Purpose

This is a progress report on the Yup’ik Language Dictionary and Processing Software project.

Summary

I have been working closely with Marie Meade, Yup’ik professor with Alaska Native Studies, to understand the morphophonological processes of suffixation in the Central Yup’ik language. I have been referencing Jacobson (1984), Reed, Miyaoka, Jacobson, Afcan, and Krauss (1977), and Mather, Meade, and Miyaoka (2002) in my research as well. This has given me an understanding of Yup’ik suffixation grammar that can be clearly defined in a set of algorithms.

This is a report on progress made on Task 1 with an outline of work to be completed for Task 2. This project is currently ahead of schedule and I expect to meet all deadlines established by the timeline given in the proposal.

Introduction

On March 23, 2010, this undergraduate research project was accepted for funding by the Office of Undergraduate Research and Scholarship, University of Alaska Anchorage. This is the first step in a larger project is to develop tools for Alaska Natives to use technology, globalization, and the information age for the protection of their language. A presentation on this project will be given in April, 2011 at the Undergraduate Research and Discovery
Symposium. The final written report for this project will be submitted to the Office of Undergraduate Research and Scholarship by June 15, 2011.

This project may be defined into five research phases or tasks, outlined here.

Task 1: Define Yup'ik word-forming grammar.
Task 2: Develop and test word-forming and spell checking algorithms.
Task 3: Develop data structures to store Yup'ik morphemes.
Task 4: Develop spell checking and morpheme definition software.
Task 5: Project finalization.

Project Status

First I will present the completed work for Task 1. Then I will discuss future work to be done, Tasks 2-5.

Completed Work

Task 1: Define Yup'ik word-forming grammar.

The Central Yup’ik language is made up of morphemes that can be broken into four categories: bases, postbases, endings, and enclitics (Jacobson, 1984). The base forms the beginning of a word, and then many (or no) postbases may be added to expand on the base. A word is then finished using an appropriate ending.

Bases may be separated into nominals (nouns), verbs, and particles, with particles being non-inflecting (Mather et al., 2002, p. 23). This means nominal and verb bases may accept postbases and endings where particles may not. As a base accepts a postbase, they form an expanded base that must continue to be suffixed with either another postbase or and ending.
Postbases may be defined into four categories: nominal-elaborating, verbalizing, verb-elaborating, and nominalizing (Mather et al., 2002, p. 33-34). Nominal-elaborating postbases (N-N) add onto a nominal base, leaving it nominal. Verbalizing postbases (N-V) add onto a nominal base, changing it into a verb. Verb-elaborating postbases (V-V) add onto a verb, leaving it verbal while nominalizing postbases (V-N) add onto a verb base, changing it into a nominal.

Endings may be easily broken into two broad categories: nominal and verbal. Nominal endings will contain information on number of objects (single, dual, plural), party of the possessors (1\textsuperscript{st} person, 2\textsuperscript{nd} person, 3\textsuperscript{rd} person), and number of possessors (single, dual, plural). For example, -ka ending means “My one N.” Verbal endings may be transitive or intransitive, indicating party information and number. For example, +’(g)amken means “I V you (singular).”

Both nominal and verb endings follow suffixation rules consistent with those of postbases, so it is a simple matter to predict how they will be added. However, these endings contain additional information that will not be covered in this report: nominal endings provide case information while Verb endings convey different moods. These details are a part of Yup’ik sentence grammar and lay outside the scope of this project. For further information, see Part II of Survey of Yup’ik Grammar Revised (Mather et al., 2002) and Endings, Yup’ik Eskimo Dictionary (Jacobson, 1984).

When adding suffixes (postbases and endings) onto bases, both the class of the base and the type of suffix must be examined (Reed et al., 1977, p. 20). The base class may be determined
by examining the ending of the base. Below is a brief summary on how the base’s class may be found (Reed et al., 1977, p. 21).

- **Class I:** Noun and verb bases ending in a single prime vowel (a, i, u)
- **Class II:** Noun and verb bases ending in two prime vowels (aa, ai, au, ii, etc.)
- **Class III:** Noun and verb bases ending in e, not preceded by a t (ne, pe, qe, etc.)
- **Class IV:** Noun and verb bases ending in te
- **Class V:** Noun bases ending with an r, preceded by one or two prime vowels
- **Class VI:** Noun bases ending with er and verb bases ending with a consonant

Types of suffixes are shown using a series of markings and punctuations that define how they may be added. These processes involve adding, dropping, and changing sequences of letters dependent upon the class of the base it is added to. The following is an abbreviated list of symbols used in defining suffixation type. For further information, see *Yup’ik Eskimo Grammar* (Reed et al., 1977, p. 22-29).

- **(g/t), etc:** These markings describe letter insertion. Some letter insertions are predictable while others must be described on a case by case basis.
- **+ :** This symbol indicates an adding type that retains the final consonant in both class V and class VI bases.
- **- :** This indicates a minus type suffix that drops the final consonant in both class V and class VI bases.
- **÷ :** This indicates a half-retaining suffix type that drops the final consonants in some cases, but retains them in others in a predictable manner.
• @: This indicates that this suffix is te affecting. The te group in class IV bases may affect the suffix, may be affected itself, or may be dropped entirely. This behavior is complex, but predictable and is expected to pose the greatest challenge in this program.

• :: This marking indicates a velar dropping type suffix. When a velar is found next to this marking and is surrounded by single vowels, one of which must be prime, it is dropped and may replace the vowels based upon a table found in *Yup’ik Eskimo Grammar* (Reed et al., 1977, p. 26).

For example, when +’(g/t)u:nga (1st person indicative intransitive verb ending) is added to verb bases: nallu- (class I, to not know), naulluu- (class II, to be sick), nere- (class III, to eat), kipute- (class IV, to buy), and yurar- (class V, to dance traditionally); we get nallunga (I don’t know), naulluugua (I’m sick), nerua (I’m eating), kiputua (I’m buying), and yurartua (I’m dancing).

The final morpheme category to be discussed is the enclitic. Enclitics are added onto the ending of a word, separated by a hyphen, and often change the meaning of the entire sentence it’s found in. Due to the small number of enclitics and their simple usage grammar, it should be a simple matter to create a data structure predefined with these morphemes.

This is just a sampling of Yup’ik word-forming grammar that will be defined in this program. Additional rules and examples will be presented as needed (or upon request) in future reports.
Future Work

Task 2: Develop and test word-forming and spell checking algorithms. Summer, 2010

For this task, I plan to reference Yup’ik writing samples using proper spelling and grammar with Yup’ik language reference materials through summer, 2010. I will use the processes outlined in Task 1 to generate guesses on the morphemes making up each word. I will start at the beginning of a word and create a list of possible bases, then, starting with my longest base guess, I will begin searching through the postbases and endings to see if I can recreate the word.

Professor Marie Meade and Interim Director Nany Furlow have both offered assistance in finding/creating Yup’ik writing samples to be used in this phase. I also plan to use vocabulary words found in Yup’ik Eskimo Grammar (Reed et al. 1977). There are lots of resources available at the UAA/APU Consortium Library that may also be used.

Task 3: Develop data structures to store Yup’ik morphemes. Fall, 2010

This task will develop the data structures that will store Yup’ik morphemes. Work on improving word-forming and spell checking algorithms will continue through this task.

Task 4: Develop spell checking and morpheme definition software. Spring, 2011

At this time, the word-forming and spell checking algorithms should be finalized. Work will begin on developing software that implements these algorithms and traverses the data structures to check spelling of Yup’ik words. Work will also begin on software that allows users to expand the digital dictionary, adding Yup’ik morphemes to the data structures.
**Task 5: Project finalization. Spring, 2011**

This task is intended to finalize software and data structures, develop demonstration programs, and compose documentation for presentations and reports.

**Conclusion**

I feel confident that Task 1 has been completed to satisfaction. Amendments may need to be made as I work through Task 2 this summer. I am currently ahead of schedule and feel confident that all tasks will be completed by the June 15, 2011 deadline. I have studied enough of Yup’ik grammar to understand the suffixation process and am looking forward to getting to work on developing solid algorithms to model this grammar in software.

The next progress report is scheduled for October, 2010 to report progress on Task 2. For further information, please contact me at esomervi@uaa.alaska.edu or visit www.camail-ellamyui.com – expected to be up by the end of May.

**References**

