Contents

1   Methodology

This work has proceeded along two parallel tracks. The first has been gathering primary field work on the data and working with published corpora in the language to uncover grammatical facts. The second is the implementation of the analysis of these grammatical facts through a computational syntactic framework. I will address my methods for each part of this separately.

1.1   Gathering data in Nuuchahnulth

Before I began my project on serial verbs and the linker, I first had to learn Nuuchahnulth to become at least conversant in the language. I did this by reading the published literature (especially Sapir & Swadesh 1939), attending language learning classes in Port Alberni (many of them with my colleague, Amie DeJong), and direct study with Adam Werle, some of which was funded through summer Foreign Language Acquisition Scholarships (FLAS). The language lessons I participated in were taught by Adam Werle and often included elders and native speakers who would assist, correct, and aid in teaching. It was through this venue that I first met fluent Nuuchahnulth elders.

In the summer of 2016, Adam and I traveled to Hot Springs Cove and collected texts from some Hesquiaht elders. On request, that data is not presented in this dissertation, but some of that work has informed my analysis, which I have confirmed with other speakers.

1.1.1   Data sources

Before I collected my own data, I looked at data from a variety of sources to generate appropriate questions. My sources were previous syntactic work on the language, especially Jacobsen (1993), Nakayama (2001), Wojdak (2003), Waldie (2004), and Woo (2007). I also relied on published corpora, especially the Nootka Texts (Sapir 1924; Sapir & Swadesh 1939, 1955; Sapir et al. 2000, 2004, 2009).

In addition to these resources, I looked at community-produced texts such as "Son of Thunderbird" and texts I received from linguists Adam Werle and Henry Kammler. Notable among these were an in-progress Bible translation Adam Werle and Sophie Billy were working on and several recordings Henry Kammler made with the late Barbara Touchie. I looked through these sources for examples of the phenomena I was looking for, annotated and cataloged them, and used some of these examples as prompts for speakers.

1.1.2 Elicitation methods

I spent January, February, and part of March of 2018 in Port Alberni working with native speakers and gathering data specifically for this dissertation. In that period of time I worked with Julia Lucas (Nuuchahnulth name *tupaat*, Ahousaht tribe, central dialect), Bob Mundy (Uclueleht tribe, Barkley sound dialect), Marjorie Touchie (Uclueleht tribe, Barkley sound dialect), Fidelia Haiyupis (Ehattesaht tribe, northern dialect), and Sophie Billy (Checkleseht, Kyuquot-Checkleseht dialect). Some of my work with Fidelia Haiyupis and Sophie Billy was funded by the Ehattesaht tribe, which has received copies of my notes and recordings. The Uclueleht tribal office has also received the notes and recordings I made with Bob Mundy and Marjorie Touchie. I have also made recordings and transcriptions available online to language learners, some of which is restricted to people who have the right password to access the folder.

When working with speakers, I tended to work two to four hours at a time and tried to structure sections in three parts: grammatical questions and elicitations, vocabulary questions and clarification questions on existing texts, and text elicitation. The purpose of this was to avoid wearing speakers out with too many grammatical questions in a row, and to collect other important data. Nuuchahnulth is not a fully-documented language (if any such thing is possible), and there are many dialect differences in the lexicon that are undocumented and unknown. Although one of my speakers did not like giving lengthy texts, I was able to collect connected, fluent texts from other speakers, which is a lasting artifact and can be used to answer questions beyond the narrow scope of my dissertation. I have approximately six hours of fluent Nuuchahnulth from that period, about two-thirds of which is currently transcribed. I have since visited my consultants again and asked follow up questions as well as collected more texts.

My methods of elicitation were: linguist-constructed sentences with preceding context verified or corrected by my consultant, summarization of short stories, finishing an incomplete narrative, requests to rephrase previous speaker utterances, direct translation from English, and description of picture stories. Elicitation was done in Nuuchahnulth when possible, but was often done in English as well. Many of these elicitation methods focused on single actors performing multiple salient actions at once. An example is: "A wolf and a dog came from the forest. One approached me and one ran back. The dog approached me. What did the wolf do?" The expression 'into the forest' in Nuuchahnulth is the verb *hitaaqƛaʔiƛ* and the word for 'run' is *kamitquk*. The purpose was to get both in a serial verb construction together. Another example is, "We are going to go camping. I want the children to help their mother. I want them to pack. I want them to carry the luggage. What should I tell them?" The purpose of this was to get a command form, which is always marked with second position inflection, with a serialized verb construction where the verbs must necessarily share the command mood. The construction would minimally have two (at least pragmatically temporally) sequential verbs and perhaps the benefactive verb to express "for your mother." Linker constructions were much more difficult to elicit in a roundabout manner like this, and so I often ended up asking directly whether I could attach it to a word, and attempted rephrases of speaker sentences with an added linker morpheme. Sometimes these took the form of "Is there a way to say this with [word with a linker attached]?"

I also attempted to elicit stories that would have a greater likelihood of illustrating serial verb or linker phenomenon, which meant settings of simultaneous action. This was not as successful as I had hoped, and I found instead that the best way to gather example constructions was through elicitation methods and gathering as much fluent text as possible. Anecdotally, staying in Nuuchahnulth for longer periods of time seemed to help more than anything else.

1.1.3 Data Collation

I collated the examples of the grammatical phenomena I was interested in. These came from a select set of stories I had previously interlinearized, from a randomly-selected subset of Nootka Texts stories, from my elicitation sessions with consultants, and from my transcriptions of elicited texts. I entered these examples into a spreadsheet that was tagged with the phenomenon that the example illustrated, and used this to help me find patterns in the grammatical data. To port this data to an implemented grammar, I simply had to export it to a comma-separated value file format and run a script that would generate a format readable by the implemented grammar (see §1.2).

1.2 Implementation through the DELPH-IN framework

My grammatical analysis has been through the DELPH-IN[1] framework, which is a computationally-implemented formalism of the head-driven phrase structure grammar (HPSG, Pollard & Sag 1994) using Minimal Recursion Semantics (MRS, Copestake et al. 2005). The implementation is built on the Grammar Matrix (Bender & Good 2010). My method for developing this grammar was first to answer a questionnaire on the Grammar Matrix webpage, generating a baseline grammar. Then from this output of type description language (TDL) formatted grammatical rules, I could then develop more complex syntactic analyses. This process includes generating type hierarchies and lexical entries as well.

In an ideal situation, I would have a grammar for each dialect, generated from a metagrammar that holds information about which grammatical rules and lexical entries belong to each dialect. However, due to time and scope constraints of the project, I simply entered different lexical entries for different dialects into the same grammar. This means that my grammar will happily mismatch morphology from different dialects, which is an incorrect overgeneration. Also due to scope constraints, my grammar also does not include a morphophonological component. This means that the second line of the IGT is what this grammar works with. That is, a sentence like *ʔuumaċukʼʷaλaḥ quʔušin* 'I am going to talk about Raven' is represented in the grammar as "ʔu-L.maċuk =!aλ=(m)aˑḥ quʔušin." The reasons for this are the theoretical separation of morphophonology from morphosyntax and the focus of the DELPH-IN machinery, which is squarely on morphosyntax and compositional semantics. To generate the surface string from the representation above would require the coding of a morphophonological analyzer (ideally a finite state transducer) which can go from the surface string to the segmented line and vice versa.

Development was done against a test suite of example sentences. These included both grammatical and ungrammatical examples. For the baseline grammar, I used simple example sentences from stories or sessions with consultants. Many of the ungrammatical examples for basic clauses were only vetted by me as ungrammatical, but I have a high degree of certainty for their ungrammaticality. For the phenomena under investigation, I used only grammatical examples from my elicitation and corpora work, and ungrammatical examples form my elicitation sessions. These came from my collated data (§1.1.3), which was loaded into a `[incr tsdb()]` database (Oepen 2001). This test suite of sentences could be run against each version of the implemented grammar and checked for changes to the parse coverage. Beyond parsing/not-parsing, each example sentence was tested for semantic faithfulness. Semantic validation has to be done manually, but regression tests allowed for parsing to be compared with previous iterations of the grammar rather than independently reverified every time the grammar changed.

[[TODO: add section about CLIMB? Example TDL??]]

---

[1]http://www.delph-in.net

Although the DELPH-IN tool set is descriptive for the purposes of both parsing and generating sentences, I focused only on parsing. My grammar as it stands has some issues with generation. These are caused by insufficient semantic constraints introduced by some morphemes (which can thus be hypothesized endlessly by the generator), and particular difficulties introduced by inflectional second position elements, explored in depth in Bender (2010). Fixing these issues in generation will require further development of the grammar.

The result of the implemented grammar is a series of files that detail the grammatical rules, the lexicon, and rules for generation. The format for most of these files is TDL, which is a series of grammatical descriptions which are equivalent to HPSG attribute-value matrices. The regression tests in `[incr tsdb()]` are also outputted to readable databases which show the resulting coverage of the grammar run over test cases. All of these materials are available at [[TODO: github repo]].

References

Bender, E. & Good, J. (2010). A grand challenge for linguistics: Scaling up and integrating models. *White Paper contributed to NSF's SBE 2020 Initiative*.

Bender, E. M. (2010). Reweaving a grammar for wambaya. *Linguistic Issues in Language Technology*, 3(1).

Copestake, A., Flickinger, D., Pollard, C. J., & Sag, I. A. (2005). Minimal recursion semantics: an introduction. *Research on Language and Computation*, 3(4), 281–332.

Jacobsen, W. H. (1993). *ADVANCES IN ROLE AND REFERENCE GRAMMAR*, chapter Subordination and cosubordination in Nootka: Clause combining in a polysynthetic verb-initial language, (pp. 235–235). Amsterdam: John Benjamins Publishing.

Nakayama, T. (2001). *Nuuchahnulth (Nootka) Morphosyntax*. University of California Press.

Oepen, S. (2001). [incr tsdb()] — *Competence and Performance Laboratory. User Manual*. Technical report, Computational Linguistics, Saarland University, Saarbrücken, Germany. in preparation.

Pollard, C. & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Sapir, E. (1924). The rival whalers, a nitinat story (nootka text with translation and grammatical analysis). *International Journal of American Linguistics*, 3(1), 76–102.

Sapir, E., Klokeid, T., Arima, E., & Robinson, K. (2000). *The Whaling Indians: West Coast Legends and Stories: Tales of Extraordinary Experience*. Canadian Museum of Civilization.

Sapir, E. & Swadesh, M. (1939). *Nootka Texts: Tales and ethnological narratives, with grammatical notes and lexical materials*. Linguistic society of America, University of Pennsylvania.

Sapir, E. & Swadesh, M. (1955). *Native accounts of Nootka ethnography*. Indiana University, Research Center in Anthropology, Folklore, and Linguistics.

Sapir, E., Swadesh, M., George, H., Thomas, A., Williams, F., Fraser, K., Arima, E., Kammler, H., Klokeid, T., & Robinson, K., Eds. (2004). *The Whaling Indians: Legendary Hunters*. Canadian Museum of Civilization.

Sapir, E., Swadesh, M., George, H., Thomas, A., Williams, F., Fraser, K., Arima, E., Kammler, H., Klokeid, T., & Robinson, K., Eds. (2009). *The Whaling Indians: West Coast Legends and Stories.* Canadian Museum of Civilization.

Waldie, R. J. (2004). Nuu-chah-nulth denominal verbs. Master's thesis, University of Victoria.

Wojdak, R. (2003). Predicative lexical suffixes in nuu-chah-nulth. *University of British Columbia Working Papers in Linguistics*, 11, 275–289.

Woo, F. (2007). *Prepositional Predicates in Nuuchahnulth.* PhD thesis, University of California, Santa Cruz.