A Morphological Analyzer for Akkadian Verbal Forms with a Model of Phonetic Transformations

François Barthélemy
Galatasaray Universitesi, Istanbul, Turkey, and
Conservatoire National des Arts et Métiers, laboratoire CEDRIC, Paris, France
barthelemy@gsunv.gsu.edu.tr

Abstract

The paper describes a first attempt to design a morphological analyzer for Akkadian verbal forms. Akkadian is a semitic dead language which was used in the ancient Mesopotamia. The analyzer described has two levels: the first one is a deterministic and unique paradigm that describes the flexion of Akkadian verbs. The second level is a non deterministic rewriting system which describes possible phonetic transformations of the forms. The results obtained so far are encouraging.

1 Introduction

Akkadian is a dead Semitic language that was spoken and written in Mesopotamia between 2800 and 0 B.C. The main difficulty in this language is that verbs take a great variety of forms that differ not only by their ending, as in Indo-European languages, but also by the beginning and by the middle part. The only constant in all the forms of a verb is the root, which is composed of three consonants. These consonants appear mixed with other consonants and various vowels. Furthermore, some of the root consonants may be "weak," meaning they do not appear in actual forms. At times, a weak consonant simply disappears, at others, it mutates to another consonant or to a vowel.

We have constructed a two-part morphological analyzer for Akkadian verbal forms. The first is a grammar, which describes the systematic morphotactics of Akkadian verbs. There is a single paradigm for all verbs, weak or strong. The second component is a transducer, which changes a given theoretical form generated by the grammar into an actual form, by applying some phonetic transformation rules (e.g., assimilation, mutation, etc.). This transducer does not rely on a phonetic theory but on a set of observed transformations. In fact, the transducer is primarily used backwards, to retrieve the theoretical original form from an actual word to be parsed.

The morphological analyzer recognizes all the verbal forms we have collected so far. On the other hand, the second level is not efficient on long forms. The paper describes the first results obtained in a work in progress. These results show that more sophisticated computational methods must be used in order to improve the efficiency.

The next section is devoted to the Akkadian language. It is followed by brief description of Akkadian verbs and how they conjugate. The morphological analyzer, its two components, and its results are presented in section 4.

2 The Akkadian Language

Akkadian is a dead language of the Semitic family. It was used as a native language in Mesopotamia and as a written language in a wider area, including the entire near-east: Egypt, Syria, Palestine, Anatolia, and Persia. The name comes from the city of Akkad, once the center of one of the oldest empires in the region. Akkadian was later the language of the Babylonian and Assyrian empires. The oldest documents written in Akkadian date back to approximately 2500 B.C., whereas the most recent ones are from the first century B.C. During the 2500 years of life, the language changed and so has traditionally been divided in several dialects using temporal and spatial criteria (Old Akkadian, Old Babylonian, Medium Babylonian, New Babylonian, Late Babylonian, Old Assyrian, Medium Assyrian, Late Assyrian).

Akkadian is written using Cuneiform signs. Most texts are written on clay tablets, though

Figure 1: An example of Cuneiform writing

a few are written on stone or metal. There are numerous documents, mostly in museum and university reserves, and new ones are discovered every year. The writing system was inherited from the Sumerian language, which was spoken and written in Mesopotamia before Akkadian appeared. Figure 1 shows an example of an Akkadian phrase in Cuneiform. The system combines logograms, syllabograms and determinatives. Most signs have both a logographic value and several syllabographic values, so the system is ambiguous. The phonetic value of a sign is a syllable made of either a single vowel (such as *a* or *i*), or a consonant followed by a vowel (*nu*, *ta*), or a vowel followed by a consonant (*ak*, *im*), or a consonant-vowel-consonant pattern (*til*, *nin*). Words are decomposed into syllables, which are written by a sign. For instance, the word *iprus* (he separated) can be decomposed into *ip–ru–us*. Note that a single *u* appears in two adjacent syllables. This system cannot write consonants which are not just before or just after a vowel. For example, *prus* is not writable. A vowel must be added somewhere. Akkadian is the only Semitic language where all vowels are written.

As Cuneiform is not very convenient for the modern writer, people interested in Akkadian represent texts with modern writing systems. The one we use for this paper and the morphological analyzer is the transcription in an extended Roman alphabet. This system is also used in the grammars, to describe the language.

3 Akkadian Verbs

Akkadian verbs take a great variety of forms, which differ in five ways:

- adjunction of a prefix
- adjunction of a suffix
- adjunction of an infix syllable or consonant within the root
- doubling of the second root consonant
- change in the vocalization of the root consonants

The factors of verbal forms are: the root of the verb; the characteristic vowels of the verb; the mode; the aspect; the gender; the number; the stem; the regional factor; the temporal factor. We now detail these factors and give some examples of their influence on morphology.

As in the other Semitic languages, verbal roots are composed of consonants. In most cases, there are three consonants in a root, though occasionally there are four. The root consonants appear in all the forms of a verb. Here are some of the numerous forms of *parasum*, to separate: *parasum, iprus, niprusam, paris, uptanarras, putarris, pitarras, ipparrasam, uŝapris, pursa, parsaku, parsatina*. One can check that every form contains the root consonants *p, r,* and *s* in this order. In the following, we will call *radicals* individual root consonants.

The Akkadian consonants are: b, g, d, w, z, h, ṭ, j, k, l, m, n, s, p, ṣ, q, r, š, t, aleph. Aleph is usually denoted by a single quote, but we prefer to write it α in this article to avoid any confusion.

Each Akkadian verb has two characteristic vowels. They are used to vocalize one or another radical in some forms. These vowels, as well as the root, comprise a lexical piece of information. There are 4 vowels in Akkadian: a, e, i, and u. Each vowel may be either long or short. Figure 2 shows examples of variations in forms due to the vowels.

There are three modes in Akkadian:

- indicative, used in independent clauses.
- subjunctive, used in dependent clauses.
- ventive, used in independent clauses to emphasize the verb's meaning

Some examples are given in figure 3.

The Akkadian language uses five aspects. As in other Semitic languages, aspects do not encode temporal information, but the status of the action.

- the imperfect denotes an action that is not accomplished. It may or may not have begun yet. It is usually translated into English by the present perfect or future tenses.
- the perfect is employed for an action just finished.

root	vowel 1	vowel 2	infinitive	imperfect	preterit
prs	a	u	parasum	iparr*as*	ipr*us*
ṣbt	a	a	ṣabatum	iṣabb*at*	iṣb*at*
pqd	i	i	paqadum	ipaqq*i*d	ipq*i*d

Figure 2: vocalic variations

	example 1	example 2	example 3	example 4
indicative	iprus	taprusi	paris	iprusu
subjunctive	iprusu	taprusi	parsu	iprusu
ventive	iprusam	taprusim	parsam	iprusunim

Figure 3: modes

- the preterit is the aspect of completed actions.
- the stative designates an atemporal state or the lasting effects of an action.
- the imperative has the same use as in English.

Some examples are given in figure 4.

As shown by the example in figure 4, gender and number are factors of the verbal form. The differences are in prefixes and suffixes. Sometimes (e.g., stative, imperative), the absence of vocalization of the third radical imposes the vocalization of the second. For instance, the stative third person masculine should be **pars*. But it is impossible to write this within the syllabic framework of Akkadian Cuneiform. A vowel *u* is therefore added.

Some verbal forms are not conjugated at any aspect: the infinitive and the participle, which act as nouns and are declined as such. The verbal adjective acts as an adjective and it is declined as well. Infinitives, participles, and verbal adjectives do not exist in the subjunctive and they have the same stems (presented below) as the conjugated forms.

Verbs are conjugated in several subsystems called *stems*. They are distinguished by prefixes, infixes, and reduplication of the second radical. There are 12 different stems classified in 5 stem groups:

- stem I is the basic one. The other stems may be described as a transformation of this one. Example: *iprus*, he separated (root *prs*, preterit).
- stem II (or, D-stem) is characterized by the reduplication of the second radical and by the prefix vowel *u*. Example: *uparris* (root *prs*, preterit).
- stem III (or, Š-stem): the š consonant prefixes the root and the prefix vowel *u* is used. Semantically, it is a causative form. Example: *ušapris* (root *prs*, preterit).
- stem III/II (or, ŠD-stem): there are both a š prefix and reduplication of the second radical. It is semantically equivalent to the stem III. Example: *ušparris*.
- stem IV (or, N-stem): the root is prefixed by an *n*. Example: *ipparis*. In this example, the *n* prefixed has changed into a *p* by an assimilation process. This is not a special case: the prefixed *n* almost always assimilates to the first radical.

Within each stem group, stems are distinguished by the presence or the absence of an infix. The infix is added after the first radical for stems of groups I and II, and after the infixed *š* or *n* for groups III, III/II, and IV. The notation for a stem is made by adding an index to the stem group number.

- no infix (index 1): all the groups have a stem without infix.
- infix *T* (index 2): stem groups I, II, and III have a stem with a *t* infixed.

pers.	imperfect	preterit	perfect	imperative	stative
3 sing.	iparras	iprus	iptaras	—-	paris
2 sing. masc.	taparras	taprus	taptaras	purus	parsata
2 masc. pl.	taparrasa	taprusa	taptarsa	pursa	parsatunu
2. fem. pl.	taparrasa	taprusa	taptarsa	pursa	parsatina

Figure 4: aspect, gender and number

- infix *Tn* (index 3): all stem groups except the III/II group have a stem with a *tn* infix.

Each stem has a specific semantics that combines with the semantics of the root to give a meaning to a verbal form. The table of figure 5 summarizes the semantics of the stems.

Though the language evolved during its 2500 year lifetime, verbal forms did not change dramatically. One typical change was the disappearance of the final *m* from the infinitive. In old Akkadian, the infinitive of the verb *prs* was *parisum* whereas in later states of the language it became *parasu*.

Akkadian also varied slightly between northern and southern Mesopotamia. The imperfect subjunctive for *parasum* is given in the figure 6 for Babylonian and Assyrian dialects.

The combination of all the factors gives a great number of different forms (more than 1000 for each verb).

There is one more factor of verbal forms that we deliberately separate from the others: it is the phonetic factor. We have already seen an effect of this factor in an example: the form IV.1 was *ipparis* instead of **inparis*. The *n* is assimilated to the following *p*. There are other examples of assimilation (ex: ṭt > ṭṭ). There are also other transformations such as dissimilation, contraction, mutation, etc.

Phonetic transformations are almost systematic for a subset of consonants called *weak consonants*. The weak consonants in Akkadian are: α, w, j. They usually do not appear at all in actual forms. Sometimes, there are traces of these consonants: there is either another consonant or a vowel that comes from the transformations occurring in the context of the weak consonant. For instance, the following transformations occur: *aw > u, ay > i, *nw > nn.* Sometimes, there is no trace whatsoever: **wiṣi > ṣi, *imnuw > imnu.* Sometimes, however, the weak consonants remain: *wašabu* (but the form *ašabu* is also attested).

N is a semi-weak consonant: it assimilates easily, but it does not disappear.

A verb with a weak consonant in its root is called a weak verb. Forms of these verbs are difficult to recognize because all the radicals are not actually in the form. For instance, the reduplication of the second radical is important to identify the stem II and the imperfect aspect. How do we recognize this form and this aspect when the second radical is weak? When the first radical is weak, it is sometimes difficult to find the relevant entry in a dictionary. Some verbs are doubly weak and there is even one verb with all three of its radical weak. Weak verbs are not rare. For instance, in 17 forms collected in a text fragment[1], there are 9 weak forms.

Here are some examples of weak verbs forms compared with their supposed original form: *jaṣip > eṣip, *iwaššab > uššab, *banaju > banu.

4 Morphological analyzer

Recognizing Akkadian verbal forms is certainly the most difficult part of Akkadian morphology. We attacked this issue first and the result is a morphological analyzer for the verbal forms. The aim of this work is to provide some help to the Akkadian learner. This aid is twofold: first, the analyzer can help students learn strong verb conjugation; second, it helps generate hypotheses about weak forms.

There are some restrictive hypotheses:

- forms are free of suffixes such as pronouns or enclitic particles. Such suffixes are quite frequent in texts.
- the analyzer is designed to analyze the Old Babylonian dialect. It should also work for some forms of other dialects, but not all of them. Most grammars describe this dialect

[1] Codex Hammurapi, items 228 to 233

	1 (no infix)	2 (infix t)	3 (infix tn)
I	basic form	reciprocal sometimes reflexive separative (motion verbs)	habitual iterative
II	factitive elativish frequentative	passive to II.1	habitual iterative
III	causative (action verbs) factitive (state verbs)	passive to III.1	iterative to III.1
III/II	identical to III	does not exist	does not exist
IV	passive to I.1 sometimes reflexive	does not exist	iterative to IV.1

Figure 5: stems semantics

gender	number	singular		plural	
		Babylonian	Assyrian	Babylonian	Assyrian
male	3	iparrasu	iparrasuni	iparrasu	iparrasuni
femel	3	taparrasu	taparrasuni	iparrasa	iparrasani

Figure 6: examples of dialectal differences

first and the other ones by the difference to this basic dialect. We have used a corpus for this dialect, namely the Hammurabi code (Szlechter, 1977).

- the length of vowels is not taken into account. Each vowel may be short or long, but the length is not always explicit in writing.

The analyzer has two levels. The first describes the complete paradigm for strong verbs without any transformation. The second describes transformations that may apply on a given form. The two-level approach of morphology is classical (Sproat, 1992). We adopted a simple model where the two levels are sequential processes with no strong interaction.

4.1 Strong verb paradigm

Conceptually, the first level of the analyzer is a finite language. We have a finite number of parameters: the root is a consonant triple and there are only a finite number of consonants. Within this domain, not all triples are confirmed roots. Each of the parameters discussed in the previous section ranges over a finite domain. If we consider all the combinations of these parameters, they are finite in number.

Though finite, the language is quite large. Enumerating all the forms is not tractable, so a grammar must be written. The natural way to describe such a language is probably a Finite State Automaton (FSA). Conceptually, our grammar of Akkadian verbal forms may be seen as an FSA, but formally, it is a Prolog Definite Clause Grammar (DCG). There are several reasons for this. First, it is a concise way to describe the FSA. A single DCG rule may implement a number of FSA transitions. Second, it gives procedures to use the FSA either for parsing or generation. Third, Prolog is convenient for computations with partial information.

This grammar is a mid-size grammar, with 162 rules in the current version. A form is described in several slices. At first, we attempted to divide forms into three parts: the prefix, the root, and the suffix. It was just too difficult to design these three parts, so we split the forms in smaller slices. There are now 9 parts:

- the personal prefix, which depends mainly on the number and gender of the subject.
- the stem prefix, which depends on the stem.
- the infix, which is placed before the first

radical if there is a stem prefix

- the first radical. This consonant never varies, but its vocalization does, depending on many factors, including the aspect, the stem, and the infix.
- the infix, which is placed after the first radical whenever there is no stem prefix.
- the second radical reduplication
- the second radical (and its vocalization)
- the third radical
- the ending (suffix), which depends on either the subject's gender and number, or on the verb's mode and aspect.

Each of these parts of a form is described using a proper non-terminal. The experiment proved that this slicing is tractable, but we believe that it is not optimal. For instance, the description of the third radical is trivial, whereas the second radical with its vowel is complex (33 rules).

The grammar in its current state implements many, but not all, of the verbal forms. The infinitive, participle, and verbal adjective are not fully implemented. More precisely, the declension, which is the nominal declension for the first two, and the adjectival declension for the latter, are not described. The other forms may all be generated by the grammar.

This grammar has been carefully tested. It is written in pure Prolog, so it is reversible, and the grammar may be used either for parsing or generation.

Currently, our grammar does not use any dictionary because we do not have any Akkadian dictionary or any Semitic root dictionary in an electronic form. We did not want to rely on non-existing resources, but the results are not as satisfactory as they would have been with a good lexical source. In parsing mode, the grammar does not actually recognize verbal forms, but gives a possible interpretation of the form. The proposed root has to be checked in a dictionary.

For a delimited corpus such as Hammurabi code, we can make a comprehensive dictionary of verbs. It is easy to interface our grammar with this lexical information. We have not yet tested whether this greatly enhances performance.

4.2 Phonetic transformations

The second level of the morphological analyzer describes the transformations that may apply on a given form. This level is not a grammar, as we are not trying to recognize a language, but to rewrite words. There is one word in input and one or several in output.

The focus of our work, designing a model of the transformations due to phonetic phenomena, is quite difficult. We started with a set of rewrite rules (given in (Ryckmans, 1960)) that we completed with other rules when required by a weak form from our corpus. These rules are simple and somehow context-free. The same rules apply on the beginning and ending of verbs. The same transformations apply on infixed t and on radical t. Neither the length of vowels nor tonic accent is taken into account. The model is therefore simplistic and it overgenerates: a rule may be applied even to some contexts where it should not.

Furthermore, it is very influenced by the set of weak forms that we have considered. Somehow, one can say that the set of rules is sufficient to give the good interpretation of all these forms, among other interpretations that are not all satisfactory. We cannot predict how the set of rules will act on other weak forms. It is likely that several other rules will be added to handle cases not yet encountered; we must consider a large set of examples.

Some transformations are very systematic (for instance the assimilation of the prefixed n for stem IV) while others are not (for instance, the dissimilation $bb > mb$). At the moment, the model does not give the probability that a rule will apply (this is a difficult computation). Since the model is non-deterministic, the application of a rule is never mandatory.

Intuitively speaking, rules are perceived as the formalization of a temporal evolution. The left-hand side of the rule represents the original form, and the right-hand side its form after the passage of time. But in our application, rules are used in the other direction. We have retrieved some attested forms from certain texts, and we want to deduce their original form, which is recognized by the Final State Automaton.

Going from an actual form to its possible prototype is difficult, mainly because the transformation process tends to shorten words. Consider the typical rule $ij > i$. If you apply it backwards, you may change any i to an ij. In fact, if most ij became i, few i come from ij. Most transformation rules have this quality.

Using the set of rule as a rewriting system is not adequate because it is does not converge – there is a termination problem. Even with only one rule $ij > i$, using it backwards would produce unbounded sequences of j. This is not only a computational drawback, it is also phonetically irrelevant.

Instead of a rewriting system, rules are used to define a transducer. Whenever rule composition seems possible, we just add this composition as a new rule to the set. The transducer has no loop and the transducing process terminates. Of course, it is a non-deterministic transducer.

We implemented the transducer in pure Prolog so that it can also be used to generate possible forms. The results obtained in generation, however, are difficult to interpret. The transformation model is too approximative to produce actual forms.

The complete code contains 47 prolog clauses.

The transducer, as it is implemented now, is not very satisfactory: it is a raw and naive implementation that we used to validate our approach. It gives some interesting results that we summarize in the next subsection.

The main problem encountered at the moment is efficiency. The transducer is non-deterministic and so generates many possible forms. For instance, a weak consonant may be inserted almost everywhere in a word. Prolog's procedural strategy results in enumerating all of the solutions. The transducing process is therefore exponential in the length of the input verbal form (this can be felt during experiments). Whereas the shorter forms (often the weak forms) are processed quickly, the computation of the complete set of solutions for the longer forms (up to 10 characters) may last several hours.

While the quality of the results of the morphological analyzer are quite satisfactory, its efficiency is not. The first level of the analyzer, namely the finite state automaton is efficient, but the transducer is not. We view several ways to solve this problem.

The first solution consists in changing the procedural way to execute the transducer, especially the way non-determinism is handled. With Prolog, the alternative solutions are found one after the other, using backtracking. An alternative solution would involve computing a single data structure to represent all the solutions, with the common parts of the different solutions shared. This would break down the complexity, since the rules encoded in the transducer apply independently on the different part of the input string. A regular expression would be the natural data structure to represent a set of strings with sharing. This form is suitable for parsing with the FSA. Parsing in this case consists in computing the intersection of two regular languages. This is a well-known operation (see for instance (Hopcroft and Ullman, 1979)).

Another idea to improve efficiency is to predict where the transducer should insert weak consonants. This could be done by a rough analysis based on consonant count.

4.3 Results

We have developed and tested the morphological analyzer using verbal forms from several sources. We collected 122 forms in (Caplice and Snell, 1988), 87 strong and 35 weak. We also used 54 forms found in the Hammurabi Code, mainly in articles 185 to 233, but also from other various articles. This is only a small subset of the verbal forms occurring in the code.

The first result is that all these forms are recognized by the morphological analyzer with the relevant interpretation. This is not a surprise, since we augmented the transducer in order to obtain the desired result. The point is: does the analyzer give wrong interpretations? It does indeed, sometimes, but its behavior is generally correct.

First, we tested 60 strong forms. On these, only two have been interpreted as possible weak verbs: *ṣabat* and *ritgum*. For *ṣabat*, three possible roots were identified: *ṣbt* (which is the correct hypothesis), $ṣ\alpha b$ and $ṣb\alpha$. The interpretations given by the analyzer for the two later are the following: the form is taken as a stative, feminine, third, person, stem I, $ṣa\alpha bat > ṣabat$ and $ṣab\alpha at > ṣabat$. This seems plausible. Concerning *ritgum*, the ambiguity comes from the t which is a radical but may be interpreted as an infix and from the m which is the mark of the ventive but can be seen as the third radical. here again, the proposed root is plausible. Surprisingly, some forms very close to the two ambiguous ones are not ambiguous.

The most ambiguous form in the data we considered is *iddu*, for which 15 roots have been computed. The right explanation of the form is **indiju* > *iddu*. There are two transformations: assimilation of the *n* and contraction of *iju*. It is a typical example of a form that is difficult to understand for the Akkadian learner. Even if there are many hypotheses, the answer given by the system may help in such a case. The help would be much better if the system had a complete Akkadian root dictionary.

The verb *alakum* which is sometimes said to be irregular (see, for instance, (Heise,)) is treated as the other verbs by our system. We followed the interpretation of its forms (radical α assimilated to the radical l) found in (Ryckmans, 1960). This gives satisfactory results.

The main weakness identified so far is in the aspect discrimination between imperfect and preterit for weak verbs with a second radical weak. For these verbs, the main difference between the two aspects, namely the second radical reduplication, is not perceptible. In that case, the vowel is significant. For instance, the verb *kanum* has a root *kwn*. The preterit is **ikwun* > *ikun* and the imperfect *ikawwun > ikan. The morphological analyzer proposes either preterit and imperfect as possible aspects for *ikun*. It is not possible to prevent the mutation of the *w* in *u* in this case, because such a mutation sometimes occurs in other contexts.

Generally speaking, the morphological analyzer gives the right solution, but also proposes other ones. These other ones are often acceptable, but sometimes, as shown by the latest example, they are not.

5 Conclusion

The work we have done so far shows that many Akkadian verbal forms can be interpreted using a single conjugation paradigm and a phonetic transformation model.

We think that our approach is a good one for Akkadian, due to the language peculiarities. Is this approach well-suited for other languages? We do not know.

The basis for our work is that we model phonetic transformations for a language with a phonetic writing. The Akkadian writing system is phonetic and syllabic. As far as we know, it is not the case of other Semitic languages. For instance, they do not transcribe vowels. The results obtained so far show that the vocalization in Akkadian breaks down ambiguity.

It is not obvious that our approach is suitable for languages other than Akkadian, for which it is quite convincing.

The work described here is in progress. We have to study the work done for the morphological analysis of the other semitic languages. We also have to search for a better way to perform the second level of the analysis.

The morphological analyzer presented in this paper could be enhanced by expressing the transformation rules more contextually and by coupling the two levels.

References

R. Caplice and D. Snell. 1988. *Introduction to Akkadian*. Biblical Institute Press, Rome, Italy.

J. Heise. http://saturn.sron.ruu.nl/~jheise/akkadian.

J.E. Hopcroft and J.D. Ullman. 1979. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.

G. Ryckmans. 1960. *Grammaire Accadienne*. Publications Universitaires de Louvain, Louvain, Belgium.

R. Sproat. 1992. *Morphology and Computation*. MIT Press, Cambridge, Massachussetts.

E. Szlechter. 1977. *Codex Hammurapi*. Pontifica Universitas Lateranensis, Rome, Italy.

6 Appendix

We did not know a previous work on Akkadian morphology by Kataja and Koskenniemi (1988) until a referee report for this workshop.

We do not have the time to make a complete comparison between their work and ours. Furthermore, the paper does not give all the details we need to make such a comparison. We can make some brief comments however:

- the two systems split the analysis in two parts: morphotactic and phonology.
- the morphotactic level in Kataja and Koskenniemi is based on a linguistic model.

It is more convincing than our morphotactic description. Our intermediate lexical representation is too written-oriented, and it is not the adequate input for the second level which deals with phonetic phenomena.

- the phonological approach seem very similar in the two works.
- Our system is deliberately non-deterministic, due to the fact that several writing of the same form may be found even in a single Akkadian text.
- the Kataja and Koskenniemi description is complete. It is not limited to verbal forms. The phonological description is said “fairly complete and tested”.
- it is not clear that the problems we encountered are solved by the other system. For instance, the ambiguity between preterit and imperfect for verbs with their second radical weak. We need the complete rule set to answer to this question, but the paper gives only examples.

The following references, given by one of the referees as relevant to our work, were not used for lack of time.

References

Kenneth R. Beesley. 1990. Finite-state description of Arabic morphology. In *Proceedings of the Second Cambridge Conference on Bilingual Computing in Arabic and English*, September 5-7. No pagination.

Kenneth R. Beesley. 1996. Arabic finite-state morphological analysis and generation. In *COLING'96*, volume 1, pages 89–94, Copenhagen, August 5-9. Center for Sprogteknologi. The 16th International Conference on Computational Linguistics.

Laura Kataja and Kimmo Koskenniemi. 1988. Finite-state description of Semitic morphology: A case study of Ancient Akkadian. In *COLING'88*, pages 313–315.

Martin Kay. 1987. Nonconcatenative finite-state morphology. In *Proceedings of the Third Conference of the European Chapter of the Association for Computational Linguistics*, pages 2–10.

George Kiraz. 1994. Multi-tape two-level morphology: a case study in Semitic non- linear morphology. In *COLING'94*, volume 1, pages 180–186.

George Anton Kiraz. 1996. Computing prosodic morphology. In *COLING'96*.