Algonquian Verb Paradigms: A Case for Systematicity and Consistency

Antti Arppe, Chris Harvey, Marie-Odile Junker, and J. Randolph Valentine

In this paper we make the case for certain types of systematicity in describing the morphology of Algonquian languages, in particular the inflectional morphology of Algonquian verbs, though these principles would apply to the documentation of any other Indigenous language. By systematicity, we mean well-organized and detailed descriptions of a language's words and their morphological structure, seeking exhaustive descriptions of all word-class paradigm-types, matched with a comprehensive lexicon incorporating paradigm-type specifications. Indeed, this is what most scholars of Algonquian languages in practice seek to produce in their linguistic documentation work. Importantly, we are not advocating for any particular notation or standard, as we know that Algonquian scholars have developed many different systems. But whatever notation is used, it is essential that it be systematic in that it can be deterministically mapped to other notations and representations.

Crucially, in this we are in effect arguing for taking the Model-Controller-View (MCV) architecture developed in computer science for user interfaces, to help us structure and organize our linguistic data (Krasner and Pope 1988; Junker and Stewart 2011). In this approach, the Model consists of the raw information that is stored in the underlying, primary database, in whatever format, as long as that

format is systematically followed and explicitly described. The View refers to how users are presented with this information, the design of which can be changed to other Views, depending on application, user need, or user group, without requiring changes in the underlying database/Model. The Controller consists of the software instructions that connect the Model to the View. By separating components in this way, we enable a rapid and robust development environment, and by using this standardized approach we can take advantage of the extensive documentation and large community of other developers following this approach.

We have found that with such an approach one can be surprisingly time-effective in creating computational morphological models that form the basis of several language technology tools to help support the revitalization and continued use of Indigenous languages (Arppe et al. 2015). Importantly, these tools and applications can reach a quality comparable to those available for majority languages. Primary among such tools are intelligent web-based dictionaries (I-DICTs), which are intelligent in that with the computational model they can link almost any inflected form with its appropriate dictionary entry as well as generate word paradigms. Such a computational model can also be adapted into a spell-checker, which, integrated into a word-processing application, can support adherence with one or more existing orthographical conventions, resulting in the production of good-quality texts where the focus can be on content and not orthography. Furthermore, one can create intelligent language training and education applications (ICALLs), which use the computational model for the dynamic generation of large numbers of learning exercises based on combining core vocabulary with finite sets of exercise templates (Antonsen et al. 2013). Finally, one can use the computational model for ongoing linguistic analysis of texts and other research.

In developing linguistic resources for Indigenous languages, we need to recognize that there are multiple audiences, with different levels of linguistic knowledge and proficiency and varying usage needs, who will want to have access to and make use of these resources. Broadly speaking, these audiences can be divided into: (1) the members of the linguistic communities in question, for whom the language is either their mother tongue or a second language important as a vehicle of cultural identity (this also includes people living away from their communities); (2) people external to the language community who are typically interested in it from the perspective of scientific study; (3) students of the language taking courses in K–12 or university contexts; and (4) nonspecialists interested in the language. In the case of endangered languages, the mix of these audiences and their needs is influenced

by the state of the endangerment of the language. For instance, different Algonquian languages and dialects exemplify a gradient continuum ranging from robust and broad use, to severely endangered and restricted use, to dormant or extinct.[1] We will exemplify these states with East Cree and Innu (robust), Plains Cree and Southwestern Ojibwe (endangered), and Mahican (dormant).

East Cree, spoken by over 12,000 people in nine communities situated in the James Bay region of Northern Quebec, is still learned by children as their first language and fluently used in schools and in the communities overall, involved in most spheres of life as an oral language. There is basic literacy among the speakers, but written communication tends to be in English or, to a lesser extent, French. Innu, spoken in Quebec and Labrador, illustrates a similar situation. In contrast, Plains Cree, spoken in Alberta and Saskatchewan, is no longer generally learned as a first language by children in the communities, and most children are exposed to their heritage language only in school in the form of weekly language and culture instruction. However, many middle-age and older community members are bilingual, having learned Plains Cree as their first language, but the situations in which Plains Cree is used in these communities is more restricted than in the case of East Cree. Moreover, due to these still-active fluent speakers having grown up in the residential school era, when Indigenous languages were not taught in any way, their proficiency is stronger in the spoken than in the written form of the language. Consequently, the fluent speakers, having less certainty of words' written forms, would benefit from spell-checking when writing their language, while heritage and other learners need information on proper pronunciation (spoken recordings of words individually and in sentential context) as well as assistance with how words are inflected, i.e., the paradigms. Furthermore, teachers and advanced learners could make good use of information on the morphological composition of words.

Southwestern Ojibwe is primarily spoken in Wisconsin and Minnesota. An assessment of speakers in 2009 (Moose et al. 2009) determined that 678 speakers remained in Minnesota and 42 in Wisconsin, the latter distributed over six communities. While access to fluent speakers is obviously very limited in these communities, it is also important to note that nearly half of Wisconsin's American Indian population resides in urban areas. Very few published materials laying out basic inflectional paradigms exist, and dictionaries such as Nichols and Nyholm (1995) follow the standard practice of providing only a couple of key inflectional forms. There is thus a great need to help independent teachers and learners with

at least the basic, core inflectional forms, if not fully enumerated paradigms of both nouns and verbs.

Mahican is an Eastern Algonquian language whose last rememberers passed away in the 1930s. The language was fairly well documented by missionaries and native speakers in the eighteenth and nineteenth centuries (mostly religious translations and wordlists survive), with some linguistic elicitation and short stories recorded in the early twentieth century. The modern Mahican community in Wisconsin has access to some of the old written sources, but there has yet to be a thorough interpretation of these sources, especially in a way that makes them accessible to Mahican people interested in the language. Centrally, this involves identifying and modeling paradigms, and presenting these alongside an online dictionary, in ways useful to both researchers and language learners (Harvey 2015), which then function as the means by which individual learning or curriculum development can begin. Accessible paradigms are critically important for dormant languages where there are no speakers to consult.

In light of the context discussed above, we discuss and exemplify the following topics in this paper: (1) the desiderata of a formal model that would best accommodate the Algonquian verb; (2) the desiderata for any standard for labeling and organizing Algonquian verbal paradigms (i.e., should one split or chunk morphemes or both?); (3) how we might best pursue consistency for the underlying primary databases across Algonquian languages, to allow for the maximal efficiency in the adaptation of applications developed for one Algonquian language to the rest, as well as for ease of comparability in language learning and linguistic research; (4) what is lexical vs. paradigmatic (i.e., which morphological processes are productive?); and (5) the desiderata on the different ways in which we can display information from the primary databases so that it is useful for the various audience types—first language speakers, heritage language learners, nonheritage language learners, instructors, linguistic researchers—and takes into account the reality of the language's relative vitality.

Templates

A few key notions are generally agreed on when describing the Algonquian verb.[2] There are four basic morphological classes that subsume certain persons and their particular grouping; and there are varying numbers of inflectional paradigms,

distributed in three orders: Independent, Conjunct, and Imperative. General characteristics of the three orders across Algonquian languages are as follows: verbs in the Independent order take personal prefixes (Conjunct and Imperative verbs do not); Conjunct verbs can undergo initial change; and the set of persons used in the Imperative is restricted.

Out of these generalizations, templates can be derived that can be used for both structuring a verb database and displaying verb conjugations for different audiences. Of the four Algonquian morphological verb classes, the Transitive Animate (VTA) verb, so called for its animate object, has the largest number of forms, which are best organized by grouping together LOCAL or I-You interactions (Speech Act Participants [SAPs] only), NON-LOCAL or third person interactions only, and MIXED person interactions involving SAP and third persons. The VTA passive has a person set that patterns like the person set of the Transitive Inanimate (VTI) verb, so called for its inanimate object, and the Animate Intransitive (VAI) verb, which takes animate subjects. Inanimate Intransitive (VII) verbs, which take inanimate subjects, only have third person inanimate subjects. VTI and VAI verbs also have relational inflection, which excludes obviative subjects, and passive or unspecified actor forms, which often have the same person affix sets as Inanimate Intransitive verbs, though not in all languages. All of the above is summarized in (1) and illustrated in Figure 1 for Innu VTA verbs.

(1) Basic Template for the four classes of Algonquian verbs:

VTA verb

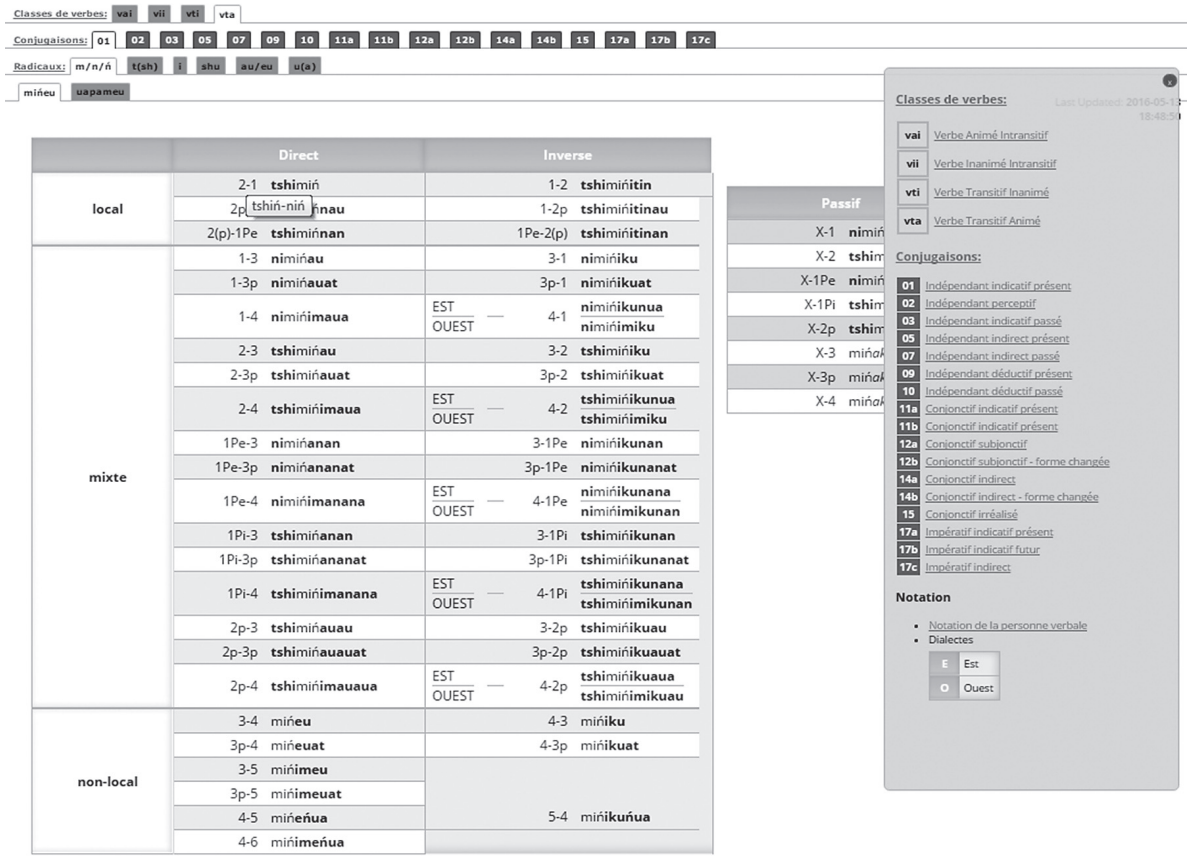
- LOCAL (I-YOU interactions, Speech Act Participants [SAP] only)
- MIXED (SAP and 3rd)
- NON-LOCAL (3rd person interactions)
- Passive set (similar to AI and TI verbs)

VTI and VAI verbs

- Regular
- Relational
- Unspecified actor sets (similar to II verbs)

VII verb

FIGURE 1. Display of an Innu VTA Verb, based on templates, and exhibiting a numbering system of subparadigms in the pop-up window (verbe.innu-aimun.ca)



Person Labeling and Notation

As one can see from Figure 1, a numbered notation is used to index person and number. Consistency is crucial. There can be many ways to DISPLAY the person marking, but we recommend that standardization across Algonquian verb databases be seriously pursued. Issues can range from choice of a notation that is ambiguous or uninterpretable to technical difficulties in importing and manipulating the data on a computer, which can result in software malfunctions or even loss of data. For example, from the perspective of cross-Algonquian comparability, if a language neutralizes number for animate obviative forms, it could be preferable to use 4(p) or 3′(p) rather than just 4 or 3′, because there are Algonquian languages where obviative singular and plural are distinguished, such as northwestern Ontario dialects of Ojibwe. Nevertheless, as long as one label (e.g., 4) is used consistently and is clearly described to represent a number-wise underspecified obviative person form, conversion to some other notation can be done with ease. Furthermore, using numbers to refer to grammatical person might well make sense to linguists but makes little sense to lay speakers.[3] Below is a short list of various person notations currently found in Algonquian Verb databases that need clear descriptions and mappings of equivalences, a task that we must leave for the future:

(2) Various Person Notations:
Inanimates: 3 or 0?
Obviative animates: 3, 3′, 3″ or 3, 4, 5, . . . ?
Obviation: ′, ″ or OBV?
Plural (vs. Singular): 22 or 2p? ; 4, 4(p) 4s or 4p ?
Inclusive-Exclusive distinction: 12, 21, 21p, 21(p), 1Pi ? / 1p or 1Pe?
Transitivity and direction—direct: 3-4, 3>4, 3→4, 3+4 ; inverse: 4→3 or 4←3?
VTI verbs: 1-0 or 1?
Passive and unspecified actor forms: X-1 or 1, X, X′, . . .

Glossing and Displaying with Templates

Once a consistent notation is adopted, glossing templates can be developed for pedagogical displays and comparative purposes. Displays can be adapted to various user types and needs. For example, in the Innu and East Cree verb conjugation applet,[4] the solution was to use a mouse-over of the abstract person number to display a gloss, established in consultation with the speakers and users, that contains

TABLE 1. VII glossing template for Innu

PRONOUN	ENGLISH SUBJECT PRONOUN	PRONOM SUJET FRANÇAIS	INNU PRONOUN
0	it	ça, il	tshekuan
0p	they	ça, ils	tshekuana
0′	it [obviative]	ça, il [obviatif]	tshekuanńu
0′p	they [obviative]	ça, ils [obviatif]	tshekuanńua

TABLE 2. VTA glossing template (21p mixed) across Innu and East Cree dialects

PRONOUN	ENGLISH SUBJECT	ENGLISH OBJECT	INNU PRONOUN	SEC PRONOUN	NEC PRONOUN
3→21p	s/he	us (you and me)	uiń→tshińanu	wî→chîyânû	wîyi→ chîynaâniu
21p→3	we (you and I)	him/her	tshińanu→uiń	chîyânû→wî	chîyâniu→ wîyi
3p→21p	they	us (you and me)	uińuau→tshińanu	wîwâu→ chîyânû	wîyiwâu→ chîyâniu
21p→3p	we (you and I)	them	tshińanu→uińuau	chîyânû→ wîwâu	chîyâniu→ wîyiwâu
21p→4	we (you and I)	him/her/them [obviative]	tshińanu→neńua (utauassima)	chîyânû→ aniyûh (utawâshimh)	chîyâniu →aniyâyiuh (ukusis-h)
4→21p	s/he or they [obviative]	us (you and me)	neńua (utauassima) →tshińaniu	aniyûh (utawâshimh) →chîyâniu	aniyâyiuh (ukusis-h) →chîyâniu

a corresponding emphatic pronoun in the Indigenous language (see Tables 1 and 2 and Figure 1). The French and English glossing templates were developed not only to generate English and French glosses of verb forms but also to check and suggest consistency with the bilingual dictionary definitions of such verbs.

Tables 1 and 2 illustrate some issues encountered in seeking to help different user types and the solutions that were adopted: How much metalanguage (terms like OBVIATIVE, PLURAL) do we use? Does the list include all possible cases, and do we give examples? For instance, consider how both emphatic and indefinite pronouns are used in the templates, how some examples are added in parentheses, and how grammatical information is indicated in brackets. Since there are no emphatic obviative pronouns, a set of remote demonstratives was used instead (gloss for

person 4, *neńua*). For inanimate subjects, the inflected obviative form of *tshekuan* was used, but for unspecified actor VAI forms, the corresponding *auen* was NOT used, as a decision was made not to offer any pronominal gloss for impersonal verbs.

Such templates can also be used to generate code to control the display of forms. The display (or View) of our verb forms can follow different orderings of pronouns, based on users' habits or preferences. The order for VTI or VAI verbs in (3a) follows what bilingual Innu-French speakers are used to from their French grammar schooling experience, while the order (3b) follows the animacy/person hierarchy.

(3) a. 1, 2, 1p, 21p, 2p, . . .

b. 2, 2p, 21p, 1, 1p, . . .

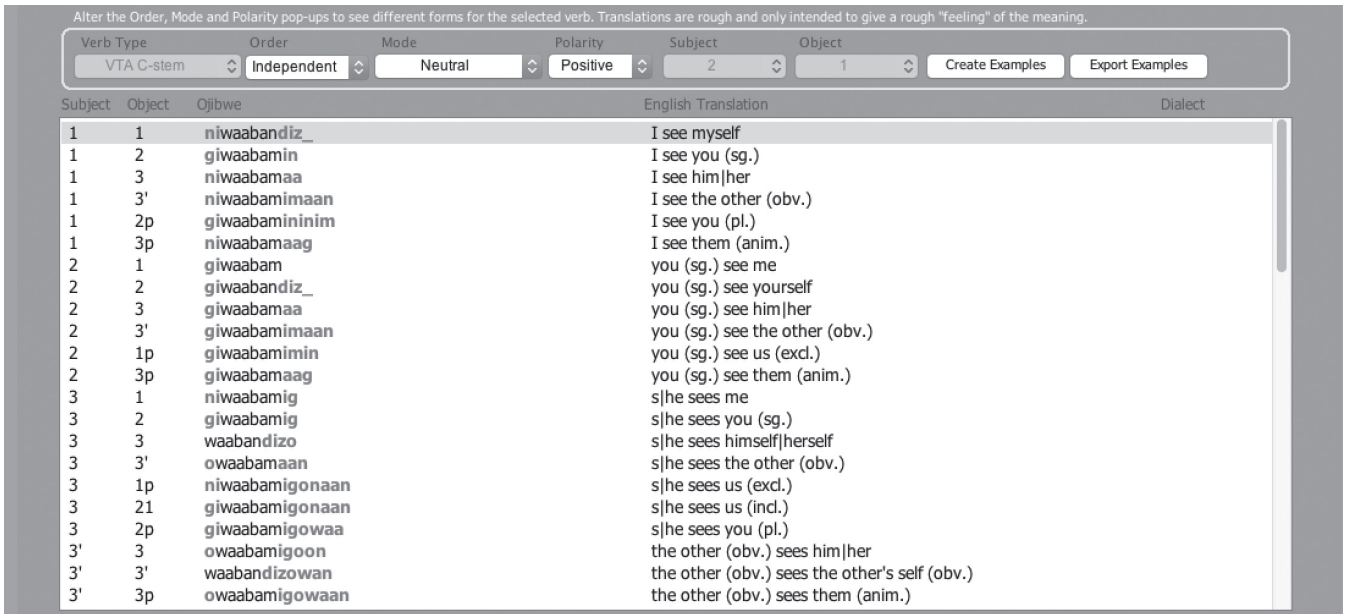
Labeling Paradigms

Labeling paradigms should be as consistent as possible, within and across Algonquian languages. One solution first proposed by MacKenzie (1980) is to adopt a numbering system based on cognate suffixes across languages and dialects. Different labels can then be applied in the displays to reflect the semantics of each paradigm in a particular dialect and the different users' preferences: those of a linguist, Indigenous teachers, and so forth. Figure 1 also illustrates such a list and its realizations in Innu. This numbering system can also be used to display paradigms economically in tabs, as in the Innu and East Cree verb applets in Figure 1, which also have a legend that links paradigm labels to web pages covering the corresponding grammar, with examples. Such an equivalence-based labeling system could be extended to include Eastern Algonquian languages and diachronic dimensions.

Generating Forms

All the examples given so far have concerned fairly vigorous languages, for which model verbs could be fully documented with many speakers, with fluent teachers and speakers as users. Generating verb forms for these languages is so far happening behind the scenes and for search engine purposes only (see next section). When the language is no longer spoken except by a few elders, the need to automatically generate and display all possible verb forms increases. While the principles of database organization stay the same, some new display issues arise.

FIGURE 2. Western Ojibwe verb paradigm (generated)



The Western Ojibwe Dictionary (Valentine and Ningewance 2009), as shown in Figure 2, offers a basic display for each verb, for which both numeric abbreviations and computer-generated English glosses (based on person templates as discussed above) are provided. Note also that, given the length of some Ojibwe inflected forms, when a form is selected, a computer-generated syllabification is provided as a pronunciation aid. Finally, for pedagogical reasons, not only is the VTA inflection provided but also the derived reflexive (VAI) verb inflection.

We can see here how the predominant user groups one has in mind will determine how to present verb paradigms. In addition to the inflectional subsets discussed above (orders, moods), polarity is added and generated, since there are distinct negative forms in all three orders. The Western Ojibwe Dictionary only displays a grammatically specified subset of inflections in its viewing area. Users can select which subcategories they want to see by using drop-down menus (top of Figure 2) or by using a help system that requires less linguistic knowledge through an interactive window to the right, by which they can specify in simple terms the grammatical context in which a form will be used (its order), its polarity (positive or negative), and its subject and object person/number/obviation features.

In order for verb paradigms to be generated like this, a number of decisions have to be made, which will influence both the documentation work and the form of the data entered in the database. A full discussion of the advantages and disadvantages of database structural types is beyond the scope of this paper (but see Junker et al. 2013), though we recommend relational databases for consistency and economy. Importantly, when lexical and morphological information is documented and stored in a well-structured and systematic way, in standard databases that linguists routinely use, such linguistic information can be converted into platform-independent, portable computational models that can in turn be packaged as software modules, e.g., as spell-checkers, which can be integrated with a word-processing application. One widely used technology for such computational models are Finite-State Machines (e.g., Beesley and Karttunen 2003). They are well-known computational data structures, are extremely fast and efficient, have a calculus allowing for powerful manipulations, allow rule-based definition of paradigms for various verb types, and are easily portable to different operating systems and platforms, and thus can be integrated with other applications. Here, one can consider the finite-state machine as another instantiation of the underlying Model, the output of which can again be adapted by a Controller to produce various Views. We have done this successfully with Plains Cree (Harrigan et al. 2016), where the computational modeling work has

been substantially facilitated by a consistently structured lexical database (the one underlying Wolvengrey 2001) which is systematically matched with descriptions of the verb paradigms (Wolvengrey 2011), building upon prior work by Wolfart (1973), Ellis (1971), and others. Combined with a Plains Cree lexical database, this computational model can be used create an intelligent dictionary (I-DICT), allowing for the generation of a variety of Views on the verb paradigms, available for any verb in the dictionary (cf. *itwêwina* n.d.)

With these considerations in mind, how should we analyze the data and represent the morphemes that constitute verb inflection?

Analyzing and Representing Morphemes: To Split or to Chunk?

One linguistic tradition in representing the morphological structure of words is maximal decomposition, so that each morphosyntactic feature is matched with some overtly observable and delineable sequence in the word (the ITEM-AND-ARRANGEMENT approach); take, for example, the the five morphemes and associated features (4a–b) from Wolvengrey (2011:56) for Plains Cree. Note that the only difference in form here is the theme sign, *-â-* vs. *-iko-* (Direct vs. Inverse), which switches which of the two referents expressed by the verb is the Actor and the Goal.

(4) a. niwîcihânânak
ni-wîcih-â-nân–ak (split)
ni-wîcih-ânânak (chunked)
1-help.VTA-DIR-1p-3p
'We (excl.) help them.'

b. niwîcihikonânak
ni-wîcih-iko-nân–ak (split)
ni-wîcih-ikonânak (chunked)
1-help.VTA-INV-1p-3p
'They help us (excl.).'

For many Algonquian languages, such splitting can be undertaken in a relatively straightforward manner for the most part, but there are word forms where this is not easy at all. For example, Nichols (1980) analyzes Southwestern Ojibwe as having 14 suffix position classes. Certain negative forms appear to show the reinsertion of morphological elements, such as *niwaabam<u>aa</u>siw<u>aa</u>naan* 'we (excl.) do not see him', which shows the TA direct theme sign /-aa/ both before and after the negative suffixes $_4$/-si/ and $_5$/-w/. An alternative item-and-arrangement approach is to treat the entire suffix complex as a unit, in the spirit of the word and paradigm

approach to morphology as exemplified in Blevins (2006; cf. also Harrigan et al. 2016). Entire suffix-complex strings were used to generate inflections in the Ojibwe dictionary illustrated in Figure 2. For learners, too, learning entire suffix complexes would seem much easier than attending to individual, sequentialized morphemes in 14 positions.

From the perspective of computational modeling, being able to describe complex word structure as minimally as possible with possibly extensive sets of rules for morpheme concatenation and for morphophonological processes was desirable early on, due to limits on computer memory. Devising such rules so that they are both complete and accurate is a time-consuming task, and in some cases simply enumerating chunks of less-regularly decomposable morpheme sequences and their associated features would be a more efficient option, and likely a psychologically more valid one as well. Moreover, the exponential increase of computer memory and processing speed has turned the chunking strategy into a viable one. Thus, we can instead present the Plains Cree forms in (3a–b) as consisting of two chunks, a circumfix-like element (made up of a prefix and suffix sequence) and the intervening stem, both associated with one or more morphosyntactic features. Because there are much fewer morpheme junctures (two in this case), one needs fewer rules to deal with potential morphophonological variation. Lexical databases often already contain such a chunked decomposition as a part of documentation work, so in order to create a computational model a linguist does not need to spend more time on devising and testing myriad rules to split these chunks further. For the VTA examples (3a–b), we can thus instead specify the Actor and Goal as first person plural (exclusive) and third person plural, or with the roles inverted, based on the entire *ni* . . . ânânak (1 . . . 1p→3p) or *ni* . . . *ikonânak* (1- . . . -3p→1p) chunks enveloping the stem *wîcih-*, without any need for further splitting.

Sometimes, chunking can even include stem or stem-final material to allow for more consistent string matching to determine stem classes. For the East Cree search engine (Junker and Stewart 2008), the verbal 'suffix' included the final stem vowel or consonant. The database (the Model) can thus include several layers of analysis, with different representations stored up (including sound files), which can be queried by different rules (the Controller) to offer different displays (the View). In Table 3 (from the database of East Cree model verbs), the third person relational dubitative form of the n-stem verb *takushin* contains multiple representations.[5] Furthermore, even if one opts for maximal chunking, such chunks can be marked with preidentified morphological splits (e.g., line [e] in Table 3), when known

TABLE 3. Southern East Cree VAI n-stem takushin relational third person independent indicative dubitative neutral

a. ᑕᐃᔑᓂᐧᐁᒉ	Word form in standard SEC syllabic spelling
b. takushinuweche	Word form in standard SEC roman spelling
c. takushinuuhche	Older spelling, converted from legacy syllabics
d. takushi-nuweche	Search Engine chunks
e. takushin-u-weche	Morpheme cuts for display: italics, bold
f. takushinw-we-ʔche	Morpheme break with underlying forms

or applicable, thus not requiring any dynamically implemented morphological decomposition.

Relationships with Dictionaries

Representations of verb paradigms are intimately linked to dictionary databases. Two basic pieces of information are essential for modeling: verb class indicated as part of speech and stem type. A number of restrictions to prevent overgenerating forms must also be stored in the lexicon. Here, we give a few examples of common problems and solutions for Algonquian languages we have worked with, and we show how modeling with dictionary databases can lead to better documentation of verb paradigms.

Number Restrictions

Some verbs only appear in the plural, which must be indicated in the dictionary database (Model), in a dedicated field, and read by the Controller to block singular forms from being generated, e.g., for numeral verbs (5a). Conversely, forms only used in the singular, like impersonal verbs (5b), also need to be marked in the dictionary. These examples are from the East Cree Dictionary (Junker et al. 2012):

(5) a. nîshuwich (VAI) stem: *i* , **pl.** 'they (anim.) are two'.

b. chimûn (vai) stem: *n*, **impersonal** 'it is raining'.

Derivational Information

We saw that the Western Ojibwe Dictionary generates all the reflexive forms of a VTA verb, but how should one treat reciprocal forms? Should these be stored in the lexicon or in the paradigms? Where do we encode 'productivity'? Dictionaries tend to have representative samples of relative root addition (6), reduplication, reciprocal and reflexive verbs, and secondary derivation processes like causatives or applicatives, but for modeling we need to be able to restrict generation rules on the level of each individual lexical entry. While some restrictions can be deduced based on pragmatic reasoning, in many cases we can discover the actual restrictions on inflectional generality/productivity only with corpus work.

(6) apû > **it**apû
'she sits' > 'she sits a certain way'

Lexicalized Forms

Lexicalized forms can be a challenge. Which grammatical category do we give to lexicalized forms such as passive (unspecified actor) forms of VAI verbs like (7b) or inflected verb forms in the Cree conjunct subjunctive like (8b)? Some guiding principles can be derived from modeling constraints, in terms of what information is minimally necessary and sufficient for a user to be able to conjugate such verbs. One solution is to create a special subtype for parts-of-speech, e.g., 'VII, impersonal' in (7b) or 'VII, subj. (VII conjunct subjunctive)' in (8b).

(7) a. makusheu (VAI) stem: *e* 's/he feasts'

b. makushânû (VII, impersonal) stem: û 'there is a feast' (East Cree Dictionary)

(8) a. uapan (VII) VII stem: *n*; CONJ. uapak*;* SUBJ. uapaki 'it is dawn, daylight'

b. uapaki (VII, subj) 'tomorrow', conjunct form of *uapan* (Innu Dictionary)

Other lexicalization patterns commonly found in Algonquian include VTA inverse forms that only take an inanimate agent, often labeled VAI in Cree dictionaries and VTAI in Ojibwe. But what is the conjugation class of these new derived

FIGURE 3. Modeling initial change for East Cree verbs

ecn	12b-N (0)	14-N (0)
ᐋᒋᐦᑎᓐ aachihtin	ᐃᔮᒋᐦᑎᐦᒡᐦ iyaachihtihchh	ᐃᔮᒋᐦᑎᓂᒀ iyaachihtinikwaa
ᐋᐦ ᐅᐦᒋ ᐅᐦᐱᓂᑭᓂᐎᐨ aah uhchi uhpinikiniwich	No example for stem: unknown	No example for stem: unknown
ᐋᐦᒋᓈᑯᓐ aahchinaakun	ᐃᔮᐦᒋᓈᑯᐦᒡᐦ iyaahchinaakuhchh	ᐃᔮᐦᒋᓈᑯᓂᒀ iyaahchinaakunikwaa
ᐋᐦᒋᓂᐦᑳᑖᐤ aahchinihkaataau	ᐃᔮᐦᒋᓂᐦᑳᑖᐦᒡᐦ iyaahchinihkaataachh	ᐃᔮᐦᒋᓂᐦᑳᑖᒀ iyaahchinihkaataakwaa
ᐋᐦᒋᐱᐦᑖᒥᑭᓐ aahchipihtaamikin	ᐃᔮᐦᒋᐱᐦᑖᒥᑭᐦᒡᐦ iyaahchipihtaamikihchh	ᐃᔮᐦᒋᐱᐦᑖᒥᑭᓂᒀ iyaahchipihtaamikinikwaa
ᐋᐦᒋᐱᔨᐤ aahchipiyiu	ᐃᔮᐦᒋᐱᔨᐦᒡᐦ iyaahchipiyichh	ᐃᔮᐦᒋᐱᔨᒀ iyaahchipiyikwaa
ᐋᐦᒋᔥᑖᐤ aahchishtaau	ᐃᔮᐦᒋᔥᑖᐦᒡᐦ iyaahchishtaachh	ᐃᔮᐦᒋᔥᑖᒀ iyaahchishtaakwaa
ᐋᐦᒋᔨᐙᐱᔨᐤ aahchiyiwaapiyiu	ᐃᔮᐦᒋᔨᐙᐱᔨᐦᒡᐦ iyaahchiyiwaapiyichh	ᐃᔮᐦᒋᔨᐙᐱᔨᒀ iyaahchiyiwaapiyikwaa
ᐋᐦᒋᔨᐙᐤ aahchiyiwaau	ᐃᔮᐦᒋᔨᐙᐦᒡᐦ iyaahchiyiwaachh	ᐃᔮᐦᒋᔨᐙᒀ iyaahchiyiwaakwaa
ᐋᐦᑯᐦᐄᐙᐤ aahkuhiiwaau	ᐃᔮᐦᑯᐦᐄᐙᐦᒡᐦ iyaahkuhiiwaachh	ᐃᔮᐦᑯᐦᐄᐙᒀ iyaahkuhiiwaakwaa
ᐋᐦᑯᐃᐦᑎᐙᐱᔅᒋᐦᑎᓐ aahkuihtiwaapischihtin	ᐃᔮᐦᑯᐃᐦᑎᐙᐱᔅᒋᐦᑎᐦᒡᐦ iyaahkuihtiwaapischihtihchh	ᐃᔮᐦᑯᐃᐦᑎᐙᐱᔅᒋᐦᑎᓂᒀ iyaahkuihtiwaapischihtinikwaa
ᐋᐦᑯᐃᐦᑎᐎᔥᑖᐤ aahkuihtiwishtaau	ᐃᔮᐦᑯᐃᐦᑎᐎᔥᑖᐦᒡᐦ iyaahkuihtiwishtaachh	ᐃᔮᐦᑯᐃᐦᑎᐎᔥᑖᒀ iyaahkuihtiwishtaakwaa
ᐋᐦᑯᐃᐦᑐᐎᐲᐦᒋᐦᑎᓐ aahkuihtuwipiihchihtin	ᐃᔮᐦᑯᐃᐦᑐᐎᐲᐦᒋᐦᑎᐦᒡᐦ iyaahkuihtuwipiihchihtihchh	ᐃᔮᐦᑯᐃᐦᑐᐎᐲᐦᒋᐦᑎᓂᒀ iyaahkuihtuwipiihchihtinikwaa
ᐋᐦᑯᓐ aahkun	ᐃᔮᐦᑯᐦᒡᐦ iyaahkuhchh	ᐃᔮᐦᑯᓂᒀ iyaahkunikwaa
ᐋᐦᑯᔨᐦᑐᐎᐲᐦᒋᔅᑎᓐ aahkuyihtuwipiihchistin	ᐃᔮᐦᑯᔨᐦᑐᐎᐲᐦᒋᔅᑎᐦᒡᐦ iyaahkuyihtuwipiihchistihchh	ᐃᔮᐦᑯᔨᐦᑐᐎᐲᐦᒋᔅᑎᓂᒀ iyaahkuyihtuwipiihchistinikwaa
ᐋᐦᒁᒋᐱᔨᐤ aahkwaachipiyiu	ᐃᔮᐦᒁᒋᐱᔨᐦᒡᐦ iyaahkwaachipiyichh	ᐃᔮᐦᒁᒋᐱᔨᒀ iyaahkwaachipiyikwaa
ᐋᐦᒁᐦᑭᑎᑖᐤ aahkwaahkititaau	ᐃᔮᐦᒁᐦᑭᑎᑖᐦᒡᐦ iyaahkwaahkititaachh	ᐃᔮᐦᒁᐦᑭᑎᑖᒀ iyaahkwaahkititaakwaa
ᐋᐦᒁᑭᒥᑖᐤ aahkwaakimitaau	ᐃᔮᐦᒁᑭᒥᑖᐦᒡᐦ iyaahkwaakimitaachh	ᐃᔮᐦᒁᑭᒥᑖᒀ iyaahkwaakimitaakwaa
ᐋᐦᒁᑯᓈᐅᑎᓐ aahkwaakunaautin	ᐃᔮᐦᒁᑯᓈᐅᑎᐦᒡᐦ iyaahkwaakunaautihchh	ᐃᔮᐦᒁᑯᓈᐅᑎᓂᒀ iyaahkwaakunaautinikwaa
ᐋᐦᒁᐱᓯᔅᑖᐤ aahkwaapisistaau	ᐃᔮᐦᒁᐱᓯᔅᑖᐦᒡᐦ iyaahkwaapisistaachh	ᐃᔮᐦᒁᐱᓯᔅᑖᒀ iyaahkwaapisistaakwaa
ᐋᐦᒁᓯᓈᑯᓐ aahkwaasinaakun	ᐃᔮᐦᒁᓯᓈᑯᐦᒡᐦ iyaahkwaasinaakuhchh	ᐃᔮᐦᒁᓯᓈᑯᓂᒀ iyaahkwaasinaakunikwaa
ᐋᐦᒁᔅᑭᑎᓐ aahkwaaskitin	ᐃᔮᐦᒁᔅᑭᑎᐦᒡᐦ iyaahkwaaskitihchh	ᐃᔮᐦᒁᔅᑭᑎᓂᒀ iyaahkwaaskitinikwaa
ᐋᐦᒁᑖᔅᑭᑎᓐ aahkwaataaskitin	ᐃᔮᐦᒁᑖᔅᑭᑎᐦᒡᐦ iyaahkwaataaskitihchh	ᐃᔮᐦᒁᑖᔅᑭᑎᓂᒀ iyaahkwaataaskitinikwaa
ᐋᐦᒁᑖᔮᐦᑖᑯᓐ aahkwaataayihtaakun	ᐃᔮᐦᒁᑖᔮᐦᑖᑯᐦᒡᐦ iyaahkwaataayihtaakuhchh	ᐃᔮᐦᒁᑖᔮᐦᑖᑯᓂᒀ iyaahkwaataayihtaakunikwaa
ᐋᐦᒁᑎᓐ aahkwaatin	ᐃᔮᐦᒁᑎᐦᒡᐦ iyaahkwaatihchh	ᐃᔮᐦᒁᑎᓂᒀ iyaahkwaatinikwaa
ᐋᐦᑖᑭᒥᐱᔨᐤ aahtaakimipiyiu	ᐃᔮᐦᑖᑭᒥᐱᔨᐦᒡᐦ iyaahtaakimipiyichh	ᐃᔮᐦᑖᑭᒥᐱᔨᒀ iyaahtaakimipiyikwaa
ᐋᐦᑖᐳᑖᐤ aahtaaputaau	ᐃᔮᐦᑖᐳᑖᐦᒡᐦ iyaahtaaputaachh	ᐃᔮᐦᑖᐳᑖᒀ iyaahtaaputaakwaa
ᐋᐦᑖᔅᑯᐱᔨᐤ aahtaaskupiyiu	ᐃᔮᐦᑖᔅᑯᐱᔨᐦᒡᐦ iyaahtaaskupiyichh	ᐃᔮᐦᑖᔅᑯᐱᔨᒀ iyaahtaaskupiyikwaa
ᐋᐦᑎᑯᑖᐤ aahtikutaau	ᐃᔮᐦᑎᑯᑖᐦᒡᐦ iyaahtikutaachh	ᐃᔮᐦᑎᑯᑖᒀ iyaahtikutaakwaa

forms? Sometimes, new paradigms have to be created to accommodate these. For example, for VAI forms lexicalized from the VTI passive in Innu, we created a new model conjugation, ending in *-kanu*, treating this form as a VAI stem.

Modeling for Accurate Documentation

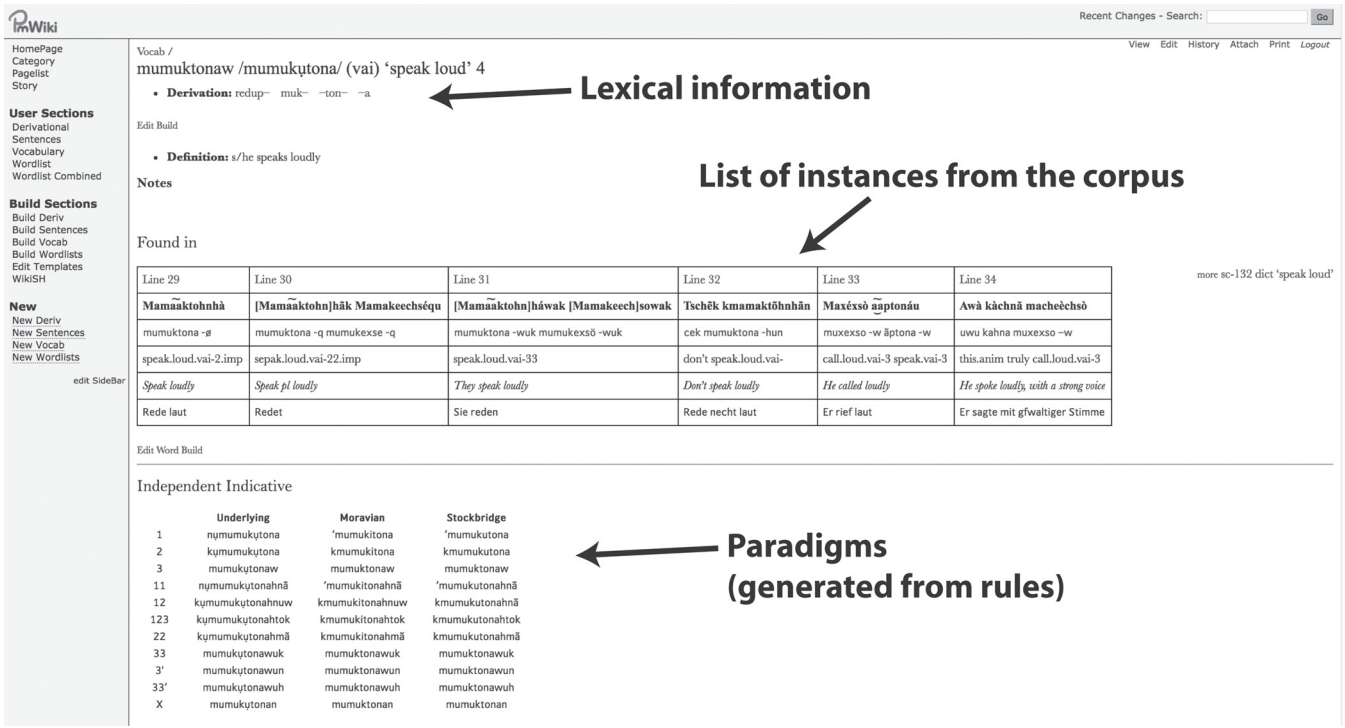
Modeling not only allows us to generate forms and build search engines, it also has the advantage of allowing us to check large amounts of real data against the model. For example, in 2007 Junker investigated the rules of initial change in East Cree. With Terry Stewart, they modeled two changed forms for each verb in the Cree dictionary, and during a workshop with elders, Junker and her Cree collaborators went through a list of over 20,000 Cree verbs, doing spot checks to verify and improve the descriptive rules for initial change (Figure 3).[6]

Linking Things Together

Many potential audiences must be able to access the paradigm/dictionary database, including language learners, educators, first language speakers, and linguists. Each of these audiences can have a DISPLAY or VIEW specific to their needs, which is generated from the same underlying database. There are a variety of ways to display the paradigms. One way familiar to many users is a wiki-driven web-based site, accessible anywhere even by means of mobile devices. A wiki, out of the box, excels at searching, linking, and tagging information from the underlying database. There are built-in tools for handling multimedia (sound, images, and video), and it is relatively easy to set up with instant online access.

A good case study is a Mahican language database developed by Harvey (2015) which combines the written corpus (interlinearized) with a dictionary database and dynamic paradigm generator—an example web page from the wiki display is in Figure 4. Any instance of a lemma or affix can be linked to its proper lexical entry page. Such a lexical entry wiki page shows several selected fields directly from the database (derivation, definition, part of speech, notes, etc.). The list of instances from the corpus is built dynamically via wiki tools. Each wiki page also has a list of tags or categories that flag potential points of interest. Here, clicking, e.g., "redupl" would extract and present a list of all reduplicative verbs in the corpus. This is particularly useful when the researcher discovers an unusual or unknown form or

FIGURE 4. Example lexical entry page in Mahican lexical database



Line 29	Line 30	Line 31	Line 32	Line 33	Line 34
Mamaa͡ktohnhà	**[Mamaa͡ktohn]hāk Mamakeechséqu**	**[Mamaa͡ktohn]háwak [Mamakeech]sowak**	**Tschĕk kmamaktōhnhān**	**Maxéxsò a͡aptonáu**	**Awà kàchnă macheèchsò**
mumuktona -ø	mumuktona -q mumukexse -q	mumuktona -wuk mumukexsŏ -wuk	cek mumuktona -hun	muxexso -w āptona -w	uwu kahna muxexso –w
speak.loud.vai-2.imp	sepak.loud.vai-22.imp	speak.loud.vai-33	don't speak.loud.vai-	call.loud.vai-3 speak.vai-3	this.anim truly call.loud.vai-3
Speak loudly	*Speak pl loudly*	*They speak loudly*	*Don't speak loudly*	*He called loudly*	*He spoke loudly, with a strong voice*
Rede laut	Redet	Sie reden	Rede necht laut	Er rief laut	Er sagte mit gfwaltiger Stimme

	Underlying	Moravian	Stockbridge
1	nụmumukụtona	'mumukitona	'mumukutona
2	kụmumukụtona	kmumukitona	kmumukutona
3	mumukụtonaw	mumuktonaw	mumuktonaw
11	nụmumukụtonahnā	'mumukitonahnā	'mumukutonahnā
12	kụmumukụtonahnuw	kmumukitonahnuw	kmumukutonahnā
123	kụmumukụtonahtok	kmumukitonahtok	kmumukutonahtok
22	kụmumukụtonahmā	kmumukitonahmā	kmumukutonahmā
33	mumukụtonawuk	mumuktonawuk	mumuktonawuk
3'	mumukụtonawun	mumuktonawun	mumuktonawun
33'	mumukụtonawuh	mumuktonawuh	mumuktonawuh
X	mumukụtonan	mumuktonan	mumuktonan

structure in the data; this can be tagged and analyzed with other examples of the same form at a later date. Finally, the verb paradigm is generated by combining principal parts from the lexical database to appropriate affix chunks, where morphophonological rules are applied just before actual display.

As seen in Figure 4, the audience here is the linguist comparing instances in the corpus to a model of the verb paradigm. Where the generated paradigms on the wiki page disagree with the attested forms, the lexeme or the model (affix chunks and morphophonology) can be immediately corrected.[7] However, if the target audience is second language learners, different display forms can be selected when the wiki page is output: the person numbers (1, 2, 3, . . .) could be replaced by Mahican pronouns (*nia*, *kia*, *naakmã*, . . .), the user could select a specific dialect, the interlinear form could instead be shown as an example sentence, and any notes could be omitted.

Conclusions

The most important consideration at the beginning of any project is the design and construction of the database in a consistent and systematic way, guaranteeing future compatibility and portability, and the ability to compare information with other linguistic databases as seamlessly as possible. Employing the database model outlined in this paper, systematic and consistent work can be easily and instantaneously tailored to a broad range of potential users. There are clear benefits to such a system for verb paradigms: researchers can test their model paradigms against a corpus; native speakers can have quick access to a source for standardized spelling; educators can plan curriculum derived from this resource; and learners can have a place to look up those verbs when they need them, in real-life situations where conversation requires an unfamiliar form.

NOTES

1. "Dormant" refers to a language with no speakers or semispeakers but for which ample documentation exists and there is a community that recognizes the language as part of their cultural heritage.
2. For influential twentieth-century models, see Ellis 1971, updated in Ellis 2016, and Wolfart 1973.

3. Also, using certain characters that have a special function in most computer code, such as an apostrophe for indicating obviation, e.g., 3', instead of the proper unicode character for 'prime' (O2B9), or symbols for indicating Actor-Goal direction that are also angle brackets, e.g., 2>3, instead of the unicode arrow 2→3, can lead to severe difficulties in importing data into a computer database, and the use of such characters should be avoided.
4. See Baraby and Junker 2011–2014 and Junker and MacKenzie 2010–2015, 2011–2015. The glossing templates owe much to discussions with Bill Jancewicz and Rand Valentine.
5. There is a clear pedagogical advantage for language learners to associate conjugations with verb classes, based on their Algonquian verb finals. For example, in the above example, a subclass with the final *-shin* 'on the horizontal' could be coded to further predict the conjugation pattern of semantically related verbs. In Ojibwe, identifying the VAI final *-ose* 'walk' can successfully predict the conjugation of a whole series of 'walking' verbs such as: *animose, aagimose, babaamose, babimose, bedose, bimose, bimwewedaawangose, bimweweyaagonewose*, etc. (see Ojibwe People's Dictionary for translations, http://ojibwe.lib.umn.edu/). We suggest that those tasks are best handled by and within dictionaries.
6. Similarly, the imperative forms of all 1,645 *u* stem AI verbs of the Innu dictionary were generated with two possible imperatives in 2014, to allow Innu editor Yvette Mollen to select the correct form, which is the test for long and short *u* stems. As a result, three categories of *u* stems were created for the database: *long u*, *short u*, and just *u* for verbs that are always in the plural where underlying length is not determinable.
7. One example is the verb suffix *-sa* (cognate with the Delaware present aspect suffix). At first, the paradigm generator did not produce this form—there were so few instances that one could not determine precisely what *-sa* means in a given sentence. During interlinearization, the suffix *-sa* was tagged wherever it appeared, and a link was automatically created. Throughout this process, the tag-link could be clicked, and all instances of *-sa* were listed in their context. Its extant functions now apparent, and with a sufficient number of instances on the corpus to be sure of the form of the suffix, it could then be added to paradigm generator. This method has been very useful in finding unpredicted forms and variation.

REFERENCES

Antonsen, Lene, Ryan Johnson, Trond Trosterud, and 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, pp. 27–38. NEALT Proceedings Series, vol. 17 / Linköping Electronic Conference Proceedings, vol. 86. Linköping: Linköping University Electronic Press.

Arppe, Antti, Lene Antonsen, Trond Trosterud, Sjur Moshagen, Dorothy Thunder, Conor Snoek, Timothy Mills, Juhani Järvikivi, and Jordan Lachler. 2015. Turning language documentation into reader's and writer's software tools. Fourth International Conference on Language Documentation and Conservation (ICDLC 4), 26 February–1 March 2015, Honolulu.

Baraby, Anne-Marie, and Marie-Odile Junker. 2011–2014. 3e éd. *Conjugaisons des verbes innus*. http://verbe.innu-aimun.ca.

Beesley, Kenneth R., and Lauri Karttunen. 2003. *Finite state morphology*. Stanford, CA: CSLI Publications.

Blevins, James. 2006. Word-based morphology. *Journal of Linguistics* 42:531–573.

Ellis, C. Douglas. 1971. Cree verb paradigms. *International Journal of American Linguistics* 37(2):76–95.

———. 2016. Verb Paradigms. In *Spoken Cree glossary.* Ottawa: Carleton University. spokencree.org.

Harrigan, Atticus, Lene Antonsen, Antti Arppe, Dustin Bowers, Trond Trosterud, and Arok Wolvengrey. 2016. Learning from the computational modeling of Plains Cree verbs. Workshop on computational methods for descriptive and theoretical morphology, Seventeenth International Morphology Meeting, Vienna, 18–21 February 2016.

Harvey, Christopher. 2015. *Wiki-generated paradigm tools*. Fourth International Conference on Language Documentation and Conservation (ICDLC 4), 26 February–1 March 2015, Honolulu.

itwêwina. N.d. *itwêwina*—Intelligent on-line dictionary for Plains Cree. Alberta: University of Alberta. http://altlab.ualberta.ca/itwewina/.

Junker, Marie-Odile, and Marguerite MacKenzie. 2010–2015. *East Cree* (*Northern Dialect*) *verb conjugation*. 4th ed. http://verbn.eastcree.org.

———. 2011–2015. *East Cree* (*Southern dialect*) *verb conjugation*. 4th ed. http://verbs.eastcree.org.

Junker, Marie-Odile, Marguerite MacKenzie, Luci Bobbish-Salt, Alice Duff, Ruth Salt, Anna Blacksmith, Patricia Diamond, and Pearl Weistche (eds). 2012. *The Eastern James Bay Cree dictionary on the web: English-Cree and Cree-English, French-Cree and Cree-French* (*Northern and Southern dialects*). http://dictionary.eastcree.org/.

Junker, Marie-Odile, and Terry Stewart. 2008. Building search engines for Algonquian languages. *Papers of the Thirty-Ninth Algonquian Conference*, ed. by Karl S. Hele and

Regna Darnell, pp. 378–411. London: University of Western Ontario Press.

———. 2011. A linguistic atlas for endangered languages: www.atlas-ling.ca. *Proceedings of the Eleventh International Conference on Education and New Learning Technologies* (*EDULEARN 11*). Barcelona.

Junker, Marie-Odile, Delasie Torkornoo, and J. Randolph Valentine. 2013. Relational databases for Cree, Innu, and Ojibwe dictionaries. Forty-Fifth Algonquian Conference, University of Ottawa, October 2013.

Krasner, Glenn E., and Stephen T. Pope. 1988. A cookbook for using the model–view controller user interface paradigm in Smalltalk-80. *Journal of Object-Oriented Programming* 1(3):26–49.

MacKenzie, Marguerite. 1980. Towards a dialectology of Cree-Montagnais-Naskapi. PhD thesis, University of Toronto.

Mailhot, José, Marguerite MacKenzie, and Marie-Odile Junker. 2013. *Online Innu dictionary*. http://www.innu-aimun.ca/dictionary.

Moose, Lawrence Leonard, Mary Moose, Gordon Jourdain, Marlene Stately, Leona Wakonabo, Eugene Stillday, Anna Gibbs, Rosemarie DeBungie, and Nancy Jones. 2009. *Aaniin Ekidong: Ojibwe vocabulary project*, ed. by Anton Treuer and Keller Papp. St. Paul: Minnesota Humanities Center.

Nichols, John D. 1980. Ojibwe morphology. PhD thesis, Harvard University.

Nichols, John, and Earl Nyholm. 1995. *Concise dictionary of Minnesota Ojibwe*. Minneapolis: University of Minnesota Press.

Valentine, J. Randolph, and Patricia N. Ningewance. 2009. Western Ojibwe dictionary. Computer application, University of Wisconsin–Madison.

Wolfart, H. Christoph. 1973. *Plains Cree: A grammatical study*. Transactions of the American Philosophical Society, n.s., 63(5). Philadelphia: American Philosophical Society.

Wolvengrey, Arok (ed.). 2001. ᓀᐦᐃᔭᐁᐧᐃᐧᣠ: ᐃᐧᑌᐃᐧᓇ / *nēhiýawēwin: itwēwina / Cree: Words*. Regina: Canadian Plains Research Center.

———. 2011. *Semantic and pragmatic functions in Plains Cree syntax*. Ultrecht, The Netherlands: LOT.