An Expanded Finite-State Transducer for Tsuut'ina Verbs

Joshua Holden[1], Christopher Cox[2], Antti Arppe[1]

University of Alberta[1] Carleton University[2] holden1@ualberta.ca, christopher.cox@carleton.ca, arppe@ualberta.ca

1. Introduction

This paper describes the development of a finite state transducer (FST) for the inflectional system of verbs in Tsuut'ina (ISO 639-3: srs), a highly endangered Dene (Athabaskan) language spoken near Calgary, Alberta, Canada. Arppe et al. (2017) outlined the creation of a prototype FST for intransitive verbs in Tsuut'ina, with only a few example lexemes. This paper describes the expansion of this FST architecture with a larger set of intransitive verbs, as well as to encompass transitive verbs of various argument structures, which vastly increases the range and complexity of the morphology that the FST needs to represent. Full verb conjugations are needed for language learning resources, but manually creating paradigm tables would be impractical. The FST automates this and responds to the Tsuut'ina community's desire for computer-based language tools that are typical FST applications, such as a morphologically enhanced online dictionary, a spellchecker, and predictive word suggestions on mobile devices. Through the integration of this FST with the GiellaLT infrastructure, all of these applications are readily derivable from the FST described in this paper.

1. Basic FST Architecture for Tsuut'ina Intransitive Verbs

Arppe et al. (2017) presented the creation of the core architecture of the FST for Tsuut'ina. Dene verbs do not use a simple, linear lexical + derivational + inflectional concatenation; rather, the three types of prefixes are interlaced (see Figure 2). The lexical tier or "verb theme" consists of a stem on the right edge, a voice/valence prefix directly to its left, and up to three prefix positions: inner lexical, a middle "areal" slot (the areal prefix being a historical agreement prefix referring to a place or situation, but today lexicalized as part of many verb themes), and an outer lexical position. See Rice (2000), Hoijer (1945), and Sapir and Hoijer (1967) for basic references on template structure.

The inflectional tier includes categories of viewpoint aspect, mood, as well as number and person agreement for subjects and objects. For the basic FST described Arppe et al. (2017), three insertion points in the verb word were necessary: 1) the outer inflectional zone directly to the right of the outer lexical prefixes (to handle the distributive plural), 2) the middle zone between the areal and the inner lexical prefixes (where third-person unspecified subject prefix *ts'i*–, and third-person plural subject *gi*– are found), and finally 3) the inner, TAMA (tense-aspect-mood-agreement) chunk zone, between the inner lexical zone and the stem-classifier combination. There are many allomorphic co-occurrence restrictions between TAMA chunks and lexical portions.

Computationally, the FST uses a 'chunking', or portmanteau, approach to the combination of aspect, mood and agreement prefixes to the left of the classifier,. This whole area is often collapsed into single syllables whose internal complexity (see Rice 2001) would be quite hard to model in a fully decomposed form. The portmanteau approach has some support from linguistic studies—see Rice et al. (2002) for psycholinguistic results involving Dene Sųłiné; Young and Morgan (1987), Faltz (1998), and McDonough (2000) for Navajo, Holden (2013) for Dene Sųłiné and Leer (1999) (inter alia) for comparison across multiple Dene languages. From the perspective of FST modelling it simply made the complexity more manageable, as only the junctures between the TAMA chunk and the surrounding morphemes are visible to the model. **Because** both the lexical and inflectional tiers of the morphology are discontinuous, we make use of three separate finite-state models for each of the three inflectional zones, as well as a fourth for the lexical tier (including the stem at the right edge and the possible preceding discontinuous lexical prefixes), all specified with the *lexc* formalism (Beesley and Karttunen, 2003). Using finite-state operations, the three inflectional component FSTs are inserted in the appropriate slots within the lexical tier (see Arppe et al. (2017, 58, ex. 2) for the master specification). The morphophonemic processes are modelled with contextual rewrite rules using *xfst* (Beesley and Karttunen, 2003). The resultant fifth FSM is then composed together with the morphological FSM. A sixth FSM is concatenated with this to link flag diacritics with morphological feature tags. Figure 2 outlines the structure of these six constituent FSMs.

The slots for inflectional morphemes within the lexical tier are indicated by a specific notation (see Figure 1 above): "." (period) stands for the inner boundary, where the TAMA chunk is to be inserted, "_" (under- score) for the middle boundary (where object agreement and third-person subjects are located), and "=" (equal sign) for the outer boundary (where the distributive plural prefix can occur). A system of flag diacritics then filters out disallowed combinations (i.e. it implements co-occurrence restrictions).

2. An Overview of Participant Marking in Tsuut'ina Verbal Morphology

A single inflected form such as (1) can index up to three event participants with pronominal markers appearing in several places in the verb-word (in boldface):

(1) moghàninischúd-áát'a
mi–oghà–**ni–nis–**s–chúd=i=át'a
3SG.IO–to–2SG.DO–1SG.SBJ:*ni*.IPFV–VV–feed.IPFV=FUT =ASRT
'I am going to feed you to it.'

In this example, inflectional markers related to event participants appear at several points: at the far left edge of the verb word, with an 3SG indirect object *mi-* marked before the incorporated postposition *oghà*– 'to'; after that postposition, with the 2SG direct object *ni*– ; and immediately before the verb stem, with a portmanteau morpheme *nis-* representing the 1SG subject form of the *ni-* imperfective paradigm with an *s-* voice-valence marker.

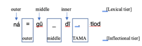

Figure 1: FST conventions for inflectional zones of *náguditlod* 's/he jumps down'

In order to extend the previously proposed FST architecture to represent non-subject event participants, then, several other linguistic facts need to be taken into consideration:

1. Object-marking patterns need to be implemented to represent both the indirect and direct object morphology seen in (1) above

2. Interactions between subject and object-marking morphology need to be represented adequately in an expanded computational model. The third-person objects have different realizations when acted on by third-person vs. non-third- person subjects. Some subject markers' positions may differ based on the presence or absence of certain object markers, such as *gi-* 3PL, which typically appears as a middle prefix (2a), but can also appear at the left edge of the word when combining with *yi-* 3D in 3PL>3SG.IO contexts (2b):

(2) a. soghàgistà
si–oghà–gi–i–s–tà
1SG.IO–to–3PL–3:*ni*.IPFV–VV–handle-animate.IPFV
'They will give something (animate) to me.'

b. giyoghàstà
gi–yi–oghà–i–s–tà
3PL 3D.IO–to–3.*ni*.IPFV–*s*.VV–handle-animate.IPFV
'They will give something (animate) to him/her/it.'

An expansion of the FST model of Tsuut'ina verbal morphology that aims to include all the attested pat- terns of verbal participant marking thus presents a potentially valuable "stress test" of the existing finite-state architecture for Dene languages, particularly in its ability to generate and recognize a much wider range of verb forms. As we note below, applying such a model to a much larger lexical sample also raises questions about current conceptions of what constitutes a lemma in morphologically rich languages such as Tsuut'ina, and of where the boundary lies between inflectional and derivational information in Dene verbal morphology.

3. Expansion of the Inflectional Tsuut'ina FST

The current expansion is meant to generate full inflectional paradigms for a comprehensive range of verbs as represented by the Onespot-Sapir glossary. This includes all transitive, oblique object and ditransitive cases in this database. We have not followed, however, a very rigorous distinction between inflection and derivation for our FST model. In particular, several aspectual categories normally seen as derivational are included in the inflectional FST either because it made the modelling itself much easier (see discussion of the transitional or inchoative and semelfactive prefixes in Section 3.2.1 and of the conative in Section 3.2.5) or in order to make the model's output fit with Tsuut'ina community language teaching traditions and priorities (see Section 3.1's discussion of the repetitive and progressive aspects). The following section 3.1 will review the language source data used for this expansion, and Section 3.2 will address the new characteristics of FST modelling of transitive verbs and the other argument structure types.

3.1. Re-elicitation and Organization of the Onespot-Sapir Glossary

The source material was an unpublished glossary collected by Edward Sapir and John Whitney-Onespot in Tsuut'ina in 1922 that contains numerous verb paradigms. In the 1990s, Tsuut'ina community linguist and speaker Bruce Starlight and collaborator Gary Donovan started transcribing and editing Sapir and Whitney-Onespot's notebooks (see Starlight et al. 2016).

Paradigms were labelled for argument structure, transitivity, aspect, and TAMA chunk subtype by co-author Holden (sometimes in consultation with co-author Cox), and the specific allomorphic combination of the stem + lexical prefixes for each TAMA value was recorded (with some temporarily excluded because their TAMA pattern was unclear, or some other factor that required verification by speakers).

The verb forms were assigned to lemmas, each of which would stand for a group of inflected wordforms. We treated repetitive and progressive aspects (linguistically derivational) as inflections to make the paradigms com- patible with Tsuut'ina community language education. The third-person singular wordform was chosen as this citation form, a convention already adopted for the intransitive model (Arppe et al., 2017).

3.2. Expansions Needed for the Onespot-Sapir Glossary

The Onespot-Sapir list contains close to 9 thousand ver- bal wordforms with many argument structure types, of which several are new to the FST model. The *lexc* excerpt below shows the argument structures and associ- ated flag diacritics, with all except the first two being novel to the expanded FST.

LEXICON Root
@U.VALENCE.IMPERSONAL@
NoDistributive;
@U.VALENCE.INTRANSITIVE@ SubjectOnly;
@U.VALENCE.TRANSITIVE@
SubjectAndDirectObject;
@U.VALENCE.OBLIQUEOBJECT@
SubjectAndObliqueObject;
@U.VALENCE.DITRANSITIVE@
SubjectDirectObjectOrObliqueObject;
@U.VALENCE.EXPERIENCER@
ObliqueObjectOnly;
@U.VALENCE.TRANSITIONAL@
SubjectAndDirectObject;

Given that inflectional material can be found in several zones of the verb, expanding the range of verbs and argument structures requires an additional inflectional prefix position as well as an increase in the range of prefixes found in existing slots.

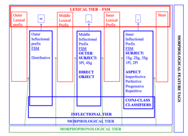

Figure 2: Core FST architecture

3.2.1. Inner zone TAMA inflections

The point directly to the left of the stem-classifier com- bination is the most complex site of inflection in Dene languages, where aspect, mood and person agreement are found. Prefix sequences here are treated as cumulative morphemes informally called "TAMA chunks".[6]

Dene languages have derivational situation aspect marking (see Rice (2000) for a further discussion) In dynamic, transitive transitional verbs such as (2b) above, the low tone is present with the *mi*– object prefix (e.g., *nà**gimì**nistà* 'I am setting them (*gimì*-, with low- tone *ì*-) down'), but absent with a full nominal complement (e.g., *tłích'áká nànistà* 'I am setting the dogs down' (no low tone)). To handle this, we added the specific argument structure "transitional transitive" in addition to distinct transitional TAMA chunk allomorphs.

3.2.2. Middle Zone Inflections

The "outer subject" prefixes *ts'i*– 'impersonal subject' and *gi*- 'third person plural subject' in this zone were already in the intransitive model in Arppe et al. (2017). This is the primary insertion point for the direct object prefixes, so adding these was a significant expansion to this zone. Flag diacritics for some of these values and the corresponding prefixes are exemplified in the *lexc* excerpt below.

LEXICON Transitive-Markers
! 1SG direct obj. (e.g., siyí?í "you saw me")
@U.OBJECTPERSON.1@
@U.OBJECTNUMBER.SG@
@P.PREFIX.MIDDLE@si Outer-Subjects;

The flag @U.DIRECTOBJECT.NOMINAL@ tells the FST to use the zero allomorph (i.e., no prefix) when an object noun is present, while the flag @U.DIRECTOBJECT.NONE@ is followed by @P.PREFIX.MIDDLE@yi, telling the FST to add the object prefix *yi*– in the absence of a direct object noun. The reflexive and reciprocal objects are handled in their own continuation lexica as well, due to a number of linguistic complexities (see *lexc* definitions below).

LEXICON Reflexive-Direct-Objects
@U.SUBJECTNUMBER.SG@
@U.SUBJECTPERSON.1@
@U.OBJECTNUMBER.SG@idi
@U.SUBJECTNUMBER.PL@
@U.SUBJECTPERSON.3@
@U.OBJECTNUMBER.PL@
@D.GI@@P.GI.ON@igidi
@U.SUBJECTNUMBER.SG@
@U.SUBJECTPERSON.4@
@U.OBJECTNUMBER.SG@its'idi Filter-Transitives;
LEXICON Reciprocal-Direct-Objects
@U.OBJECTNUMBER.PL@
@U.SUBJECTNUMBER.PL@
@U.SUBJECTPERSON.1@átłi Filter-Transitives;

3.2.3. Outer Zone Inflections

The outer insertion point only takes the *dà*– distributive prefix. While *dà*– usually pluralizes the subject, for transitive stems the distributive can refer to plurality of either the subject or the object. The second line would be marked "NoDistributive" if subject and direct object are both singular.

In any verbs whose object markers occur in the left-edge (historically oblique) inflection zone (see below), the distributive can refer to either subjects or (oblique or formerly oblique) objects. This is accounted for in a special lexicon for subjects with oblique objects, shown in the following *lexc* definitions.

LEXICON SubjectAndObliqueObject
@U.SUBJECTNUMBER.PL@
@U.OBLIQUENUMBER.PL@ Distributive;
@U.SUBJECTNUMBER.SG@
@U.OBLIQUENUMBER.PL@ Distributive;
@U.SUBJECTNUMBER.PL@
@U.OBLIQUENUMBER.SG@ Distributive;
@U.SUBJECTPERSON.4@
@U.OBLIQUENUMBER.SG@ Distributive;
@U.SUBJECTNUMBER.SG@
@U.OBLIQUENUMBER.SG@ NoDistributive;

3.2.4. Left-edge Inflections

In this expanded model, a fourth inflectional zone was necessary at the left edge of the verb word. If you recall the verb structure outlined in Section 1.2 and Section 2, Dene verb themes frequently incorporate postpositions at or toward their left edge (in the outer lexical zone). Often the postposition/ preverb is lexicalized so its meaning is not transparent. This was seen in example (1) above, where at the left edge the third-person object *mi*– is added to refer to the object of the feeding, while the normal direct object position of *s+chut* contains second-person object *ni*– 'you', resulting in a ditransitive verb. The verb "inherits" the incorporated postposition complement, resulting in double object mark- ing (if it is not a phrase). The FST assigns the flag @U.VALENCE.DITRANSITIVE@ to these cases. If the original analytical construction was a two-actant verb with an oblique object complement, the incorpo- ration of the postposition results in a direct transitive verb where the direct object inflection occurs at the left edge, rather than the standard middle position. The FST treats these as oblique object verbs with the flag diacritic @U.VALENCE.OBLIQUEOBJECT@. As with the case of the direct objects, the object prefix is suppressed when an overt nominal is present, so the FST must implement a null allomorph in these cases.

As with the transitive verbs, the 3rd person, reflexive and reciprocal were handled in separate lexica, because the same rules regarding distal/ proximal alternations and classifier changes apply.

3.2.5. Conative Paradigms

The conative con- tributes a meaning loosely glossable as 'attempt' or 'at' to the verb (e.g. 'shoot at' versus just 'shoot'). In other cases the conative is vestigial or it is hard to see what its exact contribution is, as in (3)

(3) yízi
yi– í– zi
3D CON call.IPFV
's/he is calling him/her'

Formally the conative occurs between the direct object prefixes and the third-person subject prefixes, directly in the middle of the middle zone of the FST inflectional model, so the chunking approach with TAMA will not work. Because the FST ar- chitecture was not designed to break up the middle inflectional zone, a unique flag diacritic was added to conative verbs, @U.CONATIVE.ON@, which ensured that the conative prefix was inserted in verbs that were lex- ically specified to be of this type.

3.2.6. Object Experiencer Verbs

The left-edge position is also an inflection site used to generate what we termed object experiencer verbs– single-actant predicates whose object markers refer to the semantic experiencer of an event. In *tamíyiłił* 'he/she/ it is floating' in (4), for example, the person floating is expressed by the object marker *mi-*.

(4) tamíyiłił
ta–mi–í–yi–0–łił
on 3.DO CON 3.PROG VV float
'he/she/it is floating (in one spot)'

To implement object experiencer verbs, the FST must exclude any subject other than third person and specify the proximal third-person object for third person experiencer. (There can be no distal third person marking for these themes, as the morphological subject is suppressed.) We used the flag diacritics @U.VALENCE.DO-EXPERIENCER@ and @U.VALENCE.OO-EXPERIENCER@ for direct- and oblique-object experiencer verbs, respectively, to pre- vent any subject person-number combinations other than third-person singular (with zero-marked imperfective allomorph) from being realized.

3.2.7. Restricted Argument Structures

Some verbs. for pragmatic or other reasons, had restrictions on non-third persons as either subjects or objects. For example, impersonal verbs (typically referring to agentless events such as weather or celestial situation) such as *gudìsghál* 'it is getting dark (outside)' can only take a dummy singular third-person, and no distributive marker. The flag diacritic @U.VALENCE.IMPERSONAL@ was used for such verbs to prevent other person forms. Other verbs would be pragmatically odd with first or second persons, for instance, *taánimòsh* 'it is boiling' (barring a story with anthropomorphic water or kettles). The transitive verb *íyin* 'to sing (it)', can only take third-person objects. To prevent these, sequences of flag diacritics such as @R.OBJECTNUMBER.SG@@R.OBJECTPERSON.3@ were added to require third-person singular forms.

4. Size and Performance

The current extended lexicon contains altogether 1,557 lexemes, distributed among various argument structure types and subtypes as shown in Table 1. On average, these lexemes have 1.62 suppletive verb theme vari- ants, with a median of 1.0 and a maximum of 11 theme allomorphs per lexeme (resulting from multiple allo- morphic variants for the same aspectual value). This means that some suppletive forms are missing from the Onespot-Sapir glossary, and must be elicited from flu- ent speakers.

$n_{lexemes}$	$\bar{x}_{allomorphs}$	Argument structure
842	1.50	Intransitive
630	1.83	Transitive
39	1.15	Transitive-SubjSuppr[essed]
25	1.36	ObliqueObjectExperiencer
9	1.67	ObliqueObject
3	1.33	Transitive-Conative
2	2.00	Intransitive-SubjPl[ural]Only
2	1.00	Transitive-D[irect]Obj3SgOnly
2	1.00	DirectObjectExperiencer
1	4.00	Ditransitive
1	1.00	Intransitive-SubjSuppr[essed]
1	1.00	Intransitive-Subj3[rdPerson]Only
1557	1.62	TOTAL

Table 1: Counts of lemmas and average theme allo- morphs for different argument structure types

When compiled with Foma, this entire FST is quite large, at 63.2 MB in overall size, with 1,664,399 states, 4,143,483 arcs, and more than 9×10^{18} paths (before pruning based on flag diacritics). The number of verbal wordforms that this FST covers is nevertheless finite, adding up to 1,472,669 forms in total that take 15 minutes to output using the *pairs* command in Foma. In terms of speed, this expanded FST is noticeably slow in analyzing wordforms (2min 2.57s for 1000 random word-forms, on a 2020 Macbook Pro with 32GB of RAM and a 2.3 GHz Quad-Core Intel Cor i7 CPU) but alarmingly snail-paced in generating the same 1000 wordforms (43min 6.27s). This is caused by the original design of the FST, where the flag-diacritic that constrain the acceptable strings are largely specified at the right edge of the FST, thus resulting during the FST lookup, reading the network from left-to-right, in the generation of a huge number of possible strings before encountering the limiting flag diacritics. In linguistic analysis, though, the wordform string strongly restricts possible analyses, resulting in a more acceptable but still slow speed. A possible solution would be to specify constraining flag diacritics at the left edge of the FST, at the beginning of the lexical tier, which, though feasible, would involve reconfiguring the matching flag

5. Conclusion and next steps

We have demonstrated that a working full-scale finite state model can be created for Tsuut'ina, and thus Dene languages in general, implementing all argument type of verbs, the morphologically most complex word class with a comprehensive lexicon. This expansion has it performance challenges, i.e. its relatively large size and concerningly slow speed would render real-time paradigm generation by the model impractical. However, we have envisioned solutions to these challenges.

6. Acknowledgements

This work represents one part of a long-term collaboration with partners at Tsuut'ina Nation to develop language technology in support of local language education, documentation, and revitalization initiatives. We gratefully acknowledge the leadership, support, and linguistic expertise of Bruce Starlight (Tsuut'ina Language Commissioner) and Janelle Crane-Starlight (Di- rector, Tsuut'ina Gunaha Institute) and their offices throughout this project. This work has been funded by a Partnership Grant (895-2019-1012) from the Social Sciences and Humanities Research Council (SSHRC) of Canada.

7. Bibliographical References

Antonsen, L., Johnson, R., Trosterud, T., and Uibo, H. (2013). Generating modular grammar exercises with finite-state transducers. In *Proceedings of the second workshop on NLP for computer-assisted lan- guage learning at NODALIDA*, pages 27–38.

Arppe, A., Lachler, J., Trosterud, T., Antonsen, L., and Moshagen, S. N. M. (2016). Basic language re- source kits for endangered languages: A case study of plains cree. In *CCURL 2016: Collaboration and Computing for Under-Resourced Languages: To- wards an Alliance for Digital Language Diversity (LREC 2016 Workshop), Portoroz, Slovenia*. Euro- pean Language Resource Association.

Arppe, A., Cox, C., Hulden, M., Lachler, J., Mosha- gen, S. N., Silfverberg, M., and Trosterud, T. (2017). Computational modeling of the verb in Dene lan- guages. the case of Tsuut'ina. In *Working Papers in Athabascan Linguistics ("Red Book" series)*, Fair- banks. Alaska Native Language Center.

Beesley, K. R. and Karttunen, L. (2003). *Finite State Morphology*. CSLI Publications, Stanford, CA.

Bowers, D., Arppe, A., Lachler, J., Moshagen, S., and Trosterud, T. (2017). A morphological parser for Odawa. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of En- dangered Languages*, pages 1–9.

Chen, E. and Schwartz, L. (2018). A morphological analyzer for St. Lawrence Island / Central Siberian Yupik. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Lan- guage Resources Association (ELRA).

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

Harrigan, A. G., Schmirler, K., Arppe, A., Anton- sen, L., Trosterud, T., and Wolvengrey, A. (2017). Learning from the computational modelling of Plains Cree verbs. *Morphology*, 27(4):565–598.

Hoijer, H. (1945). The Apachean verb, part I: Verb structure and pronominal prefixes. *International Journal of American Linguistics*, 11(4):193–203.

Holden, J. (2013). *Benasní-I Remember: Dene Sųłiné Oral Histories with Morphological Analysis*. Brill.

Hulden, M. and Bischoff, S. (2008). An experiment in computational parsing of the Navajo verb. *Coyote Papers*, 16.

Hulden, M. (2009). Foma: a finite-state compiler and library. In Proceedings of EACL, pages 29—32, Athens. Association for Computational Linguistics.

Hurskainen, A. (2009). Intelligent computer-assisted language learning: Implementation to Swahili. *Tech- nical reports on language technology*, 3.

Johnson, R., Antonsen, L., and Trosterud, T. (2013). Using finite state transducers for making efficient reading comprehension dictionaries. In *Proceedings of the 19th Nordic conference of computational linguistics* (NODALIDA 2013), pages 59–71.

Kari, J. (1979). Athabaskan Verb Theme Categories: Ahtna, volume 2 of Alaska Native Language Cen- ter Research Paper. Alaska Native Language Center, Fairbanks, Alaska.

Kazantseva, A., Maracle, O. B., and Pine, A. (2018). Kawennón:nis: the Wordmaker for Kanyen'kéha. In Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages, pages 53–64.

Kazeminejad, G., Cowell, A., and Hulden, M. (2017). Creating lexical resources for polysynthetic languages– the case of arapaho. pages 10–18, 01.

Lachler, J., Antonsen, L., Trosterud, T., Moshagen, S., and Arppe, A. (2018). Modeling Northern Haida verb morphology. In *Proceedings of the Eleventh In- ternational Conference on Language Resources and Evaluation (LREC 2018)*.

Leer, J. (1999). Comparative Athabaskan class notes. Unpublished manuscript CA965L1999a, Alaska Na- tive Language Archive.

Lovick, O., Cox, C., Silfverberg, M., Arppe, A., and Hulden, M. (2018). A computational architecture for the morphology of upper tanana. In *Proceedings of the Eleventh International Conference on Lan- guage Resources and Evaluation (LREC 2018)*.

McDonough, J. M. (2000). On a bipartite model of the Athabaskan verb. In Theodore B. Fernald et al., ed- itors, *The Athabaskan languages: Perspectives on a Native American language family*, number 24 in Ox- ford Studies in Anthropological Linguistics, pages 139–166. Oxford University Press, Oxford.

Moshagen, S., Pirinen, T. A., and Trosterud, T. (2013). Building an open-source development infrastructure for language technology projects. In *Proceedings of the 19th Nordic Conference of Computational Lin- guistics (NODALIDA 2013)*, pages 343–352.

Rice, S., Libben, G., and Derwing, B. (2002). Mor- phological representation in an endangered, polysyn- thetic language. *Brain and Language*, 81:473–486.

Rice, K. (2000). *Morpheme order and semantic scope*. Cambridge University Press, Cambridge.

Rice, K. (2001). Slave (Northern Athabaskan). In An- drew Spencer et al., editors, *The Handbook of Mor- phology*, pages 648–689. Blackwell, Malden, MA.

Sapir, E. and Hoijer, H. (1967). *The Phonology and Morphology of the Navaho Language*, volume 50 of *University of California Publications in Linguistics. University of California Press, Berkeley, CA / Los Angeles, CA*.

Snoek, C., Thunder, D., Lõo, K., Arppe, A., Lachler, J., Moshagen, S., and Trosterud, T. (2014). Mod- eling the noun morphology of Plains Cree. In *Pro- ceedings of the 2014 Workshop on the Use of Com- putational Methods in the Study of Endangered Lan- guages*, pages 34–42.

Starlight, B., Donovan, G., and Cox, C. (2016). From archival sources to revitalization resources: Revisit- ing the Tsuut'ina notebooks of John Onespot and Ed- ward Sapir. American Philosophical Society sympo- sium 'Translating Across Time and Space: Endan- gered Languages, Cultural Revitalization, and the Work of History', Philadelphia, PA, October 13–15, 2016.

Young, R. W. and Morgan, W. (1987). *The Navajo language*. University of New Mexico Press, Albu- querque, NM.