Program of Work: Sources of Schwa and Syllabification in Salish

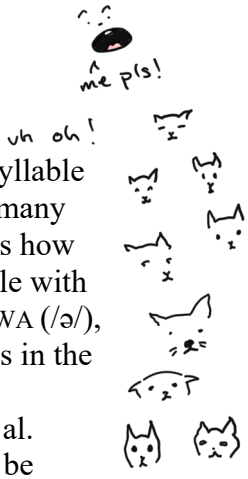
Proposed Research: My program of work explores SYLLABLE STRUCTURE across the Salish language family. Salish languages are traditionally spoken in the Pacific Northwest region of North America: at present, only half a dozen out of the original 23 still have first language speakers, but language revival efforts are under way in many Salish communities. My proposed project builds on my dissertation, entitled *The Grammar of Salish Reduplication*, but expands it to explore more general questions concerning variation in syllabification across the Salish languages.

Background: One of the primary goals in phonological theory, the subfield of linguistics dedicated to the structure and combination of sounds, is to develop general theories concerning possible sound systems in human language (as opposed to simply describing the sounds of individual languages). Investigating typologically diverse languages forms a key part of this agenda, since it enables us to establish the range and limits of allowable variation.

Cross-Linguistic Syllable Structure: Sounds within words are organized into groups called SYLLABLES. Syllable structure varies cross-linguistically. For example, sounds in the word /k.kst.tš.štt/ in Tashlhiyt Berber are organized into four syllables, which are indicated with a period (Dell & Elmedlaoui 2012:74): in other words, Tashlhiyt Berber allows syllables (and words) without vowels. However, sequences of consonants such as this could not possibly constitute words in English, where syllables are canonically organized around a vowel (Aslam & Kak 2007): for example, there are three vowel sounds corresponding to three syllables in the English word *magazine* /mæ.gə.zin/.

In spite of this variation, syllables are subject to universal principles of organization, such as the SONORITY SEQUENCING PRINCIPLE, which requires the most sonorous ("noisy") element in a syllable to be its peak, or NUCLEUS (see e.g., Parker 2011). Sounds are classified with respect to sonority: for example, the vowel /æ/ in *cat* (/kæt/) is more sonorous than the nasal consonant /m/ in mid (/mɪd/), and both vowels and nasal consonants are more sonorous than obstruent consonants, such as the /b/ in *bat* (/bæt/). English requires a sonorous nucleus, which is typically a vowel, but may also be a nasal consonant (e.g., the /n/ in *mitten* /mɪ.tn̩/). The requirement for a high sonority nucleus is not universal, however; speakers of Tashlhiyt Berber can build syllables with a low sonority consonant as the nucleus (Dell & Elmedlaoui 1985), as long as the surrounding consonants are not more sonorant. A theory of syllabification based on English alone would not be able to account for the Tashlhiyt Berber system: more generally, phonological theory must account for all patterns attested in human language without predicting the existence of impossible patterns (e.g., a language where the least sonorous sound in a syllable is the nucleus, allowing /b/ to be a nucleus in /mbm/, for example).

Salish languages are known for their unusual syllable structure (Bates & Carlson 1992): in particular, the existence of "vowelless" words (and syllables) in Nuxalk (Bella Coola) has attracted much theoretical attention (see Hockett 1955, Greenberg 1962; Cook 1994; Raimy & Idsardi 1997; Bagemihl 1998). Just as in Tashlhiyt Berber, syllables in Nuxalk allow for a low sonority nucleus, such as /ɬ/ in /p'ɬt/ 'warm' (Nater 1990:90; see also Mellesmoen 2021). However, most of the work on syllabification in Salish has focused on the most "exotic" cases, at the expense of a more balanced cross-linguistic picture. For example, unlike in Nuxalk, syllables in ʔayʔaǰuθəm (Comox-Sliammon) typically have a vocalic nucleus and do not permit long sequences of consonants: a vowel must be inserted to provide a nucleus and break up illegitimate sequences of consonants (Blake 2000). Sometimes, in fact, English consonant sequences must be broken up to conform to ʔayʔaǰuθəm syllable structure: the word /pəl.məs/, for example, is borrowed from the English word *plums* (Blake 2000:144). Just as a theory of syllabification based only on Tashlhiyt Berber would not account for English syllable structure constraints, a theory of syllabification based only on Nuxalk would not capture the observed variation in the Salish language family, because Nuxalk represents the extreme point on a continuum (see Kinkade & Czaykowska-Higgins 1998). The variety of rules and patterns in syllabification across Salish languages makes them an ideal focus of comparative research.



Schwa and Syllable Structure: Studying the distribution of vowels is a useful way to elucidate syllable structure: it provides insight into which sounds can be the nucleus of a syllable and which/how many sounds can be combined into a single syllable. The word /pəl.məs/ 'plums' in ʔayʔaǰuθəm shows how vowels must sometimes be inserted when a word is borrowed from English to make it compatible with the syllable structure of the language. Cross-linguistically, the inserted vowel is typically a SCHWA (/ə/), as in /pəl.məs/ (de Lacy 2006): this is the familiar "uh" sound represented by the last two vowels in the word *Canada* (/ˈkæ.nə.də/).

Previous work has established that schwa can be the nucleus of a Salish syllable (Shaw et al. 1999), allowing it to occur in the same peak position as any other vowel. However, schwa must be differentiated from other vowels in Salish due to its unique behaviour: it is inserted in words to provide a nucleus more often than other vowels (see e.g., Matthewson 2000), and resists changes which would result in increased prominence, size, or length (see e.g., Dyck 2004; Mellesmoen 2022). The distinction between schwa and other vowels is often described as a difference in whether the vowel is MORAIC (see e.g., Kinkade 1998, Shaw et al. 1999, Blake 2000). A mora is an abstract timing unit that most frequently corresponds to length: for example, a short vowel is typically a single mora in length, while a long vowel corresponds to two moras. Schwa in Salish is typically considered non-moraic, which means it does not have inherent length. While it can bear stress, in order to do so it must be supplemented by a coda (one or more consonants following a nucleus): the /l/ in /ḱə́l.lɪθ/ 'crooked' is doubled in ʔayʔaǰuθəm to ensure the syllable is "long enough" to be a well-formed syllable, for example (Blake 2000:127). (Compare /wá.waʔ.lɪθ/ 'baby frog' with a stressed moraic /a/: the /w/ need not be doubled.)

The hypothesis guiding this research is that understanding the distribution of non-moraic /ə/ in Salish is key to understanding cross-linguistic variation in syllable structure. However, not all cases of schwa have the same status. For example, Kinkade (1998) details four sources of /ə/ in Salish: (i) obligatory /ə/, which provides a nucleus where required (EPENTHETIC); (ii) /ə/ which is inherently part of a word (LEXICAL), (iii) /ə/ which is the reduced version of another vowel (DERIVED); and (iv) optional schwa which makes pronunciation easier (EXCRESCENT). The /ə/ in /pəl.məs/ 'plums' in Comox-Sliammon is an example of epenthesis, where the schwa provides a non-moraic nucleus to ensure the word is comprised of well-formed syllables. A lexical schwa similarly provides a non-moraic nucleus, but is part of the word itself and its position is not predictable on the basis of syllable structure alone. The shift from /eɪ/ in *Canadian* (/kə.ˈn**eɪ**.di.ən/) to /ə/ in *Canada* (/ˈkæ.n**ə**.də/) is an example of DERIVED /ə/, which is common in English as well as Salish: a derived /ə/ comes from a moraic vowel (e.g., /eɪ/), which loses a mora and reduces to /ə/ (Hammond 1997). Finally, an excrescent schwa is non-moraic and *not* a nucleus; for example, the word [čətˢtas] is pronounced with an excrescent schwa (written as ˢ), but there are only two syllables (/čət.tas/), not three (*/čə.tˢ.tas/), because the excrescent schwa does not provide a nucleus (Blake 2000:182).

While the distribution of the four types of /ə/ across the Salish language family has not been systematically investigated, there are already known differences between the languages. For example, there is no evidence for an underlying schwa in ʔayʔaǰuθəm (Blake 2000), but there are some words in St'át'imcets (Lillooet) where an underlying /ə/ may be required (see Caldecott 2009). Epenthetic, derived, and lexical /ə/ all appear as the nucleus of a syllable, but it is not clear how, or if, these vowels differ from each other phonologically or phonetically. While there are differences in place of articulation between epenthetic and lexical /ə/ for speakers of English (Davidson & Sloane 2003), there is no suspected difference between epenthetic and lexical /ə/ in other languages, such as Mohawk (Michelson 1989; Hall 2011).

A systematic comparison of differences between these types of schwa, including their distribution, is needed for a fuller picture of how /ə/ differs within and across Salish languages: this is one of the main goals of the proposed project. Some of the questions this research will answer include: How are non-/ə/ vowels and lexical /ə/ distinguished (prior to syllabification)? Do derived schwas come from moraic vowels that lose a mora, or from vowels that are deleted and replaced with an epenthetic /ə/ (i.e., is

derived /ə/ different from epenthetic /ə/)? Questions about the different types of /ə/ also tie into bigger questions in phonological theory regarding contrastive syllabification: i.e., are words stored in the lexicon with some amount of syllable structure) and if so, what types of information may be stored in the lexicon? The answer to this question has the potential to provide insight into theoretical debates about whether moras should be stored together with words (de Lacy 2020).

More broadly, differences in the distribution of vowels, especially /ə/, in Salish languages provide an ideal set of test conditions for a systematic cross-linguistic examination of syllabification. The fact that Salish languages are closely related means it is possible to ask detailed ("micro-parametric") questions about the role of particular sounds (and their sources), with fewer of the confounding variables that affect comparison between unrelated languages.

Methodology: This project has four stages: (i) surveying syllable types across the Salish language family, (ii) compiling possible phonological theories regarding syllable structure, (iii) comparing the predictions of the theories against the inventories of syllables structures within and across individual Salish languages, and (iv) developing or modifying a theory for the analysis of Salish syllables.

First Stage - Creation of a Syllable Corpus (6 months): The first stage will draw on archival resources and language documentation for all 23 Salish languages gathered throughout my dissertation research. During the first six months of my award, I will compile the relevant linguistic examples and language information into a written corpus. This will involve the selection of relevant demonstrative examples, standardization of transcription, and coding the data on the basis of key classifications (e.g., type of nucleus). The corpus will be a searchable database that can be shared with fieldworkers or researchers.

Second Stage and Third Stage – Comparing Predictions of Theory to Corpus (8 months): The second and third stage will query the phonological literature on syllable structure to identify a set of predictions which distinguish possible approaches. These predictions will be evaluated using to the corpus compiled during the first stage, and supplemented by acoustic recordings collected during my dissertation fieldwork, which includes ʔayʔaǰuθəm, Secwepemctsín (Shuswap), and St'at'imcets. This provides an opportunity to add to the existing description of these languages, which are all endangered, because acoustic recordings provide valuable information that is not reliably captured in written documentation, such as the specific acoustic correlates associated with stress or how a vowel is pronounced. Some predictions of theory, such as whether derived and epenthetic schwa are the same, can be explored more thoroughly with acoustic measurements and statistical comparisons.

While my methodology does not necessarily require additional fieldwork, I retain the ability to elicit new data on ʔayʔaǰuθəm and Secwepemctsín as a continuing member of research labs led by faculty at the University of British Columbia and the University of Alberta who hold research agreements with relevant First Nations communities. I also have well-established connections with fieldworkers collaborating with other Salish speech communities in British Columbia, who can supply linguistic examples where gaps in documentation are identified. If fieldwork data is used, it will be collected and stored in accordance with the research agreements with each community, tri-council policy (TCPS 2), and the University of Tromsø.

Fourth Stage – Analysis and Modification of Theory (10 months): The final stage of my postdoctoral research will evaluate the findings of the previous stages and determine which theoretical approach, if any, is most appropriate for the analysis of Salish syllable structure. I will either outline and implement any modifications required to the theoretical approach, or develop and propose an alternative approach. Finally, I will test the cross-linguistic validity of the theoretical approach by considering a set of cross-linguistic data from other languages, such as Tashlhiyt Berber. As part of the fourth stage, I will also focus on dissemination of knowledge by participating in relevant conferences or workshops.

Contributions: Description and analysis of endangered languages, including Salish languages, is time-sensitive and of the utmost importance. The languages which we know the least about often have the

patterns which most challenge our existing linguistic theories. There are only six Salish languages where linguistic fieldwork is possible. For the rest of the languages, the more time that passes, the greater the distance grows between the theoretical frameworks originally used to collect and analyse data and more modern approaches to phonological theory, leading to difficulties in interpretation and the possibility that important linguistic generalizations will be missed.

A consistent, explicit approach to phonological representations can allow systematic cross-linguistic comparison between languages. My knowledge of the Salish language family, coupled with a background in phonetics and phonology, will allow me to create a detailed and accurate corpus of syllable types in Salish. Issues with transcription are particularly important, since previous phonological work on Salish has sometimes employed transcription systems which have led to misunderstandings about the data. This is a problem I have highlighted in my previous work on syllables in Nuxalk (Mellesmoen 2021), where original audio recordings reveal more instances of /ə/ than are present in written examples used in the phonological literature, since the latter are based on a transcription system where predictable /ə/ is omitted (Nater 1984, 1990).

A theoretically guided, systematic cross-linguistic examination of syllable structure in Salish can also serve as a research tool to guide elicitation; linguistic analysis highlights gaps in existing documentation or where high calibre acoustic recordings may be necessary to adequately describe the language. My research can inform researchers about what to look for and how to go about looking for it.

Finally, consistent phonological representations can guide decisions about orthographic principles and practices, which are directly applicable to speech communities. For example, if predictable instances of schwa are omitted from the orthography, as in Nuxalk, then this must be paired with explicit description when /ə/ is present. In other words, pedagogical material must be written with an understanding of which information is present in the writing system, and which must be inferred from knowledge of the phonological grammar. As phonology is underrepresented in existing pedagogical material for Salish languages, this is an area where my research can have a substantial impact.

Justification of Location: While the primary focus of this program of work focuses on languages spoken in Canada, the particular theoretical expertise best suited to my project is found at The University of Tromsø (UiT). My supervisor for this project is Professor Martin Krämer, who is a phonologist and an expert on vowel epenthesis and excrescence (e.g., Krämer in press), the underlying representations of vowels (Krämer 2012), and syllable structure and phonotactics (e.g., Krämer & Zec 2020). His work covers a range of typologically distinct languages, including both understudied languages (e.g., Cocopa and Jamul Tiipay) and more well-studied languages (e.g., Italian), which means he is well-versed in cross-linguistic patterns in syllabification, and therefore an ideal choice for supervising a project on Salish syllable structure.

UiT is one of the world's best locations for theoretical work in phonology because it is the location of CASTL (The Center for Advanced Study in Theoretical Linguistics), which is a leader in generative theories of linguistics. The institution is also home to AcqVA Aurora (UiT Aurora Center for Language Acquisition, Variation and Attrition), which studies questions relevant to language revitalization. These two research centers cover a range of theoretical and experimental approaches to linguistic research. Additionally, conducting my research at UiT will allow me to build my professional network and participate in the exchange of knowledge with those working with other indigenous and underrepresented languages. For example, the department also houses Giellatekno, which is a centre for Saami language technology.

The geographical proximity to other institutions which frequently host conferences and workshops related to phonological theory in Europe is an additional benefit of this location, as I will be able to participate more easily from Tromsø. UiT will provide an intellectually stimulating location to conduct post-doctoral research, which will in turn enable me to continue my work on Salish languages in Canada.