conditional | Imperative |
|---------|------------|-------------|------------|
| 1s → 2s | -TV.pi+IDO.e6+IND.y4+1.Ø3+DL.u2+DS12A.Ø1 | -TV.pi+IDO.e6+CHD.14+1.y3+DL.u2+DS12A.Ø1 |  |
| 1 → 2 (more than 2 participants) | -TV.pi+1A.w23+IND.y4+1.Ø3+PL.iñ2 | -TV.pi+1A.w23+CHD.14+1.13+PL.iñ2 |  |
| 2s → 1s | -TV.pi+IDO.e6+IND1SG.n3+DS12A.Ø1 | -TV.pi+IDO.e6+CHD.14+1.13+SG.Ø2+DS12A.Ø1 | -TV.pi+IDO.e6+IND1SG.n3+DS12A.Ø1 |
| 1s → 3 | -TV.pi+EDO.f16+IND1SG.n3 | -TV.pi+EDO.f16+CHD.14+1.13+SG.Ø2 | -TV.pi+EDO.f16+IMP1SG.chi3 |
| 1d → 3 | -TV.pi+EDO.f16+IND.y4+1.Ø3+DL.u2 | -TV.pi+EDO.f16+CHD.14+1.13+DL.u2 | -TV.pi+EDO.f16+IMP1SG.chi3 |
| 1p → 3 | -TV.pi+EDO.f16+IND.y4+1.Ø3+PL.iñ2 | -TV.pi+EDO.f16+CHD.14+1.13+PL.iñ2 | -TV.pi+EDO.f16+IMP2SG.nge3 |
| 2s → 3 | -TV.pi+EDO.f16+IND.y4+2.m3+SG.Ø2 | -TV.pi+EDO.f16+CHD.14+2.m3+SG.Ø2 | -TV.pi+EDO.f16+IMP.Ø4+2.m3+DL.u2 |
| 2d → 3 | -TV.pi+EDO.f16+IND.y4+2.m3+DL.u2 | -TV.pi+EDO.f16+CHD.14+2.m3+DL.u2 | -TV.pi+EDO.f16+IMP.Ø4+2.m3+PL.üm2 |
| 3 → 3 | -TV.pi+EDO.f16+IND.y4+3.ng3 | -TV.pi+EDO.f16+CHD.14+3.ng3 | -TV.pi+EDO.f16+IMP3.pe3 |
| 3d → 3 | -TV.pi+EDO.f16+IND.y4+3.ng3+DL.u2 | -TV.pi+EDO.f16+CHD.14+3.ng3+DL.u2 | -TV.pi+EDO.f16+IMP3.pe3 |
| 3p → 3 | -TV.pi+EDO.f16+IND.y4+3.ng3+PL.üm2 | -TV.pi+EDO.f16+CHD.14+3.ng3+PL.üm2 | -TV.pi+EDO.f16+IMP3.pe3 |
| 3 → 1s | -TV.pi+IDO.e6+IND1SG.n3+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+1.13+SG.Ø2+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 1d | -TV.pi+IDO.e6+IND.y4+1.Ø3+DL.u2+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+1.y3+DL.u2+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 1p | -TV.pi+IDO.e6+IND.y4+1.Ø3+PL.iñ2+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+1.13+PL.iñ2+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 2s | -TV.pi+IDO.e6+IND1SG.n3+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+1.13+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 2d | -TV.pi+IDO.e6+IND.y4+2.m3+DL.u2+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+2.m3+DL.u2+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 2p | -TV.pi+IDO.e6+IND.y4+2.m3+PL.üm2+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+2.m3+PL.üm2+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
| 3 → 3 | -TV.pi+IDO.e6+IND.y4+1.Ø3+DS13A.ew1 | -TV.pi+IDO.e6+CHD.14+1.y3+DS13A.ew1 | -TV.pi+IDO.e6+IND1SG.chi3+DS13A.ew1 |
11.5 Negative imperative forms of the transitive verb *pi*-'to say (to tell)'

Negative imperative forms are infrequent. Note that a negative command can also be expressed by a negative indicative form which is marked with the suffix -a- +NRLD.a9 for non-realized situation, as shown below.

(E 277 ) [Smeets, I. 2008: 173 (82)] [21]

`amu-la-ya-y-m-i` 'you must not go' (lit: 'you will not go')

\[-iV. amu_ir+NEG. la10+NRLD.a9+IND.y4+2.m3+SG.i2\]

| Persons | Negative imperative | Translation |
|---------|---------------------|-------------|
| 2s → 1s | -TV.pi+NEG.ki10+IDO.e6+CNI.14+1.i3+SG.Ø2+DS12A.Ø1 | Don’t (you/s) tell me |
| 2d/p → 1s | -TV.pi+2A.mu23+NEG.ki10+CNI.14+IMP1SG.chi3 | Don’t (you/d/p) tell me |
| 2s/d/p → 1d | -TV.pi+2A.mu23+NEG.ki10+CNI.14+1.y3+DL.u2 | Don’t (you/s/d/p) tell us/d |
| 2s/d/p → 1p | -TV.pi+2A.mu23+NEG.ki10+CNI.14+1.y3+PL.iñ2 | Don’t (you/s/d/p) tell us/p |
| 1s → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+IMP1SG.chi3 | I don’t tell him/her/them |
| 1d → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+1.y3+DL.u2 | Let us/d not tell him/her/them |
| 1p → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+1.y3+PL.iñ2 | Let us/p not tell him/her/them |
| 2s → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+IMP2SG.nge3 | Don’t (you/s) tell him/her/them |
| 2d → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+2.m3+DL.u2 | Don’t (you/d) tell him/her/them |
| 2p → 3 | -TV.pi+NEG.ki10+EDO.fi6+CNI.14+2.m3+PL.ün2 | Don’t (you/p) tell him/her/them |
| 3 → 1s | -TV.pi+NEG.ki10+IDO.e6+CNI.14+IMP1SG.chi3+DS3A.mew1 | May he/she/they not tell me |
11.6 Flow chart diagrams of the FST analyser

FST Analyser: diagram 1

Characters definition

Suffixes incorporation

Lexicon incorporation

Morphotactic rules

Verb formation

Non-verb forms

Deployed in diagram 2

Compounding

Suffixation

Morphophonologic rules

Deployed in diagram 3

Cleaning & compiling

Fig. 9 FST Analyser flux diagram 1: general view.
Fig. 10 FST Analyser flux diagram 2: verb formation view.
Fig. 11 FST Analyser flux diagram 3: morphophonologic rules, cleaning & compiling view.
