Institut für Computerlinguistik
Universität Zürich

Herbstsemester 2010

Applying Finite-State Techniques to a Native American Language: Quechua

Lizentiatsarbeit der Philosphischen Fakultät
der Universität Zürich
Referent: Prof. Dr. Martin Volk

Verfasserin:

Annette Rios

Bachtelstrasse 32

8620 Wetzikon

Matrikelnummer 03–703–634

arios@ifi.uzh.ch

Abstract

Comprehensive finite-state morphology systems have been developed for numerous languages, nevertheless the American indigenous languages have received far less attention from the computational linguistic field than the standard European languages. For this thesis, I implemented a complete morphology system for the Andean language Quechua. Dealing with a non-standardized indigenous language of low social prestige and sparsely available resources imposes serious challenges on the development of computational linguistic tools. Nevertheless, I will show that finite-state techniques are perfectly suited to capture the relatively complex morphological structures of Quechua, once the linguistic processes determining word formation have been unravelled.

Acknowledgments

I'm grateful to many persons who helped me during the writing of this thesis. First of all, I'd like to thank my supervisor Prof. Dr. Martin Volk for his support and constructive critics.
I would also like to express my gratitude to Dr. Simon Clematide, who provided the technical support on the `xfst` implementation with a considerable amount of patience.
I'd like to thank native Quechua speaker Marisol Pillco Grajeda from Cusco, who kept answering my questions over and over again.
I am grateful to Anne Göhring for reading through the complete thesis and pointing out the remaining deficiencies.
A big thank you goes to my sister Melanie Chenoweth for proof-reading the final script.
And last but not least I want to thank my mum for the uncountable amount of hours of baby sitting, which allowed me to complete my thesis.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Background

Comprehensive systems for morphological analysis and generation have been developed for a wide range of languages, nonetheless the numerous indigenous languages of the Americas have received far less attention from computational linguists than the standard European languages. Although the characterization of the native American languages as a unity would be presumptuous, given the notable heterogeneity in this large geographical area, it is safe to say that, from a morphological point of view, these languages offer a level of complexity that exceeds by large what computational linguists deal with in European languages.

The extremely complex word formation processes observed in native American languages represent not only a special challenge for a computational approach, but may also lead to interesting insights on the possibilities and limitations of the standard methods applied for morphological analysis and generation.

As a matter of fact, almost all native languages of the Americas struggle with strong social pressure from the dominant languages (mainly English, Spanish and Portuguese), a large number having become extinct already. If the attention from the computational linguistic field towards one of these languages may help increase the social prestige of this language, then this should constitute another good reason for the commitment and dedication to an indigenous language.

An indispensable requirement for the implementation of any serious morphological tool is a comprehensive linguistic description of the language in question. This prerequisite alone restricts the possibilities to a relatively small number of indigenous languages, one of them being, by number of speakers, the largest native American language: Quechua.

Substantial theoretical knowledge of the language in question is not as essential as the availability of detailed linguistic descriptions, nevertheless it provides a good starting point for the implementation of a morphology system. Since I had already acquired a solid working knowledge of Quechua for my major subject (general linguistics) and given the relatively extensive documentation on this language family, Quechua was a natural choice for this project.

From a morphological point of view, Quechua is more complex and feature rich than most common European languages, nonetheless, its morphology is relatively simple compared to other native American languages, such as the Algonquian, Eskimo-Aleut or the Athapaskan languages, just to mention a few. Even so, Quechua imposes some serious challenges to the implementation of an analyzer and generator, while at the same time, it provides keen insights into a completely different language structure.

1.2. Outline and Objectives

The implementation of a finite-state tool for morphological analysis and generation of Quechua word forms gives rise to some interesting questions:

1. *Linguistic insights:* What kind of new linguistic insights can be achieved from an in-depth computational linguistic examination of Quechua word structures?

2. *Efficiency:* How efficient are finite-state tools for the analysis and generation of Quechua word forms?

3. *Coverage:* Is it possible to cover a satisfactorily wide range of Quechua word forms with finite-state techniques?

4. *Spell checking:* At what expense can a spell checker be implemented from the analysis and generation tools? Is spell checking with finite-state techniques feasible for Quechua? Is it efficient?

5. *Effort:* How does the effort for the implementation of Quechua finite-state tools compare to analogous systems that have been implemented for other agglutinative languages like Turkish or Finnish? How many rules does a Quechua system need? More or less than a system for Turkish?

6. *Application contexts:* What are the possible application contexts for Quechua morphology tools?

7. *Benefits:* Can natural language processing support native minority languages in their endangered usage by increasing their social prestige and awakening more persons' interest?

These questions represent the guiding principle throughout my thesis, although in the end, I may not have a concrete answer to all of them.

The thesis is structured as follows: The sections 2 and 3 provide the necessary background information to the topic. In section 2 I will present a general sketch of the Quechuan languages, containing an illustration of their distribution as well as some fascinating linguistic facts. Section 3 introduces the finite-state approach and exemplifies its usefulness for morphological analysis and generation. Section 4 contains a short description of the former programming project that lead to this thesis. Section 5 comprises a profound and detailed linguistic portrayal of the complex nature of Quechua morphology. Section 6 subsumes an evaluation for the analyzer and generator implemented for this project. Section 7 describes the enhancement of the original tools to a spell checker. Section 8 contains a detailed description of another morphology tool capable of analyzing and generating Quechua word forms, including a comparison between this system and mine. Section 9 gives an overview on two morphology systems for languages with considerable structural similarity to Quechua: Aymara and Turkish. Finally, in the last section 10, I will present the conclusions that can be drawn from the implementation of morphology tools for a language with a relatively complex and extremely flexible morphological system.

2. Quechua

Quechua is a group of closely related languages, spoken by 8-10 million people in Peru, Bolivia, Ecuador, Southern Colombia and the North-West of Argentina. Ethnologue[1] also lists some Quechua speakers for Chile. Quechua is one of the official languages of Peru and Bolivia. Peru especially, has increased efforts to provide its citizens with official information not only in Spanish, but also in Quechua and (to less extent) in some other indigenous languages like Aymara and Asháninka[2].

Although Quechua is often referred to as 'language' and its local varieties as 'dialects', Quechua represents a language family, comparable in depth to the Romance or Slavic languages (Adelaar and Muysken 2004:168). Mutual intelligibility, especially between speakers of distant 'dialects', is not always given.

2.1. Distribution of Quechuan Languages

The Quechuan Languages are divided into two main branches, Quechua I and II in terms of the Peruvian linguist Torero (1964), respectively Quechua A and B in terms of the American linguist Parker (1963). I will use Torero's labels in my thesis.

Quechua I is the more archaic group of dialects, spoken in Central Peru (see Fig. 1, blue region). It comprises a heavily fragmented dialect complex, with limited mutual comprehension between the different local varieties, although they share a number of clear common features (Adelaar and Muysken 2004:185). The origin of the Quechuan languages lies probably in this area (Cerrón-Palomino 2003).

The second branch,Quechua II, comprises all the remaining Quechua dialects:

- QIIA, spoken in Northern Peru
- QIIB, spoken in Ecuador and Colombia
- QIIC, spoken in Southern Peru, Bolivia, and Argentina

The letters A-C stand for the linguistic distance to QI, QIIA is therefore the most akin to QI, whereas QIIC is the most divergent group respective to QI.
The dialects of Quechua IIA occupy an intermediate position between Quechua I and the rest of Quechua II (Adelaar and Muysken 2004:186) (red regions in Fig. 1). The classification of these dialects is not as straightforward as it might seem: The northern dialects of Cajamarca and Ferreñafe have attributes of both Quechua

[1] http://www.ethnologue.com
[2] See for example http://www.defensoria.gob.pe/quechua/index.php.

IIB and Quechua I, whereas the dialects of Yauyos hold a similar position between the Quechua IIC and Quechua I varieties (Adelaar and Muysken 2004:186).

Quechua IIB, comprises the Ecuadorian Quechua (also called *Kichwa*), the Quechua spoken in Colombia (called *Inga* or *Ingano*) and the dialects spoken in the Peruvian departments of San Martín, Loreto and Amazonas (see Fig. 1, green region) (Adelaar and Muysken 2004:187).

Quechua IIC comprises all the remaining Quechua dialects to the south of the Quechua I group, including the dialect groups Ayacucho, Cuzco-Bolivia and Argentina (see Fig. 1, purple region). The division between Ayacucho and Cuzco-Bolivian Quechua is mainly due to the occurrence of glottalized and aspirated stops in the Cuzco-Bolivian dialects, a phonetical distinction that Ayacucho and Argentina Quechua lack. Cuzco-Bolivian Quechua itself is, however, no homogeneous group at all (Adelaar and Muysken 2004:187),(Cerrón-Palomino 2003:242-245).

The tools implemented for my thesis are designed for the Quechua IIC dialects, and within these the focus lies especially on the Ayacucho and Cuzco variants. Although Bolivian and Cuzco Quechua are traditionally considered to constitute a single dialect unity, there are substantial differences.
The most persuasive reason for the choice of QIIC for this project is the amount of available linguistic descriptions, as Quechua IIC has received more attention from linguists than the other Quechua varieties.

2.2. Linguistic Features

Quechua is an essentially agglutinative language, almost all syntactical information is expressed through suffixation. Like many other indigenous languages of the Americas, Quechua also features an inclusive and exclusive distinction in the first person plural (Cusihuamán 1976), (Soto Ruiz 2006):

ñuqa	1.Sg	'I'
ñuqanchik(Ayacucho) *ñuqanchis*(Cuzco)	1.Pl.Inclusive	'we (you inclusive)'
ñuqayku	1.Pl.Exclusive	'we (you not)'

Gender is no grammatical category, not even in pronouns.

Phonology Quechua has only three phonemic vowels, *a, i* and *u*, although *e* and and *o* occur as allophones of *i* and *u* in proximity of *q*, respectively *χ*, see Fig. 2.

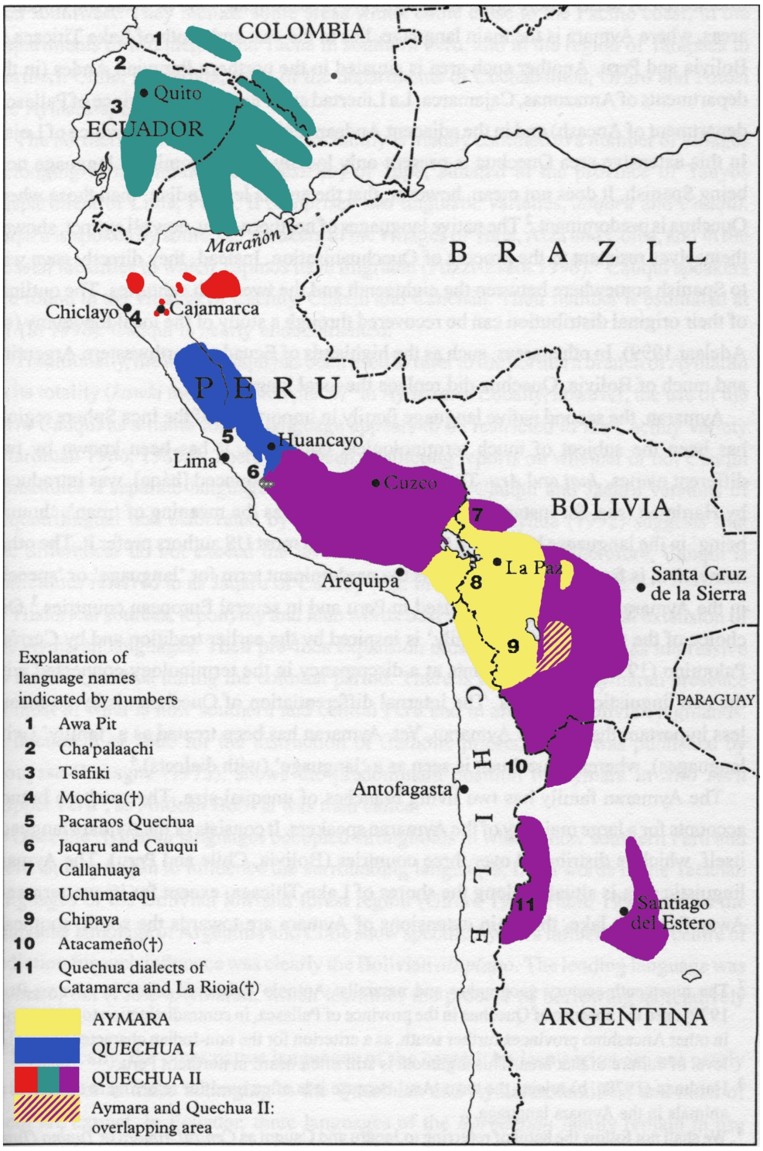


Figure 1: Map of Quechuan Dialects/Andean Languages (mid 20th century), adapted from Adelaar and Muysken (2004:169)

Table 1: Quechua IIC Consonants

		Bilabial	Labio-Dental	Dental	Alveolar	Postalveolar	Palatal	Velar	Uvular	Glottal
	plain	p (b)[2]		t (d)[2]				k (g)[2]	q	
Plosive	glottalized[1]	p’		t’				k’	q’	
	aspirated[1]	p^h		t^h				k^h	q^h	
Nasal		m		n			ɲ			
Tap or Flap				ɾ						
Fricative			(f)[2]		s	(ʃ)[3]			(χ)[4]	h
	plain					tʃ				
Affricate	glottalized[1]					tʃ’				
	aspirated[1]					$tʃ^h$				
Approximant		w[5]					j			
Lateral Appr.					l		ʎ			

[1] only in Cuzco/Bolivian Quechua
[2] only in Spanish loan words
[3] only in Cuzco Quechua, phonemic status not clear
[4] phonemic only in Ayacucho Quechua, instead of ’q’
[5] w: Voiced labio-velar approximant

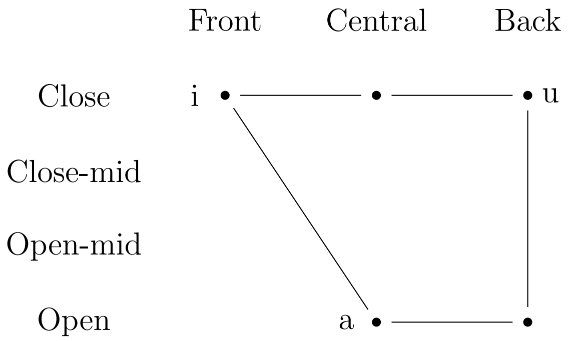
Some QI dialects also feature a phonemically distinctive vowel length.

As for the consonant inventory, Cuzco-Bolivian and Ayacucho Quechua differ, as mentioned above, with regard to glottalized and aspirated stops. Taking this circumstance into consideration, the total number of phonemic consonants amounts to a total of 25 in Cuzco-Bolivian Quechua and 15 in Ayacucho Quechua, see the phonetic chart in Table 1.

Stress is always on the penultimate syllable, in other words, stress is not fixed, but ’shifts’ with the addition of suffixes. There are a few exceptions to this general rule, mainly in connection with special emphatic utterances. These exceptions are marked by accents, e.g. with the suffix *-yá* that expresses emphasis (Dedenbach-Salazar Sáenz et al. 2002:6).

Valid syllable structures are `V`, `VC`, `CV` and `CVC`, yet under the restriction that `V` and `VC` can only occur word initially (Cerrón-Palomino 2003:256).

Figure 2: Quechua IIC Vowels



Morphology I will describe Quechua morphology in depth in section 5.

Syntax Basic word order in Quechua is SOV. While constituent order in main clauses is basically free, word order in dependent clauses follows SOV rather strictly. Syntactic constructions are generally head-final (Adelaar and Muysken 2004:207), e.g. dependent clauses usually precede the main clause.
A typologically interesting feature is Quechua's double-marking nature (Nichols 1986:72). In most constructions, both head and dependens, are morphologically marked: on the sentence level, the finite verb bears cross-reference markers for both subject and object, in addition to case-marking suffixes on the NPs. Furthermore, in possessive constructions as well the possessor as the possessum bear a genitive suffix, respectively a possessive suffix[3]:

(1)	*ñuqa*	***-pa***	*wasi*	***-y***
	I	-Gen	house	-1.Sg.Poss

'My house'.

(Dedenbach-Salazar Sáenz et al. 2002:34)

Description of event sequences is handled through so called 'clause chaining' or 'switch reference'. In these clause chaining sentences, the chained clauses are marked either as having the same subject as the finite clause (with *-spa* or *-stin*) or as having a different subject (with *-pti* in Ayacucho, *-qti* in Cuzco Quechua). I treat these chaining suffixes as nominalizers, since person marking, if present, is handled through possessive suffixes. Nonetheless, the resulting forms are not typical nouns but rather so-called converbs (Ebert 2008), as the nominal suffixes

[3] For a listing of the abbreviations used, see Appendix A.

they can bear are limited to possessive suffixes, and, only for *-spa* (same subject marker), the case suffixes *-wan* (instrumental), *-ntin* (inclusive) and possibly *-kama* (distributive/terminative). Combinations of *-spa* with these case suffixes are extremely rare though.
Consider the following Quechua sentences as examples:

(2)

Chay	*-manta*	*-pas*	*tukuy*	*machu*	*paya*	*-kuna*	*-pas*	*llamka*
Dem	-Abl	-Add	all	old.man	old.woman	-Pl	-Add	work

-na	*-n*	*-kuna*	*-ta*	*saqi*	***-spa***	*-qa*	*sapa*	*killa*	*qullqi*
-Obl	-3.Poss	-Pl	-Acc	leave	-SS	-Top	each	month	money

-cha	*-ta*	*chaski*	*-ku*	***-spa***	*sama*	*-mu*	*-y*	*-ta*	*-m*	*ati*
-Dim	-Acc	receive	-Rflx	-SS	rest	-Trs	-Inf	-Acc	-DirE	can

-nqaku.
-3.Pl.Subj.Fut

'Then all old men and old women can retire, leaving their work, and receiving some money each month.'

(Acu 2002)

(3)

Taki	***-stin***	*llamka*	*-chka*	*-nku.*
sing	-SS_Sim	work	-Prog	-3.Pl.Subj

'They work singing.'

(Dedenbach-Salazar Sáenz et al. 2002:168)

(4)

Waqa	***-pti***	*-yki*	*-qa*	*ri*	*-pu*	*-ku*	*-saq*
cry	-DS	-2.Sg.Poss	-Top	go	-Rgr_Iprs	-Rflx_Int	-1.Sg.Subj.Fut

'If you cry, I'll leave.'

(Cerrón-Palomino 2003:279)

The chained clauses in example 2 [*llamkanankunata saqi**spa**qa*] and [*sapa killa qullqichata chaskiku**spa***] have the same subject (*machu payakunapas* - old men and women) as the main clause, hence the same subject marker *-spa*.
The suffix *-stin* in example 3 differs from *-spa* only in the temporal relation it establishes between the chained clause and the main clause: *-stin* indicates, besides the identity of subjecthood, that the actions denoted by main clause and chained clause(s) occur simultaneously (Cerrón-Palomino 2003:278).
The suffix *-pti* (*-qti* in Cuzco/Bolivian Quechua) marks inequality of the subject in the chained clause from its counterpart in the main clause, as illustrated by the last sentence. This last instance of a chained clause in example 4 , nominalized by *-pti/-qti* bears a mandatory possessive suffix, indicating the subject of the chained clause. Subject markers are optional with *-spa*, whilst they never occur in combination with *-stin.*

The combination of the suffixes *-spa* and *-pti/-qti* with some ambivalent suffixes (see section 5.1 for details) is used to form subordinated clauses of various types (conditional, causal, concessive, temporal). Consider sentence 5: *-spa* or *-pti* together with the ambivalent suffix *-pas* (additive, 'also, too') yields a concessive reading, whereas in example 4, it is the combination of *-pti* with the topic marker *-qa* that implicates a conditional reading.

(5)

Qam	*-qa*	*mana*	*llamka*	***-spa***	***-pas***	*achka*	*-ta*	*-m*	*miku*	*-nki.*
you	-Top	Neg	work	-SS	-Add	much	-Acc	-DirE	eat	-2.Sg.Subj

'Although you don't work, you eat a lot.'

(Dedenbach-Salazar Sáenz et al. 2002:94)

The following section will present the technical background to the approach I used to gather some of the linguistic characteristics described in this section.

3. Finite-State Approach

Finite-state techniques are probably the most prevalent approach employed by automatic morphology systems, as their simplicity and outstanding efficiency are unequaled. Various alternative frameworks allow for a straightforward implementation of finite-state networks, e.g. the FSA Utilities Toolbox developed at the University of Groningen (NL)[4] or *foma* (Hulden 2009), another finite-state compiler, programming language, and regular expression/finite-state library that facilitates the realization of finite-state networks for automatic morphology systems. Even though *foma*, as well as the FSA Utilities Toolbox have the obvious advantage of being open source systems, I decided to use the (licensed) Xerox Finite State Tools, in short `xfst`, for my purpose, as I was already familiar with this framework. This initial choice however, does not exclude the possibility of a future re-implementation with one of the open source alternatives in order to make the system freely available to whomever might be interested.
The following two sections elucidate the theoretical background to the finite-state approach applied to describe the morphological characteristics of Quechua.

3.1. Regular Languages

A regular language, which is a special type of a formal language, is a (possibly infinite) set of words consisting of a finite set of symbols: the alphabet Σ. The set of all words over Σ is denoted as Σ^*.
Accordingly, the sequences of symbols out of Σ that make up the words of a given regular language form a subset of Σ^*. A regular language over an alphabet Σ is defined as follows (Carstensen et al. 2010:71):

- The empty language $\emptyset$ and the language that contains only the empty word $\{\epsilon\}$ are regular languages
- For each $a \in \Sigma$, the language $\{a\}$ is a regular language
- If L1 and L2 are regular languages, then
  - L1 $\cup$ L2 (union)
  - L1 $\cdot$ L2 (concatenation)
  - L1* and L2* (Kleene star)

  are also regular languages

[4] See `http://www.let.rug.nl/~vannoord/Fsa/` for more information.

3.2. Finite-State Machines

A finite-state machine (FSM) is an abstract machine that implements a regular language. Regular languages can be described formally in a concise notation: through regular expressions. Finite-state tools, such as `xfst`, offer a convenient method to compile such regular expressions into finite-state machines. Consequently, a finite-state machine, which is specified through a regular expression denoting the rules that describe a given regular language, has the capability to accept or reject a given string, according to wheter the string is a valid word form of this particular regular language. There is a further distinction into deterministic and non-deterministic finite-state machines:
In a deterministic finite-state machine, every state has exactly one transition for each possible input symbol. In a non-deterministic finite-state machine, on the contrary, the input of a particular symbol can lead to more than one transition for a given state.
A finite-state machine M=$\langle\Phi, \Sigma, \delta, S, F\rangle$ consists of (Carstensen et al. 2010:74-75):

- a set of states: Φ
- a start state: $S \in \Phi$
- a set of final states: $F \subseteq \Phi$
- a set of symbols ('alphabet'): Σ
- state transition function:
  deterministic FSM: $\delta : S \times \Sigma \rightarrow \Phi$
  non-deterministic FSM: $\delta : S \times \Sigma \rightarrow \wp(\Phi)$

The *deterministic* machine reads a given string symbol by symbol and with each symbol, it performs a transition between two states according to its internal implementation: If s is the actual state of the machine, and it reads the symbol a, the transition function $\delta(s, a)$ will lead it to the next state. If the last transition leads to a final state, the string is accepted, i.e. this particular string is a valid word of the regular language the finite-state machine specifies (Herold et al. 2007:593).
In a *non-deterministic* finite-state machine, the procedure is slightly different: The machine reads a given string symbol by symbol and with each symbol, it performs as many transitions between two states as its internal implementation allows for the given input symbol. The reading of a symbol a in a state s will lead to a set of following states $\wp(\Phi)$. If the reading of the last symbol in the input sequence leads to at least one final state, the string is accepted, which means, it is a valid word form of the regular language specified by the finite-state machine.

A graph-like representation of a finite-state machine is a network, consisting of nodes linked together with directed arc-transitions, in which final states are represented by double-lined circles (Beesley and Karttunen 2003:5). Consider Fig. 3 for an example: This deterministic machine accepts exactly two words, *dog* and *cat*.

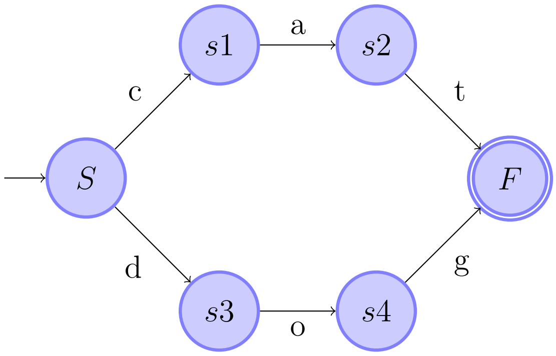


Figure 3: Finite-State Machine with L={*dog,cat*}

3.3. Finite-State Transducers

A finite-state transducer (FST) is basically an enhanced finite-state machine. As opposed to the finite-state machine, it operates on two levels, i.e. an 'input tape' and an 'output tape'. As a consequence, the scope of its applicability goes beyond the binary decision of accepting or rejecting a given string: it is able to return the corresponding output to a given input (Beesley and Karttunen 2003:11).
A finite-state transducer accordingly implements a relation between two formal languages: an upper-side and a lower-side regular language, and it literally 'transduces' strings from one language into the other. In a non-deterministic finite-state transducer, it may produce more than one possible output for a given string.
A finite-state transducer T=$\langle \Phi, \Sigma, \Gamma, \delta, S, F \rangle$ consists of (Carstensen et al. 2010:78-79):

- a set of states: Φ
- a start state: $S \in \Phi$
- a set of final states: $F \subseteq \Phi$
- a set of input symbols: Σ

- a set of output symbols: Γ
  deterministic FST: $\delta : S \times (\Sigma \cup \{\epsilon\}) \times (\Gamma \cup \{\epsilon\}) \rightarrow \Phi$
  non-deterministic FST: $\delta : S \times (\Sigma \cup \{\epsilon\}) \times (\Gamma \cup \{\epsilon\}) \rightarrow \wp(\Phi)$

See Fig. 4 for an example of a finite-state transducer that contains the relation of four word forms with Quechua *misi*, 'cat' and their respective morphological analysis (with Spanish translation of the root)[5]:

$$\begin{aligned} R = \{ \quad & gato + Dim + Acc : & misichata, \\ & gato + Dim + Gen : & misichapa, \\ & gato + Acc : & misita, \\ & gato + Gen : & misipa \quad \} \end{aligned}$$

Which side of the transducer serves as input is not predetermined, the transducer can be applied in both directions:

1. Given *misichapa* as input, applied in 'upward' direction, it produces *gato+Dim+Gen* as output.
   This is the procedure for morphological analysis.

2. Given *gato+Dim+Acc* as input, applied in 'downward' direction, it produces *misichata* as output.
   This is the procedure for generation.

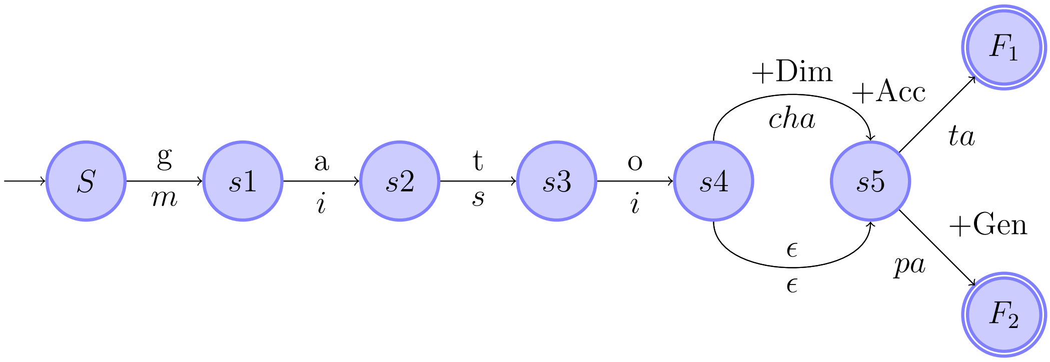


Figure 4: Finite-State Transducer

The following two sections describe the steps that were necessary in order to obtain the detailed linguistic information that is required in a finite-state approach, as well as the insights gained on Quechua morphology during this process. Section 5 illustrates the implementation of Quechua morphology with the described finite-state techniques.

[5] See Appendix A for a listing of the abbreviations used.

4. Former Project: A Quechua - Spanish Parallel Treebank

The original analyzer was implemented during a former programming project described in Rios Gonzales et al. (2009). The goal of this undertaking was to build a small Quechua - Spanish parallel treebank, containing about 230 sentences. The corpus contains the *Declaration of the Human Rights* in Ayacucho Quechua dialect, a poem in Cuzco Quechua and some information texts and the FAQ from the website of the Peruvian *Defensoría del Pueblo*[6]. The acquired knowledge about Quechua structures gained from the creation of this parallel treebank proved to be useful for the development of the Quechua word form generator in many ways.

For the syntactic annotation of the Quechua sentences, *Role and Reference Grammar* (as described in Van Valin Jr. and Polla (1997) and Van Valin (2005)) was the first choice, as the handling of complex nominal subordinated clauses with the more common phrase structures was unsatisfactory (Rios Gonzales et al. 2009:57). Another benefit RRG provides for the creation of Quechua syntax trees is the proper treatment of so called 'clause chaining' sentences (see section 2.2).

As for morphology, the analyzer built during this project was implemented with Xerox Finite State Tools `xfst` (Beesley and Karttunen 2003). The insights on Quechua suffixes achieved during this project provide the fundamental linguistic facts for the implementation of a Quechua word form generation tool, and for this reason, I will give a short overview of the procedure that led to the realization of the `xfst` analyzer. The hardest challenge was to get enough information on Quechua morphotactics for the implementation. There is no grammar that provides listing of suffix sequences or possible combinations detailed enough for a computational linguistic tool.
Nevertheless, some descriptions outline tendencies in suffix orders (Lastra (1968), Dedenbach-Salazar Sáenz et al. (2002), Soto Ruiz (1976b)), though some of them are contradictory, a circumstance probably due to differences in local varieties.

As the implementation of an analyzer and generation tool requires reliable and elaborated morphotactical knowledge, I had to find myself a method to identify the details of the morphotactical structures in Quechua word forms. Starting with the rough suffix orders in Table 2, I scanned through large amounts of Quechua texts and the comprehensive word list collected by biologist Jacob Philips[7]. Ad-

[6] The Defensoría del Pueblo is an institution that makes sure the state complies with its responsibilities for its citizens and that should also prevent the state from violating the rights of citizens.

[7] His word list is available on http://www.runasimi.de/runasimi.txt.

Table 2: Suffix Order

Nominal Root	Derivation	Possession	Case		Ambivalent Suffixes
Verbal Root	Derivation	Aspect/Tense	Person	Modality	Ambivalent Suffixes

ditionally, I had the chance to clarify some remaining obscurities with a native speaker from Cuzco (via chat).
The following description of the division into slots and their combination patterns, represents my own findings.

5. Quechua Morphology

Quechua is a strongly agglutinative, suffixing language. There are more than 130 Quechua suffixes, the exact number, as well as the spelling of the suffixes exhibits a considerable variation across dialects, even within the Quechua IIC block.

Especially the verbal forms diverge considerably: Each local variety has its own preferred suffix order (Cerrón-Palomino 2003, Cusihuamán 1976), and variation in this order is not only allowed, but can sometimes even lead to a change in meaning. This circumstance has important implications for both analysis and generation of complex Quechua word forms.

This section contains some answers to my research question number 1, formulated in section 1.2:

1: What kind of new linguistic insights can be achieved from an in-depth computational linguistic examination of Quechua word structures?

The findings on Quechua suffixes and their combination possibilities have not been described before in such detail. The need for an exact morphotactical scheme that is fine-grained enough for automatical analysis and generation of Quechua words led to insights that can not be found in any Quechua grammar:

The partitioning into slots in the nominal and verbal structures, as well as the different sub-types of Quechua roots and their particular morphotactical restrictions represent new knowledge about Quechua IIC. The following sections 5.1 and 5.2 contain an elaborate description of these findings.

5.1. Quechua Suffixes

There are five functional classes of Quechua suffixes as described in Table 3. Besides the nominalizing and verbalizing suffixes, there are many nominal and verbal derivational, respectively inflectional suffixes. Additionally, Quechua has a small set of 'ambivalent' suffixes, also called independent suffixes. These suffixes can be attached to both verbal or nominal forms, without altering the part of speech of the given word form. The position of these suffixes is at the end of the suffix sequence, their relative order is more or less fixed, dialects show some minor variation though.

The functions of the ambivalent suffixes include data source, polar question marking and topic or contrast, amongst others. In combination with interrogative expressions, these suffixes may acquire special meanings (Adelaar and Muysken 2004:209). In combination with demonstrative pronouns, the ambivalent suffixes may also take the place of conjunctions, which are virtually non-existent in Quechua, unless they are borrowed from Spanish. See Table 9 for a detailed

Table 3: Suffix Classes

1	nominalizing	V → N
	llank'a -q ,'work-Agentive' ⇒	worker
2	verbalizing	N → V
	wasi -cha-, 'house-Factitive' ⇒	to build a house
3	nominal (derivation/inflection)	N → N
	wasi -su , 'house-Augmentative' ⇒	big house
4	verbal (derivation/inflection)	V → V
	wañu -chi- , 'die-Causative' ⇒	kill
5	ambivalent	N → N, V → V
	e.g. evidentiality, topic, epistemic modality...	

listing of the ambivalent suffixes.[8]

In order to allow suffix variation, the individual suffixes must be divided into groups, according to their relative position in the word.

Three out of the five suffix classes in Table 3 require further refinement, particularly the nominal, as well as the verbal derivational/inflectional suffixes and the ambivalent suffixes (see Table 9 for the latter). I divided the nominal suffixes (N→N) and the verbal suffixes (V→V) into 7 groups according to their relative position in the word. See Tables 5 and 7 for a detailed listing of nominal and verbal suffixes.
Some of the verbal and nominal slots are 'repeatable', i.e. more than one suffix out of a group is possible, whilst others are not. If more than one suffix of a given slot is present in the word form, the relative order of these suffixes is variable, reflecting the differences between the various local varieties of the language.

There are many ambivalent roots, that can take either verbal or nominal morphology without modification (see example 6). Additionally, there are some extremely productive nominalizing and verbalizing suffixes that can change a nominal root into a verbal one, and vice versa, see also Table 4. Some of these suffixes can combine with each other (examples 7 and 8). The suffixes involved in clause

[8] The suffix *-lla* 'honorific, limitative' is missing in these scheme, as *-lla* may occur in almost every slot. *-lla* is, contrary to the rest of the ambivalent suffixes, not even restricted to the end of the suffix sequence.

chaining (see 2.2) are a special case of nominalization: These chained forms can not be re-verbalized, in fact the resulting word forms ('converbs') may only bear possessive and ambivalent suffixes (example 9).

(6)	*chinka*	*-y.*	*chinka*	*-ni.*
	loss/lose	-1.Sg.Poss	loss/lose	-1.Sg.Subj
	'My loss.'		'I lose.'	

(7)	*kachi*	*-cha*	*-sqa*
	salt	-Fact(VS)	-Perf(NS)
	'salted, salty'		

(8)	*wiña*	*-y*	*-cha*	*-ku*	*-y*
	grow	-Inf(NS)	-Fact(VS)	-Rflx	-Inf(NS)
	'to perpetuate oneself'				

(9)	*∗wiña*	*-spa*	*-cha*	*-ku*	*-y*	*(SS + VS)*
	∗wiña	*-pti*	*-cha*	*-ku*	*-y*	*(DS + VS)*

5.2. Quechua Roots

Figure 5 illustrates a simplified[9] version of the finite-state transducer implemented in the `xfst` analyzer and generator.
There are two basic sorts of roots: the ones that follow the nominal scheme and the verbal roots (*VRoot*), additionally, there is a tiny subgroup of particles that can only bear ambivalent suffixes. There is some evidence that Quechua adjectives constitute a separate word class, as they are subject to syntactical restrictions (Adelaar and Muysken 2004:208):

> The main criterion for establishing the difference is that a noun can function by itself as the subject in a sentence, whereas real adjectives can only act as subjects when followed by an element that indicates their status as an independent item; an element frequently so used is *ka-q* '(the one) that is', e.g. in *hatun ka-q* 'the (a) big one'.

Nevertheless, adjectives do not constitute a separate word class from a morphological point of view, as their morphological behaviour is identical to other nominal roots, i.e. they can bear the same suffixes. For this reason, I do not treat

[9] Nominalization through 'chaining' suffixes is missing in this figure, additionally, I did not include the ambivalent suffix *-lla*, which may ocurr in almost every slot.

Table 4: Word Class Changing Suffixes

Verbalizing Suffixes		Nominalizing Suffixes	
$N \rightarrow V$		$V \rightarrow N$	
lli	Autotransformative	*y*	Infinitive
naya	Desiderative	*ti* ∼ *li* ∼ *lu*	Characterization
na	Reubicative	*na*	Obligation
cha	Factitive	*q*	Agentive
raya	Characterization	*mpa*	Positional
ya	Transformative	*sqa*	Perfect
kacha	Simulative	’Converbs’ (chaining suffixes):	
∼ *ykacha*		*spa* *shpa*	Same Subject
		sti ∼ *stin*	Same Subject Simultaneous
		pti */qti*	Different Subject

(Appendix B contains an explanation of the linguistic terms used and some examples for illustration)

Table 5: Nominal Suffixes, Slots 1-4

Slot 1		Slot 2		Slot 3	Slot 4	
Stem Derivation		Possessor Derivation		Possessive	Number	
cha	Diminutive	*sapa*	Multiple Possessor	–	*kuna*	Plural
chika *karay* *chaq* *su*	Augmentative	*yuq* *nnaq*	Possessor Abessive			
niraq *rikuq*	Similarity					
ti *li* *liku* *yli* *lu*	Characterization					
mpa	Positional					

Table 6: Nominal Suffixes, Slots 5-7

Slot 5		Slot 6		Slot 7	
Case		Case		Case	
ntin	Inclusive	*ta*	Accusative	*kama*	Distributive
pura	Intersociative	*pa*	Genitive	*wan*	Instrumental/ Connective
nka	Distributive	*nta*	Prolocative	*rayku*	Cause
kama	Terminative	*manta*	Ablative	*puwan*	Sociative
niq	Approximative	*man*	Dative/ Illative		
pi	Locative				
paq	Benefactive				

Table 7: Verbal Suffixes, Slots 1-2

Slot 1				Slot 2	
Stem Derivation		Valency Changing Suffixes		Directionals & Reflexive	
ymana	’Rememorative’	*ysi*	’Assistive’	*ku*	’Reflexive, Intensifier’
tiya	’Simulative’	*naya*	’Desiderative’	*pu*	’Regressive, Interpersonal’
pasa	’Desesperative’	*na*	’Reciprocal’	*mu*	’Cislocative, Translocative’
rpari	’Intentional’	*chi*	’Causative’		
rqu	’Urgency’	*raya*	Perdurative		
cha	’Verbal Diminutive, childishness’				
pa	Repetitive’				
ri	’Inchoative’				
lli	’Autotransformative’				
ykacha	’Interruptive, Frecuentative’				
nya	Continuity				
yku	Affective				
paya	Multi-Repetitive				

Table 8: Verbal Suffixes, Slots 3-7

Slot 3		Slot 4		Slot 5		Slot 6	Slot 7	
Object		Aspect		Tense		Person	Modality	
wa	’1st Person Object	*sha*	’Progressive’	*rqa*	’Neutral Past’	–	*man*	’Potential’
su	’2nd Person Object’			*sqa*	’Narrative Past			

Table 9: Ambivalent Suffixes

hina ’Sim’	*puni* ’Cert’	*pas* ’Add’	*taq* ’Con/Intr’	*chu* ’Neg/Intr’	*mi*	’DE’	*iki*	’Res’
					si	’IE’	*ya*	’Emp’
		raq ’Cont’			*cha*	’Ass’	*má*	’DE-Emp’
					qa	’Top’	*sá*	’IE-Emp’
		ña ’Disc’			*ri*	’QTop’	*chá*	’Ass-Emp
					suna	’Dub’		

(dashed line - combination possible vs. normal line - combination not possible)

adjectives as a separate word class in this approach.

Besides the open word classes of adjectives and real nouns (nominal roots, *NRoot*), the nominal group comprises also the closed word classes of pronouns and particles. Out of these, only the latter can bear the nominal derivational suffixes in slot 1. Pronouns, i.e. demonstrative (*Dem*), interrogative (*Intr*) and personal pronouns (*Pers*) can not be verbalized.

There are two sorts of particles: One group (*PrtV*) allows nominal morphology and can even be verbalized via derivational suffixes, whilst the other class may only bear ambivalent suffixes (*Part*). Some borrowed particles and conjunctions of Spanish origin belong to this group as well.

In my system, the open class of nominal roots (*NRoot*) comprehends a lexicon of nominal Quechua roots, one with Quechua number words, one with indefinite nouns, and in the case of the analyzer, an additional lexicon with nominal Spanish loan words.

The verbal roots consist of a lexicon with Quechua verbal roots and, only for the analyzer, another lexicon with Spanish verbal loan words, see Table 10 for details. The nominal Quechua roots may combine with each other to build nominal compounds, whereas the verbal roots permit nominal Quechua roots and Spanish nominal loans to be incorporated as unspecific objects.

All these roots combine with the Quechua suffixes following the illustration in Fig. 5, yet there is one important exception: there are some rare cases where combinations of verbal roots with the reflexive suffix -*ku* in slot 2 seem to be treated as lexicalized units. If such a lexicalized form bears a derivational suffix from slot 1, the order may be inverse. My system recognizes these exceptions.

Pers	Personal Pronoun	Part	Particle
Intr	Interrogative Pronoun	PrtV	verbalizable Particle
Dem	Demonstrative Pronoun		

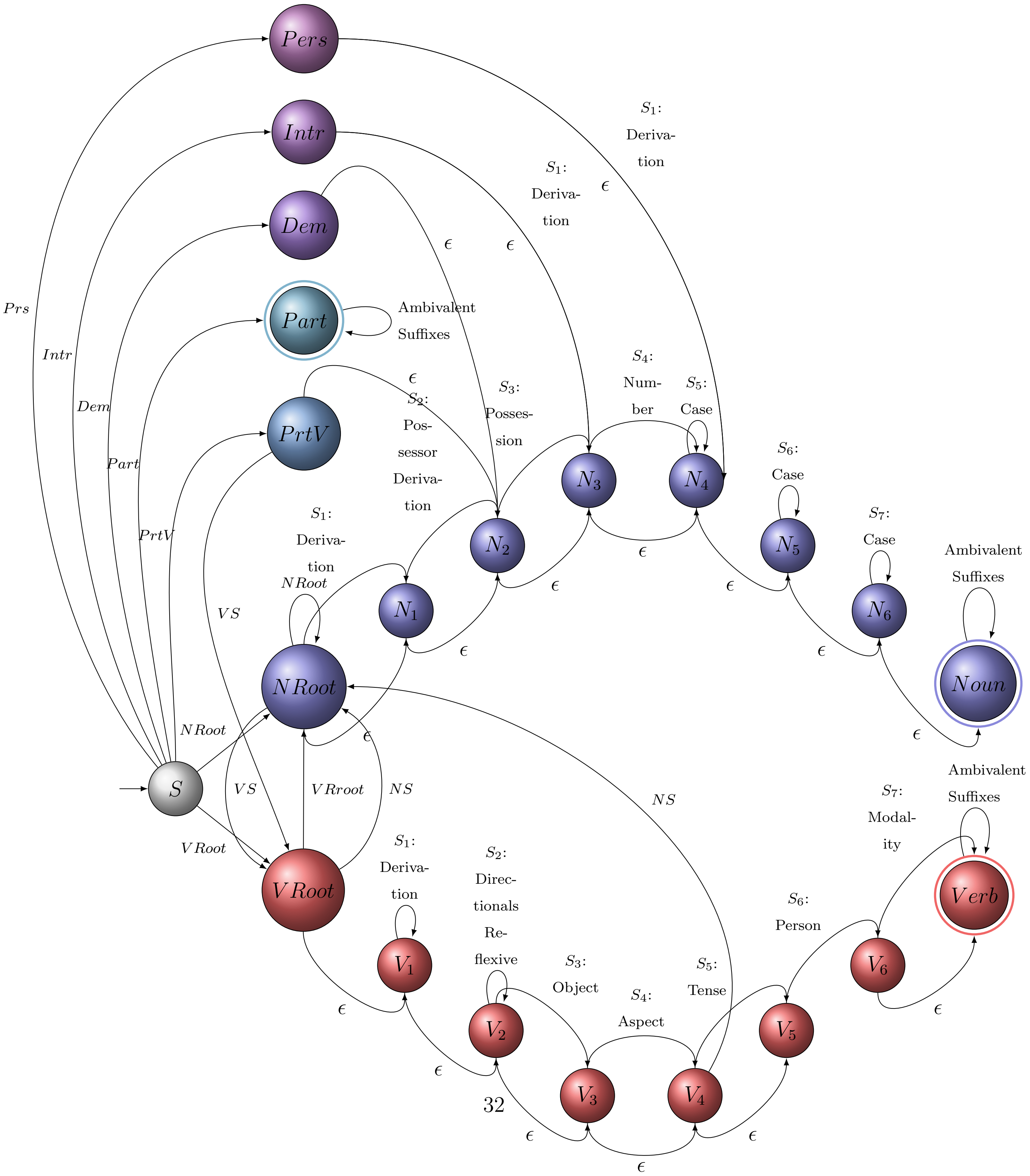


Figure 5: Simplified Quechua Suffixes Transducer

Table 10: Root Classes

Class	Subclass		Examples
Nominal	NRoot	Nominal Quechua Roots	*sach’a*, ’tree’ ; *allqu*, ’dog’
		Numbers	*huk*, ’one’ ; *chunka*, ’ten’
		Indefinite Nouns	*achka*, ’a lot’ ; *lliw*, ’all’
		Spanish loan nouns	*asinda*, ’farm’ (Sp: *hacienda*)
	Pronouns	Personal	*ñuqa*, ’I’ ; *pay*, ’he/she’
		Demonstrative	*kay*, ’that’ ; *chay*, ’this’
		Interrogative	*pi*, ’who’ ; *may*, ’where’
	verbalizable Particles		*hina*, ’like’ ; *ama*, ’not’ (Imperative)
non-verbalizable Particles		Quechua	*arí*, ’yes’ ; *icha*, ’or’
		Spanish	*sinu*, ’but’ (Sp: sino) ; *si*, ’if’
Verbal	VRoot	Verbal Quechua roots	*puri-*, ’walk’ ; *rima-*, ’speak’
		Spanish loan verbs	*distruwi-*, ’destroy’ (Sp: *destruir*)

5.3. Quechua Morphotactics

Quechua is for the most part an entirely regular agglutinative language. Nevertheless, there are some minor morphophonological irregularities that deserve consideration.
There are roughly three cases of morphophonological changes when it comes to word formation: Vowel deletion, vowel change and epenthesis. The sections 5.3.1, 5.3.2 and 5.3.3 contain a detailed description of these morphophonological behaviour. The subsequent section 5.3.4 illustrates the different types of `xfst` notations that facilitate the handling of the described morphophonological changes.

5.3.1. Vowel Deletion

Vowel deletion affects some of the ambivalent suffixes and also two of the verbal derivational suffixes. However, the circumstances that trigger this deletion are completely different.
The verbal suffixes that undergo vowel deletion are *-ykacha* (interruptive, frecuentative, discontinuative, simulative) and *-mu* (cislocative, translocative). *-ykacha* becomes *-kacha* mainly when it follows an *i*, although it may occur as *-kacha* also in other contexts, whereas *-mu* loses its vowel only if it precedes *-pu* (regressive,

interpersonal).

Finding accurate denominations for the individual Quechua suffixes is not as straightforward as it might seem. *-ykacha* illustrates this challenge in a nice way: *-ykacha* indicates that an action takes place in short intervals and (often) without care (Dedenbach-Salazar Sáenz et al. 2002:197). *-ykacha* may also confer a contemptuous connotation to the meaning of the verb (Soto Ruiz 1976b:114). Sometimes however, it conveys the meaning of a simulated action.

(10)

Allin	*-ta*	*-m*	*llamka*	*-nki*	*-qa,*	*mana*	*-m*	*llamka*
good	-Acc	-DirE	work	-2.Sg.Subj	-Top	Neg	-DirE	work

-kacha	*-na*	*-yki*	*-paq*	*-chu*	*paga*	*-chka*	*-yki*	*-qa.*
-Intrup	-Purp	-2.Sg.Poss	-Ben	-Neg	pay	-Prog	-1.Sg>2.Sg	-Top

’Work hard (lit. work well), I don’t pay you for working reluctantly.’

(Soto Ruiz 1976b:114)

-mu is another special case, as the label ’Cis_Trs’ (cis- and translocative) is counter-intuitive. Indeed *-mu* conveys the meaning of a cislocative when used with verbs of movement, i.e. a movement towards the actual point of reference, which is, except for narrative contexts, the location of the speaker. Otherwise, *-mu* adds a translocative reading to (almost) all other verbs.

(11)

Kuti	***-mu***	*-chka*	*-nku*	*-ña*	*llamka*	*-q*	*-kuna*	*-qa.*
return	-Cis_Trs	-Prog	-3.Pl.Subj	-Disc	work	-Ag	-Pl	-Top

’The workers already return.’ (towards location of speaker)

(Soto Ruiz 1976b:109)

(12)

Chay	*-qa*	*mana*	*ikhu*	*-ri*	***-m***	***-pu***	*-n*	*-chu.*
Dem	-Top	Neg	appear	-Inch	-Cis_Trs	-Rgr_Iprs	-3.Sg.Subj	-Neg

’Thereupon [the fox] did not appear.’

(Itier 1997:335)

The suffixes *-mu* and *-pu* are both in verbal slot 2, and their combination indeed occurs more frequently in inverse order. In this case, the vowel of the preceding suffix *-pu* is affected by vowel change, but see section 5.3.2.

The ambivalent suffixes that are affected by vowel deletion are the evidentiality suffixes *-mi* (direct evidentiality) and *-si* (indirect evidentiality) and the epistemic suffix *-cha* (assumptive): all three suffixes lose their vowel if they directly follow a vowel. Additionally, the suffix *-mi* becomes *-n* instead of *-m* in Cuzco and Bolivian Quechua ((Faller 2002:14), Cusihuamán (1976)).

(13) *Arí, qan -wan* ***-mi.***
yes you -Inst_Con -DirE
'Yes, with you.'

(14) *Arí, yuraq wasi* ***-m.***
yes white house -DirE
'Yes, [it's] the white house.' (Soto Ruiz 1976b:120)

(15) *Paqarin* ***-si*** *hamu -nqa.*
tomorrow -IndE come -3.Sg.Subj.Fut
'[It's said that] He will come tomorrow.'

(16) *Papa -ta* ***-s*** *tarpu -nqaku*
potato -Acc -IndE sow -3.Pl.Subj.Fut
'[It's said that] They will sow potatoes.' (Dedenbach-Salazar Sáenz et al. 2002:105)

(17) *Wasi -n -ta* ***-ch*** *ri -chka -n.*
house -3.Sg.Poss -Acc -Asmp go -Prog -3.Sg.Subj
'He's going to his house [probably].'

(18) *Atuq* ***-cha*** *miku -ru -n.*
fox -Asmp eat -Rptn -3.Sg.Subj
'The fox ate it [probably].' (Dedenbach-Salazar Sáenz et al. 2002:105)

In some varieties within Quechua IIC, the case suffix *-pa* (genitive) is affected by the same rule, it gets shortened to *-p* if preceded by a vowel. In Cuzco Quechua, *-pa* becomes *-q* under these circumstances. In Ayacucho Quechua, genitive case is always marked by *-pa*, the suffix does not undergo any morphophonological changes in this variety. In some Bolivian dialects there are still other allomorphs: *-qpa* following vowels or *-paq* following consonants. These forms can be analyzed as double genitives, my system will not recognize the latter though, as this would lead to a major confusion with the benefactive suffix *-paq*. Nevertheless, the omission of *-paq* as genitive should not result in major problems, as the ambiguous form *-paq* is rather infrequent compared to its allomorphic counterpart *-pa*. The double sequence *-qpa* may also occur to some extent in Cuzco Quechua (sometimes written as *-hpa*) (Cerrón-Palomino 2003:134).

(19) Cochabamba Quechua (Bolivia):

Jaqay	*-qa*	*Antuku*	***-qpa***	*chumpi*	*-n.*
Dem	-Top	Antonio	-Gen	belt	-3.Sg.Poss

'This one is Antonio's belt.'

(Morató Peña 1985:89)

Jesus	***-paq***	*wasi*	*-n*	*tiya*	*-n*	*í.*
Jesus	-Gen	house	-3.Sg.Poss	exist	-3.Sg.Subj	Interj

'Jesus has a house, eh?' (lit. Jesus' house does exist, eh?)

(Morató Peña 1985:90)

(20) Cuzco Quechua (Peru):

Inka	***-q***	*allpa*	*-n*	*ka*	*-sqa.*
Inca	-Gen	land	-3.Sg.Poss	be	-3.Sg.IPst

'The land was of the Incas.'

Chay	*inka*	*-nchis*	***-pa***	*pallay*	*ati*	*-sqa*	*-lla*
Dem	Inca	-1.Pl.Incl.Poss	-Gen	design	be.able	-Perf	-Hon_Aff

-n	*-ta*	*-n*	*kunan*	*ruwa*	*-sha*	*-nchis.*
-3.Poss	-Acc	-DirE	nowadays	make	-Prog	-1.Pl.Incl.Subj

'The same designs our Incas used to make, we make [them] nowadays.'

(Cerrón-Palomino 2003:391-392)

5.3.2. Vowel Change

Vowel change affects only the verbal derivational suffixes *-ku* (reflexive, intensifier), *-yku* (affective), *-rqu* ('repentino', urgency, abruptness) and *-pu* (interpersonal, regressive).

The suffixes *-mu* and *-ku* have complex meanings: The suffix *-ku* yields a reflexive reading to the action. It mainly implies that the action affects the subject itself, and in this function it contrasts with the suffix *-pu*, which indicates that the action affects someone else. Note that *-ku* and *-pu* are neutral regarding the manner of the effect: it may be positive or negative.[10]
-ku has an additional 'enhanced reflexive' reading, it may also imply that the subject performed the action with special personal interest, or with essential emotional

[10] In literature, *-pu* is often referred to as 'benefactive', which, in my opinion, is a misleading label. For this reason, I prefer the term 'interpersonal' used by Dedenbach-Salazar Sáenz et al. (2002).

commitment, for this function I use the label 'intensifier'. *-ku* can also implicate that the action is characteristic, typical for the subject (Dedenbach-Salazar Sáenz et al. 2002:178), (Soto Ruiz 1976b:108).
In combination with verbs of motion, the suffix *-pu* acquires an alternative reading: It implies that the object in motion moves back to the point of origin or that the state the verb describes, lasts for a long time (van de Kerke 1994).
Consider the following examples:

(21) with *-ku*:

Maylla	***-ku***	*-ru*	*-ni*	*-ña*	*-m*	*uya*	*-y*	*-ta*	*-qa.*
wash	-Rflx	-Rptn	-1.Sg.Subj	-Disc	-DirE	face	-1.Sg.Poss	-Acc	-Top

'I've already washed my face.' (reflexive)

(Soto Ruiz 1976b:108)

Chuqllu	*-ta*	*miku*	***-ku***	*-chka*	*-n.*
corn	-Acc	eat	-Int	-Prog	-3.Sg.Subj

'He's eating corn [with commitment].' (intensifier)

(Dedenbach-Salazar Sáenz et al. 2002:178)

Chay	*allqo*	*-qa*	*kachu*	***-ku***	*-n*	*-mi.*
Dem	dog	-Top	bite	-Int	-3.Sg.Subj	-DirE

'This dog bites'. (characterization)

(Soto Ruiz 1976b:108)

(22) with *-pu*:

Sumaq	*pacha*	*-ta*	*ruwa*	***-pu***	*-sqayki.*
beautiful	dress	-Acc	make	-Iprs	-1.Sg>2.Sg.Fut

'I'll make you a beautiful dress.' (interpersonal) (Soto Ruiz 1976b:107)

Valicha	*-qa*	*Lima*	*-ta*	*-s*	*ri*	***-pu***	*-n.*
Valicha	-Top	Lima	-Acc	-IndE	go	-Regr	-3.Sg.Subj

1. 'They say that Valicha has gone to Lima for ever.'
2. 'They say that Valicha has returned to Lima.' (regressive)

(van de Kerke 1994:235)

The suffixes *-ku* and *-pu* both change their final vowel from *u* to *a* under certain circumstances:
-ku is realized as *-ka* before *-chi* (causative) and *-mu* (cislocative, translocative), whereas *-pu* becomes *-pa* if preceding *-ku* or *-mu*. The combination *-paku* (*-pu* and *-ku*) implies that the subject attains a benefit at the expense of someone else, or it may add an occasional, provisional reading to the action (Soto Ruiz 1976b:108).

(23) with *-pu*:

Wasi	*-y*	*-mi*	*tuñi*	*-ru*	*-n,*	*chay*	*-mi*	*puñu*
house	-1.Sg.Poss	-DirE	collapse	-Rptn	-3.Sg.Subj	Dem	-DirE	sleep

-pa	*-ku*	*-chka*	*-ni*	*kay*	*-pi.*
-Iprs	-Rflx	-Prog	-1.Sg.Subj	Dem	-Loc

’My house collapsed, therefore I sleep here (temporarily).’
(Dedenbach-Salazar Sáenz et al. 2002:182)

Yapu	***-pa***	*-mu*	*-y*	*chakra*	*-n*	*-ta.*
plow	-Iprs	-Trs	-2.Sg.Imp	field	-3.Sg.Poss	-Acc

’Go plow his field.’
(Dedenbach-Salazar Sáenz et al. 2002:182)

(24) with *-ku*:

Maylla	***-ka***	*-mu*	*-y!*
wash	-Rflx	-Trs	-2.Sg.Imp

’Go wash yourself!’ (Soto Ruiz 2006:309)

The remaining two suffixes affected by vowel change are *-yku* and *-rqu*, both of them occur in free variation with their allomorphs *-yu* and *-ru*.
Etymologically, *-yku* was a directional suffix expressing ’inwards’. Except for some Quechua I dialects that still use it in this function, this directional suffix has acquired a wide range of new meanings (Cerrón-Palomino 2003:194). In QIIC dialects it implies often some kind of emotional affection, such as compassion, sympathy, concernment, pleasure, solidarity, cordiality, sorrow and much more (Soto Ruiz 2006:351), (Dedenbach-Salazar Sáenz et al. 2002:188), e.g. *rima-*, ’to speak’ vs. *rimayku-*, ’to greet’.
-yku/-yu is realized as *-yka/-ya* if preceding *-mu*, *-pu*, *-ysi* (assistive, indicating

that action was accomplished in supportive manner), *-chi* (causative), *-ri* (inchoative) and *-cha* (verbal diminutive, belittlement, endearment, childishness).[11]

(25)

Kiri	*-sqa*	*maki*	*-n*	*-wan*	*-qa*	*mana*	*-m*	*ima*	*-ta*
hurt	-Perf	hand	-3.Sg.Poss	-Inst	-Top	Neg	-DirE	something	-Acc

-pas	*hapi*	***-yku***	*-n*	*-man*	*-chu.*
-Add	grasp	-Aff	-3.Sg.Subj	-Pot	-Neg

'With a hurt hand, one can not grasp anything.'

(Soto Ruiz 2006:356)

(26)

Qu	***-yka***	*-pu*	*-y,*	*waqa*	*-ru*	*-nqa*	*-m.*
give	-Aff	-Iprs	-2.Sg.Subj.Imp	cry	-Rptn	-3.Sg.Subj.Fut	-DirE

'Give it to him, he's going to cry.'

(Soto Ruiz 1976b:106)

Similarly, the suffix *-rqu/ru* was etymologically a directional suffix expressing 'outwards'. As its counterpart *-yku*, *-rqu/ru* acquired a wide range of meanings, spanning from a recent past in Ancash Quechua (QI), over perfective aspect in Tarma Quechua (also QI), to a sense of abruptness, urgency, haste, surprise, intentness or special personal interest in QIIC dialects (Adelaar 1977), (Cerrón-Palomino 2003:194), (Dedenbach-Salazar Sáenz et al. 2002:187). *-rqu* may also convey the meaning of the action being completed 'just now' (Dedenbach-Salazar Sáenz et al. 2002:187). I decided to use the label 'repentino (Rptn) for this suffix, a term originally introduced by Cusihuamán (1976), as it captures the complex nature this suffix better than other denominations.
-rqu/ru is realized as *-rqa/ra* if preceding *-mu* (cislocative, translocative), *-pu* (interpersonal, regressive), *-ri* (inchoative), *-ysi* (assistive) and *-chi* (causative).

(27)

Ratu	*-lla*	*-m*	*tuku*	***-ru***	*-saq.*
moment	-Hon_Aff	-DirE	finish	-Rptn	-1.Sg.Subj.Fut

'I'll finish [it] quickly.'

(Soto Ruiz 1976b:106)

[11] Although not mentioned in any grammar, it seems that *-yku/-yu* also may change its vowel if preceding *-rqu/-ru*, as the following example from Soto Ruiz (1976a:103) suggests:

Qawa ***-yka*** ***-ra*** *-mu* *-y*!
look -Aff -Rptn -Cis_Trs -2.Sg.Imp
'go and look [after it]!'
(Soto Ruiz 1976a:103)

(28)

Utqay	*qawa*	***-rqa***	*-mu*	*-y*	*uywa*	*-nchik*
quickly	look	-Rptn	-Trs	-2.Sg.Subj.Imp	animal	-1.Pl.Incl.Poss

-kuna	*-ta.*
-Pl	-Acc

'Hurry, go look after our animals!'

(Soto Ruiz 1976b:106)

5.3.3. Epenthesis

Epenthesis affects all possessive suffixes, including the possessor derivation suffix *-yuq*, as well as the case suffixes *-ntin* (inclusive, sociative) and *-nta* (prolative, movement across, through). All these suffixes require the insertion of the epenthetic suffix *-ni* if preceded by a consonant.

(29) with *-ntin*:

Puka	*pikanti*	*-ta*	*kanka*	***-ntin***	*-ta*	*miku*	*-chka*	*-n.*
red	picante	-Acc	roast	-Incl	-Acc	eat	-Prog	-3.Sg.Subj

'He's eating 'puka pikanti' with roast.'

(Dedenbach-Salazar Sáenz et al. 2002:130)

Felipe	*Tomas*	***-ni***	***-ntin***	*ri*	*-chka*	*-nku*	*llaqta*	*-ta.*
Felipe	Tomas	-EP	-Incl	go	-Prog	-3.Pl.Subj	village	-Acc

'Felipe and Tomas walk together to the village.'

(Dedenbach-Salazar Sáenz et al. 2002:130)

(30) with *possessive suffix*:

Llaqta	***-nchik***
village	-1.Pl.Incl.Poss

'Our village'

Yawar	***-ni***	***-nchik***
blood	-EP	-1.Pl.Incl.Poss

'Our blood' (Soto Ruiz 1976b:73)

5.3.4. Implementation of Quechua Morphotactics with `xfst`

Replace Rules All these morphotactical irregularities are captured by `xfst` 're-place rules' in the generator, which are part of the extended regular expression notation `xfst` offers. These rules do not increase the descriptive power of regular expressions, but they represent a simple and straightforward method to define complicated finite-state relations in rule-like notation (Beesley and Karttunen 2003:132).
The most simple example of such a replace rule is `[a -> b]`, a rule that denotes the relation wherein every symbol `a` in the strings of the upper-side language is related to a symbol `b` of the lower-side language. If this network rule is applied downward to the string '`aardvark`', the output is '`bbrdvbrk`': every `a` is replaced by `b` (Beesley and Karttunen 2003:133).
As linguistic rules require more fine-grained mechanisms, it is possible to define contexts for such replace rules: `[a -> b || L _ R]`. `L` denotes the language that specifies the left context, whereas `R` denotes the language that restrains the right context for the replacement. The double-bar operator between the replacement specification and the context expressions indicates that both context expressions must match on the upper side of the relation. Contexts `L` and `R` are optional, if `L` or `R` is empty, the resulting context is treated as the universal language (any possible context is allowed) (Beesley and Karttunen 2003:135). For generation, the tags of regular suffixes are directly translated into their surface form, whereas the tags of those suffixes with irregular surface forms are handled by replace rules, in order to assure that they appear correctly in the generated string, according to the context.
Figure 6 contains the `xfst` replace rule I used to describe the morphotactical behaviour of *-pu* (interpersonal, regressive, see 5.3.2) for my Quechua generator. In `xfst` notation, the curly brackets indicate a concatenation of symbols, so `{ku}` matches the string sequence *ku*. As opposed to this, the quotation marks enclose complex multi-char symbols as in `"+Rflx_Int"`. These multi-char symbols are treated in `xfst` as a strict unit, i.e. as a single symbol. As in standard regular expressions, the | operator functions as logical OR. The rule in Fig. 6 states that the morphological tag `+Rflx_Int` must be replaced by *pa*, if the right context contains one of the suffixes defined in `trigger`, respectively by *pu* in every other context. The operator `~$A` of the second replace rule in Fig. 6 denotes the complement language of `A`, i.e. all strings that are not covered by A (see 5.3.4 for a more detailed explanation).
All instances of vowel deletion and vowel change are handled by such simple replace rules in the generator, in order to assure their correct surface forms.

```
define trigger [{ku}|{mu}|"+Rflx_Int"|"+Cis_Trs"];
define PU ["+Rgr_Iprs" -> {pa} || _ trigger];
define PU2 ["+Rgr_Iprs" -> {pu} || _ ~$[trigger]];
```

Figure 6: `xfst` Replacement Rule for *-pu* (Interpersonal, Regressive)

More Complex Replacement Rules Rules defining contexts on the lower side of the relation are also possible: the rule operator `//` indicates that the left context of the rule must match on the lower side, whereas the rule operator `\\` defines a rule where the right context must match on the lower side (Beesley and Karttunen 2003:169). Another possibility is to define both contexts to match on the lower side with the operator `\/`; I did not use this notation though, therefore I will not describe this notation in more detail.
A further enhancement to replacement rules are left arrows: `B <- A` compiles into a transducer that is the inversion of `A -> B`. Such rules are useful for modifications on the upper side, the 'tags side' of the transducer. I applied left arrow rules only for analysis, not for generation. Rules can be optional, indicated by round brackets: `(->)` respectively `(<-)`.

There is a special ambiguity in Quechua nouns with the combination of some possessive suffixes with the plural marker *-kuna*. If *-kuna* follows the possessive suffix *-yku* (1.Pl.Excl) or *-nku* (3.Pl), the sequence gets shortened to *-y-kuna* respectively *-n-kuna*. Consider that *-y* and *-n* are the corresponding possessive suffixes for first, respectively third person singular.

(31) *allqu* *-y* *-kuna*
dog -1.Poss -Pl
1. 'my dogs'
2. 'our (excl.) dogs'

(Dedenbach-Salazar Sáenz et al. 2002:41)

(32) *allqu* *-n* *-kuna*
dog -3.Poss -Pl
1. 'his dogs'
2. 'their dogs'

(Dedenbach-Salazar Sáenz et al. 2002:41)

A simple analysis would only give the singular reading as a result. In order to

```
define 3PossPl [{nku}["NPers"]["+3.Pl.Poss"]["+NN_SLOT3"]
  (<-) n["NPers"]["+3.Sg.Poss"]["+NN_SLOT3"]
    \\  _ ["--"]{kuna}["Num"]];
```

Figure 7: xfst Optional Left-Arrow Rule for Possessive-Plural Ambiguities

```
xfst[1]: apply up allquykuna
allqu[=perro][NRoot][--]y[NPers][+1.Sg.Poss][+NN_SLOT3]
                [--]kuna[Num][+Pl][+NN_SLOT4]
allqu[=perro][NRoot][--]yku[NPers][+1.Pl.Excl.Poss]
                [+NN_SLOT3][--]kuna[Num][+Pl][+NN_SLOT4]
```

Figure 8: Output from Analyzer for Ambiguous Possessive-Plural Form

add the plural reading, I defined the optional left arrow rule in Fig. 7. This rule specifies that the upper side string 'n["NPers"]["+3.Sg.Poss"]["+NN_SLOT3"]', which is the normal analysis of the third person singular possessive suffix -*n* should be replaced by the string 'nku["NPers"]["+3.Pl.Poss"]["+NN_SLOT3"]' if the right upper side context is ["--"]{kuna}["Num"], which is the analysis of the plural suffix -*kuna*. The operator \\ assures that the right context matches on the upper side, and the round brackets around the operator, (<-) indicate that the replacement is done optionally: Optionality is a must in this case, as the plural reading should be supplied without replacing the original singular reading.
As a result, the output from the analyzer for an ambiguous form like *allquykuna* looks like in Fig. 8.

Context Restriction Rules xfst offers another useful possibility to implement morphophonological information: the context restriction operator =>, that permits to define a fix context for a given string. Context restrictions are regular expressions formed according to the following template:

```
A => L _ R
```

A, L and R are regular expressions denoting languages (not relations). L and R are optional or can subsume various alternative contexts (separated by ';') (Beesley and Karttunen 2003:130). As an example, consider the context restriction rule

```
define PUPA [{pa}["VDeriv"]["+Rgr_Iprs"] =>
                _  ["--"][{ku}|{mu}]["VDeriv"]];
```

Figure 9: `xfst` Context Restriction Rule for *-pu* (Interpersonal, Regressive)

`x => a _ e`: the language denoted by this regular expression includes all strings that do not contain `x`, and additionally all strings where `x` occurs only between `a` on the left, and `e` on the right side (e.g. `a, fish, zzz, axe, ffamaxedss..`) (Beesley and Karttunen 2003:131).

Context restriction rules are useful to assure that the analyzer resolves ambiguities resulting from morphotactical irregularities correctly. Consider Fig. 9 as an illustration of the (simplified) context restriction rule that the analyzer relies on, in order to distinguish between the reading of *-pa* as repetitive or as allomorph of *-pu* (interpersonal, see 5.3.2). The rule allows the reading of the string '`{pa}`' as '`["VDeriv"]["+Rgr_Iprs"]`' only if the right context of the analysis string contains the sequence '`["--"][{ku}|{mu}]["VDeriv"]`'. In other words, *-pa* may only be analyzed as interpersonal or regressive if followed immediately by the suffix *-mu* or *-ku*. Note that this does not affect the reading as repetitive, a word form like *puri -pa -mu -n* is actually ambiguous, as *-pa* can be analyzed as repetitive or as interpersonal (although the latter would be more likely in this context).

On the other hand, the context restriction rule assures that in all other contexts, i.e. not preceding *-mu* or *-ku*, *-pa* will not get the wrong analysis as interpersonal, but only the correct one as repetitive.

Feature Restrictions A further possibility to avoid wrong output in potentially ambiguous cases is the restriction of the feature sequences on the upper side of the network. `xfst` includes a complete bundle of operators that offer a convenient way of defining complex regular expressions, that, compiled into one large regex and concatenated onto the network, facilitate the feature restrictions on the upper side.

This approach is especially advantageous for analysis, as it provides an efficient method to avoid inappropriate tag sequences. I used mainly two such operators: `~$A` and `?$A`. These are actually combinations of two operators:

> `$A` denotes the language or relation obtained by concatenating the universal language, both as a prefix and as a suffix to `A`. For example, `$[a b]` denotes the set of strings such as 'cabbage' that contain at

least one instance of '`ab`' somewhere.

(Beesley and Karttunen 2003:48)

The complement operator `~A` denotes the complement language of `A`, i.e. the set of all strings that are not in `A`, whereas `?A` denotes the language that contains `A` at most once (Beesley and Karttunen 2003:45-51).

As an illustration, consider the suffix sequence *-y -man*. Both suffixes are ambiguous, *-y* is the second person singular marker for imperative mood in finite verbs, but it is also the first person possessive suffix and the nominalizing infinitive suffix. The suffix *-man* on the other hand, has likewise a verbal and a nominal reading: as modality marker (potential) and as case suffix (dative, illative). As illustrated in example 6 in section 5.1, many roots have as well verbal as nominal readings, depending on the morphological context.
These circumstances would lead to an ambiguous interpretation of words like *taki -y -man*:

(33)	*taki*	*-y*	*-man*
	sing/song	-1.Sg.Poss/2.Sg.Subj.Imp/Inf	-Pot/Dat

1. 'My song' (dative)
2. 'to the singing' (dative)
3. '*Sing!' (potential)

The first two interpretations are valid analyses, while the last reading is nonsense (combination of potential and imperative mood). Another illustrative example is the suffix *-manta*, the ablative case suffix, which could also be analyzed as *-man -ta*: *-man* as dative or illative and *-ta* as accusative. Although this segmentation is actually etymologically plausible (Cerrón-Palomino 2003:136-137), it does not make sense for the morphological analysis to split up this sequence. In order to avoid these undesired analyses, the concerned tag sequences can be evaded by regular expressions, as illustrated in Fig. 10.

The other operator introduced above, `?$A`, could theoretically be used for most tags, as most suffixes can appear only once in a word. Nevertheless, this would be over-restrictive, as in most cases, such words will never occur in any text. For the sake of efficiency the focus lies accordingly only on sequences that might misleadingly be interpreted as a repetition of one and the same suffix.
A good example of such a case is the word *ñaña*, 'sister of a woman'. The ambivalent suffix *-ña* (discontinuative, 'already') also occurs as standalone particle. In order to prevent the analyzer from misinterpreting *ñaña* as particle *ña* followed by

```
define YMAN [~$["+2.Sg.Imp"?*"+Pot"]];
define MANTA  [~$["+Dat_Ill"?*"+Acc"]];
```

Figure 10: `xfst` Feature Restrictions

```
define Singularity [$?["+Disc"]];
```

Figure 11: `xfst` Feature Restriction for Discontinuative

the suffix -*ña*, the regular expression in Fig. 11 assures that the tag `+Disc` occurs only once in the upper side string.

5.4. Morphological Analysis vs. Generation

The morphological analyzer takes a complex Quechua word form as input and delivers its morphological components with the according tags. For the Quechua root, a Spanish translation is given instead of a morphology tag; see Table 11 for an example with *wasi* - 'house', which gets translated to the Spanish word *casa*.[12] The generation tool, on the other hand, takes a Spanish root plus morphological tags as input and builds the corresponding Quechua surface form, see Table 12 for an illustration.

5.4.1. Orthography

Requirements for a morphological generation tool differ substantially from the requirements for a morphological analyzer. For the analysis, the ambition is to cover a range of word forms as wide as possible, allowing for different orthographies, variation in suffix sequences and even Spanish loan and foreign words.
For the generation, on the other hand, it is pointless to allow different orthographies or variation in suffix order, as this would only lead to multiple parallel output variants.

There are two major contrasts in Quechua IIC written texts. The first one is a purely dialectal divergence between the Cuzco/Bolivian dialects on one side, and

[12] The morphological analyzer can be tested online at `http://kitt.cl.uzh.ch/kitt/quechua/quechua.html`.

Table 11: Example Morphological Analyzer

Input:	*wasiykikunatas*			
	Morpheme	Suffix Class	Morphology Tag	Slot
Output:	*wasi*	[=casa]	[NRoot]	
	yki	[NPers]	[+2.Sg.Poss]	[+NN_SLOT3]
	kuna	[Num]	[+Pl]	[+NN_SLOT4]
	ta	[Cas]	[+Acc]	[+NN_SLOT5]
	s	[Amb]	[+IndE]	[+AS_SLOT6]
English:	'your houses'(accusative, indirect evidence)			

Table 12: Example Morphological Generation

Input:	hablar	+Rflx_Int	+Cis_Trs	+1.Sg.Obj	+NPst	+2.Sg.Subj	+Def
Output:	*rima*	*-ku*	*-mu*	*-wa*	*-rqa*	*-nki*	*-puni*
				⇓			
				morphophonological rules			
				⇓			
				rimakamuwarqankipunim			
English:	'You definitely (came and) talked to me.'						

Table 13: Glottalization and Aspiration

simple	glottalized	aspirated
ch	ch'	chh
k	k'	kh
p	p'	ph
q	q'	qh
t	t'	th
tanta	*t'anta*	*thanta*
'together'	'bread'	'old' (things)

the Ayacucho/Argentina varieties on the other side: Cuzco/Bolivian Quechua has, like Aymara, a three way distinction of stops, whereas Ayacucho and Argentina Quechua have only simple stops, see Table 13.

Whether an author writes the word for bread as *t'anta* or just *tanta* depends accordingly on the variety he speaks.

The other point of controversy concerns an entirely conventional divergence. Quechua has three phonemic vowels: *a, i, u.* However, *e* and *o* occur as allophones in the proximity of post-velar *q.* While from a linguistic perspective it is evident that the writing of a word should consist only of phonemes, not of allophones, it seems that, maybe due to the influence of Spanish orthography, a lot of people prefer to write *i* as *e*, respectively *u* as *o* according to pronunciation.
The writing of the Quechua vowels is subject of an ongoing debate concerning the official Quechua orthography by the Academia Mayor de la Lengua Quechua (Aca 2005): The Academia uses allophones *e* and *o*, a spelling that is thoroughly rejected by most linguists. As a consequence, this spelling is not prevalent in written Quechua texts, although it is the only official orthography in Peru.

A considerable number of official texts is written according to the unified Southern Quechua orthography defined by Peruvian linguist Cerrón-Palomino (1994), which is generally considered to be linguistically more plausible than the official orthography by the *Academia Mayor de la Lengua Quechua.* Unfortunately, the book (Cerrón-Palomino 1994) was not available, otherwise this would have been the preferred spelling for the word form generation tool. As a provisional solution, I have implemented two `xfst` transducers for generation, one generates Cuzco

Table 14: Lexicon entries

		nominal roots	verbal roots
Generator	Ayacucho	1987	1165
	Cuzco	1981	1164
Analyzer		2572	1956

dialect word forms, while the other one produces Ayacucho Quechua output.

5.4.2. Lexicon

The difference in the morphological structures of the involved languages, Spanish and Quechua, has important consequences for the lexicon of the generation tool: The analyzer's lexicon consists of bare Quechua roots, no complex word forms, whereas the generator's lexicon is 'inverse', constructed from the other side: It is built on the basis of Spanish roots, which in some cases can only be translated as complex Quechua word forms. For example, Spanish has one verb to express 'to die' - *morir* and another one for the causative form 'to kill' - *matar*. Quechua has only a root for the first meaning, *wañu-*, whereas the causative meaning is derived by means of the causative suffix *-chi*, hence 'to kill' - *wañuchi-*. This is in fact a regular correspondence, nevertheless the automatic derivation of such Spanish lexical forms to complex Quechua words during generation is too difficult, as this procedure would require a deep semantic analysis of the Spanish word form. Therefore, correspondences of this type are handled directly by lexical entries.
In addition to such regular derivations, Quechua has a productive method to express non-native concepts through own material, for example the word for 'economy' (Spanish: *economía*), in Quechua becomes *qullqikamay*, consisting of *qullqi* - 'money' and *kamay* - 'create,order'. There is no formal technique to derive these innovative creations automatically, as a consequence, the generation lexicon must contain also complex word forms, not only roots.

In two cases it is preferable, or even necessary to generate 'quechuized' Spanish loan words, by adapting the orthography of the Spanish word to a Quechua conform spelling.
The first one of these two cases concerns highly ambiguous Spanish words that

Table 15: Meanings and translations of Spanish *pasar*

pasar	
-to pass over	*chinpay*
-to happen/occur	*tukuy*
-to pass by (time)	*yalliy*
-to pass somebody something	*hayway*
....	

have no one-to-one correspondence in Quechua and that are also commonly used as loan words. As an example consider the Spanish verb *pasar*, which has a lot of different meanings. Each of those meanings has its own counterpart in Quechua, but there is no single Quechua root that covers the whole range of significations expressed by *pasar*, see Table 15.

The second case of indispensable loan words relates to non-native notions that Quechua has no word for in its lexicon. As mentioned, Quechua has productive method to express non-native concepts through own material, but there are indeed some cases where only Spanish loans are used.
Especially expressions in the domain of catholic religion are usually not translated, also the names of some domestic animals that were introduced by the Spaniards and some words for new artifacts, see Table 16. Generation of these words is handled by a separate `xfst` transducer that rewrites the Spanish words in a 'Quechua-like' spelling.
In some cases, the use of Spanish loans can even lead to subtle distinctions, e.g. the Quechua native term *hampi*, 'healer' is used only for traditional healers, whereas the Spanish loan *midiku*, 'doctor' (Sp. *médico*) is applied exclusively to doctors of orthodox medicine.

Due to these circumstances, the lexicon for the generation tool differs substantially from the collection of roots originally compiled for the analyzer, see Table 14 for numbers.

Table 16: Loan Words

	Spanish	Quechua	English
religious	*casar* *dios* *iglesia*	*casaray* *dyus* *inlisya*	to marry god (christian) church
animals	*vaca* *caballo* *oveja*	*waka* *kaballu* *uwiha*	cow horse sheep
artifacts/ new notions	*carro* *computadora*	*karru* *kumputadura*	car computer

5.5. Generation

The generation process for the evaluation is handled by a Java class, that is capable of generating whole texts, not only single word forms. The default action is to generate a Quechua word out of an input string consisting of a Spanish root followed (optionally) by a sequence of morphological tags.
If generation with the normal `xfst` generator fails due to an unknown Spanish root (root is not in generator's lexicon), a Spanish loan word is generated with the `xfst` loan word generator and the suffixes are attached to this loan word.
This procedure will generate loan words as long as the tags in the input strings are valid and in a correct relative order.

6. Evaluation of the Quechua Morphological Analyzer and Generator

In this section, I will explore coverage and efficiency of my finite-state tools, as stated in the research questions 2 and 3:
2: How efficient are finite-state tools for the analysis and generation of Quechua word forms?
3: Is it possible to cover a satisfactorily wide range of Quechua word forms with finite-state techniques?
In order to find an accurate answer to these questions, I will present a small evaluation. The final conclusions about the results of this inquiry are summarized in section 6.5.

6.1. Method for the Evaluation

The applied method of choice in order to test not only the analyzer, but also the generation tool, is to analyze a given Quechua text with the analyzer and use its output to re-generate the same text.
There are a few important notes to make about the details of this procedure:
In cases of ambiguous Quechua word forms, i.e. if a word form has more than one possible analysis, I arbitrarily take the first analysis from the analyzer's output as input for the generation. The same holds true for generation: If more than one word form was generated, I arbitrarily select the first one, a proceeding that during the second part of the evaluation proved to be extremely error-prone.
In generation, multiple output alternatives occur in fact seldom, this is only the case when a Spanish word has two different meanings and as a consequence two different Quechua correspondences, e.g. Spanish *claro*, 'clear' has two meanings: 1. 'clean, pure' in Quechua *ch'uya* and 2. 'of course' in Quechua *riki*.
In cases where a Spanish word has more than two translations, the Spanish loan word is generated, see section 5.4.2.
Finally, the Unix utility `diff` provided an uncomplicated method to compare the original to the generated texts.

6.2. Texts for the Evaluation

The texts I used for the evaluation exhibit a substantial coverage of different orthographies and dialects within Quechua IIC, yet with the exception of Argentina Quechua, as I did not find a suitable text in this variety.
The following evaluation is based on these eight texts:

1. political newspaper article from 'Diario de Potosí ', ∼ 260 token

2. Acuerdo Nacional Perú, ∼ 8500 token
   (formal meeting from members of the Peruvian government, parties, NGOs and civil persons in order to exchange opinions and strengthen democracy, see Acu (2002))

3. Declaration of Human Rights, ∼ 1500 token
   (Ayacucho Quechua version)[13]

4. Article about literature in Quechua, ∼ 1000 token[14]

5. Cuzco anthem, ∼ 40 token
   (piece of poetry)[15]

6. the story *waka wawamanta* (Itier 2004:140-145), ∼ 530 token

7. a poem, ∼ 200 token

8. bible texts, ∼ 1400 token

Some of these text are translated from Spanish into Quechua, this holds especially true for the Acuerdo Nacional, the Declaration of Human Rights and the bible texts.
In particular the latter two feature a considerable amount of Spanish loan words, and in the case of the Declaration of Human Rights, also a major number of structural copies from Spanish, e.g. Spanish-like passive sentences (Quechua has no proper means of expressing passive).
The bible texts, on the other hand, contain, in addition to the loan words, also a large number of proper names. These proper names, together with the Spanish loan words, reduce the performance of the `xfst` analyzer on these texts drastically.
The analyzer indeed contains a small lexicon with the most common Spanish loan words, and additionally, I run the Spanish Tree Tagger[16] on the Quechua texts before the actual analysis was done. The tree tagger recognizes of course only Spanish words and some proper names that have the original Spanish spelling and no suffixes attached to them. The list of recognized words serves as basis for the automatical generation of a separate lexicon with Spanish foreign words, which I

[13] available from `http://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=quy`
[14] available from `http://celia.cnrs.fr/FichExt/Am/A_25_10.htm`
[15] available from `http://www.runasimi.de/qusquman.htm`
[16] Developed at the University of Stuttgart,
available from `http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/`.

accordingly attached to the original `xfst` transducer.
Following this procedure, my `xfst` transducer was able to recognize at least some of the Spanish loan and foreign words, as well as some of the proper names.
The generation tool, on the contrary, contains no loan words lexicon, as these are produced by an independent transducer, see 5.4.2. Proper names can not be generated at all, respectively they simply undergo transformation into a Quechua conform spelling, like all other unknown roots.

6.3. Most common differences

There are four common 'errors', or rather differences between the original and the generated texts:

Orthography Differences in spelling were by far the most common 'errors'. On one side, these differences are purely orthographical, like the writing of *i* as *e* and *u* as *o*.
On the other side, there is a considerable amount of dialectal variation, e.g. the writing 'bread' as *tanta* (Ayacucho Quechua) or *t'anta* (Cuzco Quechua), and even some dialectal variation in suffix forms, e.g. progressive aspect: *-chka* (Ayacucho Quechua) vs. *-sha* (Cuzco Quechua)

Translation Context As described in section 6.1, there are some cases where a Spanish root has two different meanings, and so two translations are generated. In these cases, I arbitrarily chose the first one, which sometimes lead to translations that are completely wrong in the context they appear.
For example, consider the Spanish word *contar*, which has two meanings: 1. 'to count', in Quechua *yupay* and 2. 'to narrate, to tell', in Quechua *willay.*
As a consequence of my proceeding, the word *willay* in the original text gets analyzed correctly as Spanish *contar*, but the output of the generation is 1. *yupay* and 2. *willay.* The first one is chosen, and so the generated text contains the word *yupay* instead of *willay.*

Semantics In some cases, the generation tool lacks semantic knowledge or world knowledge. This is especially true with kinship terms: Quechua kinship terms imply not only the gender of the person they refer to, as in languages like English, but also the gender of the person they are related to.
Nevertheless, the elaborate, traditional Quechua kinship terms became drastically reduced as the Spaniards introduced christianity and European-style family relations. The original distinctions between woman's relatives and man's relatives,

once important to the extremely complex structures of the Andean society, disappeared almost completely (Adelaar and Muysken 2004:235), (Zuidema 1973).

Consider the following examples:

’child’	→	*churi*	child of a man (father referring to his son/daughter)
		wawa	child of a woman (mother referring to her son/daughter)
’sister’	→	*pani*	sister of a man (man referring to his sister)
		ñaña	sister of a woman (woman referring to her sister)
similar:			
’uncle’	→	*yaya*	brother of father
		kaka	brother of mother

The generation tool does not have the ability to decide which translation is the correct one in a given context, so here again, I arbitrarily chose the first translation, a proceeding that has lead to false results in some cases.

Synonyms/Equivalent translations There are two types of equivalent translations in the generated texts:
As a consequence of omitting Quechua synonyms in the generation lexicon, the generated texts may contain a synonym instead of the original word. For example, the original text may use the word *sinchi*, ’strong, power’, its corresponding analysis being Spanish *fuerte*, the generation tool produces *kallpa* as output.
The second type of equivalent translations concerns allomorphic suffixes, e.g. the alternation in past tense of -*n* `3.Sg.Subj` with the zero morpheme -$\emptyset$. Note that these are not dialectal variations of suffix forms, but allomorphic variations that occur across dialects.

6.4. Results of Evaluation

6.4.1. Results of Analyzer Evaluation

The total number of evaluated tokens was over 13 thousand, consisting of more than 4000 distinct word forms (types), see Table 17. Out of this material, my `xfst` analyzer failed with about 2.15% of the tokens and 4.9% of the types. More than half of these failures where Spanish loan words, another common source for errors were proper nouns and misspelled words. Potential for improvement exists especially in the lexicon, as about 10% of the failures could not be analyzed due to missing lexical entries.
For a small number of words, the analysis failed because the order of the suffixes, respectively their combination was not recognized by the `xfst` transducer. Reduplication of roots is one of those cases, e.g. *hatu-hatun*, as I did not consider

Table 17: Analysis

	Token	Types
total	13629	4353
analyzed	13336 97.85%	4139 95.1%
not analyzed	293 2.15%	214 4.9%
thereof:		
span. loan word	55%	52%
proper noun	15%	17%
typos/errors	13%	15%
unknown root	9%	10%
suffix/root order	8%	6%

this possibility in my analyzer. Reduplication is however a rare phenomenon: it occurred only once in all eight texts.

6.4.2. Results of Generator Evaluation

The total number of generated words was smaller than the original number of words, as the word forms the analyzer failed to analyze could not be used as input for generation, see Table 18. In order to test the performance of the `xfst` generator, the generation process was executed without the fall back option of producing Spanish loan words for unknown lexical items.

Due to the fact that the analyzer had support of the Spanish Tree Tagger, it managed to analyze at least some of the proper nouns in the texts. In contrast, the generator had no means of producing proper names. This circumstance lead to failures in generation for 2.6% of the tokens, respectively 5.1% of the types.

The remaining word forms for which generation was unsuccessful consists of unknown words, in the overwhelming majority of cases these were terms that have no correspondence in Quechua (see Table 16) or, in some cases, Spanish function words like coordinations (*y*, 'and') or subordinations (*si*, 'if').

Table 18: Generation

	Token	Types
total (without failures of analysis)	13336	4139
generated	12729 95.4%	3884 93.9%
not generated	607 4.6%	254 6.1%
proper noun	16 2.6%	13 5.1%
unknown root	591 97.4%	241 94.9%
thereof Span. concepts/function words	543 91.9%	211 87.6%

6.4.3. Comparison Original vs. Generated Texts

About one-third of the word forms in the generated texts differed from the corresponding words in the original texts, see Table 19. In most cases, this was only due to different orthography, respectively dialectal variations. The `xfst` generator is able to produce output in Cuzco or Ayacucho dialect, both in three-vocalic writing. Nevertheless, some of the evaluated texts are written in Bolivian Quechua, whilst others employ allophonic spelling of the Quechua vowels (with *e* and *o*). This resulted in a relatively high number of divergent spellings, in 46.4% of the tokens, respectively 36.3% of the types out of the total number of different word forms.

Furthermore, the generated texts contained a considerable amount of synonyms.

In 10.2% of the divergent tokens, respectively 14.7% of the divergent types, the generated word forms are unsuitable translations in the given context, due to the arbitrary choice of the first analysis for morphologically ambiguous words.

The decision to choose arbitrarily the first generated word form lead to divergences in 8% of the tokens, respectively 13.4% of the types.

A minor number of cases concerns Spanish loan words in the original texts that in the generated texts reappear as native Quechua word forms. An even smaller

Table 19: Original vs. Generated Texts

	Token	Types
total generated:	13336	4353
total diff	3812 28.6%	1425 34.4%
orthography:	46.4%	36.3%
equiv. translation/synonym	29.2%	25%
wrong morph. analysis chosen	10.2%	14.7%
semantics/context	8%	13.4%
original SP, Gener. QUE	5.2%	8.3%
original QUE, Gener. SP	0.6%	1.1%
error in generation	0.4%	1.2%

number exhibits lacking lexicon entries in the generation tool, where the original texts contain Quechua word forms that the generator failed to produce and instead returned a quechuized Spanish loan word.

Finally, one mistake slipped in during the process of testing, concerning the Quechua verb *qallari-*, 'to begin':

qallari- constists of the root *qalla-* and the inchoative suffix *-ri*. Nevertheless, there is a special feature about *qallari-*: The root *qalla-* never occurs without the inchoative suffix. Even so, the analyzer's lexicon contains the bare root, as it should be able to split *qallari-* into its morphological components. The generator's lexicon, on the other hand, contains only the complex form, as the verb 'to begin' should always be generated as *qallari-*, not as *qalla-*.

Due to the proceeding of the test, word forms with *qallari-* became analyzed as *qalla-ri-*, 'begin-inchoative', which consequently was generated as *qallari-ri*, containing the inchoative suffix *ri-* twice.

6.5. Efficiency and Coverage

The finite-state tools implemented with `xfst` proved to be extremely efficient for analysis and generation of Quechua word forms: The lookup for the largest text in my collection[17], the *Acuerdo Nacional* took only 1.14 seconds (speed = 11184 words/second). Finite-state tools are thus capable of analyzing Quechua words at an impressive rate. The proper finite-state generation is comparable in speed if the lookup utility from `xfst` is directly used. Generating the same text, *Acuerdo Nacional*, directly with `xfst` took less than a second. Nevertheless, the java class that I used for the evaluation to handle exceptions (loan words) slows down the generation process considerably: The whole output from the analysis is filled into an array, the complete text is subsequently stepped through, word by word, in order to generate the according word form. Of course this proceeding is not optimal, but as I set it up only for testing, I did not make an effort to improve it with regard to speed.

As for coverage, the analyzer reached a remarkable coverage in my tests, although the lexicon could still be improved. The average of all processed types in the evaluated texts was 95%, while the coverage for the individual texts ranges from ∼90% (bible texts) to 99% (pacha paqariy ulla, poetry). The crucial point for the results of the analyzer is always the number of loan words from Spanish and, to less extent, the number of misspelled words and unknown proper names a given text contains.

To measure the coverage of the generator independently from the performance of the analyzer, is not as straightforward. As I had to rely on the output of the analyzer in order to test the generator, these numbers can not be considered totally irrespective of the analyzer's performance. Nevertheless, the total coverage of types reached was 93%, which is a tolerable result, given the circumstance that the generator is not able to handle Spanish loan words and proper nouns on its own.

As a final conclusion, I dare to say that the coverage of Quechua word forms reached with finite-state methods is more than satisfactory, although of course an enhancement of the lexica and a more elaborated way of dealing with Spanish loan words would lead to even better results.

[17] The original version has 10995 tokens, including all punctuation marks and special symbols.

7. A Spell Checker for Quechua

An obvious enhancement to the finite-state tools presented in the previous sections is spelling correction. The implementation of a Quechua spell checker on the basis of the generator and analyzer is straightforward and can be achieved with no difficulty through the addition of some special rules.
In this section, I aim to examine my research question number 4, formulated in section 1.2:
4: At what expense can a spell checker be implemented from the analysis and generation tools? Is spell checking with finite-state techniques feasible for Quechua? Is it efficient?
I will summarize my answers to this question at the end of this section in 7.3.

7.1. Orthography and Writing Conventions

As the generator is far more restrictive than the analyzer, I decided to use the generator's lower side ('natural language side') to provide the spell checker with the required correct word forms. As stated in section 5.4.1, there is no real 'beyond all doubt' standard orthography for Quechua. Nevertheless, a spell checker relies on some sort of 'gold standard' to check given word forms. In the absence of a better alternative, I decided to implement two separate spell checkers, one for the Cuzco and one for the Ayacucho dialect. For the Cuzco variant, I chose to adhere to a three-vocalic writing which should be consistent with the Academia's orthography for the most part, except for the use of three instead of five vowels. Additionally, I allowed alternative spellings in two cases:

1. In the official orthography, the sequence *n* or *m* followed by *p* or *k* is consequently written as *np*, respectively *nk*, e.g. *llank'a-*, 'to work' or *chinpa*, 'across, opposite'. I allow these forms to be written with *mp*, e.g. *chimpa* and, to less extent, with *mk* as in *llamk'a-*, as these forms are equally common.

2. The Academia's dictionary writes the sequence *l*, respectively *ll* followed by *q* strictly as *lq*, e.g. *qilqa*, 'letter', *qulqi*, 'money' or *salqa*, 'altitude zone (3500-4000m)'. Nevertheless, forms with *llq* are equally common, for this reason I allow alternative writings as in *qillqa*, *qullqi* or *sallqa*.

As for the Ayacucho version, I use the standard writing as in Soto Ruiz (1976a), but with three vocals. In Ayacucho, there are no alternative forms with either *n* or *m* preceding *p*, respectively *k*: only forms with *m* are acceptable, e.g. *llamka-* and *chimpa*. The same holds true for the *l/ll* alternations in front of *q*: in Ayacucho Quechua the correct spelling is always *llq* as in *qillqa*.

My spell checker is not designed to work with other dialects. It could be adapted to the Bolivian dialect as well, but probably not to Argentina Quechua, as I don't have access to detailed dialectal descriptions of this variety.

There is another important point to mention: Incorporation of unspecific objects to form complex verbal roots is non-existent in official orthography, e.g. *uywamichiy* (*uywa* - 'animal', *michiy*, 'to herd') is officially written as *uywa michiy*. Nevertheless, a considerable number of texts contain forms like *uywamichiy*, which is the reason I designed my analyzer to recognize such forms. For spell checking, I adhere to the official, separated writing, which means that a form like *uywamichy* will be corrected to *uywa michiy*.

A similar case represent the nominal compounds, as in *wasimasi* (*wasi* - 'house' and *masi* - 'fellow' → 'neighbour') or *wasipunku* (*wasi* - 'house' and *punku* - 'door'). Almost all instances of nominal compounds are written separately, as in *wasi punku*. Nevertheless, there are some nouns that tend to be attached to other nominal roots in order to build compounds, like *masi*. Indeed, these nouns appear in some grammars as suffixes, an analysis that I reject, as they are clearly standalone nominal roots. Official orthography is not consistent in this point, as formations with *masi*, *pata* ('side'), *runa* ('human being'), *suyu* ('region') and some more, are written in one word, whereas most other compounds are written separately.

As a provisional solution I decided to adhere also in this case to official orthography by the *Academia Mayor de la Lengua Quechua* (Aca 2005): My spell checker will leave compounds with the mentioned roots (*masi,suyu...*) untouched, while it will correct forms like *wasipunku* to *wasi punku*. This is however an unsatisfactory approach, as indeed a large number of toponyms and also proper names of mythological figures consist of exactly such compound forms, written together (e.g. Ayacucho: *aya* - 'soul' and *kuchu* - 'corner' → 'the corner of souls'[18]).

7.2. An Error Metric: Edit Distance

The formal comparison of two strings requires an error metric that facilitates the measuring of how much two strings deviate from each other. The edit distance represents such a metric:

[18] The etymology of this name is not completely clear, yet one theory is that the Incas proceeded so violently when they conquered in the 15th century the empire of the Chanca (at this time the rulers in this region), that the place was given this name.

> 'The **edit distance** between two strings measures the minimum number of unit editing operations of **insertion**, **deletion**, **replacement** of a symbol and **transposition** of adjacent symbols that are necessary to convert one string into another.'
> (Oflazer 1996:74)

The four edit operations *deletion, insertion, replacement* and *transposition* can be implemented as replacement rules (see section 5.3.4). In order to control the maximum number of such operations, a complex symbol like `"<DEL>"` is inserted after each operation, the number of these special tags in a modified word form reveals the edit distance to the original (correct) word form in the transducer.

7.3. Implementation of Edit Operations

Figure 12 shows the implementation of the first three edit operations. In a first step, the symbols, which a Quechua word form may contain, are defined (`letters`). The next section contains the rules for *deletion, insertion* and *replacement.* Deletion and insertion are restricted to contexts other than the beginning of a word. The last section in Fig. 12 contains the feature restriction `editDistance`: A valid word form may contain maximum one tag resulting from one of the edit operations defined above, so in this example the allowed edit distance is 1. In order to increase the maximum edit distance, the same feature restriction rule can be optionally attached to the existing one (the operator `.o.` stands for composition): `[ [$?[editOperations]] .o. ([$?[editOperations]]) ]`.

The transposition of two adjacent symbols is missing in Fig. 12. As for every possible interchangement a single rule is required, I automatically generated a separate file containing all these interchangement rules and tried to append them to the transducer. Unfortunately, this results in a network that is too large to compile.
A trade-off less intense for compilation is to allow transposition only for adjacent letters on the (Latin American) keyboard. Still the resulting network is very large and a lookup therefore takes a long time.

The actual spell checking is handled by a simple Perl script that in a first step calls `xfst` to lookup the word form in a recognizer, containing only the lower side of a restrictive analyzer. Note that this is not the same analyzer used for normal morphological analysis: The normal analyzer is designed to cover a wide range of possible spellings and dialect variations, whereas the analyzer for spell checking should recognize only word forms written correctly in the defined standard writing.

```
define Voc [a|i|u];
define Cons [b|c|d|f|g|h|j|k|l|m|n|ñ|p|q|r|s|t|v|w|y|z|’];
define Letters [Voc|Cons];

######################################################################

define DEL [Letters (->) "<DEL>" || .#. ?+ _ ];
define INSERTION [[..] (->) Letters "<INSERT>"|| .#. ?+ _ ];
define REPLACE [Letters (->) Letters "<REPLACE>"];

######################################################################

define editOperations ["<DEL>"|"<INSERT>"|"<REPLACE>"|"<INTERCHANGE>"];

define editDistance [$?[editOperations]];

define cleanup [ "<DEL>" ->""] .o.[ "<INSERT>" -> ""]
               .o. [ "<REPLACE>" -> ""]  .o. ["<INTERCHANGE>" -> ""];
```

Figure 12: `xfst` Edit Operations

I used the lower side of the Ayacucho, respectively Cuzco Quechua generator for this purpose, as the generator is far more restrictive than the original analyzer. The lexicon for spell checking is not the same as the normal generator's lexicon, more accurately, it is a revised version of the analyzer's lexicon (see section 5.4.2 for the differences between the two lexica).

In order to speed spelling correction up for light cases (orthographic variations, morphotactically wrong suffix forms), which I expect to occur more frequently than other misspellings, I compiled an additional analyzer specially for these cases. This analyzer is basically the same as the one I use to check if a word form is correct, but not as restrictive: alternative suffix forms and 5-vocalic spellings are tolerated and corrected, but no edit operations are allowed[19]. Inserted preliminary to the actual spell checkers this additional transducer increases speed for the most simple (and most common) cases of misspelled words.

The procedure for spell checking is as follows (see the UML activity diagram in Fig. 16 as an illustration):

If the recognizer accepts the word form, it is a correctly spelled word and the spell check is complete. If the recognizer rejects the word form, the Perl script calls the enhanced analyzer, if the word form is correct (edit distance = 0), but written in a deviating orthography, the enhanced analyzer will return the morphological analysis to the Perl script, which in turn calls the generator in order to produce the spelling corrections. If the enhanced analyzer does not recognize the word form, the Perl script calls the spell check transducer compiled with edit distance 1. If the word form lies within edit distance 1 from a correct word form, all the possible analyses will be returned to the Perl script, which in turn feeds those as input to the normal generator. The word forms which the generator produces with this input are finally printed out as spell checked suggestions. If this fails, the Perl script will call a second spell checker, compiled to recognize words within edit distance 2. The rest of the procedure is identical as before with edit distance 1: the Perl script uses all possible analyses as input for the generator, which finally produces the correction suggestions. In order to avoid errors resulting from synonyms or contextual errors as described in section 6.3, the spell checker's generator uses the same lexicon as the spell checker's analyzer. The lexica for both spell checkers are cleaned up versions from the original analyzer's lexicon: synonyms or other words with the same Spanish translation (e.g. Sp. *hijo* - 'son,daughter' as *wawa* or *churi*, see 6.3) have numbers and so get unambiguously generated to the originally intended input word. See Figures 13 - 15 for examples.

[19] I do not count alternative suffix forms or the writing of *e* instead of *i*, respectively *o* instead of *u* as edit operations: As my spell checker is restricted to forms within the edit distance of 2, words written in 5-vocalic orthography would fall beyond this threshold if they contain more than two instances of *e/o*, and so could not be corrected.

```
Edit Distance = 0:

1. allquchallaykis (correct), time 0m0.127s:

$ echo 'allquchallaykis' | perl spellcheck.pl
ok

2. allqosi, o instead of u, -si after vowel should be -s, time 0m0.121s:

$ echo 'allqosi' | perl spellcheck.pl
        allqus

3. 'uywamichiy', incorporation: uywa (animal) -michiy (herd), time 0m0.392s

$ echo 'uywamichiy' | perl spellcheck.pl
        uywa michiy
```

Figure 13: Spell Check for Edit Distance = 0

In a first attempt, I allowed within the edit operation of replacement each letter to be replaced by any other letter. The resulting network compiles, but is huge, and therefore slow: a lookup with a misspelled word of edit distance 2 took more than three minutes on a Dual-Core AMD 64bit work station.
Considering this circumstance, I implemented a second version where I limited the possible replacement of letters to adjacent letters on the Latin American keyboard, just as described above for the transposition. The resulting network is much faster, (15 seconds for the same word form as before), but still not fast enough for real usage.

To conclude this section about the implementation of a spell checker for Quechua with finite-state tools, the research question 4 can now be answered.
The implementation of a Quechua spell checker via analyzer and generator can be achieved with little effort. Nevertheless, the resulting networks are large and therefore not only heavy to compile, but also too slow to be used in a real application. Simplification of the replacement and interchangement edit operations led to a notable improvement in performance, but for real usage the spell checker is still not efficient enough.

```
Edit Distance = 1:

with allqu (dog):

1. Insertion:
alllqu (l inserted), time 0m1.908s

$ echo 'alllqu' | perl spellcheck.pl
        allqu

2. Deletion:
allq (u deleted), time  0m1.963s:

$ echo 'allq' | perl spellcheck.pl
        allaq
        alliq
        allqa
        allqu

with wasi (house):

3. Interchangement:
wsai (interchangement of a and s), time 0m1.863s:

$ echo 'wsai' | perl spellcheck.pl
        wasi

4. Replacement:
wadi (replacement of s with d), time 0m1.958s:

echo 'wadi' | perl spellcheck.pl
        wawi
        wasi
        wari
```

Figure 14: Spell Check for Edit Distance = 1

```
Edit Distance = 2:

with wasiykichismi (wasi ’house’, -ykichis ’2.Pl.Poss’, -mi ’DirE’
                    = ’your (Pl) house’):

1. Deletion of s, Replacement of i with d, 0m15.930s:
echo ’waiykichismd’ | perl spellcheck.pl
        wawiykichismi
        waqiykichismi
        wakiykichismi
        wasiykichismi
        watiykichismi
        waniykichismi
        wariykichismi

2. Interchangement of a and s, Insertion of i, 0m15.292s:
echo ’wsaiykichiismi’ | perl spellcheck.pl
        wasiykichismi
```

Figure 15: Spell Check for Edit Distance = 2

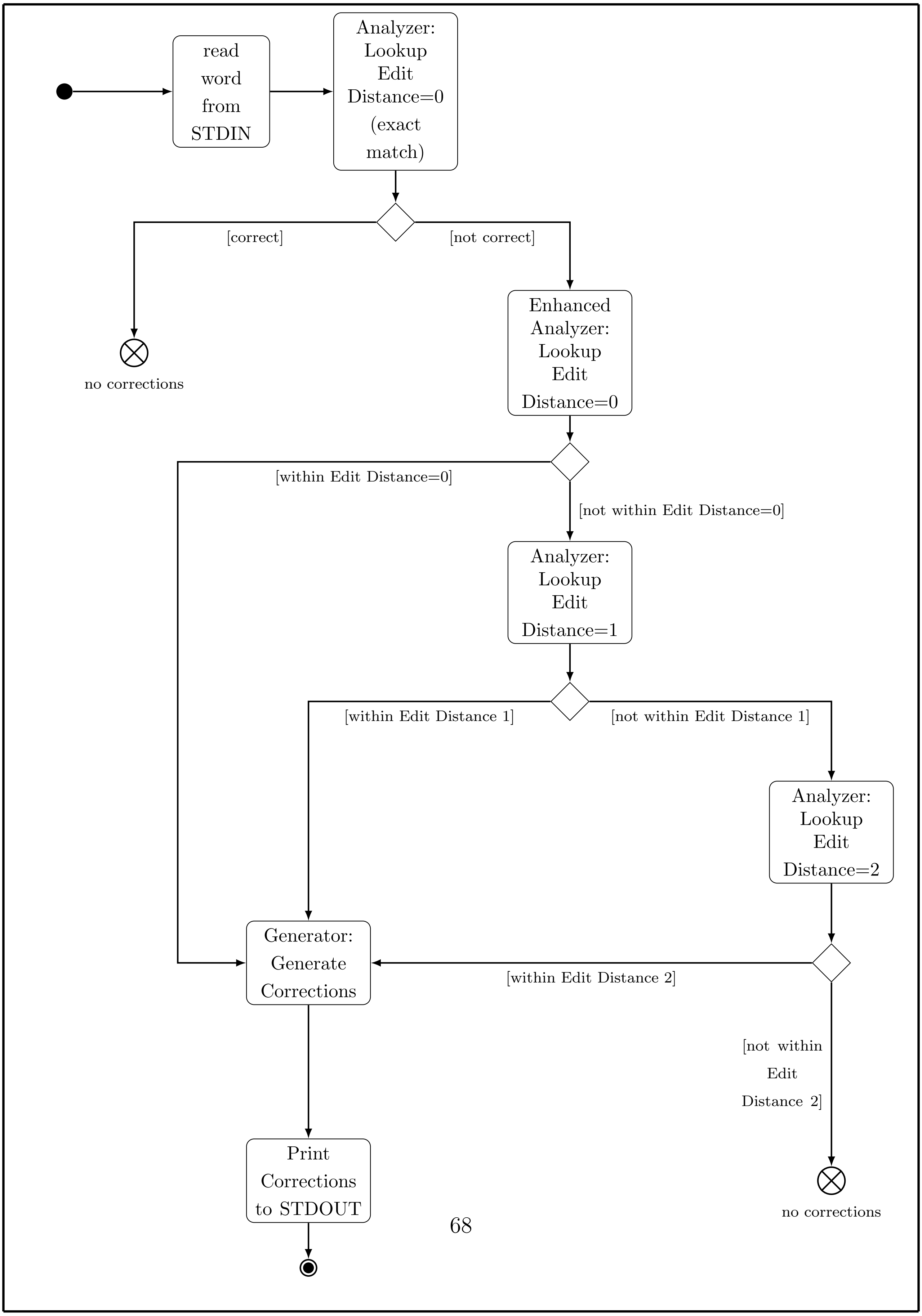

read word from STDIN
Analyzer: Lookup Edit Distance=0 (exact match)
[correct]
[not correct]
no corrections
Enhanced Analyzer: Lookup Edit Distance=0
[within Edit Distance=0]
[not within Edit Distance=0]
Analyzer: Lookup Edit Distance=1
[within Edit Distance 1]
[not within Edit Distance 1]
Analyzer: Lookup Edit Distance=2
Generator: Generate Corrections
[within Edit Distance 2]
[not within Edit Distance 2]
Print Corrections to STDOUT
no corrections

Figure 16: Spell Checker - UML Activity Diagram

8. AntiMorpho - Another Quechua Morphology Tool

AntiMorpho is a morphology tool capable of analyzing and generating word forms from Spanish and Quechua, developed by M. Gasser at Indiana University[20](Gasser 2008). I have examined only the Quechua part of this system (version 1.1).

The kernel of the AntiMorpho system consists of finite-state transducers augmented with grammatical constraints in the form of feature structure descriptions. Additionally, AntiMorpho has a user-friendly interface written in Python. The indigenous language tools included in AntiMorpho are especially designed for application in computer-assisted language learning (CALL) contexts. Due to this circumstance, the Quechua analyzer and generator do not cover as many word forms as my system, but in turn deliver a very detailed analysis. The meaning of the roots is rendered by various Spanish translations, although segmentation into root and suffixes is not always made. See Fig. 17 for an example analysis of the Quechua word form *yachachikuq*:

(34)

Yacha	*-chi*	*-ku*	*-q*
know	-Caus	-Rflx	-Ag

'Learner, Student'
lit. 'the one who makes himself know'

The actual root is *yacha-*, yet AntiMorpho analyses *yachachi* or even *yachachiku* as one root. This proceeding indeed makes sense in a CALL application context, as the Spanish translations for the derived forms differ substantially, and it may be more convenient for a language learner to be presented with full translations instead of a detailed morphological analysis.
From a linguistic point of view, there are some major discrepancies concerning some of the labels, e.g. the past forms of indirect evidentiality, formed with the suffix *-sqa*, bear the label "pluscuamperfecto" (pluperfect). In Andean Spanish the pluperfect has acquired an evidential quality (Escobar 1997), probably due to the influence of Quechua and Aymara. Consider the following example as an illustration of the evidential use of the Spanish pluperfect in Andean Spanish:

(35)

...	*según*	*dice*	*que*	***había***	***aparecido***	*por*	*ahí*	*...*	*dos*
...	according.to	say	that	had	appeared	by	there	...	two

señores	*una*	*señora*	*y*	*un*	*señor*	
people	a	lady	and	a	man	

[20] available from `https://www.cs.indiana.edu/~gasser/software.html`

’as it is said that [there] had appeared around there ... two people a lady and a man’

(Escobar 1997:865)

Nevertheless, the label ”pluperfect” on a past tense form expressing indirect evidentiality may be misleading to Spanish speakers of other dialects, as the Quechua form with *-sqa* has nothing to do with the category of pluperfect in the traditional linguistic sense: A finite Quechua verb form with *-sqa* refers to an action in the past the speaker has not directly experienced, but instead come to know through someone else. The pluperfect, on the contrary, establishes a temporal relation to some reference point in the past, indicating that the denoted action took place before that reference point.
Another example for an inappropriate label is ’augmentativo’ for the nominal suffix *-sapa* (’multi-possessor’), consider the following example:

(36)	*wasi*	*-su*		*wasi*	*-sapa*
	house	-Aug		house	-MPoss
	’big house’		vs.	’owner of many houses’	

-sapa is not an augmentative suffix, it derives the possessor of a noun, or, more accurately ’the possessor of many’.

In addition to morphological analysis, AntiMorpho offers the possibility to generate word forms. Due to the fact that the input for generation requires the specification of grammatical features in AntiMorpho specific notation, which I am not familiar with, I renounced to test the generator and focused my attention to the comparison of the AntiMorpho analyzer and my Quechua analyzer.
Some important notes have to be made about the AntiMorpho Quechua analyzer:

- input needs to be in Cuzco dialect
- AntiMorpho recognizes only three-vocalic writing
- input has to be in small letters only
- special characters and punctuation marks are not recognized
- AntiMorpho does not handle numbers
- very limited handling of Spanish loan words

```
>>> l3.anal_word('qu','yachachikuq')
Palabra: yachachikuq
CG: substantivo, raíz: <yachachikuq>
 español: el que se hace enseñar, alumno, el que aprenden, estudiante
CG: agentivo, raíz: <yachachi>
 español: educar, enterar, diciplinar, manifestar, demostrar, capacitar,
 instruir, hacer saber, enseñar, anunciar, explicar, acostumbrar
CG: agentivo, raíz: <yachachiku>
 español: instruir, cualificar
CG: agentivo, raíz: <yachachi>
 español: educar, enterar, diciplinar, manifestar, demostrar, capacitar,
 instruir, hacer saber, enseñar, anunciar, explicar, acostumbrar
 Derivación: limitativo
```

Figure 17: Output of AntiMorpho for the word *yachachikuq* - 'Learner,Student'

In order to compare my Quechua analyzer and the AntiMorpho analyzer, I scanned the internet for a new text according to the specified characteristics (Cuzco dialect, three-vocalic writing, low number of Spanish words). I decided to use parts of the text "modulo 1" from the charity organization CARE's homepage[21]. This text contains amongst others also the declaration of the human rights. I removed this part, as I had already used the declaration of human rights for the evaluation of my own system. As the AntiMorpho analyzer does not handle special characters, punctuation marks, upper-case letters and numbers, I performed two evaluations, one on the original text and another one on a "cleaned" version of the original text, containing no upper-case letters, numbers, punctuation marks and special characters at all. Table 20 contains the results of this evaluation.

[21] `http://www.care.org.pe/`, the text is available on `http://www.care.org.pe/pdfs/cinfo/libro/EDU_003_modquechu.pdf`

Table 20: ANTIMORPHO vs. my analyzer

			ANTIMORPHO analyzer	my analyzer
Original Text:				
total tokens:	4344	analyzed:	2883	4163
		%	66.4%	95.8%
		failed:	1461	181
		%	33.6%	4.2%
total types:	1849	analyzed:	1046	1705
		%	56.6%	92.2%
		failed:	803	144
		%	43.4%	7.8%
Cleaned Text:				
total tokens:	3009	analyzed:	2416	2828
		%	80.3%	94%
		failed:	593	181
		%	19.7%	6%
total types:	1581	analyzed:	1244	1452
		%	78.7%	91.8%
		failed:	337	129
		%	21.3%	8.2%

9. Morphology Systems for Similar Languages

In this section, I will search for an answer to my research question number 5 from section 1.2:
5: How does the effort for the implementation of Quechua finite-state tools compare to analogous systems that have been implemented for other agglutinative languages like Turkish or Finnish? How many rules does a Quechua system need? More or less than a system for Turkish?
I will summarize my conclusions about this point at the end of this section in 9.3.

9.1. Aymara Morphological Analysis and Generation

The term *Aymara* is actually ambiguous: it is used for the language family as a whole, as well as for the largest and best known member of this family.
The Aymaran languages are divided into two branches, a northern and a southern branch, which are separated by a considerable geographical distance (Adelaar and Muysken 2004:264). The only surviving languages of the northern branch are the highly endangered Jaq'aru (number of speakers: 725) and the nearly extinct Kawki (number of speakers: 11)(Adelaar and Muysken 2004:612). Consider the fact that these numbers date back to a book originally published 10 years ago.
The language Aymara itself is the unique remaining member of the southern branch (also Aimara, Aymará), which is spoken in the Peruvian and Bolivian Altiplano and also in the North of Chile (see Fig. 1, yellow region). It outranks the other Aymaran languages by large in number of speakers (> 2 millions)(Adelaar and Muysken 2004:612).

The morphology tool implemented 'in an exercise that got out of control' by Beesley (2003) relies heavily on the Aymara grammar by Hardman et al. (1988) and Hardman (2001), which describe the Aymara dialect of the department of La Paz in Bolivia. An original prototype was implemented already in 1988/89, ten years later, in 1998 the system was completely rewritten using Xerox Finite State Technology(Beesley 2003:20).
The lexicon for the system is written completely in `XML`, each entry for a root contains the corresponding English and Spanish translations and word class glosses in both languages. The lexicon comprises 360 roots (Beesley 2003:20).
Suffixes are stored in `XML` files, with Spanish and English glosses. A Perl script finally 'translates' the `XML` dictionaries into the `lexc` language, which in turn is compiled into a finite-state transducer.

Aymara is an agglutinative language, and its structure is remarkably similar to Quechua for the most part. Nevertheless, the morphophonological processes that

take place in word formation are far more complex than in Quechua.
Aymara syllables are always closed; this circumstance implies that a given suffix, attached to a root or another suffix, always directly follows a vocal. Some Aymara suffixes exhibit a special feature concerning the preceding vowel. More precisely, some suffixes have an inherent tendency to delete or lengthen the preceding vowel, whilst others have no effect on the surface form of the preceding morpheme.
Additionally, 'vowels at the end of suffixes must be marked as strong (resisting deletion), weak (deleting if *any* suffix follows), or neutral (being deleted, lengthened or left alone depending on the morphophonology of the suffix that follows)' (Beesley 2003:21),(Beesley 2000). Other morphophonological changes include assimilations, optional shortening of lengthened vowels and methathesis (Beesley 2003:22).

Due to the small lexical resources of the system, many word forms will not be recognized. As a fall back option, the Aymara system uses a morphological 'guesser': an auxiliary transducer that attempts to detect all possible analyses of a word form based on a phonologically possible Aymara root (Beesley 2003:24).

Besides the complexity of the language itself, a morphology tool for Aymara has to deal with challenges similar to my Quechua system:

- Orthography of a given text may differ from the official standard orthography (*Alfabeto Único*)
- Texts will inevitably contain Spanish loan words, consequently a morphological analyzer needs an instrument to handle those
- Gaps and fuzzy areas in linguistic descriptions may impede a 'complete' implementation of grammatical details (note that Aymara has received less attention from linguists than Quechua, publications are not as numerous)

Nevertheless, there are more benefits of an Aymara morphology tool than a modest contribution to the promotion of this Andean language, as some of the insights gained during implementation 'flow back to improve the more traditional paper grammars' (Beesley 2003:26),(Beesley 2000).

9.2. Turkish Morphological Analysis

Turkish represents, with its approximately 50 million speakers[22], the largest member of the Turkic language family. Computational work on Turkish may offer interesting insights for my own system, as it has many structural similarities with

[22] According to `http://www.ethnologue.com/show_language.asp?code=tur`, July 2010.

Quechua: Turkish is a strongly agglutinative, suffixing language, it uses nominalization to build subordinated clauses, has double marked possessive constructions, and it also marks evidentiality in past tense verb forms (Bußmann 2002:714-715). The main difference in word structure, as compared to Quechua, lies in the complex morphophonological mechanisms that take place in the process of Turkish word formation: two-dimensional vowel harmony, vowel deletion, and context-sensitive realizations of certain consonants (Oflazer 1994:4-5).
Oflazer (1994) describes the implementation of a two-level morphological analyzer for Turkish with the `PC-KIMMO` environment. This original analyzer was later reimplemented with `xfst`, as described in Oflazer (1996).
The Turkish morphology system analyzes a given Turkish lexical form into a sequence of feature-value tuples, as opposed to the string of morpheme glosses my Quechua system produces (Oflazer 1996:81).

According to Clements and Sezer (1982:215-216), 'there are two well defined types of word harmony systems. In symmetrical systems, roots do not alternate, and alternating affixes agree with the category of the nearest non-alternating vowel. In asymmetrical systems, both roots and affixes alternate; moreover, alternating forms assimilate to a single ("dominant") value of the harmony category if and only if a non-alternating morpheme of that value appears in the word.'

As the complex and manifold morphotactics of the Turkish language represent a special challenge for computational processing, I will describe these phenomena in more detail. Turkish vowel harmony, in conformity with Clements' definition, is symmetrical, and it is two-dimensional: it consists actually of two intersecting vowel harmony systems, one involving the feature of backness (`[±front]`) and the other involving the feature of roundness (`[±rounded]`) (Clements and Sezer 1982:215). All vowels in a Turkish word form agree in backness, as suffix vowels alternate according to the category of the root vowel. On the other hand, only the high vowels agree in rounding with the preceding vowel, whether high or not; non-high vowels show no alternation in rounding (Clements and Sezer 1982:215). See Table 21 for a detailed listing of the Turkish vowels according to the relevant categories. Table 22 illustrates some examples of Turkish word formation in conformity with vowel harmony.

The intrusion and consequent assimilation of a large number of loan words from Arabic, Persian and to a lesser extent from other languages, has resulted in word forms which do not adhere to the above mentioned principles of vowel harmony. Moreover, orthography of loan words may diverge from their actual pronunciation, a circumstance that may also lead to 'exceptional' vowel sequences in a given word (Oflazer and Inkelas 2006).

Table 21: Turkish Vowels

	`[+front]`		`[-front]`	
	`[+rounded]`	`[-rounded]`	`[+rounded]`	`[-rounded]`
High	ü	i	u	ı
Low	ö	e	o	a

Table 22: Turkish Vowel Harmony (Clements and Sezer 1982:216)

	Nom.Sg	Gen.Sg	Nom.Pl	Gen.Pl
		-Vn	-lVr	-lVr-Vn
’rope’	*ip*	*ip-in*	*ip-ler*	*ip-ler-in*
’girl’	*kız*	*kız-ın*	*kız-lar*	*kız-lar-ın*
’face’	*yüz*	*yüz-ün*	*yüz-ler*	*yüz-ler-in*
’stamp’	*pul*	*pul-un*	*pul-lar*	*pul-lar-ın*
’hand’	*el*	*el-in*	*el-ler*	*el-ler-in*
’stalk’	*sap*	*sap-ın*	*sap-lar*	*sap-lar-ın*

Another challenging morphophonological process is the voicing assimilation of certain consonants, which, depending on their context, have voiceless or voiced surface forms (Oflazer 1994:5).
All these special morphotactical realizations need to be covered by a set of two-level rules. The Turkish system described contains 31 two-level rules that implement the mentioned morphotactic phenomena, such as vowel harmony and consonant changes across morpheme boundaries. Additionally it comprehends about 150 rules that assure the accurate realization of the complex morphotactics by enforcing long-distance feature sequencing and co-occurrence constraints (Oflazer 1996:81). Note that this extensive collection of two-level rules outnumbers by large the small set of rules implemented in my Quechua tools.[23]
In addition to this extensive handling of the complex morphotactics, the Turkish morphology system comprehends a number of separate lexicons for different word classes and some special cases (about 28,000 root entries)(Oflazer 1996:80).

Oflazer (1996) describes the implementation of an error-tolerant spell-checker for Turkish with `xfst`. Error-tolerant recognition allows for the identification of strings that deviate mildly from any string in the underlying finite-state recognizer. This approach requires an error metric that facilitates the measuring of how much two strings deviate from each other. The edit distance represents exactly such a metric (see section 7).

Most spell checkers for languages like English rely heavily on large lists of valid word forms. For agglutinative languages like Turkish or Quechua, this approach is not viable, as the number of possible word forms is far too large. Therefore, the application of finite-state methods to spell correction in Turkish may provide valuable insights for Quechua spell checking.

The original Turkish analyzer provided the basis for the spell checker, yet it had to be simplified, as the generation of glosses and the possibly multiple analyses of a surface form are redundant in spell checking. For this reason, the authors removed all empty transitions and analyses with the same surface form from the original transducer via the Xerox tool `isfm`. The resulting recognizer has only surface symbols (Oflazer 1996:86-87).
I will not go into the details of the algorithm for error-tolerant finite-state recognition described in Oflazer (1996), but the comparative evaluation of the author reveals an average correction time comparable to word list based spell checking for languages like English, Dutch, German and even Finnish. Especially the latter provides a good illustration why a finite-state spell checker for morphologically complex languages is superior to a word list based approach: the average correc-

[23] The analyzer contains 7 rules, the generator contains 85 rules.

tion time for Finnish misspelled word forms with the word list method ranges up until nearly three times as long (with a threshold of edit distance=1) as the average correction time for Turkish with the finite-state method (Oflazer 1996:85-86).

9.3. Effort of Implementation: A Comparison to Similar Systems

As for the effort, a system for Aymara has to deal with the same challenges as a system for Quechua: Lack of linguistic descriptions that are detailed enough for the implementation of a morphology system, scarce written text resources, divergent orthographies, respectively common ”misspellings” as compared to the official orthography. These difficulties, especially the poor availability of written texts and linguistic descriptions, are probably even more pronounced for Aymara than for Quechua.

The level of complexity of a morphology tool for Aymara is comparable to the one for my Quechua systems, although Aymara has a slightly more elaborated morphophonology that requires a sophisticated set of rules to assure that word forms are treated correctly.

As for Turkish, the initial situation is completely different: Availability of written texts and comprehensive grammars and lexica is not a problem, nor are there competing orthographies. On the other hand, Turkish has a complex two-dimensional vowel harmony system, and the effort to cover all the special morphophonological processes that take place in word formation is far beyond the requirements for a language like Quechua.

As a consequence, it is safe to say that the effort in building a morphology system for Quechua with finite-state methods is comparable to languages like Aymara and Turkish: A system for Aymara faces the same obstacles and might even be slightly more complicated, whereas a Turkish has the advantage of good resources, while it needs to deal with a complex morphophonology. Given these conditions, I estimate the effort for a Turkish system to be more or less equal to my Quechua system. Of course the effort for the two systems is unevenly distributed: While the main effort in Quechua lies on the detection of the exact word formation processes (suffix combinations and restrictions, types of roots, etc), in Turkish, these processes are already described in grammars, yet its complex morphophonology needs considerably more attention for a computational approach.

10. Conclusions

I have presented the implementation of a comprehensive finite-state analyzer as well as a generator for the Southern Quechua dialects. Based on these tools, I implemented, with little effort, two spell checkers: one for Cuzco Quechua and one for Ayacucho Quechua.

The main challenge for these achievements was to figure out the linguistic, especially the exact morphotactical details needed for analysis and generation. The lack of linguistic descriptions that are accurate enough for a computational approach, as well as the absence of large, reliable lexica is what makes working with an indigenous language like Quechua so demanding, as compared to the generally well-described languages in Europe.

Even given these adverse circumstances, the finite-state approach proved to be perfectly well suited to capture the characteristics of Quechua morphology. The analyzer reaches a satisfactory coverage: 95% with 8 texts of completely different domains, reaching from legal texts to poetry and extracts from the bible. An additional advantage is the high efficiency of finite-state networks, that allows to process even large texts within a few seconds.

10.1. Resolved Research Questions - A Summary

1: What kind of new linguistic insights can be achieved from an in-depth computational linguistic examination of Quechua word structures?

Quechua IIC morphotactics have been described in detail only for specific local varieties, yet the finite-state approach requires a level of generalization and abstraction that can not be found in any conventional grammar: The partitioning into slots in the nominal and verbal structures, as well as the different sub-types of Quechua roots and their particular morphotactical restrictions represent new knowledge about Quechua IIC.

2: How efficient are finite-state tools for the analysis and generation of Quechua word forms?

The efficiency in analysis and generation of Quechua word forms reached by the finite-state tools is outstanding: The `xfst` tools are capable of handling even large input texts within a matter of seconds. An additional advantage of the finite-state approach is its simplicity, as it allows for a straightforward implementation once the linguistic facts are established.

3: Is it possible to cover a satisfactorily wide range of Quechua word forms with finite-state techniques?

The small evaluation presented in section 6 shows that finite-state techniques are ideally suited to cover a wide range of possible word forms in a morphologically rich agglutinative language like Quechua. Nevertheless, the major challenge to every morphology system for Quechua are the numerous loan words, which make a 100% coverage appear highly unrealistic.

4: At what expense can a spell checker be implemented from the analysis and generation tools? Is spell checking with finite-state techniques feasible for Quechua? Is it efficient?
The implementation of a Quechua spell checker on basis of the `xfst` analyzer and generator can be achieved with little effort. Minor corrections of deviating orthographies or morphotactically wrong suffix forms is extremely efficient. Nevertheless, correction of misspelled words within an edit distance of two is too slow to be applied in a real application, as the enhanced networks are extremely large.

5: How does the effort for the implementation of Quechua finite-state tools compare to analogous systems that have been implemented for other agglutinative languages like Turkish or Finnish? How many rules does a Quechua system need? More or less than a system for Turkish?
The effort in building a morphology system for Quechua with finite-state methods is comparable to languages like Aymara and Turkish, although the main effort is unequally distributed: The main challenge for the implementation of morphology systems in Quechua and Aymara is the disentanglement of the exact morphotactics. In Turkish, on the contrary, these processes are already described in grammars, and the main challenge lies in the correct handling of the complex morphophonology. As a consequence, a system for Turkish requires a substantially larger number of rules than a system for Quechua.

10.2. Open Research Questions

The last two of my research questions presented in section 1.2 have remained unanswered so far:
6: What are the possible application contexts for Quechua morphology tools?
7: Can natural language processing support native minority languages in their endangered usage by increasing their social prestige and awakening more persons' interest?

As for the first question, at least two application contexts have been described in this thesis: Spell checking and support for language learners (CALL, see section 8 about L^{3}MORPHO). In addition to these applications, the Quechua finite-state tools may be useful for preprocessing tasks, such as lemmatization or stemming. Furthermore, the Quechua analyzer could easily be converted into a tokenizer with

a few additional tags that keep track of the actual state: If the machine "knows" its internal state when it reaches the final state for a given word form, then it is straightforward to provide the information if an analyzed string is a nominal or a verbal word form.
The Quechua morphology tools might also be useful in machine translation contexts, indeed the generator will be applied in the upcoming SQUOIA project (see next section for more details).
I actually plan to rewrite all my tools using *foma*, as described in Hulden (2009), instead of `xfst`, for the following reasons:

1. *foma* has a built-in function to define a threshold of edit operations for a given network. Due to this advantage, spell checking should be considerably faster than with `xfst`.

2. *foma* is open source, and would therefore allow me to make my tools freely available to everyone who is interested.

If spell checking with *foma*, as expected, reaches a reasonable speed, my spell checker could be used in a real application.
A further possible enhancement would be "rewrite" tools, to adapt the writing of a given text to another dialect, e.g. 'translate' an Ayacucho Quechua text to Cuzco Quechua, or even Bolivian Quechua.

As for the last research question, number 7, it is difficult to find an accurate answer, as it is almost impossible to measure the direct effect of a language technology system on the social situation of a discriminated language. Nevertheless, it was one of my main goals to share my fascination for this exotic indigenous language, and of course I hope that, at some point, there will be some kind of benefit from my tools for native Quechua speakers.

10.3. Outlook - The SQUOIA Project

In 2011 the SQUOIA project, funded by the Swiss National Science Foundation (SNSF) will be started at the University of Zurich. The project investigates the role of parallel treebanks for building and tuning hybrid machine translation systems. The focus will lie on two language pairs: Spanish to Quechua, and Spanish to German translation.
In a first step we will build parallel treebanks for both language pairs. As text resources for Quechua are scarce, the translation of the corresponding Spanish texts into Quechua will be commissioned to a translator. This approach permits us a relatively free choice of texts for our treebanks, as we are not limited to texts

available in all three languages.

Once the parallel treebanks are complete, we will derive weighted transfer rules from them. The idea is to build in a first step two purely rule-based machine translation systems out of these rules. In a second step, we will compute statistical language models of the two target languages, Quechua and German, that will allow us to rank translation hypotheses. For this purpose we rely on two relatively large monolingual corpora for the target languages. Finally, we plan to enhance our rule-based systems with these statistical components in order to build two hybrid translation systems.

The main research interest lies on the comparison on one side between the rule-based and the hybrid approach, and on the other side on the contrast between the two language pairs: What are the differences between the development process for a MT system for a typologically relatively close language pair (Spanish - German) with easily available resources, versus a system for a typologically diverse language pair (Spanish - Quechua), one of them being a minority language of poor prestige and scarce resources?

We hope that in working with these contrasting language pairs, we will come to insights on how and to what extent the individual conditions of the languages in question do actually have an effect not only upon the development process of MT systems, but also on their performance.

References

2002. *Perú Suyupa Hatun Rimanakuynin, Acuerdo Nacional nisqa* . Lima, Perú: Empresa Peruana de Servicios Editoriales S.A.Segraf - Editora Perú.

2005. *Diccionario: Quechua - Español - Quechua, Qheswa - Español - Qheswa: Simi Taqe*, 2 edition. Cuzco, Perú: Academia Mayor de la Lengua Quechua.

Adelaar, W. F. H.
1977. *Tarma Quechua: Grammar, Texts, Dictionary* . Amsterdam: The Peter de Ridder Press.

Adelaar, W. F. H. and P. Muysken
2004. *The Languages of the Andes*, Cambridge Language Surveys. Cambridge University Press.

Beesley, K. R.
2000. A note on phonologically conditioned selection of verbalization suffixes in Aymara. Technical report, Xerox Research Centre Europe.

Beesley, K. R.
2003. Finite-State Morphological Analysis and Generation for Aymara. In *Proceedings of the Workshop of Finite-State Methods in Natural Language Processing*, 10th Conference of the European Chapter of the Association for Computational Linguistics, Pp. 19–26.

Beesley, K. R. and L. Karttunen
2003. *Finite State Morphology.* CSLI Publications.

Bußmann, H.
2002. *Lexikon der Sprachwissenschaft*, 3 edition. Stuttgart: Kröner.

Carstensen, K.-U., C. Ebert, C. Ebert, S. Jekat, R. Klabunde, and H. Langer, eds.
2010. *Computerlinguistik und Sprachtechnologie: Eine Einführung* . Springer.

Cerrón-Palomino, R.
1994. *Quechua sureño, diccionario unificado quechua-castellano, castellano-quechua.* Lima: Biblioteca Nacional del Perú.

Cerrón-Palomino, R.
2003. *Lingüística Quechua*, 2. edition. Centro de Estudios Regionales Andinos Bartolomé de Las Casas (CBC).

Clements, G. N. and E. Sezer
1982. Vowel and Consonant Disharmony in Turkish. In *The structure of phonological representations. Part II*, H. v. d. Hulst and N. Smith, eds., Pp. 213–255. Foris Publications.

Cusihuamán, A. G.
1976. *Gramática Quechua: Cuzco-Collao*, Gramáticas referenciales de la lengua quechua. Lima: Ministerio de Educación.

Dedenbach-Salazar Sáenz, S., U. von Gleich, R. Hartmann, P. Masson, and C. Soto Ruiz
2002. *Rimaykullayki - Unterrichtsmaterialien zum Quechua Ayacuchano*, 4. edition. Berlin: Dietrich Reimer Verlag GmbH.

Ebert, K. H.
2008. Forms and Functions of Converbs. In *From Siberia to Ethiopia: Converbs in a Cross-Linguistic Perspective*, K. H. Ebert, J. Matissen, and R. Suter, eds., volume 20 of *Arbeiten des Seminars für Allgemeine Sprachwissenschaft*, Pp. 1–33. ASAS.

Escobar, A. M.
1997. Contrastive and Innovative Uses of the Present Perfect and the Preterite in Spanish in Contact with Quechua. *Hispania*, 80(4):859–870.

Faller, M. T.
2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua.* PhD thesis, Stanford University.

Gasser, M.
2008. Computational morphology and the teaching of indigenous languages. In *First Biennial Symposium on Teaching Indigenous Languages of Latin America.*

Hardman, M. J.
2001. *Aymara*, Languages of the World. Munich: LINCOM EUROPA.

Hardman, M. J., J. Vásquez, and J. de Dios Yapita
1988. *Aymara: Compendio de estructura fonológica y gramatical.* La Paz: ILCA.

Herold, H., B. Lurz, and J. Wohlrab
2007. *Grundlagen der Informatik. Praktisch - Technisch - Theoretisch.* München: Pearson Studium.

Hulden, M.
2009. Foma: a Finite-State Compiler and Library. In *EACL Demo*, Pp. 29–32. The Association for Computer Linguistics.

Itier, C.
1997. El zorro del cielo: Un mito sobre el origin de las plantas cultivadas y los intercambios con el mundo sobrenatural. *Bulletin de l'Institut français d'études andines*, (26):307–346.

Itier, C.
2004. *Karu Ñankunapi*, volume 20 of *Biblioteca de la Tradición Oral Andina*, 2. edition. Cuzco: Centro de Estudios Regionales Andinos Bartolomé de Las Casas (CBC).

Lastra, Y.
1968. *Cochabamba Quechua Syntax*. Paris: Mouton The Hague.

Morató Peña, L.
1985. *Quechua Boliviano: Curso Elemental, Tomo 1*, 1 edition. Cochabamba, Bolivia: Instituto de Idiomas "Tawantinsuyu".

Nichols, J.
1986. Head-Marking and Dependent-Marking Grammar. *Language*, 62(1):56–119.

Oflazer, K.
1994. Two-level Description of Turkish Morphology. *Literary and Linguistic Computing*, 9(2).

Oflazer, K.
1996. Error-tolerant Finite State Recognition with Applications to Morphological Analysis and Spelling Correction. *Computational Linguistics*, 22(1).

Oflazer, K. and S. Inkelas
2006. The Architecture and the Implementation of a Finite State Pronunciation Lexicon for Turkish. *Computer Speech and Language*, 20(1).

Parker, G. J.
1963. Clasificación genética de los dialectos quechuas. *Revista del Museo Nacional, Lima*, (32):241–252.

Rios Gonzales, A., A. Göhring, and M. Volk
2009. A Quechua-Spanish Parallel Treebank. In *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7)*, F. van Eynde, A. Frank, K. De Smedt, and G. van Noord, eds., Pp. 53–64. Landelijke Onderzoekschool Taalwetenschap.

Soto Ruiz, C.
1976a. *Diccionario Ayacucho-Chanca*, Diccionarios de consulta de la lengua quechua. Lima: Ministerio de Educación.

Soto Ruiz, C.
1976b. *Gramática Quechua: Ayacucho-Chanca*, Gramáticas referenciales de la lengua quechua. Lima: Ministerio de Educación.

Soto Ruiz, C.
2006. *Quechua - Manual de Enseñanza*, volume 4 of *Lengua y Sociedad*, 3 edition. Lima: IEP Instituto de Estudios Peruanos.

Torero, A.
1964. Los dialectos quechuas. *Anales Científicos de la Universidad Agraria, Lima*, (IV):446–478.

van de Kerke, S.
1994. Mismatches Between Affix Order and Interpretation: Quechua -chi, -mu and -pu Revisited. In *Language in the Andes*, P. Cole et al., eds., Pp. 231–245. University of Delaware.

Van Valin, R. D. J.
2005. *Exploring the Syntax-Semantics Interface.* Cambridge University Press.

Van Valin Jr., R. D. and R. J. L. Polla
1997. *Syntax - Structure, Meaning and Function*, Cambridge Textbooks in Linguistics. Cambridge University Press.

Zuidema, R. T.
1973. Kinship and Ancestorcult in three Peruvian Communities. Hernandez Principe's Account of 1622. *Bulletin de l'Institut français d'études andines*, 2(1):16–33.

A. Abbreviations

A	Abl	Ablative	*-manta*	B	Ben	Benefactive	*-paq*
	Abss	Abessive	*-nnaq*				
	Acc	Accusative	*-ta*				
	Add	Additive	*-pas*				
	Aff	Affective	*-yku*				
	Ag	Nomen Agentis	*-q*				
	Amb	Ambivalent Suffix					
	Aprx	Approximative	*-niq*				
	AS	Ambivalent Suffix					
	Asmp	Assumptive	*-cha/-ch*				
	Asp	Aspect Suffix					
	Ass	Assistive	*-ysi*				
	Aug	Augmentative	*-su*				
	Autotrs	Autotransformative	*-lli*				
C	Cas	Case Suffix		D	Dat	Dative	*-man*
	Caus	Causative	*-chi*		Def	Definitiveness (Certainty)	*-puni*
	Cert	Certainty	*-puni*		Dem	Demonstrative Prn.	
	Char	Characterization	*-li/-ti..*		DE	Direct Evidence	*-mi/-m*
	Cis	Cislocative	*-mu*		Des	Desiderative	*-naya*
	Cis_Trs	Cis-/ Translocative	*-mu*		Desesp	Desesperative	*-pasa*
	Con	Connective	*-taq*		Dim	Diminutive	*-cha*
	Concr	Concretization	*-na*		DirE	Direct Evidence	*-mi/-m*
	Cond	Conditional			Disc	Discontinuative	*-ña*
	Conec	Connective Postposition	*ima*		Dist	Distributive	*-nka/-kama*
	Con	Connective	*-wan/-taq*		DS	Different Subject	*-qti/-pti*
	Con_Inst	Connective/ Instrumental	*-wan*		Dub	Dubitative	*-suna*
	Cont	Continuative	*-raq*				
	Cont	Continuity	*-nya/-miya*				
	Contr	Contrastive Postposition (’or’)	*icha*				
E	Emph	Emphatic		F	Fact	Factitive	*-cha*
	Excl	Exclusive			Frec	Frecuentative	*-kacha*
	EP	Epenthetic Insertion	*-ni*		Fut	Future Tense	

G	Gen	Genitive	*-q/-pa*	H	Hon	Honorific	*-lla*
I	IE	Indirect Evidence	*-si/-s*				
	Ill	Illative	*-man*				
	Inch	Inchoative	*-ri*				
	Incl	Inclusive (case)	*-ntin*				
	IndE	Indirect Evidence	*-si/-s*				
	Inf	Infinitive	*-y*				
	Inst	Instrumental	*-wan*				
	Int	Intention	*-rpari*				
	Int	Intensifier	*-ku*				
	Intr	Interrogative	*-chu/-taq*				
	Intrup	Interruptive	*-kacha*				
	Intsoc	Inter-Sociative	*-pura*				
	IPst	Past of Indirect Evidence	*-sqa*				
	IPrs	Interpersonal	*-pu*				
K	Kaus	Cause (’because’)	*-rayku*	L	Lim	Limitative	*-lla*
M	MPoss	Multi-Possessor	*-sapa*		Loc	Locative	*-pi*
	Mod	Modality					
	MRep	Multi-Repetitive	*-paya*				
N	Neg	Negation		O	Obl	Obligation	*-na*
	NPst	Neutral Past	*-rqa*		Obj	Object	
	NPers	Nominal Person Suffix (Possessive)					
	NRoot	Nominal Root					
	NRootES	Nominal Root of Spanish Origin					
	NRootNUM	Numeral Nominal Root					
	NRootCMP	Compound Nominal Root					
	NS	Nominalizing Suffix					
	Num	Number					
	NumOrd	Ordinal Numeral					

P	Part	Particle		Q	QTop	Topic in Questions	*-ri*
	PartES	Particle of Spanish Origin					
	Perdur	Perdurative	*-raya*				
	Perf	Perfect	*-sqa*				
	Pl	Plural	*-kuna*				
	Posi	Positional	*-mpa*				
	Poss	Possessive					
	Pot	Potential	*-man*				
	Prn	Pronoun					
	PrnInterr	Interrogative Pronoun					
	Prog	Progressive	*-sha/ -chka*				
	Proloc	Prolocative	*-nta*				
	Purp	Purpose	*-na*				
R	Rel	Relational	*-n*	S	Sg	Singular	
	Rflx	Reflexive	*-ku*		Sim	Similarity	*-hina*
	Rflx_Int	Reflexive/ Intensifier	*-ku*		Sml	Simulative	*-tiya*
	Rem	Rememorative	*-ymana*		Sml	Simulative	*-kacha*
	Rep	Repetitive	*-pa*		SS	Same Subject	*-spa*
	Res	Resignation, Implicitness	*-iki*		SS_Sim	Same Subject-Simultaneity	*-stin*
	Reub	Reubicative	*-na*		Soc	Sociative	*-puwan*
	Rgr	Regressive	*-pu*		Subj	Subject	
	Rgr_Iprs	Regressive/ Interpersonal	*-pu*				
	Rptn	'Repentino', Precipitation, Unexpected Action	*-rqu*				
	Rzpr	Reciprocal	*-na*				

T	Term	Terminative	*-kama*	V	VDeriv	Verbal Derivational Suffix	
	Tns	Tense Suffix			VDim	Verbal Diminutive, Childishness	*-cha*
	Top	Topic	*-qa*		VPers	Verbal Person Suffix	
	Trs	Translocative	*-mu*		VRoot	Verbal Root	
	Trsf	Transformative	*-ya*		VRootES	Verbal Root of Spanish Origin	
					VRootCMP	Verbal Root with incorporated NRoot	
					VS	Verbalizing Suffix	

B. Linguistic Notions: Examples

Most Quechua suffixes are not limited to their basic function, but have a number of additional cases of application. As a complete listing of use cases for every suffix would go beyond the scope of this thesis, the examples given are for the most part restricted to the basic function of the suffixes in question. This is especially true of the evidential suffixes, which present a puzzling complexity of use cases, see Faller (2002) for a detailed description.

A:

(37) Ablative *-manta*: ’from’ (nominal suffix, case):

Hinastin	*hawa*	*-llaqta*	***-manta***	*-n*	*Lima*	*-man*	*-qa*	*runa*
everywhere	outside	-village	-Abl	-DirE	Lima	-Ill	-Top	man

llimp’a	*-mu*	*-n.*
migrate	-Cis	-3.Sg.Subj

’People migrate to Lima from everywhere in the outside villages (provinces).’

(Cusihuamán 1976:132)

(38) Abessive *-nnaq* (’Privative’): ’without’ (nominal suffix, case):

wawa	*-nnaq*
child	-Abss

’childless’

(Adelaar and Muysken 2004:217)
(adapted from Pacaraos dialect to Cusco Quechua)

(39) Additive *-pas/-pis*: ’too, also, and’ (ambivalent suffix):

Ñuqa	***-pas***	*ri*	*-saq*	*-mi.*
I	-Add	go	-1.Sg.Subj.Fut	-DirE

’I will go too.’

(Soto Ruiz 1976a:130)

(40) Affective *-yku*: ’affection, respect, concernemnt, cordiality’ (verbal derivational suffix, see also section 5.3.2):

Haywa	***-yka***	*-mu*	*-wa*	*-y*	*kachi*	*-cha*	*-ta.*
hand	-Aff	-Cis	-1.Obj	-2.Sg.Imp	salt	-Dim	-Acc

'Could you please give me the salt.'

(Soto Ruiz 1976a:106)

(41) Agentive/nomen agentis/agent noun *-q*: 'the one who does' (nominalizing suffix):

Qarawi	*taki*	***-q***	*-kuna*	*-ta*	*-m*	*maska*	*-chka*	*-ni.*
qarawi	sing	-Ag	-Pl	-Acc	-DirE	look.for	-Prog	-1.Sg.Subj

'I'm looking for qarawi singers.'

(Soto Ruiz 1976a:138)

(42) Approximative *-niq*: 'near, about there' (nominal suffix, case):

Karu	***-niq***
far	-Appr

'a little further'

(Cusihuamán 1976:231)

(43) Assumptive *-cha*: 'assumption, doubt, possibility' (ambivalent suffix):

Ramun	***-cha***	*ruwa*	*-nqa,*	*pay*	*-mi*	*yacha*	*-n.*
Ramón	-Asmp	do	-3.Sg.Subj.Fut	he	-DirE	know	-3.Sg.Subj

'Ramón will probably do it, he knows [how to do it].' (Soto Ruiz 1976a:138)

(44) Assistive *-ysi*: 'do sth in a helping, supporting manner' (verbal derivational suffix):

Llamka	***-ysi***	*-saq*	*wawqi*	*-y*	*-ta.*
work	-Ass	-1.Sg.Subj.Fut	brother	-1.Sg.Poss	-Acc

'I will help my brother working.' (Soto Ruiz 1976a:107)

(45) Augmentative *-su* ∼ *-chaq* ∼ *-chachaq* ∼ *-karay* ∼ *-chika*: 'big, large' (nominal derivational suffix):

hatu	***-chaq***	∼	*hatu*	***-chachaq***
big	-Augmen	-	big	-Augmen

'enormous, very big'

(Cusihuamán 1976:226)

(46) Autotransformative *-lli*: 'to take on a characteristic, an attitude' (verbal derivational and verbalizing suffix):

Hucha	***-lli***	*-ku*	*-y*
sin	-Autotrs	-Rflx	-Inf

'to sin' (Cusihuamán 1976:196)

Uma	***-lli***	*-ku*	*-y*
head	-Autotrs	-Rflx	-Inf

'to take responsibility' (Cusihuamán 1976:196)

B:

(47) Benefactive *-paq*: 'for, in favour of, also: purpose, temporal determination' (nominal suffix, case):

Tullu	*-ta*	*apa*	*-chka*	*-ni*	*allqu*	***-paq****.*
bone	-Acc	bring	-Prog	-1.Sg.Subj	dog	-Ben

'I am bringing the bone for the dog.' (benefactive)
(Dedenbach-Salazar Sáenz et al. 2002:64)

Para	*-y*	***-paq***	*ka*	*-chka*	*-n.*
rain	-Inf	-Ben	be	-Prog	-3.Sg.Subj

'It's going to rain at any minute now.' (temporal)
(Dedenbach-Salazar Sáenz et al. 2002:64)

Yanu	*-ku*	*-na*	*-n*	***-paq***	*yaku*	*-ta*	*asta*	*-chka*	*-n.*
cook	-Rflx_Int	-Purp	-3.Sg.Poss	-Ben	water	-Acc	carry	-Prog	-3.Sg.Subj

'She bringing water for [her] cooking.' (purpose)
(Dedenbach-Salazar Sáenz et al. 2002:64)

C:

(48) Causative *-chi*: 'make do sth' (verbal derivational suffix):

yachay	know
*yacha**chiy***	teach
wañuy	die
*wañu**chiy***	kill

(49) Certainty *-puni*, see Definitiveness

(50) Characterization *-ti* ∼ *-li* ∼ *-liku* ∼ *-yli* ∼ *-lu*: 'take characteristics of sth, be like' (nominal derivational suffix):

China	***-ti***
feminine	-Char

'coward, womanish'

(Cusihuamán 1976:232)

Mancha	***-li***
fear	-Char

'anxious, afraid'

(Cusihuamán 1976:233)

(51) Cislocative/Translocative *-mu*: (verbal directional suffix):
'hither' - with verbs of movement/weather verbs:

Kallpa	***-mu***	*-chka*	*-n*	*warma.*
run	-Cis	-Prog	-3.Sg.Subj	boy

'The boy is running here.' (towards speaker)

(Dedenbach-Salazar Sáenz et al. 2002:180)

Qayna	*punchaw*	*ancha*	*-ta*	*para*	***-mu***	*-rqa*	*-ø.*
previous	day	very	-Acc	rain	-Cis	-NPst	-3.Sg.Subj

'Yesterday it was raining a lot (here).'

(Dedenbach-Salazar Sáenz et al. 2002:180)

'movement away from reference point' - with other verbs:

Llamka	***-mu***	*-saq*	*Felipe*	*-wan.*
work	-Trs	-1.Sg.Subj.Fut	Felipe	-Con

'I will work with Felipe.' (somewhere else, not here)

(Dedenbach-Salazar Sáenz et al. 2002:181)

(52) Concretization *-na* (nominalizing suffix), see Obligation

(53) Connective/Contrastive *-taq*: 'and, also, but' (ambivalent suffix):

Kunan	*-qa*	*eskuyla*	*-ta*	*-n*	*ri*	*-pu*	*-saq;*	*paqarin*
today	-Top	school	-Acc	-DirE	go	-Iprs	-1.Sg.Subj.Fut	tomorrow
-taq	*icha*	*-qa*	*tayta*	*-y*	*-ta*	*yanapa*	*-saq.*	
-Con	but	-Top	father	-1.Sg.Poss	-Acc	help	-1.Sg.Subj.Fut	

'Today I'll go to school, but tomorrow I will help my father.'

(Cusihuamán 1976:252)

Qusqu	*-pi*	*-qa*	*llank'a*	*-ni*	***-taq***	*estudia*	*-ni*	***-taq***
Cuczo	-Loc	-Top	work	-1.Sg.Subj	-Con	learn	-1.Sg.Subj	-Con
ima	*-n.*							
also	-DirE							

'In Cuzco I work and also study.'

(Cusihuamán 1976:253)

(54) Continuative *-raq*: 'still' (ambivalent suffix):

Kawsa	*-chka*	*-n*	***-raq***	*-mi.*
live	-Prog	-3.Sg.Subj	-Cont	-DirE.

'He's still alive.' (lit. 'He's still living)

(Dedenbach-Salazar Sáenz et al. 2002:118)

(55) Continuity *-nya* ∼ *-miya*: 'continuously, constantly' (verbal derivational suffix):

kuti	***-nya***	*-y*
return	-Cont	-Inf

'to return again and again'

(Cusihuamán 1976:202)

qhawa	***-miya***	*-y*
watch	-Cont	-Inf

'to observe constantly'

(Cusihuamán 1976:202)

D:

(56) Definitiveness/Certainty *-puni*: 'sure, definitively, absolutely, no doubt' (ambivalent suffix):

Piña	*-ku*	*-n*	***-puni***	*-m.*
be.upset	-Rflx	-3.Sg.Subj	-Def	-DirE

'He's upset, no doubt.' (Dedenbach-Salazar Sáenz et al. 2002:118)

(57) Direct Evidence, Ayacucho: *-m/-mi*, Cuzco: *-n/-mi*: 'speaker experienced/witnessed proposition' (ambivalent suffix, data source):

Pilar	*-qa*	*t'anta*	*-ta*	***-n***	*mikhu*	*-rqa*	*-n.*
Pilar	-Top	bread	-Acc	-DirE	eat	-NPst	-3.Sg.Subj

p = 'Pilar ate bread.'
evidentiality = speaker saw that p

(Faller 2002:18)

Mana	***-n***	*muchila*	*-y*	*-pi*	*-chu*	*ka*	*-sha*	*-n.*
Neg	-DirE	backpack	-1.Sg.Poss	-Loc	-Neg	be	-Prog	-3.Sg.Subj

p = 'It is not in my backpack.'
evidentiality = speaker "infers" from not having seen it in the backpack that it is not in the backpack

(Faller 2002:19)

(58) Desiderative *-naya*: 'demand to do sth, desire, long for sth' (verbal derivational suffix, verbalizing suffix):

Puñu	***-naya***	*-wa*	*-chka*	*-n.*
sleep	-Des	-1.Obj	-Prog	-3.Sg.Subj

'I feel sleepy.' (lit. it urges me to sleep)

(Dedenbach-Salazar Sáenz et al. 2002:156)

yaku	***-naya***	*-wa*	*-n.*
water	-Des	-1.Obj	-3.Sg.Subj

'I am thirsty.' (lit. it urges me for water)

(Dedenbach-Salazar Sáenz et al. 2002:156)

(59) Desesperative *-pasa*, occurs according to Cusihuamán (1976) only with *yuya-* (to remember, to think) (verbal derivational suffix):

yuya	***-pasa***	*-y*
remember	-Desesp	-Inf

'to worry, to be contemplative' (Cusihuamán 1976:202)

(60) Discontinuative *-ña*: 'already, yet, change of state' (ambivalent suffix):

Miku	*-chka*	*-n*	***-ña***	*-m.*
eat	-Prog	-3.Sg.Subj	-Disc	-DirE

'He's already eating.'

(Dedenbach-Salazar Sáenz et al. 2002:118)

(61) Distributive *-nka*, *-kama* (when in nominal Slot 7): 'each, every' (nominal suffixes, case):
-nka: only in quantified contexts with verbs that express a notion of sharing or receipt

Kimsa	*tanta*	***-nka***	*apa*	*-mu*	*-nqa.*
three	bread	-Dist	bring	-Cis	-3.Sg.Subj.Fut

'He will bring three breads for every one.' (Dedenbach-Salazar Sáenz et al. 2002:191)

-kama: Distributive in nominal Slot 7, Terminative in nominal Slot 5

Allin	*wasi*	*-yuq*	***-kama***	*-m*	*kay*	*runa*	*-kuna*	*-qa.*
good	house	-Poss	-Dist	-DirE	this	man	-Pl	-Top

'These men, each of them has a good house.' (lit. Each of these man is a good-house-owner)

(Soto Ruiz 1976a:81)

(62) Dubitative *-sina* ∼ *-suna*: 'doubt, it seems that' (ambivalent suffix): Derived form of *-chu -s hina*, Interrogative + IndE + Postposition *hina* ('so, like, similar'). *-sina* occurs in free variation with *-suna* and conveys a meaning similar to assumptive *-cha* (Faller 2002:173).

Mana	***-suna***	*ruwa*	*-sqa*	*-y*	*-qa*	*allin*	*-chu.*
Neg	-Dub	make	-Perf	-1.Sg.Poss	-Top	good	-Neg

'It seems that what I'm doing is not good.'
(lit. my-done is not good)

(Cusihuamán 1976:247)

F:

(63) Factitive -*cha*: ’make sth, form sth’ (verbalizing suffix):

Wasi	***-cha***	*-ni.*
house	-Fact	-1.Sg.Subj

’I build a house.’

(Dedenbach-Salazar Sáenz et al. 2002:126)

(64) Frecuentative -*ykacha*: see Interruptive

H:

(65) Honorific/Limitative -*lla*: ’only, just, also: affection, sympathy, respect’ (ambivalent suffix):

Ñuqa	***-lla***	*-m.*
I	-Aff	-DirE

’only me’

Imayna	***-lla***	*-taq*	*ka*	*-chka*	*-nki?*
how	-Aff	-Intr	be	-Prog	-2.Sg.Subj

’How are you?’ (friendly)

(Dedenbach-Salazar Sáenz et al. 2002:43)

I:

(66) Illative -*man* (also as Dative): ’to, towards’ (nominal suffix, case):

Wanta	***-man***	*ri*	*-chka*	*-ni.*
Huanta	-Ill	go	-Prog	-1.Sg.Subj

’I am going to Huanta.’

(Soto Ruiz 1976a:79)

(67) Inchoative -*ri*: ’begin to’ (verbal derivational suffix):

Ña	*-n*	*irqi*	*-cha*	*-qa*	*saya*	***-ri***	*-pa*	*-ku*	*-sha*
Already	-DirE	child	-Dim	-Top	stand.up	-Inch	-Inter	-Rflx	-Prog
-n	*-ña.*								
-3.Sg.Subj	-Disc								

'The child already begins to stand up.'

(Cusihuamán 1976:210)

(68) Indirect Evidence/Reportative *-si*: 'it's said that, speaker has not experienced/witnessed proposition or was not in conscious state' (ambivalent suffix, data source):

Marya	*-qa*	*yacha*	*-y*	*wasi*	*-pi*	***-s***	*ka*	*-sha*	*-n.*
Marya	-Top	know	-Inf	house	-Loc	-IndE	be	-Prog	-3.Sg.Subj

p = 'Marya is in school.'
evidentiality = speaker was told that *p*

(Faller 2002:189)

Ñuqa	*musqu*	*-ni*	*-y*	*-pi*	***-s***	*rima*	*-rqa*	*-ni.*
I	sleep	-EP	-1.Sg.Poss	-Loc	-IndE	speak	-NPst	-1.Sg.Subj

p = 'I spoke in my sleep.'
evidentiality = speaker was told that *p*

(Faller 2002:190)

(69) Instrumental *-wan* (also used as Connective): 'with, by, also: causee in causative constructions' (nominal suffix, case):

Ramun	*-cha*	***-wan***	*-mi*	*ri*	*-saq.*
Ramón	-Dim	-Con	-DirE	go	-1.Sg.Subj.Fut

'I'll go with Ramón.'

(Soto Ruiz 1976a:79)

Ichuq	*maki*	*-n*	***-wan***	*-mi*	*qillqa*	*-n.*
left	hand	-3.Sg.Poss	-Inst	-DirE	write	-3.Sg.Subj

'He writes with his left hand.'

(Soto Ruiz 1976a:79)

Parqu	*-chi*	*-ni*	*chakra*	*-ta*	*Pedro*	***-wan.***
irrigate	-Caus	-1.Sg.Subj	field	-Acc	Peter	-Inst

'I have the field irrigated by Peter.' (causee)

(Adelaar and Muysken 2004:216)

(70) Intentional *-rpari* ∼ *-ypari* (verbal derivational suffix), at least 3 meanings:

1. 'do sth on purpose, deliberately'
2. 'do sth by mistake, accidentally, abruptly'
3. 'completive' (after Cerrón-Palomino (2003:262))

Tarpu	***-rpari***	*-kipa*	*-rqu*	*-ysi*	*-chi*	*-sqa*	*-yki*	*-puni*
sow	-Int	-Rep	-Rptn	-Ass	-Caus	-Fut	-1.Sg>2.Sg	-Def

'I will make them help you sow again for sure.'
(Bolivian Quechua, suffix *-kipa* borrowed from Aymara)

(Cerrón-Palomino 2003:261)

(71) Interpersonal/Regressive *-pu*: 'action affects other person than subject, movement back to point of origin (only with verbs of movement), state lasts for a long time/forever' (verbal suffix, see also section 5.3.1):

Sumaq	*pacha*	*-ta*	*ruwa*	***-pu***	*-saq.*
beautiful	dress	-Acc	make	-Iprs	-1.Sg.Subj.Fut

'I'll make you a beautiful dress.' (interpersonal)

(Soto Ruiz 1976a:107)

Hatuku	*-y*	*-kuna*	*-qa*	*unay*	*-ña*	*-n*	*wañu*
grandparent	-1.Sg.Poss	-Pl	-Top	long.ago	-Disc	-DirE	die
-ka	***-pu***	*-nku.*					
-Rlfx	-Rgr_Iprs	-3.Pl.Subj					

'My grandparents have died long ago.' (state)

(Cusihuamán 1976:215)

(72) Interruptive/Frecuentative/Simulative *-ykacha* (verbal derivational suffix, verbalizing suffix, see also section 5.3.1):

1. 'do sth without attention, repeat action in short intervals'
2. 'simulated action, pretend'

Frecuentative:

¿Imanasqa	*-taq*	*chay*	*runa*	*-kuna*	*kay*	*-kuna*	*-pi*	*puri*	***-kacha***
why	-Intr	this	man	-Pl	that	-Pl	-Loc	walk	-Frec
-nku,	*icha*	*suwa*	*ka*	*-nku?*					
-3.Pl.Subj	or	thief	be	-3.Pl.Subj					

'Why do these men keep sneaking around here, are they thieves?'

(Soto Ruiz 1976a:114)

Simulative:

qhari	***-kacha***	*-y*
man	-Sim	-Inf

'Pretend to me manly, pretend to be courageous'

(Cusihuamán 1976:199)

(73) Intersociative *-pura*: 'among' (nominal suffix, case):

Warmi	***-pura***	*rima*	*-chka*	*-nku.*
woman	-Intsoc	speak	-Prog	-3.Pl.Subj

'The women speak among themselves.'

(Soto Ruiz 1976a:80)

L:

(74) Limitative *-lla*, see Honorific

(75) Locative *-pi*: 'in, at (also temporal), by (means of transport)' (nominal suffix, case):

Wamanga	***-pi***	*-m*	*yacha*	*-ni.*
Huamanga	-Loc	-DirE	live	-1.Sg.Subj

'I live in Huamanga.'

(Soto Ruiz 1976a:81)

Chaki	*-lla*	***-pi***	*-n*	*eskuyla*	*-ta*	*-qa*	*ri*	*-pu*	*-ni.*
foot	-Lim	-Loc	-DirE	school	-Acc	-Top	go	-Rgr_Iprs	-1.Sg.Subj

'I go to school by foot.'

(Cusihuamán 1976:133)

M:

(76) Multi-Possessor *-sapa*: 'owner of many, much (also figuratively)' (nominal derivational suffix):

Wasi	***-sapa***	*runa*	*hamu*	*-chka*	*-n.*
house	-MPoss	man	come	-Prog	-3.Sg.Subj

'The man who owns many houses is coming.' (Soto Ruiz 1976a:142)

piki	***-sapa***	*allqu*
flea	-MPoss	dog

'flea afflicted dog'

(Soto Ruiz 1976a:142)

yuyay	***-sapa***
mind	-MPoss

'intelligent'

(Cusihuamán 1976:226)

O:

(77) Obligation/Purpose/Concretization *-na*: 'derives abstract nouns from verbal roots, may also convey the meaning of obligation or even purpose' (nominalizing suffix):

Concretization:

mihu	***-na***
eat	-Concr

'food'

picha	***-na***
sweep	-Concr

'broom'

puklla	***-na***
play	-Concr

'toy'

(Cusihuamán 1976:221)

Obligation:

P'acha	*-y*	*-mi*	*t'aqsa*	*-ku*	***-na***	*-y*	*ka*	*-sha*
cloth	-1.Sg.Poss	-DirE	wash	-Int	-Obl	-1.Sg.Poss	be	-Prog

-n.
-3.Sg.Subj

'I have to wash my clothes.'

(Cusihuamán 1976:221)

Purpose:

Willa	*-wa*	***-na***	*-yki*	*-paq*	*-mi*	*hamu*	*-ra*	*-ni.*
tell	-1.Obj	-Purp	-2.Sg.Poss	-Ben	-DirE	come	-NPst	-1.Sg.Subj

'I came for you to tell me.' (lit. 'For your-telling-me I came')

(Soto Ruiz 1976a:155)

P:

(78) Perdurative *-raya*:
verbal suffix: 'continuous state' (allomorph *-nraya*), some transitive verbs become intransitive with *-raya*,
verbalizing suffix: 'Characterization' (verbal derivational suffix, verbalizing suffix):

Kuchi	*-qa*	*koral*	*-pi*	*-raq*	*-mi*	*wichqa*	***-raya***	*-chka*
pig	-Top	pen	-Loc	-Cont	-DirE	close	-Perdur	-Prog
-n.								
-3.Sg.Subj								

'The pig is still locked up in the pen.'
(verbal suffix, intransitivization of *wichqa-* 'to close sth, to enclose sth')

(Soto Ruiz 1976a:112)

puka	***-raya***	*-y*
red	-Char	-Inf

'to blush' (verbalizing suffix)

(Cusihuamán 1976:198)

(79) Perfect *-sqa*: 'action in completed state, perfect, also: place where an action is realized' (nominalizing suffix):

qispi	***-sqa***	*wasi*
finish	-Perf	house

'completed house'

(Soto Ruiz 1976a:136)

¿Riqsi	*-nki*	*-chu*	*yacha*	***-sqa***	*-y*	*-ta?*
know	2.Sg.Subj	-Inter	live	-Perf	-1.Sg.Poss	-Acc

'Do you know where I live?'

(Soto Ruiz 1976a:137)

Ama	*-n*	*chay*	*yacha*	***-sqa***	*-yki*	*-ta*	*qunqa*	*-nki*
Neg.Imp	-DirE	this	know	-Perf	-2.Sg.Poss	-Acc	forget	-2.Sg.Subj

-chu.
-Neg

'Don't forget what you have learnt.' (lit. Don't forget this your knowing)

(Cusihuamán 1976:225)

Contrast with Concretization nominalizer *-na* (note that both examples can also mean 'the house where I work'):

Llamka	***-na***	*-y*	*wasi*
work	-Concr	-1.Sg.Poss	house

'the house where I will work'

Llamka	***-sqa***	*-y*	*wasi*
work	-Perf	-1.Sg.Poss	house

'the house where I worked'

(Soto Ruiz 1976a:137)

(80) Positional *-mpa*: 'be in position of' (nominal derivational suffix, nominalizing suffix):

saya	***-mpa***
stand.up	-Posi

'vertical, standing'

(Cusihuamán 1976:236)

kinra	***-mpa***
side	-Posi

'recumbent, reclined'

(Cusihuamán 1976:236)

(81) Potential/Condicional *-man* (verbal suffix, modality):

Mana	*para*	*-qti*	*-n*	*-qa*	*Qusqu*	*-ta*	*ri*	*-y*	***-man***
Neg	rain	-DS	-3.Sg.Poss	-Top	Cuzco	-Acc	go	-1.Sg.Subj	-Pot

-mi.
-DirE

'If it wasn't raining, I would go to Cuzco.'

(Cusihuamán 1976:178)

Amaláy	*apa*	*-mu*	*-y*	***-man***	*ka*	*-ra*	*-ø*	*rutu*
hopefully	carry	-Cis	-1.Sg.Subj	-Pot	be	-NPst	-3.Sg.Subj	cut

-na	*-y*	*-ta.*
-Concr	-1.Sg.Poss	-Acc

'If only I had brought my sickle.' (Soto Ruiz 1976a:103)

(82) Progressive *-sha/-chka/-sa/-siya* (dialectal variations, verbal suffix, aspect): Bolivian Quechua:

Mama	*-yku*	*t'anta*	*-ta*	*ruwa*	***-sa***	*-n.*
mother	-1.Pl.Excl.Poss	bread	-Acc	make	-Prog	-3.Sg.Subj

'Our mother is making bread.'

(Morató Peña 1985:75)

Cuzco Quechua:

Para	*-qa*	*chaya*	*-ku*	***-sha***	*-lla*	*-n*	*-mi.*
rain	-Top	come	-Rflx_Int	-Prog	-Lim	-3.Sg.Subj	-DirE

'It is still raining.'

(Cusihuamán 1976:255)

Ayacucho Quechua:

Kay	*-pi*	*-qa*	*ña*	*-m*	*tarpu*	*-ku*	***-chka***	*-nku*
That	-Loc	-Top	already	-DirE	sow	-Rflx_Int	-Prog	-3.Pl.Subj

-ña.
-Disc

'They are already sowing here.'

(Soto Ruiz 1976a:111)

Q:

(83) Question Topic *-ri* (*-rí* with emphasis): 'marks the topic element in questions, indicates that the question has a connection to the preceding conversation, implicates that speaker is interested in keeping the conversation going' (Cusihuamán 1976:238), (ambivalent suffix):

Yu,	*wawqi,*	*may*	*-ta*	*ri*	*-sa*	*-nki*	***-ri.***
hey	brother	where	-Acc	go	-Prog	-2.Sg.Subj	-QTop

'Hey, brother, where are you going?'

Llaqta	*ukhu*	*-ta*	*ri*	*-sa*	*-ni,*	*qan*	***-rí**.*
village	inside	-Acc	go	-Prog	-1.Sg.Subj	you	-QTop

'I'm going to the village centre, and you?'

(Morató Peña 1985:51)

R:

(84) Relational *-n*: 'attached to nominal roots, indicates that the resulting noun is part of, or relates to a "whole", an entity' (nominal derivational suffix):

maki	***-n***
hand	-Rel

'hand'

hatun	*-ni*	***-n***
big	-EP	-Rel

'major portion, the bigger part of sth'

(Cusihuamán 1976:234)

(85) Reflexive/Intensifier *-ku*: 'reflexive, personal interest, special involvement of speaker, action is characteristic, typical for subject' (verbal suffix, see also section 5.3.1):

Qayllu	*-y*	*-ta*	*k'utu*	*-ru*	***-ku***	*-ni.*
tongue	-1.Sg.Poss	-Acc	bite	-Rptn	-Rflx	-1.Sg.Subj

'I've bitten my tonge. (Reflexive)

(Cusihuamán 1976:212)

Asnu	*-kuna*	*-qa*	*hayta*	***-ku***	*-nku*	*-m.*
donkey	-Pl	-Top	kick	-Int	-3.Pl.Subj	-DirE

'Donkeys [usually] kick.' (Characterization)

(Soto Ruiz 1976a:108)

(86) Rememorative *-ymana*: 'act of mediation, reflectation, contemplation' (verbal derivational suffix):

yuya	***-ymana***	*-y*
think	-Rem	-Inf

'to do an examination of conscience, to meditate'

(Cusihuamán 1976:203)

(87) Resignation *-iki* (only in combination with *-mi, -si, -cha* ⇒ *-miki, -siki, -chiki*, unbound form is *riki*, ambivalent suffix):
-miki: 'excuse, apology, also: self-evidence, obviousness'

Chay	*rumi*	*-ta*	*-qa*	*mana*	*-m*	*quqa*	*-ri*	*-y*	*-ta*	*ati*
this	stone	-Acc	-Top	Neg	-DirE	lift	-Inch	-Inf	-Acc	be.able

-n	*-man*	*-ku*	*-chu,*	*hatun*	*-su*	***-m***	***-iki*.**
-3.Pers	-Pot	-Pl	-Neg	big	-Aug	-DirE	-Res

'This stone can not be lifted, as it is very big.' (Soto Ruiz 1976a:125)

-siki: 'emphasizes indirect evidence, renders proposition more impersonal'

Llumpay	*-ta*	***-s***	***-iki***	*rabya*	*-chi*	*-sqa*	*-ø*	*hina*	*-pti*
much	-Acc	-IndE	-Res	rage	-Caus	-IPst	-3.Sg.Subj	be.so	-DS

-n	*-si*	*panya*	*-ru*	*-sqa*	*-ø.*
-3.Sg.Poss	-IndE	punish	-Rptn	-IPst	-3.Sg.Subj

'[They say that] he$_1$ punished him because he$_2$ made him furios.'
(lit. 'He made [him] furios, being like that, he punished [him]')

(Soto Ruiz 1976a:125)

-chiki: 'reduces uncertainty of conjecture with-*cha*, may also imply resignation'

Mana	***-ch***	***-iki***	*llamka*	*-y*	*-ta*	*tari*	*-ra*	*-ø*	*-chu.*
Neg	-Ass	-Res	work	-Inf	-Acc	find	-NPst	-3.Sg.Subj	-Neg

'He probably did not find work.'

(Soto Ruiz 1976a:125)

(88) Reubicative *-na*: 1. 'take position of sth, put in position' 2. 'take away sth' (verbalizing suffix):

k'uchu	***-na***	*-y*
corner	-Reub	-Inf

'to put in a corner'

(Cusihuamán 1976:195)

rap'i	***-na***	*-y*
leave	-Reub	-Inf

'to strip the leaves off' (Cusihuamán 1976:195)

(89) Regressive *-pu*, see Interpersonal

(90) Repetitive *-pa*: 'again, repeatedly', often: 'action repeated in order to improve sth' (verbal derivational suffix):

Kuta	***-pa***	*-yku*	*-y*	*chay*	*-ta,*	*chamra*	*-lla*	*-raq*	*ka*
grind	-Rep	-Aff	-2.Sg.Imp	this	-Acc	grain	-Lim	-Cont	be
-chka	*-n.*								
-Prog	-3.Sg.Subj								

'Grind this again, it's still granulous.'

(Soto Ruiz 1976a:104)

(91) Repentino *-rqu* ∼ *ru*: 'urgency, abruptness, haste, surprise, intentness, priority' (verbal derivational suffix, see also section 5.3.2):

Apa	***-ra***	*-ysi*	*-y,*	*llumpay*	*llasaq*	*-mi*	*qipi*	*-n*
carry	-Rptn	-Ass	-2.Sg.Imp	much	heavy	-DirE	bundle	-3.Sg.Poss
ka	*-chka*	*-n.*						
be	-Prog	-3.Sg.Subj						

'Help [him] carry, his bundle is very heavy.' (Soto Ruiz 1976a:106)

Achakáw,	*chaki*	*-y*	*-ta*	*misk'a*	***-ru***	*-ku*	*-ni!*
ouch	foot	-1.Sg.Poss	-Acc	bump	-Rptn	-Rflx	-1.Sg.Subj

'Ouch, I bumped my foot [against a stone]!'

(Cusihuamán 1976:208)

(92) Reciprocal *-na*, often in combination with *-ku* (Reflexive): 'mutal action' (verbal derivational suffix):

Paykuna	*-qa*	*kuya*	***-na***	*-ku*	*-nku*	*ancha*	*-ta*	*-m.*
they	-Top	love	-Rzpr	-Rflx	-3.Pl.Subj	much	-Acc	-DirE

'They love each other very much.'

(Soto Ruiz 1976a:112)

Wallpa	*-kuna*	*tupsa*	***-na***	*-ku*	*-ru*	*-nku.*
hen	-Pl	peck	-Rzpr	-Rflx	-Rptn	-3.Pl.Subj

'The hens peck each other.'

(Soto Ruiz 1976a:112)

S:

(93) Similarity *hina/niraq* ∼ *niray*: 'so, as, like, similar' (status as ambivalent suffix or postposition not clear; I treat them as postpositions, nevertheless, the analyzer will also recognize them as suffixes):

Qaqa	***hina***	*kapka*	*-m*	*kay*	*tanta*	*-qa.*
rock	Sim	hard	-DirE	this	bread	-Top

'This bread is hard as rock.'

(Dedenbach-Salazar Sáenz et al. 2002:189)

suyt'u	***niray***
cone	-Sim

'sharp-pointed' (Cusihuamán 1976:231)

(94) Simulative *-tiya* ,also *-ykacha/-kacha* see Interruptive: '1. simulated action, act as if, 2. repeat an action in order to overcome some sort of opposition or obstacle, 3. do sth with reluctance' (verbal derivational suffix):

Llamka	***-tiya***	*-y*	*-lla*	*llamka*	***-tiya***	*-chka*	*-nku*	*qilla*	*warma*
work	-Sml	-Inf	-Lim	work	-Sml	-Prog	-3.Pl.Subj	lazy	boy
-kuna	*-qa.*								
-Pl	-Top								

'[These] lazy boys only pretend to be working.'

(Soto Ruiz 1976a:113)

Utulu	*millpu*	***-tiya***	*-chka*	*-n,*	*chuñu*	*kunka*	*-n*	*-pi*
cock	swallow	-Sml	-Prog	-3.Sg.Subj	chuño	throat	-3.Sg.Poss	-Loc
takya	*-ru*	*-n.*						
stop	-Rptn	-3.Sg.Subj						

'The cock tries to swallow repeatedly, a *chuño* got stuck in his throat.' (*chuño* = freeze-dried potato)

(Dedenbach-Salazar Sáenz et al. 2002:198)

(95) Sociative *-puwan* ∼ *-piwan*: ’(together) with, and, too’, special form of Instrumental (*-wan*)

Kay	***-puwan***	*chahay*	***-puwan***	*-mi*	*chakra*	*-yku*	*-qa.*
this	-Soc	that	-Soc	-DirE	field	-1.Pl.Excl.Poss	-Top

’This [one] and [one] that are our fields.’

(Cusihuamán 1976:134)

T:

(96) Terminative *-kama* (nominal suffix, case):
1. ’until, as far as, up to’ (spatial and temporal)
2. in combination with *-na* (Concretization, Obligation): ’while, in the meantime, until’
3. Distributive, see Distributive

Wata	***-kama***	*-m*	*mana*	*hamu*	*-nqaku*	*-chu.*
year	-Term	-DirE	Neg	come	-3.Pl.Subj.Fut	-Neg

’They won’t come until next year.’

(Soto Ruiz 1976a:81)

Sama	***-na***	*-y*	***-kama***	*-m*	*pay*	*puklla*	*-ku*	*-chka*	*-n.*
rest	-Concr	-1.Sg.Poss	-Term	-DirE	he	play	-Rflx_Int	-Prog	-3.Sg.Subj

’He plays while I’m resting.’

(Soto Ruiz 1976a:81)

(97) Topic marker *-qa*: ’marks the topic in the discourse; in contrast with the evidentials, *-mi* and *-si*, and the epistemic suffix *-cha*, which are often attached to the focal element and occupy the same morphological slot as *-qa*’ (ambivalent suffix):

Mana	*-n*	*rit’i*	*-mu*	*-n*	*-chu*	***-qa***	*Chinchero*	*-ta*	***-qa***.
Neg	-DirE	snow	-Cis	-3.Sg.Subj	-Neg	-Top	Chinchero	-Acc	-Top

’[But] it does not snow in Chinchero.’

(Cusihuamán 1976:238)

Pi	*-chá*	*alkalde*	***-qa***	*hayku*	*-nqa*	*kunan*	*wata,*	*í?*
who	-Asmp_Emph	mayor	-Top	enter	-3.Sg.Subj.Fut	now	year,	eh

'Who will come in as mayor this year, eh?' (Cusihuamán 1976:238)

(98) Transformative *-ya*: 'to become, to turn into' (verbalizing suffix):

Llumpay	*-ta*	*wira*	***-ya***	*-chka*	*-ni.*
much	-Acc	fat	-Trsf	-Prog	-1.Sg.Subj

'I'm becoming very fat.'

(Soto Ruiz 1976a:139)

-Wawa	*-yki*	*-qa*	*runa*	***-ya***	*-ru*	*-sqa*	*-ña*	*-m.*
son	-2.Sg.Poss	-Top	man	-Trsf	-Rptn	-Perf	-Disc	-DirE

-Ari	*-yá,*	*kay*	*-qa*	*-ya*	*qunqa*	*-y*	*-ta*	*hatun*	***-ya***	*-ru*	*-n.*
yes	-Emph	this	-Top	-Emph	forget	-Inf	-Acc	big	-Trsf	-Rptn	-3.Sg.Subj

'-Your son has already become a man.'
'-Yes, all of a sudden he has grown up so much.' (lit. become so tall)

(Soto Ruiz 1976a:139)

(99) Translocative *-mu*, see Cislocative

V:

(100) Verbal Diminutive *-cha*: 'do sth in a childish manner, also: courtesy, esteem, intimacy' (verbal derivational suffix):

¿Ña	*-chu*	*rima*	***-cha***	*-chka*	*-n*	*-ña*	*wawa*	*-cha*	*-yki?*
Already	-Intr	speak	-Vdim	-Prog	-3.Sg.Subj	-Disc	child	-Dim	-2.Sg.Poss

'Does your child already speak?'

(Soto Ruiz 1976a:105)

Anya	***-cha***	*-ku*	*-chka*	*-n*	*uña*	*allqu*	*-cha.*
bark	-Vdim	-Int	-Prog	-3.Sg.Subj	baby	dog	-Dim

’The puppy is barking.’

(Soto Ruiz 1976a:105)

Lebenslauf

Personalien

Name	Annette Rios
Adresse	Bachtelstrasse 32 8620 Wetzikon
Email	arios@ifi.uzh.ch
Geburtsdatum	15.01.1981
Kinder	Naira, geboren 2002

Ausbildung

seit 2003	Universität Zürich Allgemeine Sprachwissenschaft Computerlinguistik Informatik
1996 - 2001	Kantonsschule Zürcher Oberland Typus D

Anstellungen

seit 2008	Hilfsassistentin am Institut für Computerlinguistik
2002	Büroaushilfe, Crédit Suisse, Human Resources
2001 - 2002	Auslandaufenthalt (Peru, Ecuador, Kolumbien)