Computational Modeling of Verbs in Dene Languages: The Case of Tsuut'ina

Antti Arppe[1], Christopher Cox[2], Mans Hulden[3], Jordan Lachler[1], Sjur N. Moshagen[4], Miikka Silfverberg[3], and Trond Trosterud[4]
[1]University of Alberta, [2]Carleton University, [3]University of Colorado Boulder, [4]University of Tromsø

1 Introduction

The objective of this paper is to present a general computational architecture for practically modeling the complex morphology of verbs in Dene languages, using the finite-state approach with Tsuut'ina as a case example, to serve as a basis for various software applications that can be used for the revitalization and continued maintenance of these languages.

2 Morphological Structure of the Dene Verb

The morphological structure of verbs in Dene languages is considered to be about as complex as it can get among the languages of the world. However, the overall structure, with outer (*disjunct*), inner (*conjunct*), and stem 'zones' of verb (cf. Kari 1989) is not generally, and thus computationally, difficult to model. What is seen as most challenging primarily concerns the extensive morphological fusion in subject-aspect inflection immediately preceding the stem (cf. K. Rice 2005: 404–407), a span which "is at least historically a concatenation of morphemes before the verb stem" (K. Rice 2005: 404). As Keren Rice (2005: 405) delicately puts it, "[t]he morphophonemics of this span of the verb is complex."

In the following section 2, we will discuss in further detail the morphological structure of Tsuut'ina verbs, and how their lexical and inflectional elements can be generally analyzed and modelled. In this, we will contrast the templatic and paradigmatic approaches to representing the structure of Tsuut'ina verbs, in particular concerning inflectional elements in the inner inflectional region immediately preceding the stem in (the conjunct zone), since this will form the basis of the computational modeling solution that we will present in Section 3.

2.1 Lexical elements

Traditional morphological accounts of Dene verbs may present well over ten prefixal positions preceding a single-syllable stem, with complex morphophonological processes applying at each of the multiple morpheme junctures (see Table 2.2 below). Moreover, the combinations of some these prefixal elements with stems can be seen to represent distinct lexical meanings, while other prefixes have an inflectional character. For instance, in Tsuut'ina the simple stem *-tsiy* corresponds to 'cry' (in the imperfective aspect), whereas the prefix-stem sequence *ts'á-* + *-zíd* means 'wake up', and the triplicate discontiguous sequence of prefixes and a stem *nà-* + *gu-* + *di-* + *-tłod* is understood the denote 'jump down'. It is crucial to note that these prefix-stem combinations are often quite arbitrary, and together make up the basic meaning of the verb; when prefixes play such a lexical meaning role, we will refer to them as *lexical prefixes* below. The prefixes that appear in these constructions may bear some resemblance to other prefixes that have an identifiable lexical meaning, or may only ever appear in collocation with a certain stem. In addition, since morphemes from anywhere in the verb can be co-opted into these arbitrary lexical combinations, it often happens that inflectional morphemes appear between lexical prefixes at different points within the verb form, among the various lexical prefixes. Nevertheless, the locations of the lexical prefixes in the pre-stem template are mostly well-known, in positions we will denote as the *inner*, *middle*, and *outer* boundaries, as seen in Table 2.1. We use the term *"middle"* lexical boundary to indicate where inflectional morphemes indicating so-called 'outer' subjects (fourth person singular and third person plural) and direct objects appear (slots 6-7 in Table 2.2 below). With this one exception, the "*outer*" and "*inner*" lexical boundaries otherwise correspond with the traditional disjunct and conjunct zones, respectively (cf. Kari 1989).

Type of combinatorics	**Outer prefixes**	**Middle prefixes**	**Inner prefixes**	**Stem**
No lexical prefixes	-	-	-	*-tsiy* 'cry'
Inner lexical prefix + stem	-	-	*di-*	*-tł'á* 'run'
Middle lexical prefix + stem	-	*gu-*	-	*-náh* 'speak'
Outer lexical prefix + stem	*ts'á-*	-	-	*-zíd* 'wake up'
Middle + inner lexical prefixes + stem	-	*gu-*	*di-*	*-tìs* 'walk with a cane'
Outer + inner lexical prefixes + stem	*tsí-*	-	*di-*	*-tł'á* 'run away'
Outer + middle lexical prefixes + stem	*k'à-*	*gu-*	-	*-nát* 'finish talking'
Outer + middle + inner lexical prefixes + stem	*nà-*	*gu-*	*di-*	*-tłod* 'jump down'

Table 2.1. Possible combinations and locations of lexical prefixes and stems in Tsuut'ina.

Furthermore, in many cases the individual stems, and in some cases also the prefix-stem combinations, can have multiple allomorphs associated with the 4-5 aspects/moods, as is exemplified in Table 2.2. As can be seen, for the imperfective, perfective and progressive aspect forms of e.g. *-tsiy* 'cry', it only has a set of allomorphic stems but no lexical prefixes, whereas for the repetitive aspect, not only is there a fourth allomorphic stem but also lexical prefix at the outer inflectional boundary. Nevertheless, such stem allomorphs as well as the possibly allomorphic lexical prefixes and their positions at the outer, middle, and inner inflectional boundaries are well known, which is demonstrated in Table 2.2. below for a number of lexical meanings besides *-tsiy*, with a variety of combinations of lexical prefixes and stems.

Table 2.2. Examples of the allomorphy of stems and lexical prefixes, corresponding to the 4-5 aspects/moods for a selection of lexical meanings, *-tsiy* ‘cry’, *di-* + *-tł’á* ‘run’, *gu-* + *-náh* ‘speak’, *ts’á-* + *-zíd* ‘wake up’, *tsí-* + *di-* + *-tł'á* ‘run away’, *nà-* + *gu-* + *di-*+ *-tłod* ‘jump down’ (here referred to by their imperfective aspect forms). The numbers 0-12 refer to the templatic slots illustrated in Table 2.3 below in Section 2.2.

DISJUNCT: 12-9		**CONJUNCT: 8-1**				**STEM: 0**
Outer Lexical Prefix 11–9[1]	Outer Inflectional Chunk 8	Middle Lexical Prefix 7	Middle Inflectional Chunk 7-6	Inner Lexical Prefix 5	Inner Inflectional Chunk 4-1	STEM/ (ASPECT) 0
- - - ná		- - - -		- - - -		tsiy (IPFV) tsày (PFV) tsíł (PROG) chish (REP)
- - - -		- - - -		di di di di		tł'á (IPFV) tł'ò (PFV) tł'áł (PROG) tł'ásh (REP)
- - - ná		gu gu gu gu		- di - di		náh (IPFV) nát (PFV) náł (PROG) násh (REP)
ts'á ts'á ts'á ts'áná		- - - -		- - - -		zíd (IPFV) zid (PFV) ził (PROG) zhiizh (REP)
tsí		-		di		tł'á (IPFV)

[1] This table does not represent inflection on incorporated postpositions (slot 12 in the morphological template given below), but can trivially be extended to incorporate these patterns using the same computational methods discussed in the following sections.

tsí tsí nátsí		- - -		di di di		tł'ò (PFV) tł'áł (PROG) tł'ásh (REP)
nà nà nà nàná nìná		gu gu gu gu gu		di di di di di		tłod (IPFV) tłot (PFV) tłíł (PROG) tłiizh (REP) tłiizh (REP)

While the lexical components of a Tsuut'ina verb can be described in a relatively straight-forward fashion as described above, we yet need to deal with the inflectional morphology that can be seen to intervene with the lexical components at three boundaries. Two alternative approaches for dealing with this are discussed below in Section 2.2.

2.2 Inflectional elements

2.2.1 Templatic morphology

Traditionally, Dene linguistics has relied on a morphological template and extensive morphophonemic rules to handle the complexities of this part of the verb. In most such traditional approaches to Dene morphology, a verb like this would be modelled with a morphological template - a slot-and-filler system for abstract morphological parts, illustrated for the basic Tsuut'ina verb *gámíł* 'they are swimming along' in (Table 2.3). In particular, the syllable immediately preceding the stem would be analyzed as constituting a morphological complex, typically representing a combination of 4-6 distinct morphemes, including (non-third) subject person and number, aspect, mode, conjugation class (i.e. *si*-, *ni*-, *yi*-, *ì*-, or none, representing one of the historical Dene aspect/conjugation class markers), and voice/valence (i.e. the historical voice/valence markers or "classifiers" that fuse phonologically with the preceding prefixes), with substantial morphophonological interactions yielding us the actual forms.[2]

[2] The status of *ì*- as an aspect/conjugation marker in Tsuut'ina remains unclear. Unlike the rest of the markers in this set, *ì*- co-occurs with other aspect/conjugation markers (e.g., *ìsissìl* 'I got hot', containing both *ì*- and *si*-), which makes its distribution similar to some qualifier prefixes. Unlike other qualifier prefixes, however, *ì*- fuses in phonologically unpredictable ways with certain aspect/conjugation markers (e.g., *ìzìl* 'he/she/it got hot', where, unlike all other 3S *si*- forms in Tsuut'ina, contains no overt reflex of *si*- itself).

Table 2.3: Tsuut'ina templatic analysis of *gámíł* 'they are swimming (progressive)' (after Cook 1984: 126, 278–281). The inner prefixes encompass positions 1–5, the middle prefixes 6–8, and the outer prefixes 9–12.

#	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
Name	Incorporated PP	Adverbial	Iterative (*ná-*)	Incorporated stem	Distributive (*dà-*)	Object	Deictic subject	Thematic	Aspect (Mode)	Aspect	Local subject	Classifier	STEM
Morpheme(s)	-	-	-	-	-	-	*gi-*	-	-	*yi-*	-	-	*-míł*
Feature(s)	-	-	-	-	-	-	3PL SBJ	-	-	yi-PROG	-	-	swim-PROG

Applying a template like this to analyze or generate inflected verb forms is non-trivial: it often requires extensive knowledge not only of the relevant morphemes and their co-occurrence restrictions, but also of their relative ordering and possible morphophonemic adaptations that could affect their form (see Rice 2000, McDonough 2003, for recent critiques of the traditional templatic approach to analysis). In this case, as illustrated in (1a-c) from Cook (1984: 278–281), either breaking down or building up an actual inflected verb form like (1c) *gámíł* requires being able to get at (1a) its abstract parts, and (1b) the abstract rules needed to fuse these parts together, and neither of these are necessarily all that easy to produce on the fly.

(1a)		gi-	yi-	míł	**Abstract parts**
		they	he,she,it:PROG	swim:PROG	
					Abstract rules:
(1b)	1.	gi-	**yi- → á-**	míł	(*ɣi*-Augmentation)
	2.	gi- → g-	á-	míł	*i*-Deletion)
(1c)		gámił			**Actual word**
		'they are swimming'			

Not only is this kind of structure more difficult to model computationally (requiring many more morphophonemic adaptations, and hence much more testing to make sure these kinds of rules work as expected in all but only those cases where they should apply), but it is not clear that general-purpose morphophonemic rewrite rules will necessarily do the trick in a reasonably manageable fashion. The analytical and practical challenges posed by the relative opacity of Dene verbal morphophonology have often been noted in the wider literature. Kari (1975: 331), commenting on the "notorious" nature of morphophonological alternations in Dene verbs, recalls that "[i]n concluding his review of Hoijer's *Navaho Phonology*, Harris remarked, 'In addition to all the regular or partially regular alternations (in Navajo), there are so many individual replacements that a set of general rules for deriving phonemic forms, like those devised for Tubatulabal by Swadesh and Voegelin and for Menomini by Bloomfield seems impossible here.'" Difficulties like this are thus not exceptional, and provide additional motivation for exploring other means of modelling this central aspect of Dene verbal structure.

2.2.2 Paradigmatic approaches

An alternative approach to maximal decomposition/splitting in the analysis of inner inflection is to treat that combination of theoretically posited complex sequences of abstract morphemes and associated features rather as pre-composed *chunks* or *portmanteaux* (sometimes referred to as the T-A-M-A, or tense-aspect-mode-agent, chunk). This is also supported by psycholinguistic evidence suggesting that word complex structure is mentally represented and processed as larger chunks (see e.g. S. Rice, Libben & Derwing, 2002, for related results involving Dene Sųłiné). The practical advantage from both a general and computational linguistic modeling perspective is that chunking significantly reduces the amount of morphophonemic processes one has to model, as only junctures between the chunked zones need to be treated. This reduction in complexity is particularly valuable in treating the inflectional prefixes that immediately precede the verb stem, whose morphophonological composition is not always predictable (K. Rice 2001). Indeed, there are precedents for such as a chunking approach within the Dene linguistic tradition, e.g. Young & Morgan (1987), Faltz (1998), McDonough (2000) for Navajo, and Leer (1999, *inter alia*) for comparative Na-Dene.

Therefore, in the case of the same example as above (1c), rather than try to derive a form like *gámíł* by parts and rules, we essentially just classify this verb in our lexicon as inflecting like any other verb of the *yi*-Progressive type, and then look up the proper third person plural prefix from a table that gives us the appropriate allomorph. In other words, we do not try to derive these parts on the fly: we look up the chunks, which allows us to sidestep all of the complicated morphophonemics that we would otherwise have to painstakingly model.

In support of taking this approach, it is crucial to note that in fact there are not a prohibitively large number of inflectional patterns to choose from for this "inner" inflection. In the case of Tsuut'ina, verbs generally fall into one of only ten classes, with different allomorphs within each pattern depending on what kind of prefixes appear before them (the columns here) and what voice-valence markers they come with (the rows here): Ø, *s*-, *i*-. Table 2.3 presents an example of one of these classes, summarizing the patterns of allomorphy that appear with *yi*-progressive verbs under different prefixation conditions.

Table 2.3: Inner inflection patterns for *yi*-progressive verbs.

Type	No prefix			Inner or middle prefix			Outer prefix		
Basic	"Be crying": (tsáł)			"Be laughing": (di-dlùw)			"Be becoming like that": (xa=nááł)		
	1S		yistsáł	1S		diyisdlùw	1S		xayisnáł
	2S		yítsáł	2S		diyídlùw	2S		xayínáł
	3S		yitsáł	3S		dádlùw	3S		xayináł
	1P		yaàtsáł	1P		diyaàdlùw	1P		xayaànáł
	2P		yastsáł	2P		diyasdlùw	2P		xayasnáł
	3P		gátsáł	3P		gidádlùw	3P		xagánáł
	4		ts'átsáł	4		ts'idádlùw	4		xats'ánáł
With *s*-	"Be tickling O": (wùsh)			"Be almost swallowing O": (di-nìł)			"Be taming, training O": (á=díł)		
	1S		yiswùsh	1S		diyisnìł	1S		áyisdíł
	2S		yíswùsh	2S		diyísnìł	2S		áyísdíł
	3S	O	yáswùsh	3S	O	dásnìł	3S	O	áyisdíł
	1P		yaàwùsh	1P		diyaànìł	1P		áyaàdíł
	2P		yaswùsh	2P		diyasnìł	2P		áyasdíł
	3P	O	gáswùsh	3P	O	gidásnìł	3P	O	agásdíł

	4		ts’áswùsh	4		ts’idásnìł	4		áts'ásdíł
With *i-*	“Be running along”: (tł’áł)			“Be starting to run along”: (didi-tł’áł)			“Be running along home”: (ná=tł’áł)		
	1S		yistł’áł	1S		didiyistł’áł	1S		náyistł’áł
	2S		yíítł’áł	2S		didiyíítł’áł	2S		náyíítł’áł
	3S		yiitł’áł	3S		didáátł’áł	3S		náyiitł’áł
	1P		yaàtł’áł	1P		didiyaàtł’áł	1P		náyaàtł’áł
	2P		yastł’áł	2P		didiyastł’áł	2P		náyastł’áł
	3P		gáátł’áł	3P		gididáátł’áł	3P		nágáátł’áł
	4		ts’áátł’áł	4		ts’ididáátł’áł	4		náts’áátł’áł

Importantly, it should be noted that this approach does not preclude other active phonological or morphophonological processes affecting these verbs from being modelled separately, without requiring this paradigm-based approach to be set aside altogether. In some Dene languages (e.g., Navajo and Tahltan, among others; cf. Shaw 1991, McDonough 2003), for instance, additional phonological changes would be needed to represent well-known cases of coronal harmony that may affect the forms taken by the prefix combination allomorphs given in tables like the one above (e.g., *zààs yiswùsh* ‘I am tickling the baby’ being realized as *zààs yi**sh**wùsh* in Tsuut’ina).[3] Not only is making a separation between the semi-predictable inflectional patterns summarized in these tables and the more regular processes of phonological adaptation attested elsewhere in the language potentially worthwhile for overall analytical clarity, but the latter changes can also be treated straightforwardly with the computational tools discussed below. In conclusion, it is the paradigmatic, chunking approach described immediately above that we adopt in our computational model of Tsuut’ina, the implementation of which we discuss below in Section 3.

[3] Although coronal harmony is active in Tsuut’ina, the Tsuut’ina orthography does not conventionally show harmonized [-anterior] forms except when they appear in stems (e.g., /tsʰ/ → /tʃʰ/ in ***tsiy*** ‘cry.IPFV’ vs. **ch**ish ‘cry.REP’), which makes this less of an issue for computational modelling than with some other Dene languages.

3 Computationally modeling the Dene Verb

Finite-state machines (FSMs) (see e.g. Beesley & Karttunen 2003) have become one standard way for computationally modeling the morphological structure of words in natural languages. There are currently several open source implementations of FSM compilers, e.g. *xfst* (Beesley & Karttunen 2003), *foma* (Hulden 2009) and *hfst* (Lindén et al. 2011). The key advantages of FSMs are many: they have well-known computational properties, having been tested and honed over several decades of theoretical and practical development, they are extremely fast and use efficient computational data structures, they have a calculus for powerful manipulations which is in particular useful for our Tsuut'ina solution, they allow for a rule-based definition of paradigms for various verb subtypes (however one wants to define these), thus not requiring large corpora from which to learn such rules - that is impossible in the case of most endangered languages such as the Dene ones, with scant resources - they are portable to the main operating systems currently in use, and on the basis of the factors mentioned before they allow for easy integration with other software applications, for instance as spell-checking modules within word-processors, morphologically "intelligent" electronic dictionaries, and "intelligent" computer-aided language-learning applications (Trosterud 2004, 2006; Arppe et al. 2015; Antonsen et al. 2013; Johnson et al. 2013). As their disadvantages one might note that FSMs are only as good as the available linguistic descriptions, and some may consider them too powerful in that the rules allow for the recognition and generation of complete paradigms, where all individual cells have not been verified by speakers, and it is not a given that all the forms of the paradigm make sense for all verbs belonging to some paradigm type - nevertheless, this is a general problem with any rule-based system, and in practice impossible to resolve when the number of paradigms cells number in the hundreds if not thousands.

For many languages with a rich morphology, a FSM is based on one layer declaring allowable morpheme/feature sequences (i.e. morphotactics), coupled with a set of morphophonological rules applied at the morpheme junctures. In the case of Dene verbs, extending on work by Hulden and Bischoff (2008) for Navajo, we present a somewhat variant approach, where the morphological/morphotactic structure of Dene words is understood to consist of not a single but two interleaving layers, (1) one representing the lexical elements, including the stem combined together with potentially discontiguous lexical prefixes, and (2) the other the intervening inflectional elements.

Crucially, this builds upon (1) knowing the exact location of the various lexical prefixes, so that the intervening inflectional chunks can be appropriately incorporated, as well as (2) specifying the allowable inflectional patterns (chunks) associated with each lexical construct. This is implemented by defining, with the *lexc* component of the FST formalism, four FSMs, (1) one for the lexical level[4] and (2) one for each intervening inflectional chunks (at the inner, middle, and outer boundaries)[5], with (3) co-occurrence constraints among the two levels (1) and (2), determining which subset of morpheme chunks are allowable in the three inflectional slots, implemented using flag diacritics in the *lexc* component of the FST formalism.[6] The slots for inflectional morpheme chunks at the inner, middle, and outer boundaries are indicated in the lexical tier FSM by special characters which cannot occur in the morphemes, being '.' (period) for the inner boundary, '_' (underscore) for the middle boundary, and '=' for the outer boundary. These four component FSMs are then composed together with the help of finite-state calculus into one FSM representing the entire morphotactics by replacing the special characters indicating the inner, outer and middle boundaries in the lexical FSM with the entire corresponding inflectional FSMs.[7] The resulting morphological FSM is next composed together with a fifth FSM representing morphophonological processes applying at the morpheme chunk boundaries, expressed by a set/series of rewrite rules using *twolc* component (or the rewrite rule component) of the FST formalism,[8] which is further concatenated with a sixth FSM defined with the *lexc* component for linking the morphological features, incorporated in the flag diacritics within the morphological FSM, with morphological feature tags, which will be outputted in word form analysis and inputted for word form generation.[9] The overall structure of the six constituent

---

[4] https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/stems/verbstems.lexc

[5] https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/affixes/verb_inner_affixes.lexc
https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/affixes/verb_middle_affixes.lexc
https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/affixes/verb_outer_affixes.lexc

[6] A reviewer notes that the use of flag diacritics might also allow models such as this to capture certain finer-grained linguistic generalizations about morphemic co-occurrence that are common to many Dene languages, such as the observation that *si*-paradigm verbs do not occur with reflexes of the historical **nə*- perfective marker that appears in other (*ni*- and *yi*-) perfective paradigms. We concede that generalizations such as this concerning the distribution of individual inflectional morphemes are not readily represented in the current model, which groups several conjunct-zone inflectional markers together into sets of distinctive surface-form 'chunks' that are essentially agnostic as to their possible morphemic constituents, and acknowledge that flag diacritics would provide one means of representing these distributional observations in a more intricate morphological model. We would only note that these generalizations are not required to produce a fully functional model of Tsuut'ina verbal inflection – any surface phonological 'irregularities' caused by the presence or absence of a marker like **nə*- are implicitly captured in the forms of the inflectional chunks themselves. Moreover, these generalizations would not necessarily offer any net reduction in complexity if introduced into the model, given the intricacy of the attendant morphophonological adaptation rules required to produce observed surface forms. For the purposes of computational modelling, it is enough for models such as this to use flag diacritics to select the appropriate inflectional chunk allomorphs to insert into a given lexical item, depending on the prefixation conditions under which that inflectional chunk will then appear (i.e., whether preceded by an inner, middle, and/or outer prefix). This is not to say that further analysis of the morpheme structure of individual inflectional chunks were not possible; see Young & Morgan (1987) for one example of templatic analyses being provided for inflectional forms in an otherwise paradigm-based model. Other generalizations (e.g., concerning co-occurrence patterns involving outer prefixes and particular inflectional paradigms) can typically be captured in these models through separate lexical entries, or through higher levels of abstraction in lexical definitions implemented in the formalisms discussed below.

[7] https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/verb_lexicon.xfscript.in

[8] https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/phonology/srs-phon.twolc

[9] https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/affixes/verbtags.lexc

FSMs representing the lexical, inflectional and morphophonological tiers as well as the morphological features is presented in Figure 3.1, and the master specification of the composition or concatenation of the six constituent FSMs in (2). The full source code for our Tsuut'ina FSM can be found on-line within the *Giella* infrastructure (Moshagen et al. 2013): `https://victorio.uit.no/langtech/trunk/startup-langs/srs/src/morphology/`

(2) The master specification of the composition and concatenation of the six constituent FSMs to create our computational model of Tsuut'ina verbs.[10]

```
read regex Stems .o. MarkPrefixes .o. InsInner .o. InsMiddle .o.
InsOuter .o. "." -> InnerAffixes , "_" -> MiddleAffixes , "=" ->
OuterAffixes ;
twosided flag-diacritics
define Grammar ;
read regex Grammar Tags ;
```

[10] `Stems`, `InnerAffixes`, `MiddleAffixes`, `OuterAffixes`, and `Tags` refer to FSMs that have been compiled from the corresponding *lexc* files, whereas `MarkPrefixes`, `InsInner`, `InsMiddle`, and `InsOuter` refer to rewrite rules that have been defined earlier in the master compilation specification code (2).

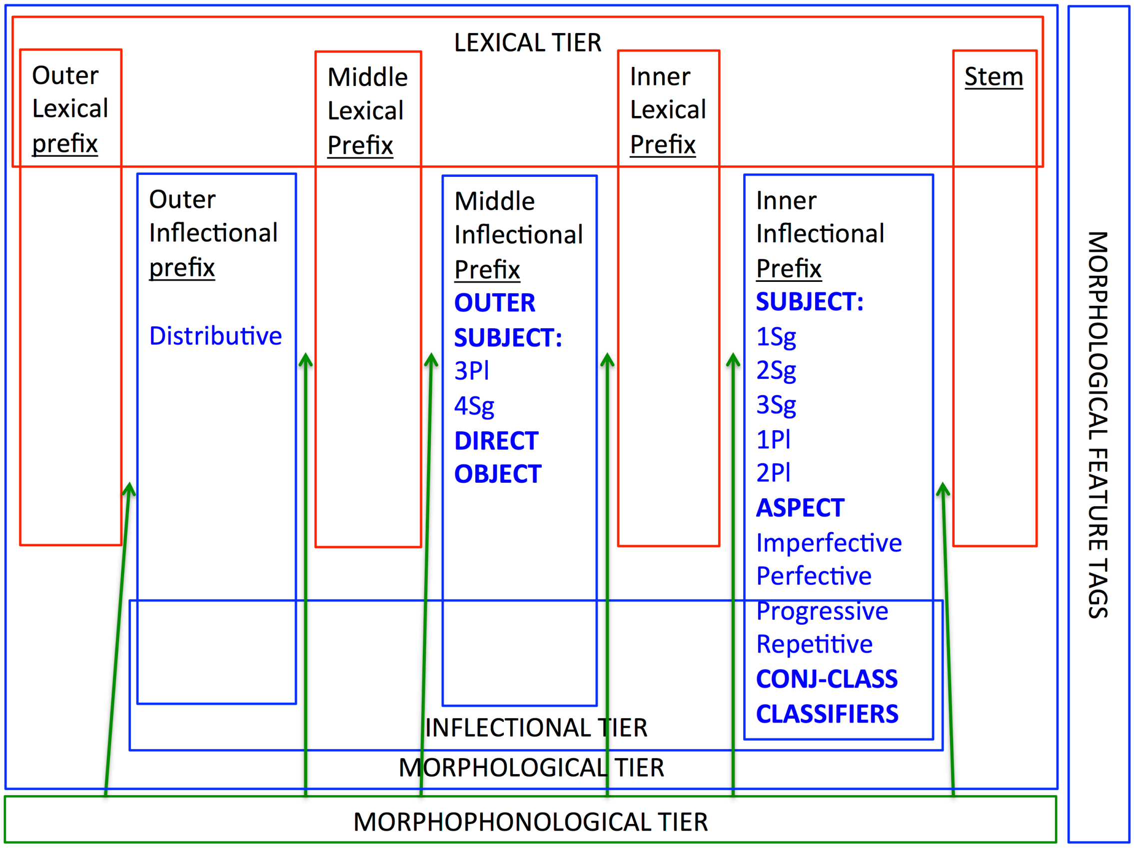

Figure 3.1. The overall structure of six components of the Tsuut'ina Finite-State Machine.

Figure 3.2 illustrates how the lexical tier is populated with descriptions of possible combinations of stems and inner, middle, and outer lexical prefixes, including any allomorphy concerning the stems and the lexical prefixes for the different types of aspect, while the inflectional tier enumerates all the subclasses of morpheme chunks in the inner (representing morphological features for the inner subject and aspect as well as incorporating conjugation class markers and classifiers: currently 66 subclasses, with another nine attested in documentation and awaiting implementation in this model), middle (outer subject and direct object), and outer inflectional slots (distributive). For instance, for the intransitive verb *itsiy* 'cry' within the lexical tier, we find for the imperfective aspect the stem *-tsiy* without any lexical prefixes, the stem *-tsày* without any lexical prefixes for the perfective aspect, the stem *-tsił* without any lexical prefixes for the progressive aspect, and the combination of an outer lexical prefix and stem *ná-*$_{\text{outer}}$ + *-chish* for the repetitive aspect. Each definition of a lexical allomorph (combination of stems and lexical prefixes) on the lexical tier is the linked with a particular subclass of the paradigm in the inner, middle and outer slots in on the inflectional tier. For instance, the imperfective aspect forms of *itsiy* 'cry' combine the stem *-tsiy* (which has no lexical prefix) with the `0-Imperfective-no-prefix` set (*0-Ipfv-no* in Figure 3.2 below), the perfective aspect forms combine the stem -

tsày with the `yi-y-Perfective-no-prefix` set (*0-Pfv-no*), the progressive aspect forms combine the stem *-tsił* with `yi-Progressive-no-prefix` set, and the repetitive aspect forms combine the outer prefix + stem *nà-...-chish* with the `0-Imperfective-outer-prefix` set (*0-Ipfv-outer*), similar to the imperfective forms. The middle inflectional prefix morphemes denoting outer subjects are the same for all the paradigm types, as is the case for the outer inflectional prefix denoting distributive. Even though the inner inflectional prefix slot does not represent the outer, third person subjects, for some paradigm subclasses the inner inflectional slot may contain some phonological elements in conjunction with a middle inflectional prefix. Further below, Figure 3.3 illustrates the set of individual morpheme chunks in the inner inflectional slot applicable within the `0-Imperfective-no-prefix` (*0-Ipfv-no*) subclass, required e.g. with the imperfective stem *-tsiy* for (which has no lexical prefixes) *itsiy* 'cry'. How all of the above is specified using the *lexc* component of the FST formalism is detailed below in Sections 3.1 and 3.2; further examples are provided in Cox et al. (2016; slides 36-47)

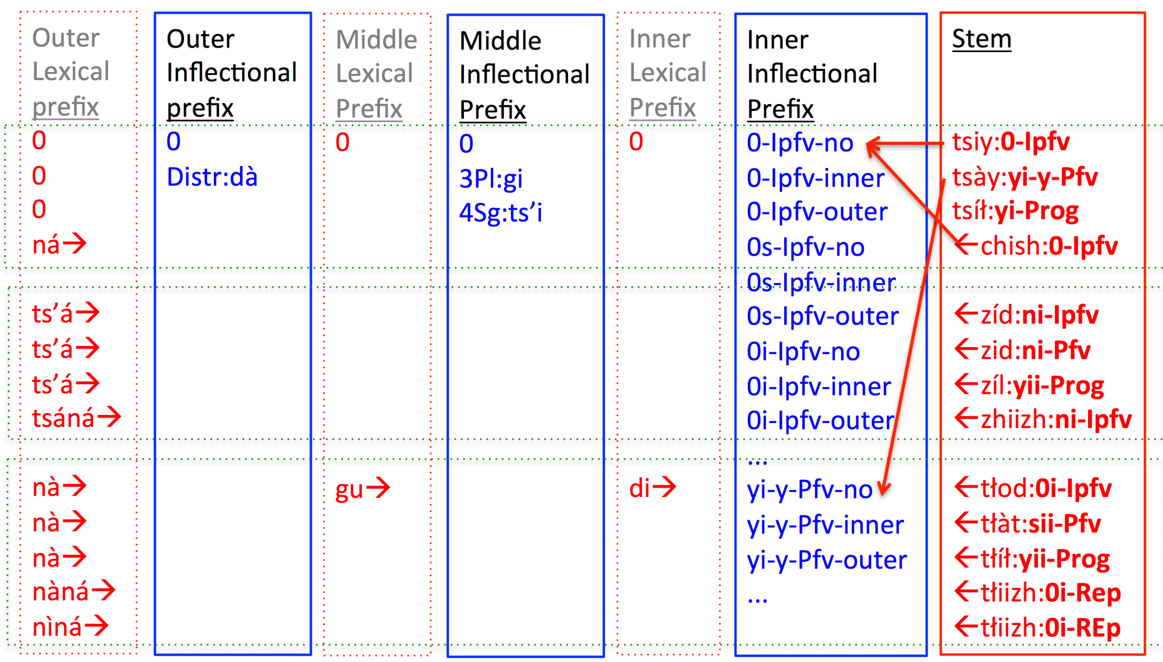

Figure 3.2. An illustration of how the disjoint lexical tier (stem + lexical prefixes) are interleaved with the outer, middle, and inner morphological slots of the inflectional tier, as well as a sample of the labels of TAMA morpheme chunk paradigm subsets for the inner inflectional slots, using as examples the various lexical prefixes and stems for three lexical meanings *-tsiy* 'cry', *ts'á-* + *-zíd* 'wake up' and *nà-* + *gu-* + *di-* + *-tłod* 'jump down', and their allomorphy and associated TAMA morpheme chunk paradigm subsets for the different aspects/moods, and the shared morphemes the middle and outer inflectional slots; the arrows indicate the applicable TAMA morpheme chunk paradigm subsets of *-tsiy*.

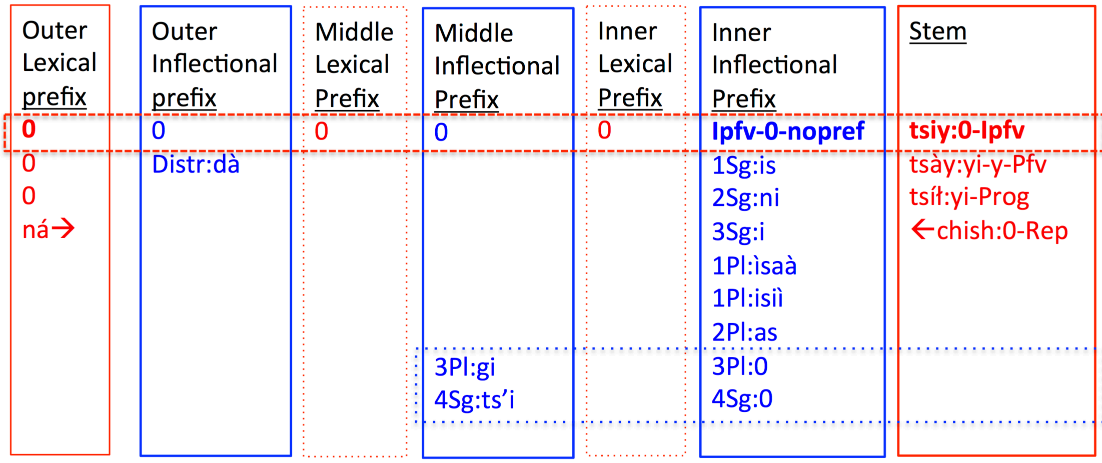

Figure 3.3. An illustration indicating the TAMA morpheme chunk paradigm subset covering inner subject forms (*0-Imperfective ~ 0-Ipfv*), applicable for the imperfective aspect of the stem *-tsiy* 'cry', which has no lexical prefixes, and thus uses the no-prefix variant (*Ipfv-0-nopref*).

3.1 Lexical tier

Within the lexical tier of our computational model of Tsuut'ina verbs, we need to specify (1) the zero to three lexical prefixes that in conjunction with a stem together constitute a verb construction with a specific lexical meaning; (2) the possible allomorphy concerning not just the stems but their combinations with lexical prefixes (where the lexical prefixes and stem can both vary), associated with the five aspects/moods a Tsuut'ina verb can represent; and (3) the subclass of the inflectional paradigm, primarily concerning the inner inflectional prefix slot that we treat as a morpheme chunk, which is applicable for each lexical prefix + stem allomorph (i.e. for each aspect/mood) of a Tsuut'ina verb.

The lexical tier consists of one FSM specified using the *lexc* component of the FST formalism, which is made up of three sequential parts, or continuation lexica: the first step `Verbstems` specifies all the aforementioned three pieces of information (1-3) in a linguist-friendly format (see table 3.1 below), the second step consists of a set of parallel, alternative continuation lexica converting the specification concerning (3) into "computer-friendly" FST-internal flag-diacritics which allow for linking the lexical tier FSM constraints with the inflectional tier (see table 3.2 below), and the third and final step, the shared continuation lexicon `Verbsuffixes`, simply terminates the lexical tier FSM. Further details on how all this is implemented in practice is presented below.

In order to deal with allomorphy, we have to select some inflected form as the lemma with which we can link the various stem + lexical prefix allomorphs together. Here, we have opted to use the *Imperfective Mood Third Person Singular* form as such a lemma, e.g. *itsiy* 's/he/it cries' for ‘cry’, *ts’ázíd* ‘s/he/it wakes up’ for ‘wake up’, *tsídiitł'á* ‘s/he/it runs away’ for ‘run away’, and *nàgudiitłod* ‘s/he/it jumps down’ for ‘jump down’.

For indicating where the one to three lexical prefixes (or none) occur at the inner, middle, and outer inflectional boundaries, as discussed above in the beginning of Section 3 and in Section 2.1, in a verb construction, we have specified a notation using special characters which should not occur in any of the actual morphemes, with ‘.’ (period) designating the inner boundary, ‘_’ (underscore) the middle boundary, and ‘=’ the outer boundary, already mentioned above. This notation is applied for each allomorphic combination of lexical prefixes and stems. Within the *lexc* formalism, the aforementioned lemma, which we supplement with an English gloss in [brackets] as the underlying form on the left-hand side in the notation, is linked with whatever allomorphic surface realizations of lexical prefixes and stems, on the right-hand side in the notation, apply for each aspect/mood.

For instance, for the imperfective forms of *nàgudiitłod* ‘jump down’, which has all three lexical prefixes, *nà-* at the outer boundary, *gu-* at the middle boundary, and *di-* at the inner boundary, and the stem *-tłod*, we can represent this structure as `nàgudiitłod[jump-down]:nà=gu_di.tłod` (cf. Table 2.1 above) Similarly, the perfective underlying combination of lexical prefixes and stem for this verb is represented as `nàgudiitłod[jump-down]:nà=gu_di.tłòt`, the progressive as `nàgudiitłod[jump-down]:nà=gu_di.tłíł`, and the two variants for repetitive as `nàgudiitłod[jump-down]:nàná=gu_di.tłiizh` and `nàgudiitłod[jump-down]:nìná=gu_di.tłiizh`. In contrast, for the imperfective forms of *itsiy* ‘cry’, which has only the stem and no lexical prefixes, we can represent this structure as `itsiy[cry]:tsiy`. In our implementation, empty boundary positions need not be explicitly declared, but that we generate these automatically as part of the composition process (as noted above in master compilation code 2). Table 3.1 below presents further examples in the first step of the lexical tier coded in the continuation lexicon `Verbstems` of how the allomorphy for combinations of lexical prefixes and stems for various lemmas as well as the associated continuation classes with the paradigm subsets are represented within the *lexc* formalism.

Table 3.1. Examples of the first step in the lexical tier FSM coded in the continuation lexicon `Verbstems`, specifying allomorphy for combinations of lexical prefixes and stems and the associated paradigm subsets for the different aspects/moods of Tsuut'ina verbs *itsiy* 'cry', *ts'ázíd* 'wake up' and *tsídiitł'á* 'run away' and *nàgudiitłod* 'jump down'.[11]

```
LEXICON Verbstems
...
itsiy[cry]:tsiy                         INTR-0-IPFV;
itsiy[cry]:tsày                         INTR-yi-y-PFV;
itsiy[cry]:tsíł                         INTR-yi-PROG;
itsiy[cry]:ná=chish                     INTR-0-IPFV; ! (REP)
...
ts'ázid[wake-up]:ts'á=zíd               INTR-ni-IPFV;
ts'ázid[wake-up]:ts'á=zid               INTR-ni-PFV;
ts'ázid[wake-up]:ts'á=ził               INTR-yi-PROG;
ts'ázid[wake-up]:ts'áná=zhiizh          INTR-ni-IPFV; ! (REP)
...
tsídiitł'á[run-away]:tsí=di.tł'á        INTR-0i-IPFV;
tsídiitł'á[run-away]:tsí=di.tł'ò        INTR-sii-PFV;
tsídiitł'á[run-away]:tsí=di.tł'áł       INTR-yii-PROG;
tsídiitł'á[run-away]:nátsí=di.tł'ásh    INTR-0i-IPFV; ! (REP)
tsídiitł'á[run-away]:tsí=di.tl'ò        INTR-yii-POT; ! (POT)
...
nàgudiitłod[jump-down]:nà=gu_di.tłod    INTR-0i-IPFV ;
nàgudiitłod[jump-down]:nà=gu_di.tłòt    INTR-sii-PFV ;
nàgudiitłod[jump-down]:nà=gu_di.tłíł    INTR-yii-PROG ;
nàgudiitłod[jump-down]:nàná=gu_di.tłiizh INTR-0i-IPFV ; ! REP
nàgudiitłod[jump-down]:nìná=gu_di.tłiizh INTR-0i-IPFV ; ! REP
```

In the overall compilation process, the simple, single-character lexical prefix markers as evident above are replaced by flag diacritics which govern in part which variant of a paradigm subtype is applicable in the three inflectional slots. In specific, the outer lexical prefix marker '=' is replaced with the flag-diacritic `@P.PREFIX.OUTER@`, the middle lexical prefix marker '_' with `@P.PREFIX.MIDDLE@`, and the inner lexical prefix marker '.' with `@P.PREFIX.INNER@`. If there are no lexical prefixes, none of these three flag-diacritics are incorporated for a verb

[11] Note that the repetitive aspect/mood forms (marked in the comment section of the *lexc* code, initiated by an exclamation mark '!' as `REP`) for the verbs presented here use imperfective morpheme chunks in the inner inflectional slot. In Tsuut'ina, as in some other Dene languages, repetitive forms do not have a distinct set of inflectional markers like those associated with imperfective, perfective, progressive, and potential verb forms (cf. Axelrod 1993 for arguments that similar forms in Koyukon should be treated as 'super-aspects', essentially layering on top of other aspectual markers). Repetitive forms can therefore reuse the inflectional markers defined for one of these other aspects/moods, while still specifying the distinctive stem forms needed here.)

construction in the lexical tier FST. How these are used in integrating the inflectional tier will be discussed in the section 3.2 below.

As noted above in section 2.2, the morpheme chunk in the inner inflectional prefix TAMA slot represents a number of morphological features, namely all singular as well as first and second person plural subjects (known as the inner subject), aspect, mood, conjugation class (i.e., *si-, ni-, yi-*, or none, representing one of the historical Dene aspect/conjugation class markers), and voice/valence (i.e., the historical voice/valence markers or "classifiers" that fuse phonologically with the preceding prefixes). These bundles of features can be combined into, and represented by, single labels, such as *0-Imperfective*, *yi-y-Perfective* and *yi-Progressive* which are applicable for the various moods/aspects of *itsiy* 'cry'. It is these labels which are used in `Verbstems` to specify continuation lexica that determine which subclasses of TAMA morpheme chunks in the paradigm are applicable for a verb construction in each aspect/mood. For instance, as can be seen in Table 3.1 above, in the case of *itsiy* 'cry' the inner inflectional morpheme chunk subset `INTR-0-IPFV` is specified as the TAMA continuation lexicon for the imperfective forms with the stem *-tsiy*, `INTR-yi-y-PFV` for the perfective forms with the stem *-tsày*, `INTR-yi-PROG` for the Progressive forms with the stem *-tsíł*, and `INTR-0-IPFV` for the repetitive forms with the combination of an outer prefix and stem *nà-...-chish*.

As the second step on the lexical tier, these TAMA continuation lexica are used to specify three flag diacritics concerning valence (`@U.VALENCE.X@`), aspect (`@U.ASPECT.X@`), and the TAMA chunk subtype (`@U.TAMA.X@`) (with '`X`' in the flag-diacritic notation denoting a variable value), which act as the actual constraints on which paradigm subtypes specified on the inflectional tier are applicable in the inner inflectional slot when implemented as an FSM. As exemplified in the *lexc* code snippet below (Table 3.2), in the case of the aforementioned aspect/mood continuation lexica for the imperfective forms of the intransitive verb *itsiy* 'cry' with the lone stem *-tsiy*, in addition to designating intransitive valence with `@U.VALENCE.INTRANSITIVE@` and imperfective aspect with `@U.ASPECT.IPFV@`, the TAMA subtype is set as `@U.TAMA.%0@` (the '`%`' character needed to escape the zero character '`0`' that otherwise refers to an empty transition in *lexc* code). Finally, as the third step in the lexical tier, the continuation lexicon `Verbsuffixes` , which is shared by all the preceding continuation lexica in step two setting valence, aspect and TAMA class, simply signifies the end of the specifications of the requirements concerning the inflectional tier, concluding the lexical tier FSM with the character reserved for this purpose in *lexc* code, the hash-mark "`#`' (Table 3.3).

Table 3.2. Examples of the second-step continuation lexica on the Lexical tier converting the information on the paradigm subset applicable for a combination of lexical prefixes and stems in FST-internal flag-diacritics.

```
LEXICON INTR-0-IPFV
@U.VALENCE.INTRANSITIVE@@U.ASPECT.IPFV@@U.TAMA.%0@ Verbsuffixes;

LEXICON INTR-yi-y-PFV
@U.VALENCE.INTRANSITIVE@@U.ASPECT.PFV@@U.TAMA.yi-y@ Verbsuffixes;

LEXICON INTR-yi-PROG
@U.VALENCE.INTRANSITIVE@@U.ASPECT.PROG@@U.TAMA.yi@ Verbsuffixes;
…
```

Table 3.3. The third, final-step continuation lexicon in the specification of the lexical tier FST

```
LEXICON Verbsuffixes
# ;
```

3.2 Inflectional tier

On the inflectional tier of our computational model, the possible individual morphemes or morpheme chunks applicable for each of the three inflectional slots, i.e. at the inner, middle, and outer boundaries, linked with the lexical tier as defined above, are first specified as three separate FSMs using the *lexc* formalism. The core structure and content of these three *lexc* specifications are the same, associating morphemes or morpheme chunks with flag-diacritics representing the one or more associated morphological features, e.g. `@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.SG@is` for *is-* which is the first person singular subject morpheme chunk for an intransitive verb (which belongs to the *0-Perfective* inflectional subtype and has only a stem and no lexical prefixes in the case of the perfective aspect).

Nevertheless, the *lexc* description for the inner inflectional slot that is the most extensive since it needs to specify all the possible paradigm subtype variant chunks that are applicable in this slot, depending on the combination of the lexical prefixes and stem on the lexical tier for each aspect/mood. Furthermore, the *lexc* description for the inner inflectional contains a bit more structure, since for each such paradigm subtype in the inner inflectional slot, there are three variants: one for when there are no lexical prefixes, one for when there is a lexical inner prefix (and possibly but not necessarily lexical middle and/or outer prefixes as well), and one for when there is a lexical outer prefix, but no lexical inner or middle prefix. Thus, the *lexc* code for the inner inflectional slot consists of a sequence of two continuation lexicon types. The first step quite simply enumerates the continuation lexica for all the general paradigm subsets[12], identified with labels of the format TAMA/SUBCLASS-MOOD/ASPECT, e.g. `0-IPFV`, `yi-y-PFV` or `yi-PFV` in the case of the *itsiy*. These paradigm subsets are referred to in the *lexc* code for the lexical tier, and the resultant FSM, not as conventional continuation lexica, but through flag-diacritics defining the valence, aspect and TAMA subtypes (of the form `@U.VALENCE.X@`, `@U.ASPECT.X@`, and `@U.TAMA.X@` in Table 3.2 above), with which the first step continuation lexica in the inner inflectional FSM must agree. These valence, aspect and TAMA flag-diacritics, combined further by flag-diacritics indicating the type of lexical prefix, if any (detailed below), then direct to the second step consisting of the continuation lexica specifying the actual TAMA morpheme chunks applicable for the three variants.[13]

In the case when there are no lexical prefixes, the second-step continuation lexicon is indicated by the suffix `NoPrefix` added to the general, first-step continuation lexicon name, while the existence of at least an inner lexical prefix is indicated with the suffix `InnerPrefix`, and the existence of an outer or middle but no inner lexical prefix is indicated with the suffix `OuterPrefix`. As noted above, which of these three variant continuation lexical is applicable is determined by a flag-diacritic, i.e. `@D.PREFIX@` in the case of no lexical prefixes, `@R.PREFIX.INNER@` in the case of an inner (and possibly middle and outer) lexical prefix, `@R.PREFIX.MIDDLE@` in the case of a middle (and possible outer but no inner) lexical prefix, and `@R.PREFIX.OUTER@` in the case of only an outer lexical prefix but no middle or inner ones. These flag-diacritics in the morphological tier finite-state transducers need to match with the corresponding flag-diacritics which are automatically added to the lexical tier FSM based on the inflectional boundary markers '.', '_', and '=', as part of the overall compilation process. Thus, `@R.PREFIX.INNER@` in the inner inflectional FSM requires `@P.PREFIX.INNER@` in the lexical tier FSM, and likewise `@R.PREFIX.MIDDLE@` requires `@P.PREFIX.MIDDLE@`, and `@R.PREFIX.OUTER@` requires `@P.PREFIX.OUTER@`. When there are no lexical prefixes for a verb construction, that form has none of the three aforementioned flag-diacritics in the lexical tier FSM, thus agreeing with the flag-diacritic `@D.PREFIX@` in the inner inflectional FSM.

For instance, the continuation lexicon `0-IPFV` , in the first step in the inner inflectional FSM, referring to the paradigm subtype *0-Imperfective*, applicable for both the imperfective and repetitive aspect forms of the verb *itsiy* 'cry', can direct either to the `0-IPFV-AFF-`

---

[12] Currently, we have specified 22 such paradigm subsets for imperfective verbs, with some more attested in documentation and awaiting implementation in this model.

[13] Since we currently specify 22 paradigm subsets for imperfective verbs, this results in altogether 66 continuation lexica to cover the three possible variants of morpheme chunk sets.

NoPrefix, 0-IPFV-AFF-InnerPrefix or 0-IPFV-AFF-OuterPrefix[14] continuation lexica in the second step in the in inner inflectional FSM. As the imperfective forms of *itsiy* have only the stem *-tsiy* without any lexical prefixes, in that case the second-step continuation lexicon which will be selected is 0-IPFV-AFF-NoPrefix. In contrast, for the repetitive forms with the outer-prefix-stem combination *ná-...chish*, the second-step continuation lexicon that will be selected is 0-IPFV-AFF-OuterPrefix.

Table 3.4. First steps in the *lexc* specification of the inner inflectional tier FST, exemplified here by the three variants of inner inflection for *0-perfective* verbs, based on what type of, if any lexical prefixes the verb construction consists of.

```
LEXICON 0-IPFV
@U.ASPECT.IPFV@@U.TAMA.%0@@D.PREFIX@          0-IPFV-AFF-NoPrefix ;
@U.ASPECT.IPFV@@U.TAMA.%0@@R.PREFIX.INNER@    0-IPFV-AFF-InnerPrefix ;
@U.ASPECT.IPFV@@U.TAMA.%0@@R.PREFIX.MIDDLE@   0-IPFV-AFF-InnerPrefix ;
@U.ASPECT.IPFV@@U.TAMA.%0@@R.PREFIX.OUTER@    0-IPFV-AFF-OuterPrefix ;
…
```

Table 3.5 below presents the full set of morpheme chunks for the second-step continuation lexica LEXICON 0-IPFV-AFF-NoPrefix and the LEXICON 0-IPFV-AFF-InnerPrefix with their inner inflectional paradigm subsets. As can be seen, when there is no lexical prefix, the morpheme chunk is associated here with flag-diacritics representing the relevant morphological features, e.g. @U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.SG@ for first person singular manifested as *is-*, @U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.SG@ for second person singular manifested as *ni-*, and so forth. As an important detail, we can notice that when an inner lexical prefix is linked with a stem, the morpheme chunks may become somewhat different, e.g. while first person singular is manifested as *s-* instead of *is-*, the second person singular remains *ni-*. Moreover, the aforementioned flag-diacritics representing the morphological person and number features can be linked with morphological feature tags, which will be outputted in word form analysis, and inputted for word form generation (implemented in the sixth FSM referred to in the beginning of Section 3; the details of this morphological tagger FSM are left outside the scope of this article).

[14] The element 'AFF' in the names of these continuation lexica indicates that they contain the actual morpheme chunks, as is evident in Table 3.5 below.

Table 3.5. Second steps in the *lexc* specification of the inner inflectional tied FST, which contains the actual TAMA morpheme chunks and the associated morphological features represented by flag-diacritics.

```
LEXICON 0-IPFV-AFF-NoPrefix
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.SG@is    #;
@U.SUBJECTPERSON.2@@U.SUBJECTNUMBER.SG@ni    #;
@U.SUBJECTPERSON.3@@U.SUBJECTNUMBER.SG@i     #;
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.PL@isaà  #;
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.PL@isiì  #;
@U.SUBJECTPERSON.2@@U.SUBJECTNUMBER.PL@as    #;
@R.SUBJECTPERSON.3@@R.SUBJECTNUMBER.PL@      #;
@R.SUBJECTPERSON.4@@R.SUBJECTNUMBER.SG@      #;
...
LEXICON 0-IPFV-AFF-OuterPrefix
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.SG@s     #;
@U.SUBJECTPERSON.2@@U.SUBJECTNUMBER.SG@ni    #;
@U.SUBJECTPERSON.3@@U.SUBJECTNUMBER.SG@      #;
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.PL@saà   #;
@U.SUBJECTPERSON.1@@U.SUBJECTNUMBER.PL@siì   #;
@U.SUBJECTPERSON.2@@U.SUBJECTNUMBER.PL@s     #;
@R.SUBJECTPERSON.3@@R.SUBJECTNUMBER.PL@      #;
@R.SUBJECTPERSON.4@@R.SUBJECTNUMBER.SG@      #;
```

The specifications for the middle and outer inflectional tier FSMs consist of only one step in their *lexc* code (exemplified for intransitive verbs in Tables 3.6 and 3.7 below), associating the possible morphemes for those slots with the flag-diacritics representing the corresponding morphological features, as well as unifying with the corresponding lexical prefix markers, `@U.PREFIX.MIDDLE@` and `@U.PREFIX.OUTER@`, respectively. Thus, in the case of the middle inflectional slot marking "outer" subjects, we can have the morpheme *gi-* corresponding to a third person plural subject, and *ts'i-* for the fourth person singular subject. And in the case of the outer inflection slot, we can have the morpheme *dà-* marking a distributive form, restricted to only plural subjects with the flag-diacritic `@U.SUBJECTNUMBER.PL@`. As with the inner inflectional FSM, the flag-diacritics representing morphological features incorporated in the middle and outer inflectional FSMs can be linked with morphological feature tags as part of the output or input of the entire computational model.

Table 3.6. Extract of *lexc* specification of the middle inflectional slot FSM.

```
LEXICON Root
@U.SUBJECTPERSON.3@@U.SUBJECTNUMBER.PL@@U.PREFIX.MIDDLE@gi   #;
@U.SUBJECTPERSON.4@@U.SUBJECTNUMBER.SG@@U.PREFIX.MIDDLE@ts'i #;
                                                             #;
...
```

Table 3.7. Extract of the *lexc* specification of the outer inflectional slot FSM.

```
LEXICON Root
@U.VALENCE.INTRANSITIVE@@U.SUBJECTNUMBER.PL@@P.DISTRIBUTIVE.ON@@U.PREFIX.OUTER@dà  #;
@U.VALENCE.INTRANSITIVE@@U.SUBJECTNUMBER.PL@@P.DISTRIBUTIVE.OFF@   #;
...
```

As noted earlier, the specifications of the lexical tier, as well as the applicable morphemes or morphemes chunks in the inner, middle, and outer slots on the inflectional tier, as described above, are first compiled into four separate FSMs. Then, the three inflectional tier FSMs are inserted using finite-state calculus within the lexical tier FSM, by replacing the special characters '.', '_', and '=' marking the inner, middle and outer inflectional boundaries with the entire corresponding inner, middle, and outer inflectional FSMs (in addition to continuing to mark these slots with the corresponding flag-diacritics `@P.PREXIX.INNER@`, `@P.PREXIX.MIDDLE@`, and `@P.PREXIX.OUTER@`). Crucially, the constraints between the verb constructions consisting of zero-three lexical prefixes + stems (and their allomorphy for the different aspects/moods) and the morpheme chunks allowable in the inner inflectional TAMA slot are maintained by the flag-diacritics incorporated in these four constituent morphological FSMs, which become part of the overall morphological FSMs as part of the composition process.

4 Applications

Computational morphological models such as the one we have described above, while useful in and of themselves for linguists specializing in a particular language, enable the creation of several software applications that have usefulness for a much broader end-user audience of native speakers and learners of an Indigenous language such as Tsuut'ina, supporting its continued use on digital platforms (Arppe et al. 2016). Prime examples of such applications are intelligent electronic dictionaries, spell-checkers, and computer-aided language learning applications (Arppe et al. 2015). In this, we have come to make use of the *Giella* language technology development infrastructure (Moshagen et al. 2013), originally created by the Giellatekno[15] and Divvun[16] research and development teams at the University of Tromsø for developing a suite of such applications for the Sámi languages, since this infrastructure has been designed to allow for as seamless an integration as possible of a computational model for a language, using FST technology as described above, as software components providing the various linguistic functionalities within the aforementioned end-user applications.

In our experience, one of the greatest benefits, as expressed by native speakers of Indigenous languages, is found by being able to augment an electronic dictionary with the computational

[15] `http://giellatekno.uit.no/index.eng.html`
[16] `http://divvun.org/`

morphological model, creating an "intelligent" dictionary (I-DICT), which is able to both recognize and generate inflected word forms (Johnson et al. 2013). It has long been noted that, for languages with rich inflectional morphology such as Tsuut'ina and other Dene languages, an electronic dictionary that is morphologically simplistic can be extremely cumbersome to use, especially for speakers, learners and others lacking many years of linguistic training. With the help of a computational morphological analyzer, however, users may input any inflected form of a word, and be redirected to the appropriate lemma as well as be provided with information on the morphological features of the word-form (see Figure 4.1 below of a mock-up *Gunáhà*, an intelligent on-line dictionary we are developing for Tsuut'ina, to be eventually accessible via `http://altlab.ualberta.ca/gunaha/`). Furthermore, the computational morphological model can also be used in reverse to generate individual word-forms expressing some desired combination of morphological features, or various selections of word-forms as paradigms of varying extents (e.g. core word-forms or the full inflectional paradigm, see Figure 4.1). Moreover, and intelligent on-line dictionary can be linked with a text collection that has been linguistically analyzed using the computational model, allowing for searching the use of a lemma in its various inflected forms within naturally contexts (Junker and Luchian 2007; Junker and Stewart 2008).

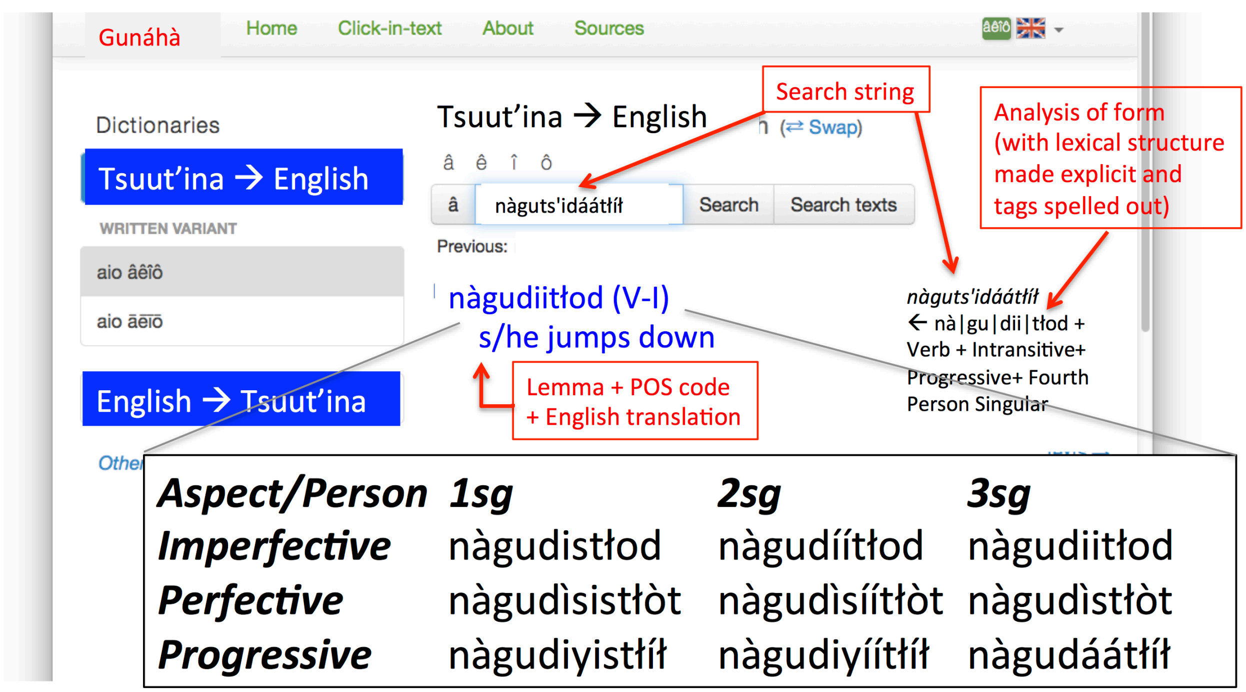

Figure 4.1. A mock-up on an Intelligent dictionary search for the Tsuut'ina intransitive verb form *nàguts'idáátłíł*, the progressive fourth person singular form of the lexical entry *nàgudiitłod* 's/he jumps down', including a sample paradigm generated for this verb.

Furthermore, a computational morphological model defines the set of well-formed words, and therefore can be adapted into a spell-checkers integrated within a word processing application (see Figure 4.2). Such a tool can be helpful for literacy programs and speed the creation and proofing of high-quality texts in the language. Where communities are attempting to promulgate a particular written form of the language as standard, such a spell-checker can help in the codification and enforcement of those standards. However, it is important to note that a computational model and the subsequent spell-checkers can be adapted for multiple dialects and orthographical standards, if the Indigenous language communities so wish. Finally, it is a short leap from the applications described above to classroom applications as well. Foremost among these are intelligent computer-aided language learning (or I-CALL) applications (Antonsen et al. 2013). The combination of a lexicon, a computational morphological model and a relative small set of simple grammatical exercise templates can allow for the creation of an essentially infinite number of language drills of various types.

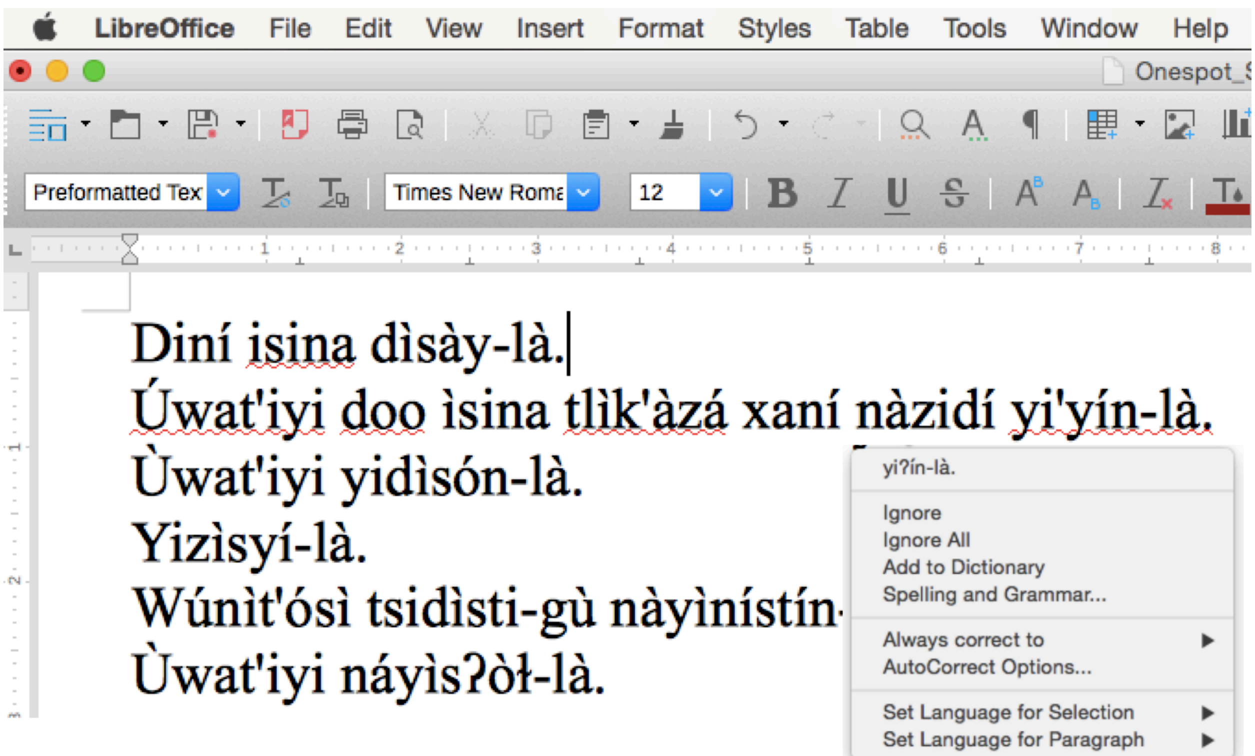

Figure 4.2. An example of a currently working demonstration version of a spell-checking module for Tsuut'ina, based on our computational morphological model and integrated with LibreOffice, providing a suggestion for the incorrectly spelled Tsuut'ina word *yi'yín-là* (using an apostrophe instead of a glottal stop character 'ʔ', and thus underlined by a red squiggly line), for which the correct form is *yíʔín-là*. Other recognized typos in the Tsuut'ina passage above include: (1) missing tone marking: *isina*, *doo*; (2) wrong tone marking: *Úwat'iyi* (Ú<Ù); and (3) plain letter used instead of the diacritic one: *tlìk'àzá* (l<ł).

5 Conclusions

In this paper we have presented a general architecture and demonstration case for computationally modeling Tsuut'ina, using the finite-state transducer paradigm, designed so that non-computational linguists should be able to augment and adapt it with a reasonable amount of training, and which at the same time should be computationally efficient. The development of the current demonstration version of the computational model took only several days of collaboration by field linguists and computational linguists (i.e. the authors of this paper). However, one must recognize that this built upon decades of linguistic documentation work by linguists and Elders as well as other native speakers of Tsuut'ina. Next, our intention is to substantially expand the current demonstration version in terms of vocabulary and morphology (adding new parts-of-speech such as nouns). Moreover, we hope that this Tsuut'ina model can be an example and inspiration for the creation of similar computational models for other Dene languages. Importantly, all the software tools, components and applications presented in this article are open-source – the primary cost involved is that of the time and effort of the people who want to further develop this computational model for Tsuut'ina, or start developing ones for other Dene languages.

Acknowledgements

The comments of the reviewing editors to an earlier version of this manuscript are greatly appreciated. Furthermore, the work presented herein is the result of two Computational Workshops on Indigenous Languages (CWIL), focusing on Dene languages in general and Tsuut'ina in particular, organized at the University of Alberta in June and October 2015, and was first presented at the Dene Languages Conference in Yellowknife, NT, in June 2016 – we want to thank all workshop and conference participants for their input, in particular Sally Rice. We also gratefully acknowledge the contributions of the Office of the Tsuut'ina Language Commissioner, Tsuut'ina Nation, to the long-term documentation and revitalization of the Tsuut'ina language. Finally, we are grateful for the funding provided by the Social Sciences and Humanities Research Council (SSHRC), Canada, through a Partnership Development Grant (890-2013-0047) and by the Kule Institute for Advanced Study (KIAS), University of Alberta, through a Research Cluster Grant, as well as support from the Helsinki Finite State Technology Project, University of Helsinki.

References

Antonsen, Lene, Ryan Johnson, Trond Trosterud & Heli Uibo. 2013. "Generating modular grammar exercises with finite-state transducers." Proceedings of the second workshop on NLP for computer-assisted language learning at NODALIDA 2013. *NEALT Proceedings Series 17 / Linköping Electronic Conference Proceedings* 86: 27–38.

Arppe, Antti, Jordan Lachler, Trond Trosterud, Lene Antonsen & Sjur N. Moshagen. 2016. "Basic Language Resource Kits for endangered languages: A case study of Plains Cree." *Workshop on Collaboration and Computing for Under-Resourced Languages (CCURL)*, Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia, 23 May 2016. URL: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-CCURL2016_Proceedings.pdf

Arppe, Antti, Lene Antonsen, Trond Trosterud, Sjur Moshagen, Dorothy Thunder, Conor Snoek, Timothy Mills, Juhani Järvikivi & Jordan Lachler. 2015. "Turning language documentation into reader's and writer's software tools." *4th International Conference on Language Documentation and Conservation (ICDLC)*. Honolulu, HI, 26 February–1 March 2015. URL: http://altlab.artsrn.ualberta.ca/wp-content/uploads/2015/10/ICDLC_150228.pdf

Axelrod, Melissa. 1993. *The Semantic of time: Aspectual categorization in Koyukon Athabaskan*. Studies in the Anthropology of North American Indians. Lincoln, NE / London: University of Nebraska Press.

Beesley, Kenneth R. Beesley & Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.

Cox, Christopher, Mans Hulden, Miikka Silfverberg, Jordan Lachler, Sally Rice, Sjur N. Moshagen, Trond Trosterud & Antti Arppe. 2016. "Computational modeling of the verb in Dene languages. The case of Tsuut'ina." *Dene Languages Conference*, Yellowknife, Northwest Territories, Canada, 6–7 June 2016. URL: http://altlab.artsrn.ualberta.ca/wp-content/uploads/2016/07/Tsuutina-FST-DLC-160607C.pdf

Hulden, Mans & Shannon T. Bischoff. 2008. "An experiment in computational parsing of the Navajo verb." In Shannon T. Bischoff & Mans Hulden (eds.), *Coyote Papers, Vol. 16. Special issue on Navajo Language Studies*, 101–118. Tucson, AZ: University of Arizona Press.

Cook, Eung-Do. 1984. *A Sarcee grammar*. Vancouver, BC: University of British Columbia Press.

Faltz, Leonard. 1998. *The Navajo verb. A grammar for students and scholars*. Albuquerque, NM: University of New Mexico Press.

Hulden, Mans. 2009. "Foma: a finite-state compiler and library." In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL): Demonstrations Session*, 29–32. Athens: Association for Computational Linguistics.

Johnson, Ryan, Lene Antonsen & Trond Trosterud. 2013. "Using Finite State Transducers for Making Efficient Reading Comprehension Dictionaries." In Stephan Oepen, Kristin Hagen & Janne Bondi Johannessen (eds.), *Proceedings of the 19th Nordic Conference of Computational Linguistics* (NODALIDA 2013), Oslo University, Norway, May 22–24, 2013. NEALT Proceedings Series 16: 59–71. http://www.ep.liu.se/ecp/085/010/ecp1385010.pdf

Junker, Marie-Odile & Radu Luchian. 2007. “Developing web databases for Aboriginal language preservation.” *Literary and Linguistic Computing* 22: 187–206.

Junker, Marie-Odile & Terry Stewart. 2008. “Building search engines for Algonquian languages.” In Karl S. Hele & Regna Darnell (eds.), *Papers of the 39th Algonquian Conference,* 378-411. London: University of Western Ontario Press.

Kari, James. 1975. “The disjunct boundary in the Navajo and Tanaina verb prefix complexes.” *International Journal of American Linguistics* 41(4), 330–345.

Kari, James. 1989. “Affix positions and zones in the Athapaskan verb complex: Ahtna and Navajo.” *International Journal of American Linguistics* 55(4): 424–454.

Leer, Jeff. 1999. “Comparative Athabaskan class notes.” Unpublished manuscript CA965L1999a. Fairbanks, AK: Alaska Native Language Archive.

Lindén, Krister, Erik Axelson, Sam Hardwick, Miikka Silfverberg & Tommi Pirinen. 2011 “HFST - Framework for Compiling and Applying Morphologies.” In Cerstin Mahlow & Michael Piotrowski (eds.), *Proceedings of Second International Workshop on Systems and Frameworks for Computational Morphology (SFCM)*, 67-85. Berlin / Heidelberg: Springer.

McDonough, Joyce. 2000. “On the bipartite model of the Athabaskan verb.” In Theodore B. Fernald & Paul R. Platero (eds.), *The Athabaskan languages: Perspectives on a Native American language family*, 139–166. Oxford University Press.

McDonough, Joyce M. 2003. *The Navajo Sound System*. Dordrecht: Kluwer.

Moshagen, Sjur, Tommi Pirinen & Trond Trosterud. 2013. Building an open-source development infrastructure for language technology projects. *NEALT Proceedings Series 2013*, Vol. 16: 343-352. doi: http://www.ep.liu.se/ecp/085/031/ecp1385031.pdf

Rice, Keren. 2001. “Slave (Northern Athabaskan).” In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology,* 648–689. Malden, MA: Blackwell.

Rice, Keren. 2005. ”A typology of good grammars.” *Studies in Language* 30(2): 385–415.

Rice, Sally, Gary Libben & Bruce Derwing. 2002. “Morphological representation in an endangered, polysynthetic language.” *Brain and Language* 81: 473–486.

Shaw, Patricia A. 1991. “Consonant harmony systems: The special status of coronal harmony.” In Carole Paradis & Jean-François Prunet (eds.), *The special status of coronals: Internal and external evidence*, 125–157. San Diego: Academic Press.

Trosterud, Trond. 2004. “Porting morphological analysis and disambiguation to new languages.” In: Julie Carson-Berndsen (ed.), *First Steps in Language Documentation for Minority Languages: Proceedings of the SALTMIL Workshop at LREC 2004,* 90–92. Lisbon: ELRA.

Trosterud, Trond. 2006. “Grammatically based language technology for minority languages.” In Anju Saxena & Lars Borin. (eds.), *Lesser Known Languages of South Asia*, 293–316. The Hague: Mouton de Gruyter.

Young, Robert W. & William Morgan. 1987. *The Navajo language.* Albuquerque, NM: University of New Mexico Press.