Modelling the Reduplicating Lushootseed Morphology with an FST and LSTM

Jack Rueter
University of Helsinki
first.last@helsinki.fi

Mika Hämäläinen
Metropolia University of Applied Sciences
first.last@metropolia.fi

Khalid Alnajjar
Rootroo Ltd
first@rootroo.com

Abstract

In this paper, we present an FST based approach for conducting morphological analysis, lemmatization and generation of Lushootseed words. Furthermore, we use the FST to generate training data for an LSTM based neural model and train this model to do morphological analysis. The neural model reaches a 71.9% accuracy on the test data. Furthermore, we discuss reduplication types in the Lushootseed language forms. The approach involves the use of both attested instances of reduplication and bare stems for applying a variety of reduplications to, as it is unclear just how much variation can be attributed to the individual speakers and authors of the source materials. That is, there may be areal factors that can be aligned with certain types of reduplication and their frequencies.

1 Introduction

A significant proportion of the world's languages face the threat of endangerment to varying degrees. This endangered status poses certain constraints on the extent to which modern NLP research can be conducted with such languages. This is due to the fact that many endangered languages lack extensive textual resources that are readily accessible online. Furthermore, even with available resources, there is concern about the quality of the data, as it may be influenced by various factors such as the author's level of fluency, accuracy of spelling, and inconsistencies in character encoding at the most basic level (see Hämäläinen 2021).

Reduplication appears in many languages of the world (Raimy, 2000). While full reduplication is observed as a repeated word form, partial reduplication is associated with extensive variety both regular and irregular. This paper focuses on a finite-state description of the partial reduplication patterns found in the Lushootseed language forms (lut) and (slh). The most predominant forms of reduplication in Lushootseed are distributive (Distr) and diminutive (Dim), which can, in fact, appear in tandem, but there are restrictions delimiting their use (see Broselow 1983, Bates 1986, Urbanczyk 1994). In addition to Distr and Dim, however, we also find a third and slightly less frequent random or out of control distributive (OC) (see Bates et al. 1994, Urbanczyk 1996).

The base of these three types of reduplication can be found in the initial two to three phonemes of the word root most often referred to with the notation C_1VC_2, but the authors of this paper will surround the vowel with parentheses to indicate the possibility of its absence: $C_1(V)C_2$ and thus accommodate the radical CC mentioned in (Beck 1999:24; Crowgey 2019: 39, 42).

The radical consist of simple and compound letters alike, e.g., q́ʷ, gʷ, ƛ’, all of which add to the issues of facilitating the extensive variation in Lushootseed reduplication. First, the concept of compound letters involved in regular reduplication segments is a very import part of finite-state description for Lushootseed. Although the 46 phonemes canonize the extensive alphabet, they create their own demands on the description.

Our facilitation of Lushootseed reduplication with a finite-state machine[1] is based on the use of a five-place holder segement concatenated directly before the radical. We number these right-to-left away from the radical {p5}{p4}{p3}{p2}{p1} where the odd-numbered place holders represent consonants, and the even-numbered ones vowels. The system is set up so that the place holders {p3}{p2}{p1} are used with Distr, Dim and OC reduplication, whereas the more remote place holders {p5}{p4} are used to deal with Distr + Dim combinations. Albeit, theory sees the distributive losing the third phoneme due to a principle of antigemination (see Broselow 1983: 326–329, and Urbanczyk 1994: 515) referencing also (Hess

[1] Our code is published in https://github.com/giellalt/lang-lut

Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP), pages 40–46
July 14, 2023 ©2023 Association for Computational Linguistics

1967: 7) and (Snyder 1968: 22). We have assumed the absence of geminates and therefore have left them out of the equation. Perhaps, further studies will require their addition to our finite-state description of reduplication in permeating the Lushootseed vocabulary.

2 Related work

Several different methods are currently in use to model morphology of endangered languages computationally. In this section, we will covers some of the existing rule-based, statistical and neural approaches. Our method embraces the rule-based tradition because machine-learning based methods rely on a lot of annoated data we currently do not have for Lushootseed.

In the rule-based research, morphology has mainly been modelled using a finite-state transducer (FST) using one of several technologies such as HFST (Lindén et al., 2013), OpenFST (Allauzen et al., 2007) or Foma (Hulden, 2009). Such an approach has been successful in describing languages of a variety of different morphological groups such as polysynthetic languages (e.g. Plains Cree (Snoek et al., 2014), East Cree (Arppe et al., 2017) and Odawa (Bowers et al., 2017)), agglutinative languages (e.g. Komi-Zyrian (Rueter et al., 2021), San Mateo Huave (Tyers and Castro, 2023), Skolt Sami (Rueter and Hämäläinen, 2020), Sakha (Ivanova et al., 2022) and Erzya (Rueter et al., 2020)) and fusional languages (e.g. Akkadian (Sahala et al., 2020) and Arabic (Shaalan and Attia, 2012)).

For statistical approaches, Tang (2006) has done research on English morphology by an approach that comprises two interrelated components, which are morphological rule learning and morphological analysis. The morphological rules are acquired by means of statistical learning from a list of words. On another line of work, Kumar et al. (2009) has developed a machine learning technique that utilizes sequence labeling and kernel methods for training, which enables the model to effectively capture the non-linear associations between various aspects of the morphological features found in Tamil.

With the emergence of UniMorph (McCarthy et al., 2020), which continues to include only partial morphological descriptions of each language, a great deal of neural based research has emerged to conduct morphological analysis. The typical models that are used are LSTM (Matteson et al., 2018; Akyürek et al., 2019) and Transformer (see Kodner et al. 2022) based models.

3 Materials and methods

The materials used for this paper come from the Lushootseed dictionary of Bates et al., 1994 and language learning binders by Zalmai Zahir and Peggy kʷiʔalq Ahvakana (Book 1 dᶻixʷ First, Book 2 dəgʷi You, Book 3 s.ʔəɬəd Food, Book 4 ʔalʔal House) as well as a binder of transcriptions to recordings from the University of Washington archives received in 2003 on the Muckleshoot Reservation.

The method involves a mnemonic descriptive approach, implemented for a decidely deterministic machine and human-friendly solution – if there is such a thing. To this end, we adhere to a three-phoneme segment approach to Lushootseed description and simply start with the labeling 123. Here ‹1› indicates the first consonant of the radical (root), ‹2› the vowel (which seems to be absent/latent in at least a few roots), and ‹3› the second consonant. We then introduce a series of five ordered place holders to precede the root.

The insertion of place holders is convenient in this finite-state description if they come before the root. Although there are numerous segments of regular morphology, inserting a series of five place holders immediately before the root can be seen as just another step in regular concatenation. Here it might be mentioned that theoretic distinctions between inflection and clitics do not come before consideration for orthographic practices (cf. Beck 2018).

The five place holder, numbering away from the first three letters of the root is set so the odd numbers correlate with the consonants and the even numbers with the vowels. Thus, {p3} correlates with *kʷ*, {p2} with *a*, and {p1} with *t*

{p5} {p4} {p3} {p2} {p1} kʷ a t

kʷatač: kʷatač ‘climb’

s‹kʷatač: skʷatač ‘mountain’

s ‹ {p5}:0 {p4}:0 {p3}:kʷ {p2}:a {p1}:0 kʷ a t a č :

skʷakʷətač ‘mountains’

s ‹ {p5}:0 {p4}:0 {p3}:kʷ {p2}:a {p1}:0 kʷ a:0 t a č :

skʷakʷtač ‘hill’

With this as a point of departure, we can then enumerate four predominant tendencies, one – total reduplication, one – partial to the left, two partial to the right. First, total reduplication is 123123, which is extremely regular and typically distributive in meaning. Second, comes the diminutive with extensive variation: 1213, 12123, 1i13, 1i123, 1iq13. Third, and less frequent in the materials are 123

4 FST models

The finite-state description of Lushootseed involves several layers of experience. It addresses issues involving orthography, morphophonology, concatenation and symmetric tagging for subsequent machine readability. The orthography, which is canonized by the language's reduplication patterns, uses lower-case letters with multiple diacritics, as no precomposed letters are available for nearly half of the alphabet. The concatenative morphology, which with the exception of the possessive person marking strategy, is symmetric but involves abbreviated or short-hand forms for some consecutive morphemes. The variation in multiple reduplication patterns appears to be partially monolectic or geographic in nature, but there is definitely also breathing room for variation in where individual derivations are used. In general, both preposed and postposed affixing is present, and, in particular, there is asymmetry in the possessive person marking strategy. For language-independent comparison, we use flag diacritics in our models, which allows us supersegmental concatenation and facilitates regular tagging practices for use in downstream language technology, even work with Python libraries.

4.1 Orthography

Although there are established keyboard layouts provided on official language-community sites [2], there are other keyboards, which may include nonstandard diactritic and letter combinations, that are visibly present on the net and in easily accessible language materials. This has meant the establishment of *spellrelax* files to allow for recognizing non-word internal single right quotation mark, instead of a combining comma above diacritic, for example, or even small letter L with middle tilde ‹U+026B› in place of small letter L with belt ‹U+026C›.

4.2 Concatenation and Tagging

Reduplication has been dealt with as a problematic feature in earlier descriptions of the languages where it is regarded as nonlinear (see Urbanczyk 1996). Our solution has been to introduce a segment of five place holders that facilitate copying values directly to predefined positions. As our concatenation in compilation reads right-to-left, memory retention is minimized to the three phonemes before the place holder series {p5}{p4}{p3}{p2}{p1}. If these place holders are to be used, the machine has already seen the reduplication trigger, which appears left of the word stem.

The relatively mnemonic triggers have been named according to relative position in the radical model C_1VC_2, i.e., 123. Thus, the distributive reduplication $C_1VC_2C_1VC_2$ is labeled distr_trigger_123123. Analogically, the diminutive reduplications $C_1VC_1C_2$, $C_1iC_1C_2$, $C_1iʔC_1C_2$, $C_1iʔC_1VC_2$ are represented by the triggers dim_trigger1213, dim_trigger1i13, dim_trigger1iq13, dim_trigger1iq123, respectively. OC reduplication (out of control, random) in $C_1VC_2VC_2$ is represented by OC_12323.

The reduplication *g*^{w}*aadg*^{w}*ad* in *lə =bə =lə cu–g*^{w}*aadg*^{w}*ad* (source Beck 2018: example 13) 'talking' could be illustrated as $C_1VVC_2C_1VC_2$, i.e., trigger_122123. The underlying use of our placeholders, however, would show the following transformation

{p5}:1 {p4}:2 {p3}:0 {p2}:2 {p1}:3 1 2 3
{p5}:g^{w} {p4}:a {p3}:0 {p2}:a {p1}:d g^{w} a d

Reduplication triggers are accompanied by diacritic flags, which make it possible to position tags in the output. Flag diacritics are also used to address the symmetrical tagging of prefixes after the lemma, on the one hand, and to disallow simultaneous tagging for two possessive markers, on the other.

5 Current state

Presently the lexicon is extremely small. It contains 110 verbs and 283 nouns, which might explain the low coverage rate of 70%, i.e., 1822 unrecognized tokens out of a total of 6186 tokens in the test corpus.

The two-level model has 31 rules governing reduplication copying patterns in the place holders and vowel loss or permutation in the root. The vowel system has be complemented by vowels with acute and grave accents, which might be useful in pedagogical use of the language model, and in work with language variation across the continuum of the language community.

[2] https://tulaliplushootseed.com/software-and-fonts/

source	target
ł u l ə č ̓ i č ̓ č ̓ i č ̓ ə l p y a q i d	N Pl Nom
a d d ə x ʷ t u b u ʔ q ʷ ə x ʷ	N Sg Nom RemPst Ptc PxSg2 Clt
b ə a d d ə x ʷ t u b u ʔ b u ʔ q ʷ	N Pl Nom Anew RemPst Ptc PxSg2

Table 1: Examples of the training data

tag	Anew	Clt	Hab	Irr	Pl	Ptc	PxPl1	PxPl2	PxSP3	PxSg1	PxSg2	RemPst	Sg
precision	0.77	0.96	1.00	0.98	0.94	0.91	0.90	0.89	0.80	0.83	0.92	0.81	0.87
recall	0.97	0.77	0.89	0.97	0.95	0.89	0.79	0.55	0.61	0.90	0.91	0.99	0.82
F1-score	0.86	0.86	0.94	0.98	0.94	0.90	0.84	0.68	0.69	0.87	0.91	0.89	0.84

Table 2: Per tag results of the neural model

The lexc continuation lexica number at 135. These continuation lexica provide coverage for regular nominal and verbal inflection, which utilizes a mutual set of morphology controlled partially with flag diacritics.

6 Neural Extension

No matter how extensive an FST transducer is, it still cannot cover the entire lexicon of a language. For this reason, we also experiment with training neural models to do morphological analysis based on the FST transducer described in this paper. The goal is not to replace the FST we have described in this paper, but to develop a neural "fallback" model that can be used when a word is not covered by the FST.

We follow the approach suggested by Hämäläinen et al. (2021), we use the code that has been made available in UralicNLP (Hämäläinen, 2019). This approach consists of querying the FST transducer for all the possible morphological forms for a given lemma. For a given input, the FST will thus produce all possible inflections and their morphological readings.

We limit our data to nouns only, and we use a list of 214 Lushootseed nouns to generate all the possible morphological forms for. This way, we produce a dataset consisting of around 756,000 inflectional form-morphological reading tuplets. This means that we have an average of 3536 inflectional forms for each lemma. We split this data into 70% training, 15% validation and 15% testing. The test data has words that are completely unseen to the model in the training data. This means that in the testing, the model needs to analyze based on lemmas and word forms it has not seen before even in a partial paradigm.

For the model itself, we use a Python library called OpenNMT (Klein et al., 2017) and use it to train an LSTM based recurrent neural network architecture with the default settings of the library. The task is defined as a character-level neural machine translation problem where each word form are split into characters separated by a white-space in the target side and the morphological readings produced by the FST are split into separate morphological tokens. Examples of the training data can be seen in Table 1.

The overall **accuracy of the model is 71.9%**. This is measured by counting how many full morphological readings the model predicted correctly for each word form in the test corpus. The results per morphological tag can be seen in Table 2. These results exclude the *N* (noun) tag and *Nom* (nominative) tag because all morphological forms had those tags in the dataset.

7 Discussion and Conclusions

In order to further test the accuracy of our Lushootseed description, more test data and descriptions of regular inflection will be needed. The challenge is to continue with the outline given for an inflectional complex (see Lonsdale 2001) and define what can actually be described as regular.

More time will be required to model more recent reanalyses of the morphological complexes. This means we may need to establish whether a six-placeholder segment is required to aptly describe Lushootseed reduplication and put our description in line with a hypothesis of antigemination.

The idea of describing morphological complexes as series of aligned clitics is very interesting (see Beck 2018). This will actually provide fuel for future work with syntax, since most of the semantic information is already present in the word roots where the clitics conglomerate.

Limitations

The FST does not yet have an extensive coverage of the Lushootseed vocabulary, so it does not work on all domains of text. Also, writing an FST takes a lot of time and requires special knowledge of the language. The neural model is limited to nouns only, but it can work on out-of-vocabulary words unlike the FST, however, we have only tested its accuracy using the words that are known to the FST, which means that words that follow very different inflection patterns will, most likely, not be analyzed correctly. Furthermore, the neural model was not trained on derivational morphology, which means that word derivations might also result in erroneous predictions.

Ethics statement

When dealing with an endangered language it is important to make sure that the research also contributes to the language community. This is the reason why we open-source our FST and neural model. We also work on data that has been given to us by speakers of Lushootseed with the intention of us working on building morphological descriptons and tools for the language. This means that we are not conducting our research with no regard to the language community.

Acknowledgments

This research is supported by FIN-CLARIN and Academy of Finland (grant 345610 *Kielivarojen ja kieliteknologian tutkimusinfrastruktuuri*).

References

Ekin Akyürek, Erenay Dayanık, and Deniz Yuret. 2019. Morphological analysis using a sequence decoder. *Transactions of the Association for Computational Linguistics*, 7:567–579.

Cyril Allauzen, Michael Riley, Johan Schalkwyk, Wojciech Skut, and Mehryar Mohri. 2007. Openfst: A general and efficient weighted finite-state transducer library: (extended abstract of an invited talk). In *Implementation and Application of Automata: 12th International Conference, CIAA 2007, Praque, Czech Republic, July 16-18, 2007, Revised Selected Papers 12*, pages 11–23. Springer.

Antti Arppe, Marie-Odile Junker, and Delasie Torkornoo. 2017. Converting a comprehensive lexical database into a computational model: The case of East Cree verb inflection. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 52–56, Honolulu. Association for Computational Linguistics.

Dawn Bates. 1986. An analysis of lushootseed diminutive reduplication. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society (1986), pp. 1–13.*

Dawn Bates, Thom Hess, and Vi Hilbert. 1994. *Lushootseed Dictionary*. University of Washington Press. Seattle and London. Bates, Dawn (ed.).

D. Beck. 1999. Words and prosodic phrasing in lushootseed narrative. *In Hall, T. A. and Kleinhenz, U., editors, Studies on the Phonological Word, pages 23–46.*

David Beck. 2018. Aspectual affixation in lushootseed: A minor reanalysis. In *Wa7 xweysás i nqwal'utteníha i ucwalmícwa: He loves the people's languages. Essays in honour of Henry Davis*. UBC Occasional Papers in Linguistics.

Dustin Bowers, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2017. A morphological parser for odawa. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 1–9, Honolulu. Association for Computational Linguistics.

Ellen Broselow. 1983. Salish double reduplications: Subjacency in morphology. *Natural Language Linguistic Theory 1(3).*

Joshua Crowgey. 2019. *Braiding Language (by Computer): Lushootseed Grammar Engineering*. University of Washington. A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy,.

Mika Hämäläinen. 2019. Uralicnlp: An nlp library for uralic languages. *Journal of open source software.*

Mika Hämäläinen. 2021. Endangered languages are not low-resourced! *Multilingual Facilitation.*

Mika Hämäläinen, Niko Partanen, Jack Rueter, and Khalid Alnajjar. 2021. Neural morphology dataset and models for multiple languages, from the large to the endangered. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 166–177, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.

Thom. Hess. 1967. *Snohomish Grammatical Structure*. Unpublished Ph.D. dissertation. University of Washington.

Mans Hulden. 2009. Foma: a finite-state compiler and library. In *Proceedings of the Demonstrations Session at EACL 2009*, pages 29–32, Athens, Greece. Association for Computational Linguistics.

Sardana Ivanova, Jonathan Washington, and Francis Tyers. 2022. A free/open-source morphological analyser and generator for sakha. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5137–5142, Marseille, France. European Language Resources Association.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. Association for Computational Linguistics.

Jordan Kodner, Salam Khalifa, Khuyagbaatar Batsuren, Hossep Dolatian, Ryan Cotterell, Faruk Akkus, Antonios Anastasopoulos, Taras Andrushko, Aryaman Arora, Nona Atanalov, Gábor Bella, Elena Budianskaya, Yustinus Ghanggo Ate, Omer Goldman, David Guriel, Simon Guriel, Silvia Guriel-Agiashvili, Witold Kieraś, Andrew Krizhanovsky, Natalia Krizhanovsky, Igor Marchenko, Magdalena Markowska, Polina Mashkovtseva, Maria Nepomniashchaya, Daria Rodionova, Karina Scheifer, Alexandra Sorova, Anastasia Yemelina, Jeremiah Young, and Ekaterina Vylomova. 2022. SIGMORPHON–UniMorph 2022 shared task 0: Generalization and typologically diverse morphological inflection. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 176–203, Seattle, Washington. Association for Computational Linguistics.

Arun Kumar, V Dhanalakshmi, RU Rekha, KP Soman, S Rajendran, et al. 2009. Morphological analyzer for agglutinative languages using machine learning approaches. In *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, pages 433–435. IEEE.

Krister Lindén, Erik Axelson, Senka Drobac, Sam Hardwick, Juha Kuokkala, Jyrki Niemi, Tommi A Pirinen, and Miikka Silfverberg. 2013. Hfst—a system for creating nlp tools. In *Systems and Frameworks for Computational Morphology: Third International Workshop, SFCM 2013, Berlin, Germany, September 6, 2013 Proceedings 3*, pages 53–71. Springer.

Deryle. Lonsdale. 2001. A two-level implementation for lushootseed morphology. *Papers for ICSNL 36 (Bar-el, L., L. Watt, and I. Wilson, eds.). UBCWPL 6:203– 214.*

Andrew Matteson, Chanhee Lee, Youngbum Kim, and Heuiseok Lim. 2018. Rich character-level information for Korean morphological analysis and part-of-speech tagging. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2482–2492, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Arya D. McCarthy, Christo Kirov, Matteo Grella, Amrit Nidhi, Patrick Xia, Kyle Gorman, Ekaterina Vylomova, Sabrina J. Mielke, Garrett Nicolai, Miikka Silfverberg, Timofey Arkhangelskiy, Nataly Krizhanovsky, Andrew Krizhanovsky, Elena Klyachko, Alexey Sorokin, John Mansfield, Valts Ernštreits, Yuval Pinter, Cassandra L. Jacobs, Ryan Cotterell, Mans Hulden, and David Yarowsky. 2020. UniMorph 3.0: Universal Morphology. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3922–3931, Marseille, France. European Language Resources Association.

Eric Raimy. 2000. *The phonology and morphology of reduplication*. de Gruyter.

Jack Rueter and Mika Hämäläinen. 2020. FST morphology for the endangered Skolt Sami language. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 250–257, Marseille, France. European Language Resources association.

Jack Rueter, Mika Hämäläinen, and Niko Partanen. 2020. Open-source morphology for endangered mordvinic languages. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 94–100, Online. Association for Computational Linguistics.

Jack Rueter, Niko Partanen, Mika Hämäläinen, and Trond Trosterud. 2021. Overview of open-source morphology development for the Komi-Zyrian language: Past and future. In *Proceedings of the Seventh International Workshop on Computational Linguistics of Uralic Languages*, pages 29–39, Syktyvkar, Russia (Online). Association for Computational Linguistics.

Aleksi Sahala, Miikka Silfverberg, Antti Arppe, and Krister Lindén. 2020. BabyFST - towards a finite-state based computational model of ancient babylonian. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3886–3894, Marseille, France. European Language Resources Association.

Khaled Shaalan and Mohammed Attia. 2012. Handling unknown words in Arabic FST morphology. In *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, pages 20–24, Donostia–San Sebastián. Association for Computational Linguistics.

Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2014. Modeling the noun morphology of Plains Cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 34–42, Baltimore, Maryland, USA. Association for Computational Linguistics.

Warren Snyder. 1968. *Southern Puget Sound Salish Texts, Place Names, and Dictionary*, volume 9. Sacramento; Sacramento Anthropological Society.

Xuri Tang. 2006. English morphological analysis with machine-learned rules. In *Proceedings of the 20th Pacific Asia Conference on Language, Information and*

Computation, pages 35–41, Huazhong Normal University, Wuhan, China. Tsinghua University Press.

Francis M. Tyers and Samuel Herrera Castro. 2023. Towards a finite-state morphological analyser for san mateo huave. In *Proceedings of the Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 30–37, Remote. Association for Computational Linguistics.

Suzanne Urbanczyk. 1994. Double reduplication in parallel. *Proceedings of the June 1994 Prosodic Morphology Workshop. Utrecht.*

Suzanne Urbanczyk. 1996. Morphological templates in reduplication. *University of Massachusetts/University of British Columbia.*