Graduate Institute of Applied Linguistics

Thesis Approval Sheet

This thesis, entitled

An Automated Template Approach to Morphology and Syntax Description for Ayta
Abellen (using FLEx and PAWS)

written by

Roger Stone

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts

with major in

Applied Linguistics

has been read and approved

by the undersigned members of the faculty

of the Graduate Institute of Applied Linguistics.

_____

Dr. Gary Simons (Mentor)

_____

Dr. Paul Kroeger

_____

Dr. Michael Boutin

_____

Date signed

An Automated Template Approach to Morphology and Syntax Description for Ayta
Abellen (using FLEx and PAWS)

By

Roger Stone

Presented to the Faculty of
the Graduate Institute of Applied Linguistics
in partial fulfillment of the requirements
for the degree of

Master of Arts
with major in
Applied Linguistics

Graduate Institute of Applied Linguistics
June 2010

THESIS DUPLICATION RELEASE

I hereby authorize the Graduate Institute of Applied Linguistics Library to duplicate this thesis when needed for research and/or scholarship.

Agreed: _____
(student signature)

Refused: _____
(student signature)

ABSTRACT

An Automated Template Approach to Morphology and Syntax Description for Ayta
Abellen (using FLEx and PAWS)

Roger Stone
Master of Arts
with major in
Applied Linguistics

The Graduate Institute of Applied Linguistics, June 5, 2010

Supervising Professor: Dr. Gary Simons

This thesis can be called applied computational linguistics for the Ayta Abellen
language of the Philippines. FieldWorks Language Explorer (FLEx) was used to model
morphology while Parser and Writer of Syntax (PAWS) was used to model syntax. These
tools both generate computer readable files that can be used for automated parsing of
morphology and syntax. The tools also generate automated linguistic sketches in which
the data are presented as conventionally structured linguistic descriptions.

This thesis not only evaluates the effectiveness of these tools when used for
computational modeling of an Austronesian language but also considers the effectiveness
of the resulting template-based automated descriptions for morphology and syntax. The
morphological parser reached a 99.8% parsing rate while the syntactical parser was able
to parse 75-80% of natural language sentences. The automated descriptions of
morphology and syntax needed further editing and refinement but were helpful in
providing a good starting point for linguistic description.

ACKNOWLEDGEMENTS

Contents

Chapter 1

INTRODUCTION

The research described in this thesis tests the usefulness of current software tools for linguistic analysis, namely FieldWorks Language Explorer (FLEx) for morphological analysis and Parser and Writer for Syntax (PAWS) for syntactic analysis. Both tools use built-in templates to automatically generate editable linguistic descriptions. These built-in templates also generate formal rules that a parsing engine can use to test the adequacy and robustness of the linguistic description with empirical data. This research evaluates the effectiveness of the linguistic descriptions and parsers using data from Ayta Abellen, an Austronesian language of the Philippines.

Ayta Abellen is one of five Ayta negrito languages which belong to the Sambalic subgroup of the Austronesian language family. According to Stone (2009), the most closely related language to Ayta Abellen is Botolan Sambal. The speakers of Ayta Abellen are located in the mountainous western part of Tarlac province on Luzon; the language area also extends into Zambales province. Ayta Abellen is an endangered language. While the Philippine Commission on Indigenous Peoples reports the size of the ethnic group as 29,963,[1]

---

[1] Available online at: http://www.ncip.gov.ph/ethno_region.php.

due to intermarriage with lowland Filipinos, the number of fluent speakers of the language is much lower with estimates ranging from 3,500 in Stone (2009) to 6,000 in Nitsch (2009).

Little research on the Ayta Abellen language has been published although (Nitsch 2009, 1998a, 1998b) has produced drafts of phonology, morphophonemics, and grammar papers. Stone (2008) describes clause topics and Stone (2009) includes Ayta Abellen as one of eight languages in a comparative study of Sambalic languages.

These papers form only a small part of an "adequate" description of the language. The analyzed data that does exist is not in the form of traditional, published, peer-reviewed linguistic papers, making it inaccessible to the world of academic research. This raises the question of how to automate a process for structuring the data and how to make it accessible in order to document this lesser known language.

The process of documenting the lesser known languages of the world is dramatically different now than just thirty years earlier when computers were not commonplace. A field linguist in those days might have collected words on 3" x 5" cards and developed their own organizational strategies for data management. Analyzing the data involved flipping through cards and developing theories about how the morphology and syntax of the language worked. Organizing the data meant gathering observations about the language in a notebook or typing a grammar paper on a typewriter.

Now, after thirty years of linguistic software development by SIL and others, it is possible to utilize powerful computational tools in doing the analysis. The FieldWorks

Language Explorer, or FLEx,[2] is an example of a program designed specifically for the average field linguist who unknowingly uses a "stealth-to-wealth" approach to linguistic discovery and description as described in Black and Simons (2006). As the linguist analyzes words in an interlinear text collection, information is collected and automatically stored regarding the morphology of the language.

The program then helps the linguist describe the morphological data in terms of inflection, derivation, and compounding using a built-in template to generate a draft morphology sketch for the language. With the data already organized, the linguist then can edit the draft morphology sketch to turn it into a publishable morphological description.

FLEx focuses on morphology. Another program called PAWS (Parser and Writer of Syntax)[3] deals with syntax. This program also utilizes a "stealth-to-wealth" strategy as it asks the linguist a series of questions about the language and elicits example data which is then organized and output as a draft syntax sketch. Here also the linguist can edit the draft to turn it into a publishable syntactic description of the language.

Both of these tools output structured linguistic information in an XML (Extensible Markup Language) format which allows the information to be re-used in a variety of ways. The data can be output in a morphology sketch, publishable dictionary, interlinearized texts, and so on. This is also significant because it enables linguistic formatting standards to be maintained from one linguistic paper to another and it also simplifies the formatting of data

---

[2]Available online at: http://www.sil.org/computing/fieldworks/flex/.
[3]Available online at: http://www.sil.org/computing/catalog/show_software.asp?id=85.

types specific to linguistics.[4] The XML format is also cross-platform and interoperable meaning that it can be processed on most hardware and a variety of software technologies. This is a significant improvement over a traditional method of writing a linguistic paper in the format of one word processor and only being able to share the information with another linguist who has the same word processor.

This thesis uses these tools for analysis and description of the Ayta Abellen language. I aim to show not only how these tools can be used for computational modeling of an Austronesian language but also to consider the effectiveness of the resulting template-based automated descriptions for morphology and syntax.

---

[4]This is especially true when using the XLingPaper standard available online at: http://www.sil.org/~blacka/xlingpap/index.htm.

Chapter 2

PROBLEM

Sapir (1921) has said that "All grammars leak" meaning that as much as we want to believe that all published grammatical descriptions are watertight, there is always data that the linguist has not seen or has chosen to ignore (residue). The description may adequately explain what is happening in a carefully selected set of data or the description may adequately account for the majority of the data. But still there is residue that is not accessible to the reader. How can the reviewer of a grammatical description ever obtain access to the residue, or ever be certain that the description would actually work for a totally new text gathered from a native speaker, or even that it would even work for the whole data corpus the linguist was studying? A new kind of linguistic analysis and description is needed.

2.1 Morphology

The real test of a description of the morphology of a language is whether it can be used to parse any word in the language appearing in a natural text while at the same time rejecting ungrammatical wordforms. The ultimate goal is that the description could be empirically verified by implementing it in an automated parser that would correctly parse

any wordform occurring in natural text. FLEx enables such synchronization between description and automatic parsing as it uses the same underlying data to generate both a morphology sketch and an automated parser for testing the accuracy of the analysis on natural texts.

The goal of morphological research is not only to parse all the words of a language into their constituent morphemes, but also to account for the whole structure of the wordform. Ideally, Philippine language features like voice/focus and aspect could be linked to the morphemes that bear them and then percolate up through the parsing process to become a part of the description of the whole wordform. This information can then be used by syntactic parsers that would operate on the analyzed wordforms. This is especially important for Philippine languages where voice affixes relate to the semantic roles of the nouns in the sentence. A syntactic parser that will work on the output of the morphological parser would need this feature information also, not just the final syntactic category of the wordform.

FLEx is the only software package that integrates a morphological parser, lexicon, and analysis tools with the ability to automatically generate structured morphology sketches. Unfortunately, even though FLEx has been available for a few years, uptake has been slow among Philippine linguists. Of those that are using it for dictionary purposes, none have made a serious attempt at a computational description of the morphology. Because of the degree of similarity between affixation systems of Philippine languages, having an

empirically tested morphology model for one Philippine language should prove helpful for other Philippine language researchers and thus simplify their task.

FLEx automatically generates a sketch of the morphology based on the information in the database. This tool has the potential for aiding both linguists and linguistic consultants in identifying problems and inconsistencies in analysis as well as being a basis for a publishable morphological description. To date, however, the parsing tool has not yet been used for checking analyses and the morphology sketch has not yet been used for describing a Philippine language. Research identifying the strengths and weaknesses of using this method of analysis and description can help clarify its usefulness for other Austronesian languages as well.

2.2 Syntax

The real test of a syntax description is whether it can account for the syntax of any grammatically correct sentence in that language. When prose-only descriptions of the syntax of a language are written, it is hard to ascertain what proportion of the language is being accounted for. If, however, the syntax description can be formalized in a way that a computer can use it for parsing natural language, it may be possible to state what portion of the syntax is accounted for by the description.

The team at the Center for Language Technologies led by Dr. Rachel Roxas of De La Salle University in Manila is seeking to develop computational tools and ultimately

machine translation systems to and from Tagalog. Members of this team have said in Alcantara and Borra (2008a) that there has been "minimal work done on the development of a computational grammar for the Filipino language for the development of robust and industrial strength natural language analysis and technologies." Many approaches have been proposed and prototypes have been developed for automatic generation of formal grammars from various text collections, but as of this time, all of the final results have been labeled "insubstantial" by Borra.[1]

The lack of a computational grammar for Tagalog is a significant hindrance to NLP research in the Philippines. The same problem exists for other Philippine languages, especially the minority languages. But because of the similarities of syntax of many Philippine languages, it is hypothesized that a computational grammar for any Philippine language could contribute insights into the development of computational models for other Philippine languages. Thus, a computational grammar for any Philippine language could be used as a starting point for developing computational grammars for other Philippine languages. However, as of this time, no such published grammar exists in a current linguistic formalism.

The PAWS Starter Kit could be a significant step forward toward this goal of developing a computational grammar for Philippine languages. If the developer's claim is really true that 75% of the syntax of any language can be modeled in formal rules generated

---

[1]Personal communcation (Nov. 11, 2008).

by just answering the questions in the PAWS Starter Kit, then this would be a significant step forward for all those doing research in this area. The problem is that no one has attempted to use the system for any Philippine language. Determining the percentage of phenomena handled by the automatically generated rules and performing an analysis of what went wrong and right in the syntax parsing process would be most helpful not only to the developers of PAWS, but also to Philippine linguists considering whether to use it.

The syntax sketch that PAWS generates would seem to be a helpful tool for linguistic description, but it has not been used for a syntactic description of any Philippine language. This thesis provides an initial syntactic description of a Philippine language using the PAWS generated sketch. Discovering the areas where the linguist needs to adjust the description will be the most helpful in evaluating the usefulness of this tool.

Chapter 3

PURPOSE

The first purpose of this research is to develop a complete morphological parsing system for Ayta Abellen that can serve as a model for Philippine linguists wanting to use FLEx for analyzing affixation. Hopefully, this will encourage more linguists to use the powerful tools found in FLEx.

A second purpose is to evaluate the morphology sketch that FLEx outputs. By obtaining feedback from international linguistic consultants who have worked with Philippine languages, I was able to determine in what way this kind of sketch description will be helpful for linguists trying to organize their Philippine language data and also to make suggestions for its improvement.

A third purpose is to be the first test case of using the PAWS program with a Philippine language. Regular interaction with the developer of PAWS has facilitated the further development of this tool in relation to Philippine languages. The benefit, however, is not just for the developers of the tool, but ultimately for other Philippine linguists who will be able to determine if this would be a useful tool for syntactic description.

A fourth purpose is to evaluate the syntax sketch that PAWS outputs in terms of its descriptive adequacy for a Philippine language and to make suggestions for its improvement.

The fifth, and possibly the most important purpose of this research is to make available primary field data for the little known language of Ayta Abellen.

Chapter 4

LITERATURE REVIEW

This chapter reviews previous work done on parsing with a special focus on Philippine languages and parsing tools available for field linguists. Literature of a more general nature is included where it contributes in a theoretical or practical way to the software tools or methodologies used. Section 4.1 looks at morphological parsing, while section 4.2 looks at syntactic parsing.

4.1 Morphological parsing

Grimes (1983) developed the PARADIGM computer program which was a tool for hypothesizing affix position charts based on wordforms with affix breaks indicated. The program was developed for use on some of the earliest personal computers with the goal that it could be used by field linguists.

Koskenniemi (1983), in looking at the problem of the complex morphology of Finnish, developed a two-level model of morphological parsing where phonological rules are finite state transducers with an input stream and an output stream; these represent phonological processes operating on two levels, an underlying level and a surface level.

Software developers at SIL International implemented Koskenniemi's model on a PC and called it PC-KIMMO. While Antworth's description of the program (Antworth 1990) uses parsing examples from Tagalog, the parser was not simple enough for the average linguist to use and it has not been used with Philippine languages.

Another early PC morphology parsing system was AMPLE as described in Weber and others (1988). While originally developed for Quechua languages (Weber and Mann 1979), the project received input from Colombian and North American Indian languages and was adjusted to be a more generic morphology tool. The tool has been used in many parts of the world, but it is not totally reliable in determining word categories for languages having both prefixes and suffixes due to the difficulty in defining the order of application of affixes. A second problem with AMPLE was that when a linguist had worked through the whole process of developing a parsing system for a given language, there was nothing to show a linguistic consultant other than a few cryptic computer files. The linguist's system might work, but the information could not be presented to other linguists in an understandable way.

A dictionary tool for field linguists called Shoebox[1] was extended to be able to do automated morpheme breaking and interlinearization of texts. It lacked the capacity to aid in the description of word formation or morphophonemic processes but it was quite simple to use. Over time, this tool became the default choice of many field linguists for both

---

[1]Available online at: http://www.sil.org/computing/shoebox/. The Unicode compatible version is known as Toolbox and is available online at: http://www.sil.org/computing/toolbox/downloads.htm.

dictionary making and parsing. The FLEx parser was designed in response to the realization that the AMPLE and KIMMO morphological parsers were not meeting the needs of field linguists. These parsers, although more powerful, were not as easy to use as Shoebox and the majority of potential users chose to do parsing using the much less powerful tool.

XAMPLE, the default FLEx parser, was designed specifically to address these problems. Foundational to addressing the need for the parser to be easy for field linguists to use was the notion of using a simple item-and-arrangement morphology model that would be understandable for graduates of most linguistic training programs. The XAMPLE parser in FLEx, then, is built on the notion that there are three different processes for word formation: inflection, derivation, and compounding (Black and Simons 2006). A FLEx user is only required to understand these processes in analyzing the morphology and FLEx automatically generates the rules needed for the XAMPLE parser. Thus, the user is not required to learn how to write rules specific to the parser (as is the case with PC-KIMMO and AMPLE).

In the field of natural language processing much work has been done on further developing the two-level processing model initially proposed by Koskenniemi. Much of this has been centered around the work of Beesley and Karttunen (2003). The finite-state morphology theory proposes that both concatenative and nonconcatenative strings can be modeled with finite-state machines.

Much of the research on automated morphological parsing has focused on European languages beginning with Finnish (Koskenniemi 1983), and including languages like Croatian (Ćavar 2008), Hungarian (Prószéky and Kis 1999) and Turkish (Sak et al. 2009). Many researchers have worked on French parsing systems, including Porter (1980). Research on Asian languages has focused on Indonesian (Uliniansyah et al. 2002) and Indonesian (Hammarström 2009).

In the Philippine context, work on automated morphological parsing systems has been done for Cebuano and Tagalog. TagMA, standing for Tagalog Morphological Analyzer (Fortes 2004) is an implementation of optimality theory. This system was tested on 1,600 Tagalog verbs, yielding a 96% rate of correct analysis. Bonus (2004) developed TagSA (Tagalog Stemming Algorithm). While this had a lower rate of correct results (85%), it should be noted that Bonus tested his system on natural texts rather than pre-selected wordforms. Nelson (2004) developed a two-level parser for Tagalog that can export to XML format. Cena (2006) has developed a Tagalog parser which he terms "not pretty" but which has high parsing success rate.

4.2 Syntactic parsing

Because I am focusing on the use of syntax parsing tools for field linguists, a full review of mainframe parsing research is outside the scope of this paper. One computational framework which was developed for use on mainframes and received some attention from

field linguists was Augmented Transition Networks (ATN) which have been described by Woods (1970) and Winograd (1971). Grimes (1975) and others encouraged field linguists to use ATNs as a way of checking their syntactic analyses. This was the framework used by Errington (1979) for analyzing the Philippine language Cotabato Manobo before it was tested on a mainframe computer.

In the 1980s a new approach to syntax parsing was emerging as various frameworks were developed such as LFG (lexical functional grammar), FUG (functional unification grammar), DCG (definite-clause grammars), and GPSG (generalized phrase structure grammar). An underlying element in each of these was the concept of feature unification and all of these frameworks in a way influenced the development of PATR-II which Shieber (1986) calls "the simplest of the unification-based formalisms." In the late 1990s the PATR-II formalism was implemented for personal computers with PC-PATR (McConnel 1995). This is the parser that PAWS is built on.

Numerous NLP (Natural Language Processing) researchers are working on the problem of syntax parsing for Tagalog. Giganto (2004) developed PinoyMMT which has a syntactic parser component for both Tagalog and Cebuano. However, since the system was only tested on 16 sentences, the system is really just a prototype. Another syntax parsing system for Tagalog using LFG was developed by Borra and others (2007) for use in a machine translation system between English and Tagalog. The need for "the formal grammar

for languages involved" was cited as a problem. The LFG framework was also used for work on Tausug verbal sentences by Manguilimotan (2007).

The lack of computational grammars is a recurrent theme in papers by Philippine NLP researchers like Alcantara and Borra (2008b). The lack of such resources is cited as a significant limiting factor in the ability to develop "substantial" machine translation systems.

Chapter 5

METHODOLOGY

This chapter explains how I used FLEx and PAWS to describe the morphology and syntax of Ayta Abellen. Before examining the details of the programs, it is helpful to examine the process as a whole in order to see how they fit together into one processing stream with multiple outputs. In (1) the oblong shapes indicate user input, the slanted rectangles symbolize computer programs, and the regular rectangles denote outputs intended as inputs to a computer program. Finally, the rectangles with wavy lines represent outputs intended for a human to read.

(1)

Natural texts	User analyses	Answers to questions
		Selected examples

FieldWorks Language Explorer (FLEx)

PAWS Starter Kit

| Dictionary | Morphology Sketch | Parsing lexicon | Syntax Sketch | Syntax grammar (PSRs) |

Sentences to parse

XAMPLE Morphology Parser

Morphologically parsed sentences

PC-PATR Syntax Parser

Syntactically parsed trees

Word parse report

Sentence parse report

Section 5.1 explains how the lexicon and word grammar are modeled in FLEx by showing how the various word formation processes are handled. It also describes the process I went through to set up FLEx to accurately parse words. Section 5.2 lists the areas of syntax covered by the PAWS Starter Kit and briefly describes how I tested the phrase structure rules (PSRs) using PC-PATR.

5.1 Morphology

The morphological parsing was done using the parser built into FieldWorks Language Explorer. A lexicon of over 3,500 roots was compiled. Inflectional morphology was modeled using FLEx's inflectional templates. For each grammatical category, FLEx allows the user to specify the position of the inflectional affixes relative to the stem. The columns of the template can be labeled and each slot is marked as optional or required for a wordform of that category. The template itself is a formalism of the grammar while the inflectional affixes that populate the template are part of the lexicon. The allomorphs of the inflectional affixes are also stored in the lexicon. A screen shot showing a portion of a verb template is given below in (2).

(2) **Affix Templates**

Template Name   Eng AV inflection template
Tag

Description   Eng This is the inflection template for AV verbs.
Tag

Table

aspect-	AV voice-	STEM
aN- CONT	-om- AV	
m- CTPLT	^0- AV	
n - PFV		
-in- PFV		

Derivational morphology is also modeled in the lexicon. For each derivational affix, the user specifies what category the affix attaches to and what category it changes the root or stem into, as seen in the dialog box given in (3). Allomorphs of derivational affixes are also specified in the lexical entry.

(3)

Edit Grammatical Info.	✕
Lexical Entry: pag-₁	

Senses: GER

Grammatical Info.

Affix Type:	Attaches to Category:	Changes to Category:
Derivational ▾	Verb ▾	Noun ▾

OK	Cancel	Help
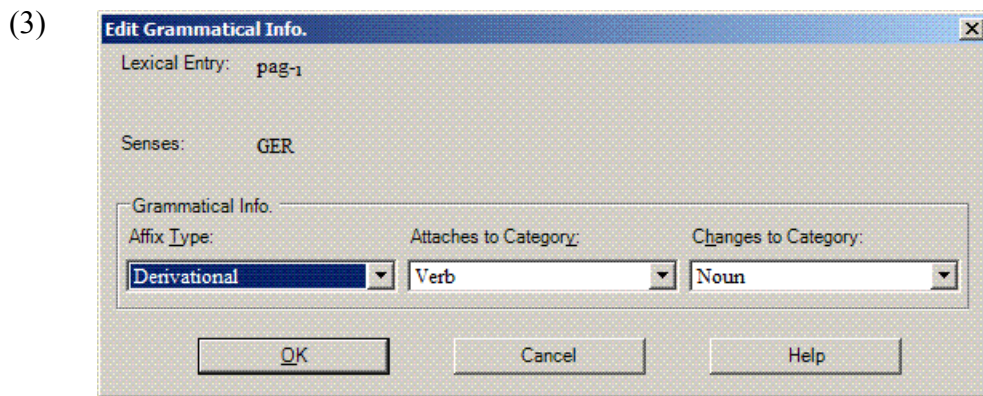
The XAMPLE parser in FLEx pulls information from several sources in order to parse a wordform. It looks in the lexicon to find information about lexemes, allomorphs, and stem formation (derivation) while it looks in the grammar to find information about morpheme ordering and whether inflectional affixes are required or optional. The parser can be run on either individual words with resulting output as in (4) or on interlinear texts.

(4)

^0-	i-	byay
m	i	biyay
CTPLT	CV	give
Category = V; Slot = aspect	Category = V; Slot = non-AV pre	Category = V

Stone (2004) has previously suggested an analysis for Ayta Abellen verbal inflection morphology. The research for this thesis began by entering this analysis into the FLEx database and then testing it on native authored texts. Adjustments were made as new natural

texts were parsed in the Interlinear Texts section of the program. Each of these adjustments to the analysis was stored in the FLEx program so that successive parsing efforts improved over time. When the parser was successful in analyzing any new word for which the stem was already in the lexicon, the analysis was deemed adequate and the automatically generated morphology sketch was saved.

The morphology sketch as output by FLEx was in XML format. It was further edited using the XML editor XMLMind.[1] The XML format of the sketch conforms to the XLingPaper XML standard for linguistic papers (Black 2009).

Discussion of the accuracy of the resulting morphological parser is found in 6.3 while the morphology sketch generated by FLEx can be found in Appendix B.

5.2 Syntax

The PAWS Starter Kit (Black and Black 2009) is a tool for analyzing and modeling the syntax of a language. The tool does not do analysis in the same sense as some other NLP tools which aim to extract information about the syntax of a language by examining large text corpora. Rather, it is an expert system that gives the user explanations about various syntactic structures before asking the user to make choices about how the language works with respect to those structures, such as whether certain elements are separate words or affixes in the language and where an element occurs with respect to the head of the phrase.

---

[1]Available online at: http://www.xmlmind.com/xmleditor/download.shtml.

The system is adjustable in content in that follow-up questions depend on the answers to the initial questions. If the user answers that a certain feature is not present in the language, no further questions are asked about that feature in PAWS. If a feature exists, the system requests sample data to illustrate the construction.

For example, the section on Adjectives and Adjective Phrases begins with a description of adjective phrases with English examples. Then comes the screen shown in (5) which asks if the language has words or phrases that can modify qualitative adjectives, and provides a box for entering examples. This is followed by a question concerning the position of these modifiers of qualitative adjectives.

(5)   Key your examples here:

```
tobat a mayadet.
mayadet a tobat.
```

Does your language have any words or phrases which can modify adjectives?
- No
- Yes - only certain degree words, though
- Yes - both certain degree words and manner adverb phrases
    Where do these modifiers occur with respect to the adjective?
    - Before the adjective
    - After the adjective
    - On either side of the adjective, but not both sides at the same time

The end of each section provides information about which lexical entries need to be marked in the lexicon with specific grammatical features. These are the features that are

written into the phrase structure rules which are generated for PC-PATR, the syntax parser.

For example, PAWS instructs the user to mark motion verb stems with the feature "motion"

in the lexicon. PC-PATR implements a unification model which enables features to percolate

up through the morphological parsing process so that they may become features of the

phrase structures that are referenced in rules. Unification then enables PC-PATR to

disambiguate wordforms for which the morphological parser produces multiple parses with

different syntactic features.

When the user has finished answering the syntax questions in PAWS and marking

the features in the lexicon, initial testing can be done for each PAWS section (e.g. Adjective

Phrases) using the sample data supplied by the user (which is automatically exported into

test files). PC-PATR requires a front-end morphological parser like AMPLE or XAMPLE

to parse the wordforms into component morphemes and also to determine the grammatical

category and syntactic features of each wordform. It then uses phrase structure rules stored

in the grammar file (.grm) generated by PAWS in order to perform syntactic parsing on the

morphologically parsed input text. The rules in the grammar file use general phrase structure

rules that are constrained with unification constraints.[2] A sample of the format used for rules

by PC-PATR can be seen in (6) where a topic phrase (TopicP) is defined as a topic marker

(TopicM) followed by a determiner phrase (DP). The three lines that follow define the

constraints on the arguments, namely, that the DP head must be minus for the feature *focus*

---

[2]Cheryl Black has written an article detailing a PATR implementation of Government & Binding theory using the X-Bar theory of phrase structure. It is available online at: http://www.sil.org/SILEWP/1997/006/.

(6c), minus for *wh* (6e), and minus for *temporal* (6e). The last two lines define features

which through unification percolate up to the whole construction, namely, that the head of

the TopicP is the head of the DP phrase (6f) and that TopicP becomes plus for the feature

topic-marked (6g).

(6)   a. *rule {Topic Phrase}*
         b. *TopicP = TopicM DP*
         c. *<DP head type focus> = -*
         d. *<DP head type wh> = -*
         e. *<DP head type temporal> = -*
         f. *<TopicP head> = <DP head>*
         g. *<TopicP head type topic-marked> = +*

When the PAWS process was completed, the PSR grammar generated by PAWS

and the lexicon exported from FieldWorks were used by PC-PATR to perform syntax parsing

on natural texts. Test sets of forty-nine and sixty-four sentences from Ayta Abellen natural

language stories were used for testing the coverage of the grammar that the PAWS Starter

Kit generated. Using two different testing methods that are detailed in 7.3 I show that the

claim that 75% of the syntax of any language can be modeled in formal rules generated by

PAWS is generally true for Ayta Abellen. A sample parse tree for one of the sentences used

in the test is shown below.

(7)

```
                                    S_1
        _____/    _____
    TopicP_2+                                                    IP_11
   ___/      \___                                                  |
TopicM_3+       DP_6+                                            I'_12
 _/    \_         |                                                |
Dem_4+ Linker_5+ D'_7+                                           VP_13
  |        |       |                            _____/|_____
Yabayin    a      NP_8+                      V_14+          DP_15               DP_19
 this     LNK      |                           |             |                   |
                 N'_9+                       antibeen       D'_16               D'_20
                   |                        CONT bark OV      |                   |
                 N_10+                                      NP_17               NP_21
                   |                                         |              ___/    \___
                maambal,                                  Pron_18        N'_22+      Pron_24
              python comma                                  |              |           |
                                                           lan          N_23+         ko
                                                     3PLGenPoss GEN.Lnk    |         1SgGen
                                                                         aho
                                                                         dog
```
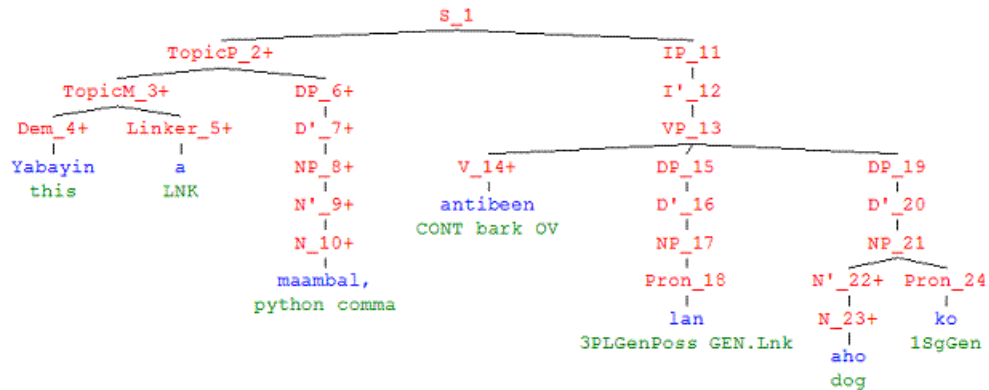
(8)   a. *Yabayin a   maambal, an-tibe-en*
            this   LNK python   CONT-bark-OV

           la=n   aho ko.
           3PL.GEN=GEN dog 1SG.GEN
           'This python, my dogs are barking at it.'

The second part of the output of PAWS was a syntax sketch in XML format. The sketch makes statements about the syntax of the language based on the user's answers to the questions in PAWS, and then includes the sample data supplied by the user as illustrative examples.

The PAWS Starter Kit covers seventeen areas of syntax, including:

(9)   a. *Overall Typology*
       b. *Quantifiers and Quantifier Phrases*
       c. *Adverbs and Adverb Phrases*
       d. *Adjectives and Adjective Phrases*
       e. *Simple and Possessed Nominal Phrases*
       f. *Proper Names*
       g. *Pronouns*
       h. *Pre-/Post-positional Phrases*
       i. *Basic, Single-clause Sentences*
       j. *Complement Clauses*
       k. *Questions*
       l. *Relative Clauses*
      m. *Adverbial Clauses*

n. *Negative Constructions*
o. *Coordination Constructions*
p. *Focus and Topic Constructions*
q. *Exclamations and Greetings*

When testing confirmed that the answers to the syntax questions in PAWS were

accurate, the resulting syntax sketch was output in XML format and refined using an XML

editor. The resulting sketch is found in Appendix E.

Chapter 6

MORPHOLOGY RESULTS

This chapter examines morphology primarily from a computational perspective. In other words, the focus is on describing Ayta Abellen affixes in a manner that enables a computer program (in this case FLEx) to automatically parse words and tag the component morphemes with labels based on grammatical functions and semantic features. The goal is to develop inflectional templates for inflectable grammatical categories and to describe derivational affixation. These, then, should be integrated in such a way that the word formation process can be displayed for humans to read as well as being the basis of computational parsing.

In 6.1, I explain the various affix sets in Ayta Abellen and provide a rationale for which sets are inflectional and derivational. In 6.2, I describe four characteristics of the language that were difficult to model, namely null derivation, partially inflected verbs, reduplication, and null morphemes. In 6.3, I discuss the details of the results achieved by using the morphological parser in FLEx. In 6.4, I detail two structures that are currently not handled by the XAMPLE parser and in section 6.5 I evaluate the usefulness of the human readable morphological sketch automatically generated by FLEx.

6.1 Ayta Abellen morphological parsing using XAMPLE

When considering the morphology of Philippine languages, the primary interest is in verbal morphology. The bulk of the morphological complexity in Philippine languages occurs in the verbal affixes. Ayta Abellen is no exception. While nouns are inflected for number, the bulk of the morphological complexity is seen in the verbal affixes.

Various approaches have been taken in order to account for the verbal affixes of Philippine languages. Most Philippine languages have verbal affixes like *ka-*, *pag-*, *paki-*, and *paka-*, which have been analyzed in many ways as to how these affixes and the voice affix sets interact with each other and the root to form verbs. Usually, descriptions focus on the semantics of affixes or clause syntax with less emphasis placed on morphotactics; i.e., how verbal affix sets are arranged relative to the root and to each other. This study is primarily interested in the morphotactics. Emphasis is placed on determining the nature and positions of the verbal morphemes so that a computer can be programmed to parse wordforms into their component morphemes.

In computational morphological parsing one is not merely trying to describe affixes the way a traditional grammar would but rather to define the word formation process in such a way that the parser can use it to analyze wordforms. Ultimately, one must be able to properly set up the inflectional template system in FLEx and define how the derivational affixes function. To do this, it is necessary to know two basic pieces of information: (1) the

number of sets of affixes, and (2) the nature of the affixes, whether they are inflectional or derivational.

First, in relation to determining the number of affix sets it was necessary to account for the information preceding the stem in order to determine whether actor voice forms like those for *habi* 'speak' in (10) have one prefix or two.

(10) a. *naghabi*
          n-   pag-   Ø-   habi
          PFV   DUR   AV   speak
          'spoke'

      b. *maghabi*
          m-   pag-   Ø-   habi
          CTPLT   DUR   AV   speak
          'will speak'

      c. *ampaghabi*
          aN-   pag-   Ø-   habi
          CONT   DUR   AV   speak
          'speaking'

      d. *nakahabi*
          n-   paka-   Ø-   habi
          PFV   AV   APT   speak
          'was able to speak'

      e. *makahabi*
          m-   paka-   Ø-   habi
          CTPLT   APT   AV   speak
          'will be able to speak'

      f. *ampakahabi*
          aN-   paka-   Ø-   habi
          CONT   APT   AV   speak
          'is able to speak'

g. *nakihabi*
   n-　　　*paki-* *Ø-* *habi*
   PFV　　 APT　AV　speak
   'requested to speak to'

h. *makihabi*
   m-　　　*paki-* *Ø-* *habi*
   CTPLT　　APT　AV　speak
   'will request to speak to'

i. *ampakihabi*
   aN-　　　*paki-* *Ø-* *habi*
   CONT　　　APT　AV　speak
   'is requesting to speak to'

One analysis option would be a portmanteau parsing strategy where no parsing cuts are made before the stem and a longer list of affixes is generated with multiple grammatical functions for each affix. The result of choosing a portmanteau analysis would be separate *nag-*, *mag-*, *ampag-*, *naka-*, *maka-*, *ampaka-*, *naki-*, *maki-*, and *ampaki-* prefixes. Doing this for all verbal affixes would create a long list of affixes but would make computational parsing simple. In spite of the simplicity of this approach I would like to give two reasons why preference is not given to the portmanteau analysis. First, it is clear which parts of the verbal wordforms mark aspect and which mark mode. It is not the case that the feature and the form have become so fused that one cannot see which part of the wordform controls which feature. Second, it is descriptive parsimony to have a smaller set of affixes rather than the large set of affixes produced by the portmanteau analysis.

A second piece of evidence is that verbs are not required to take what I am calling mode affixes, as can be seen in (11).

(11)   a.   *mideng*
      m-   *Ø-* *ideng*
      CTPLT AV stand
      'will stand'

    b.   *anhabien*
      aN-   *habi* *-en*
      CONT   speak OV
      'is saying [it]'

    c.   *haglapan*
      Ø-   *haglap* *-an*
      CTPLT   help   GV
      'will help'

So, I conclude that there are two separate sets of affixes that precede the root, one

for aspect (12) and one for mode (13).

(12)

Aspect	Affix
Contemplated	*m-/Ø*
Continuous	*aN-*
Perfective	*n-/-in-/iN-*

(13)

Mode	Affix
Plural action mode	*paN-*
Plural action detransitive mode	*pangi-*
Habitual action mode	*pi-*
Aptative mode	*paka-*
Durative mode	*pag-*
Stative mode	*ka-*[1]
Request mode	*paki-*
Reciprocal mode	*pay-*

Regarding Philippine voice, actor voice is often unmarked as in (14a) but is sometimes marked with *-om-* as in (14b) or *om-* as in (14c). Non-actor voice is most often marked with suffixes as in (14d) and (14e) but can be marked with a prefix for conveyance voice as in (14f).

(14)   a.   *magadal*
          m-   pag- Ø-   adal
          CTPLT   DUR   AV   study
          'will study'

      b.   *komodang*
          Ø-   -om- kodang
          CTPLT   AV   walk
          'will walk'

      c.   *ominom*
          Ø-   om-   inom
          CTPLT   AV   drink
          'will   drink'

---

[1]There also is a *ka-* affix in the stative inflectional template. The *ka-* form has two different functions. The derivational form is part of the stem that is inflected, as in words like *pinag-ka-lamo* 'companioned with' and *nag-ka-main* 'came to have'. The inflectional form is fused with voice marking to make a portmanteau of voice and mode, as in *am-paka-habi* 'is able to speak' and *ang-kai-lako* 'is being sold'. The various inflected forms can be found in (34).

d. *tawayan*
 Ø-   *taway* *-an*
 CTPLT   taste   GV
 'will taste'

e. *anhabien*
 aN-   *habi* *-en*
 CONT   speak OV
 'is saying [it]'

f. *ibyay*
 Ø-   *i-*   *byay*
 CTPLT   CV   give
 'will give [it]'

While the placement of the voice marking affixes varies in position, they do form a third set of affixes as seen in (15).

(15)

Voice	Affix
Actor	*Ø/-om-*
Patient	*-om-*
Object	*-en*
Conveyance	*i-*
Goal	*-an*

So, there are three sets of affixes, marking aspect, mode, and voice. The second question concerns whether each of these sets are inflectional or derivational. Traditionally, affixation has been divided into two categories, inflectional and derivational. For each affix or group of affixes, a set of criteria is used to determine if the affixes are primarily inflectional or derivational. Even with the most basic sets of criteria, making a choice for Philippine

language affix sets has not been easy, and linguists have arrived at various positions. Most Philippine language dictionaries have labeled the voice or "focus" affixes as inflectional as in Newell and Poligon (1993) and McKaughan and Al-Macaraya (1996). On the other hand Starosta (2002) and Himmelmann (1991) have argued that these affixes are all derivational, while Starosta and others (1982) have argued that certain voice affixes, namely *i-* and *-an*, are derivational while others are inflectional.

FLEx assumes that inflectional and derivational processes are basic to the word formation process. So, FLEx forces the user to determine whether affixes are inflectional or derivational. For my analysis, I am using the following list of nine characteristics of inflectional vs. derivational affixation as found in Kroeger (2005).

(16)

	Derivational	**Inflectional**
Category-changing	often	generally not
Paradigmatic	no	yes
Productivity	limited and variable	highly productive
Type of meaning	often lexical	often purely grammatical
Semantic regularity	often unpredictable	regular
Restricted to specific syntactic environments	no	yes
Position	central (near root)	peripheral (near edges of word)
Portmanteau forms	rarely	often
Repeatable?	sometimes	never

Both the aspect and voice affixes exhibit characteristics of being inflectional. They only attach to verb stems and thus do not change categories. While it could be argued that

the root in (17) is really a noun meaning 'thought' and that the voice affix *-en* is thus causing

a change of categories, I prefer to view the change of categories as null derivation that is

happening before the inflection of the stem. This seems a cleaner analysis than saying that

sometimes the voice affixes are inflectional and sometimes derivational. This is discussed

in more detail in 6.2.1.

(17) *ihipen*
      Ø-   *ihip   -en*
      CTPLT think OV
      'will   think about [it]'

They can be organized into paradigms as in (12) and (15). These affixes are highly

productive in the sense that they can attach to a wide range of verb stems. Some stems like

habi 'speak' can be inflected for all voices and aspects as in (18).

(18) a. *maghabi*
          m-   pag- Ø- habi
          CTPLT   DUR AV speak
          'will speak'

      b. *habien*
          Ø-   *habi   -en*
          CTPLT speak OV
          '[it] will be said'

      c. *habian*
          Ø-   *habi   -an*
          CTPLT speak GV
          'will speak to'

      d. *naghabi*
          n-   pag- Ø- habi
          PFV   DUR AV speak
          'spoke'

e. *hinabi*
   -in-   *habi*
   PFV.OV   speak
   '[it] was said'

f. *hinabian*
   -in-   *habi*   *-an*
   PFV   speak   GV
   'said to'

g. *ampaghabi*
   aN-   *pag-*   *Ø-*   *habi*
   CONT   DUR   AV   speak
   'speaking'

h. *anhabien*
   aN-   *habi*   *-en*
   CONT   speak   OV
   '[it] is being said'

i. *anhabian*
   aN-   *habi*   *-an*
   CONT   speak   GV
   'speaking to'

The type of meaning associated with these affixes is primarily grammatical. In (18) it is also evident that there is semantic regularity with regular changes in meaning for goal and object voice forms.

With regard to the question of whether the voice and aspect affixes are restricted to specific syntactic environments, it can be seen in (19) that the choice of voice marking on the verb is directly related to the syntactic structure of the sentence. So, it can be said that these are restricted to specific syntactic environments.

(19)   a.   *M-ag-Ø-habi*   *ya=y*   *Pabling kanan ahawa na.*
          CTPLT-DUR-AV-speak   3SG.NOM=NOM   Pabling   DAT   wife   3SG.GEN
          'Pabling will speak to his wife.'

     b.   *Ø-Habi-en*   *nan   Pabling kanan ahawa na.*
          CTPLT-speak-OV   GEN   Pabling   DAT   wife   3SG.GEN
          'Pabling will tell [it] to his wife.'

     c.   *Ø-Habi-an*   *nan   Pabling ye   ahawa na.*
          CTPLT-speak-GV   GEN   Pabling   NOM   wife   3SG.GEN
          'His wife Pabling will speak to.'

Aspect affixes are always word initial and so they are clearly peripheral. Voice

affixes are most often word final or word initial (as with conveyance voice *i-*) and so they

are more appropriately labeled peripheral even though AV *-om-* infixes do occur with

consonant initial roots.

With regard to portmanteau forms, the absence of an explicit object voice marker

for perfective aspect verbs as in (18e) is evidence for a portmanteau of aspect and voice.

While it might be argued that there is simply a null object voice marker here, comparing

this example to (20) we see that this also occurs with conveyance voice constructions.

(20)   *In-lantad*   *na=y*   *anti*
      CV.PFV-make.known   3SG.GEN=NOM   present

      ha   nakem   na.
      DAT   inner.being   3SG.GEN
     'He made known what was in his soul.'

Neither aspect or voice affixes are repeatable in the same word. So, on the basis of

these criteria for inflectional affixation, I label both the aspect and voice affix sets as

inflectional.

Now let's consider the mode affixes as in (13). Applying these criteria again to these affixes I will begin by looking at the question of whether these affixes derive different grammatical categories. It might be argued that it is these affixes that are causing nouns like *angin* 'wind' to change into verbs as in (21). But examples where this happens without mode affixes like for *anag* 'termite' in (22) show that this is not the case. I will argue in 6.2 that this is really null derivation happening before the application of the mode affixes.

(21) *N-ang-Ø-angin   ya=n   makhaw ihtew ha   dagat.*
     PFV-PL-AV-wind   3SG.NOM=LNK strong   there DAT sea
     'The wind blew hard there on the sea.'

(22) *Hiyay tapih yo,   ang-anag-en   ana.*
     TM   clothes 2PL.GEN CONT-termite-OV now
     'Your clothes are now being eaten by termites.'

Regarding the question of whether the mode affixes form a paradigm, Kroeger (2005) has defined a paradigm as "a set of forms which includes all of the possible values for a particular grammatical feature." There is not a standard set of values for Philippine-type language "mode" and it would be difficult to define mode (in the sense I am using it here) as a grammatical feature. So, it can be said that the mode affixes do not form a paradigm.

The mode affixes are productive, being able to attach to most verbal roots. But even still, there are limitations on the roots to which they can be affixed and these limitations seem to come from the interaction of the semantics of the stem with the affix. For instance, the *paki-* affix (meaning polite request) can be attached to many verbal stems. But it can never be attached to a verbal root with lexical meaning relating to something normally

occurring accidently as in (23). So, it would more appropriate to label their productivity as variable and thus more characteristic of derivation.

(23)   a.   **makiangin*
          m-   paki- angin
          'request to have wind'

    b.   **makiampag*
          m-   paki- ampag
          'request to fall'

The type of meaning associated with the mode affixes is more semantic than grammatical. These affixes do not help define the roles and relations of words in the sentence as the voice affixes do and they do not correlate with a standard grammatical category as the aspect affixes do. Instead, these affixes give information about how the action takes place, for example habitually (*pi-*), reciprocally (*pay-*), as a request (*paki-*), with multiple actions (*paN-*), etc. This would seem to be more of a semantic rather than grammatical function, although the existence of the plural action detransitive mode affix (*pangi-*) could be cited as evidence against the claim of them being purely semantic meaning.

The mode affixes are not restricted to specific syntactic environments like the voice and aspect affixes. The mode affixes can occur with different aspect and different voice verbs as in (24)-(27). There is not a connection between the syntactic structure of the sentence and the type of mode affix applied to the verb.

(24)   *N-ag-Ø-habi*   *ya=n*   *oman ye*   *Ben.*
       PFV-DUR-AV-word 3SG.NOM=LNK again NOM Ben
       'Ben spoke again.'

(25) *Ang-i-paki-ingalo   ko   kanyo,   ka-katongno.*
CON-CV-REQ-mercy 1SG.GEN 2PL.DAT PL-brother
'I am pleading with you, brethren.'

(26) *M-ang-kemt-en   mi   hila   ha   mabakil.*
CTPLT-PL-overrun-OV 1PL.EX 3PL.NOM DAT mountain.slope
'We were overrunning them in the mountain slopes.'

(27) *Am-pay-habi-an   la   no   hinya=y*
CONT-REC-word-GV 3PL.GEN as-to what=NOM

  labay Ø-habi-en   nin   yatew.
  want   CTPLT-word-OV GEN that
'They are discussing together about what that means.'

The position of the mode affixes in relation to the voice and aspect affixes is central, near the root. By choosing to separate the mode affixes from aspect and voice I am saying that these are not portmanteau forms. The mode affixes are not repeatable in the sense of being able to have two of the same affix in the same word but they are repeatable in the sense that there are words where more than one mode affix can occur adjacent to another one in the same word as in (28).

(28) a. *nakipaglaban*
        n-   paki- pag- Ø- laban
        PFV   REQ DUR AV fight
        'join others to fight against'

    b. *ampakapagtaka*
        aN-   paka- pag- Ø- taka
        CONT   APT   DUR AV amaze
        'being amazed'

So, in terms of the criteria in (16), these affixes are derivational.

Now, with the information about (1) the number of affix sets, and (2) their nature (whether inflectional or derivational), it is possible to implement the solution in FLEx. In addition to marking the aspect and voices affixes as inflectional in the lexicon and the mode affixes as derivational, an inflectional template was built to place the aspect and voice affixes in the correct relative order and to indicate whether they are optional or obligatory.[2] In the first attempt to implement the analysis, an attempt was made to place all the verbal affixes into one inflectional template as seen in (29) where column headers marked with parentheses denote optionality while unmarked columns are required.

(29)

Aspect		(Voice1)		Stem	(Voice2)	
m-/Ø-	'CTPLT'	*i-*	'CV'		*-an*	'GV'
aN-	'CONT'	*-om-*	'AV'		*-en*	'OV'
n/in-/-in-	'PFV'	*Ø-*	'AV'			

This inflectional template is descriptive and simple, and using it I was able to parse a very high percentage of Ayta Abellen words. In fact, with this arrangement, over 97% of words on a wordlist of over 600 words from natural text were parsed correctly. But, the pursuit of training the parser to handle the other 3 percent led me to try various ways of implementing the analysis with multiple templates.

---

[2]The initial analysis that I entered into FLEx before testing with the parser was based on Stone (2004) where I extended the analysis of Ayta Abellen verbal morphology originally done by Nitsch (2009) by incorporating insights from Rubino (1998) on Ilokano and Antworth (1979) on Botolan Sambal.

Because of null aspect and voice morphemes, I had to make the aspect and voice columns obligatory. But this caused pseudo verbs like *labay* 'want' to invoke these null forms and generate the incorrect parse in (30) instead of (31).

(30) *Ø-Ø-labay*
      CTPLT-AV-want
      'will want'

(31) *labay*
      want
      'want'

By making the voice column obligatory, I needed then to make a split between prefix and suffix marked voice inflections. The result was separate templates for prefix marked voice as in (32) and suffix marked voice as in (33).

(32) **Actor, Patient, and Conveyance voice inflection template**

Aspect	Voice prefix	Stem
aN- 'CONT'	*Ø-* 'AV'	
-in- 'PFV'	*i-* 'CV'	
m- 'CTPLT'	*-om-* 'AV '	
n - 'PFV'	*-om-* 'PV'	

(33) **Object and Goal voice inflection template**

Aspect	Stem	(Voice suffix)
aN- 'CONT'		*-an* 'GV'
-in- 'PFV'		*-en* 'OV'
m- 'CTPLT'		
n - 'PFV'		

A separate template was needed to handle stative verbs which have unique forms for each voice as in (34). There is also a stative verb prefix *ka-* that is derivational as seen in (13).

(34) **Stative verb inflection template**

Aspect	Stative voice	Stem	(Stative voice suffix)
aN- 'CONT' *-in-* 'PFV' *m-* 'CTPLT' *n -* 'PFV'	*ka-* 'STA OV' *ka- -an* 'STA GV' *kai-* 'STA CV' *paka-* 'STA AV'		*ka- -an* 'STA GV'

Most Philippine languages have a small set of verbs that either cannot be inflected or take limited inflection. Some call them pseudo verbs and label them adjectives as in Schachter and Otanes (1972) for Tagalog. Others, like Rubino (1998) for Ilokano, call them "root verbs" that are "morphologically simple".

A pseudo verb subcategory of verb was created with its own template as in (38) in order to handle these partially inflectable roots. By marking these roots in the lexicon as being of the category pseudo verb, making the suffix column optional in the pseudo verb inflectional template, and having a null aspect marker of the category pseudo verb, the examples in (35), (36), and (37) are all parsed correctly.

(35) *main*
      EXT
      'there is'

(36) *main-an*
      EXT-GV
      'there will be'

(37) *m-ag-ka-main*
CTPLT-DUR-STA-EXT
'there will be'

(38) **Pseudo verb inflection template**

Stem	(Voice suffix)
	-an 'GV'
	-en 'OV'

Words like *hinabi* 'said [it]' in Ayta Abellen, and its Tagalog equivalent *sinabi* 'said [it]' pose a unique problem in that there appears to be no voice marker, even though it is clear that these are object voice constructions. A variety of solutions exist for this problem in Tagalog. Wolfenden (1961) showed the Tagalog verb formation process in a flowchart where there is an *-in-* for "action begun" and an *-in* for objective voice, but didn't explain how the *-in* gets dropped when they both should occur in the same wordform. Schachter proposed to solve this problem for Tagalog with a rule for "the addition of *ni-* or *-in-* plus the deletion of the suffix *-in*" (Schachter and Otanes 1972). Kroeger (1993) seems to imply null marking of object voice by placing the OV label at the end of example words.

In the FLEx implementation described here, this problem of *-in-* portmanteau is handled by the inflectional template in (41). Without this template, a word like *hinabi* would be inflected using the AV/PV/CV template which would give the incorrect parse in (39) rather than the correct parse in (40).

(39) **Ø-h<in>abi*
AV<PFV>-word
'said'

(40)   *h<in>abi*
<PFV.OV>word
'said'

(41)   **Perfective Aspect, Object/Conveyance Voice inflection template**

Aspect/voice	Stem
-in- 'PFV, OV' *in--* 'PFV, OV' *iN-* 'PFV, CV'	

Finally, there are forms like (42) that get inflected before taking a derivational occupational prefix. This is a problem that may be somewhat unique to Ayta Abellen. This was handled by using the "requires more derivation" option in FLEx and the template seen in (43).

(42)   *māng-<om>onin*
OCCUP<AV>live
'resident'

(43)   **Prolonged action inflection template**

Om prefix	Stem
-om- 'AV '	

Why the need for six templates? Are some of these templates just computational tricks to make verbs parse? Not really. The last three templates represent information that is sometimes overlooked in traditional grammars, or for which special rules and exceptions are written to explain this less systematic data. Because I was intent on making every word parse, I was forced to deal with even lower frequency wordforms. While the approach

modeled here may or may not prove to be the best way of organizing the information about affixation, it does handle all the data in a transparent manner.

6.2 Ayta Abellen morphological parsing challenges

In this section I discuss the four most difficult problems that needed to be addressed with parsing Ayta Abellen words. In 6.2.1, I discuss null derivation and how to model it in FLEx. In 6.2.2, I discuss the problem of the need for null morphemes in the analysis. In 6.2.3, I mention the problems of syllable and root reduplication. In 6.2.4 I discuss the problem of derivation taking place after inflection.

6.2.1 Null derivation

In English it is said that we can "verb" any noun. In Ayta Abellen also, many noun roots can also be expressed as verb forms. For example, there is the root word *anag* 'termite'. This can be labelled as a noun root. But the root can be affixed to form stems that function either as a verb as in (44) or as an adjective as in (45).

(44) *Hiyay tapih   yo,   ang-anag-en   ana.*
      TM   clothes 2PL.GEN   CONT-termite-OV   now
      'Your clothes are now being eaten by termites.'

(45) *Ang-k<om>onin hila=yna   ha   ma-anag   a   baey.*
      CONT<AV>dwell   3PL.NOM=now   DAT ADJ-termite LNK   house
      'They are now living in a termite infested house.'

The question for developing a computational model of the lexicon is where to show the derivation taking place. Is it the case that the verbal affixes *aN-* or *-en* are deriving a

verb from a noun? We have already concluded, however, that these affixes are inflectional rather than derivational. But even if we did want to say it is one of these affixes that is causing the change, we would have to choose which one is causing the change, which would be difficult to do. This is a case of null derivation or conversion, where a derivational process with a null surface form is converting the root from one grammatical category to another before inflectional affixes are applied.

As I looked at the data I realized that there were many examples like *ang-anagen* 'being eaten by termites' where there were no derivational mode affixes in the surface forms but where there seemed to be some kind of derivation taking place. There are at least three solutions to this problem: (1) Add senses with all possible grammatical categories for each root so that affixes will always be able to attach to a sense with the correct grammatical category. (2) Define most of the voice affixes to have both an inflectional and derivational sense. (3) Define null morphemes that attach to the root and convert the root into a stem with the correct grammatical category. The first solution is labor intensive and doesn't really model what is happening in the language, being more of a work around to make the program parse the wordforms. The second solution blurs the distinction between inflection and derivation, saying that these affixes can basically function either way. This option would basically rely on the mode derivational affixes to derive verbs from nouns and would say that when a mode derivational affix is missing, the voice affix can function as a derivational affix to cause the conversion. The third option shows the category conversion as a separate

derivational process and thus leaves the voice and aspect affixes as solely inflectional. This option is also computationally more efficient, using a minimum number of entries in the lexicon.

In looking at the possible conversions between the three open category grammatical categories (N, V, Adj) there were three that were deemed not possible and thus not modeled. They are: Adj→N, V→N, and V→Adj. The three that were modeled are: N→V, N→Adj and Adj→V. The noun to verb null morpheme is constrained by the requirement of each slot on the inflectional templates. This is constrained by the fact that one of the inflectional templates must apply to any V and that all templates involve at least one required non-null morpheme. The adjective to verb morpheme is also constrained by the fact that verbs generally require affixation. An example of how the parser handles null derivation is shown in the screenshot for *ang-anagen* 'being eaten by termites' in (46).

(46)

Parse of **ang-anagen.**

Result

This word parsed successfully. The following are the sequences of allomorphs that succeeded:

ang--	^0-	anag	-en
aN	^0	anag	en
CONT	N>V	termite	OV
Category = V; Slot = aspect	From category = N; To category = V	Category = N	Category = V; Slot = non-AV suff

This use of null derivational affixes is more efficient than trying to put every instance of noun and verb as senses in the lexicon. For example, the entry *habi* 'word' is marked as

a noun. The parser, however, can recognize a verbal form without having verb as a grammatical category in the entry, as in (47), because of the null derivational affix. It can also recognize the verbalized form *pomoti* from the adjective *poti* 'white' as in (48).

(47)

-in- in PFV Category = V; Slot = aspect	^0- ^0 N>V From category = N; To category = V		habi habi language Category = N	-an an GV Category = V; Slot = non-AV suff

(48)

^0- m CTPLT Category = V; Slot = aspect	-om- om AV Category = V; Slot = AV voice	^0- ^0 Adj > V From category = Adj; To category = V	poti poti white Category = Adj

6.2.2 Null morphemes

Developing a parsing model for Ayta Abellen would be simpler if every inflection feature were always marked. But as was mentioned in 6.1, sometimes aspect and voice are unmarked.

For the actor voice data in (49a), voice is not marked but aspect is marked. There are, however, actor voice verbs such as with the root *lateng* 'arrive' that are marked for voice but not marked for aspect as in (49b), or marked for both aspect and voice with perfective and continuous aspect, as in (49c) and (49d). To account for this actor voice data I needed to posit that there are two forms of actor voice, one being unmarked and the other being marked with *-om-*. For aspect, contemplated aspect is marked with *m-* but has an allomorph of *Ø-*.

(49) a. *m-ag-adal*
      m-   *pag-* *Ø-* *adal*
      CTPLT   DUR AV study
      'will study'

    b. *lomateng*
      Ø-   *-om-* *lateng*
      CTPLT   AV   arrive
      'will arrive'

    c. *linomateng*
      -in-   *-om-* *lateng*
      PFV   AV   arrive
      'arrived'

    d. *anlomateng*
      aN-   *-om-* *lateng*
      CONT   AV   arrive
      'arriving'

Looking at goal voice verbs, voice is consistently marked with the *-an* suffix but aspect is unmarked for contemplated aspect as in (50a) while marked with a prefix for continuous aspect in (50b) and an infix for perfective aspect in (50c). Here again, I need to conclude that there is a null (Ø-) contemplated aspect prefix.

(50) a. *habian*
      Ø-   *habi* *-an*
      CTPLT speak GV
      'will speak to'

    b. *anhabian*
      aN-   *habi* *-an*
      CONT   speak GV
      'is speaking to'

c. *hinabian*
   -in-   *habi* *-an*
   PFV   speak GV
   'spoke to'

For the case of perfective aspect with conveyance and object voice, it is less obvious how to model this as there appears to be a portmanteau of features with one affix marking both perfective aspect and conveyance voice in (51) and one affix marking perfective aspect and object voice in (52).

(51)   *In-habi*   *na=yna=n*   *peteg* *kanla.*
         PFV.CV-say 3SG.GEN=now=LNK true   DAT.3PL
         'He really told [it] to them.'

(52)   *Yatew ye*   *h<in>abi*   *nan* *babayi.*
         that   NOM   <PFV.OV>-say GEN woman
         'That is what the woman said.'

Even after splitting the inflectional templates into separate templates for different voice patterns, it was still difficult to handle this with null morphemes alone. For object voice there would be a potential null morpheme on both sides of the root (CTPLT aspect, Object voice). This would allow for overgeneration and since this only happens with perfective aspect verbs, it was decided to add another inflectional template as shown below.

(53)   **Perfective Aspect, Object/Conveyance Voice inflection template**

Aspect/voice	Stem
-in- 'PFV, OV'	
in-- 'PFV, OV'	
iN- 'PFV, CV'	

6.2.3 Reduplication

Another challenge was the existence of reduplication both at the syllable and root

levels. FLEx handles this easily by allowing the user to specify a reduplicated

consonant-vowel segment, or a reduplicated vowel as an allomorph as in (54) where V^1

means that the vowel is reduplicated from the stem.

(54)
Allomorphs	
Affix Allomorph	Ayt [V^1]
Morph Type	prefix
Environments	/ # _ [V^1]

FLEx also allows the user to specify a reduplicated root as an allomorph as in (55).

(55)
Allomorphs	
Affix Allomorph	Ayt [...]
Morph Type	prefix
Environments	

6.2.4 Derivation on inflected stems

Ayta Abellen has a limited set of words where a stem inflected with *-om-* for actor

voice takes the derivational prefix *māng-* 'OCCUP' which converts the stem to a noun.

Examples can be seen in (56) and (58). The same thing occurs for the gerundivizer *pāng-*

as seen in (58).

(56) *māngkomonin*
māN-   -om- konin
OCCUP   AV   reside
'resident'

(57) *māngomodang*
      māN-   komodang
      OCCUP   walk
      'walker around (rebel)'

(58) *pāngomodang*
      pāN-   -om- kodang
      GER   AV   walk
      'walking'

While derivation normally precedes inflection, for this small set of AV forms it appears that derivation is occurring after inflection. Fortunately this can be modeled in FLEx by adding another template and selecting the option that it "requires more derivation" as shown in (59).

(59)

Table	⊙ om prefix-		⊙STEM
	-om- AV		
Note	To edit the template table: right-click on STEM, on a slot name, or on a morpheme.		

Requires more derivation ☑

6.3 Evaluating the parser

Before attempting automated parsing I did manual parsing in FLEx using words in 10 native authored texts (2,996 words). After finishing this, I began to experiment with the inflectional template options and eventually ran some initial tests with the automated parser in FLEx on six datasets.

These initial tests, however, were done on running text (with many repetitions of the same wordforms) and not on a wordlist. So the 96 percent parsing rate was high but not representative of how the parser would handle all the wordforms in the language. In the

next phase of development I used a test dataset comprised of a wordlist from the 10 native

authored texts. There were 765 distinct wordforms and while 83 words failed to parse, 58

of these were due to text problems such as borrowed words, proper names, and spelling

errors. With these not included in the calculation, the result was that 3 percent of wordforms

failed to parse due to a problem in the description of the morphology as shown in (60).

(60) **Training set**

Word (%)	Parse Failure (%)	Description
641 (84%)		correct parse
41 (5%)		multiple parses (both correct and incorrect)
83 (11%)		failed to parse
	21 (3%)	morphology problem
	4 (.5%)	lexical problem (missing root, category, or allomorph)
	58 (7.5%)	text problem (spelling error, borrowed words)
765	83	Total

Another aspect to consider in evaluating the parser is overgeneration. This is seen

in the 41 words which did parse correctly but also had an extraneous, incorrect parse. The

degree of overgeneration is shown in (61) and the total overgeneration rate was 5.3%.

(61)   **Training set overgeneration**

Words	# of extra incorrect parses
21 (2.7%)	1 extraneous parse
10 (1.3%)	2 extraneous parses
2 (.3%)	3 extraneous parses
7 (.9%)	4 extraneous parses
1 (.1%)	6 extraneous parses
41 (5.3%)	Total

After testing this set, I realized the need to rework the inflectional templates, namely to make each slot required and split into multiple templates for different voice configurations. I also added a template to handle the portmanteau forms seen in (41) and the partially inflected roots mentioned in 6.1. After completing these adjustments the parser now handles all words except for those structures mentioned in 6.4.

My final task was to develop the lexicon and word grammar to the point that they would produce a correct parse for every word in a wordlist of 3,730 words generated from written texts. This wordlist is basically a composite picture of all available committee reviewed written texts in the language. Because of this review, there were fewer incorrect wordforms than in the previous natural texts wordlist. Even so, the parser's accuracy had reached the point where it was serving as a spell checker. The only words that would not parse were due to limitations with the FLEx parser and are detailed in 6.4.

6.4 Structures not covered with XAMPLE

There are two word structures in Ayta Abellen that are not adequately covered with the XAMPLE parser. The first problem, that of coincident elision and assimilation at morpheme boundaries can be modeled, although in an unsystematic way. The second, however, that of root reduplication with both prefixes and suffixes cannot be modeled with the current XAMPLE parser.

The first problem is morphophonemic assimilation and suppletion of the initial root phoneme as in (62). The current solution is to add allomorphs to each root where this process can apply. So, in the case of the root *haglap* 'help', I have also added an allomorph *aglap*. This is by no means optimal and doesn't describe adequately the phonological processes at work here. It is also not possible to automatically generate these allomorphs because this process does not occur with every root entry.

(62) *managlap*
      m-   *paN-* *haglap*
      CTPLT   PL   help
      'will help'

In section 6.2.3 I showed an example of how full root reduplication with prefixes is handled. Unfortunately, when there are both prefixes and suffixes the current XAMPLE model is inadequate as noted in the documentation: "The way we are modeling full reduplication in FieldWorks Language Explorer, the root must be at one end and then any affixes (including the reduplication morpheme) must either be all prefixes or be all suffixes"

(Black 2006). So, the Ayta word *magmotawmotawan* cannot be handled by the parser.

There were eight other words in the list that would not parse for the same reason.

(63)   *magmotawmotawan*
m		pag	motaw	motaw	an
CTPLT		DUR	fool	fool	GV

'will be foolish about'

Both of these problems, however, are addressed in the experimental version of the

Hermit Crab parser that has been incorporated into FLEx 3.0. The Hermit Crab parser has

the capacity to handle phonological processes and so the elision/assimilation problems will

be handled with rules rather than allomorphs. The Hermit Crab parser also handles root

reduplication in a different manner and will not have the limitation seen in the example

above.

A third problem is that of when CV reduplication occurs on an inflected root as in

(64). The parser looks at the root for determining if there is a match for the reduplication

pattern. But in this case the CV reduplication pattern matches the inflected CV pattern rather

than the pattern at the beginning of the root. This is the only example of its kind in the data.

It is not parsed by the XAMPLE parser and it is hypothesized that this will be a problem

for the Hermit Crab parser also.

(64)   *pāndodomaya*
pāN	do	om	daya
PL.PA.GER	EMPH	PV	blood

'bleeding'

6.5 Evaluating the sketch

The morphology sketch that was automatically generated by FLEx is given in full below as Appendix B. Before examining the morphology sketch that FLEx outputs it is helpful to look at a diagram of what Simons and Black (2009) call Third Wave Publishing. The basic idea is that knowledge about the morphology and grammar of the language is put into a data structure where it can be automatically repurposed for varyious output requirements such as academic papers, computational parsing rules, computational lexicon, print dictionary, and so on. This is vastly different than Second Wave Publishing where a distinct file must be created by the author for each intended output. The model that Simons and Black present in (65) proposes "information structuring" of linguistic information as opposed to word processing.

(65)
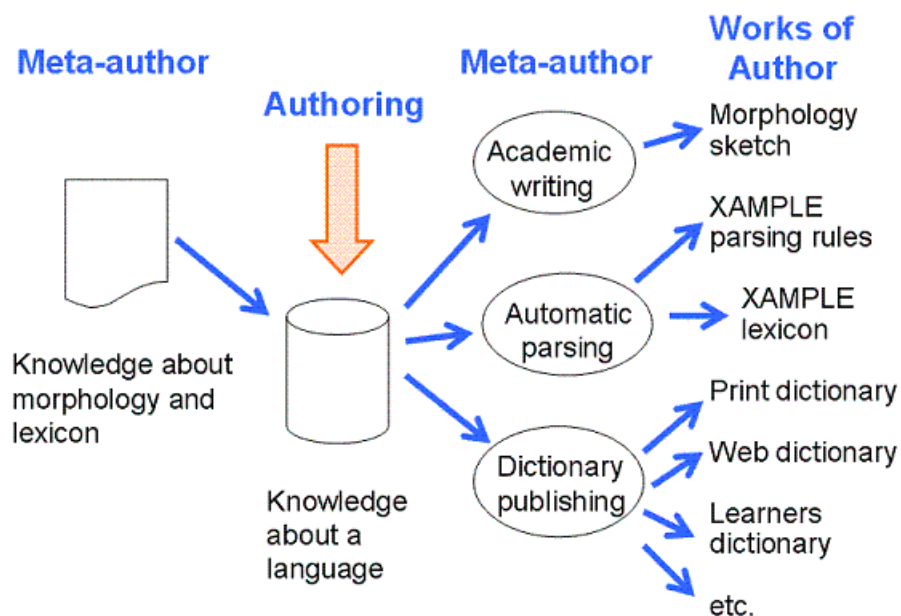
When considered in light of the process of Third Wave Publishing, the FLEx

morphology sketch is an automated output of structured information whereby reusable

content (HTML and XML format) is generated from an underlying knowledge source. The

morphology sketch is generated with the click of a button and can then be printed, uploaded

to the web, or searched for content. This sketch is an example of the kind of document that

can be automatically produced from a structured information source like FLEx.

So computationally, the morphology sketch fits the Third Wave Publishing paradigm

but how useful is it from the standpoint of linguistic description? Black and Simons (2006)

note that the motivation for the development of the FLEx morphology sketch was to enable

linguists who have done a computational implementation of their analyses to be able to

present the description in a manner that can be read and understood by other linguists. This

had not been the case with AMPLE users who came up with morphology descriptions that

could only be understood and processed by a computer.

The morphology sketch output by FLEx is general enough to be understandable to

linguists. In personal communication during 2009, Barlaan confirms that the FLEx

morphology sketch could constitute the basis of one's morphological analysis of the surface

structure but goes on to say that for Philippine languages, attention to semantics of affixes

is crucial. The original morphology sketch output by versions of FLEx before 3.0 only

showed the basic gloss for each example lexeme displaying the corresponding definition

from the lexicon. In response to review of this thesis the definition field was added to the

morphology sketch, giving linguists a place to include some of the grammatical and semantic information necessary for describing Philippine languages.

In perusing morphological descriptions of other Philippine languages it is apparent that a great deal of attention is given to describing the syntax and semantics of what is called the "focus" or voice system of Philippine languages. FLEx, as an expert system, has included "voice/focus in Philippine-type languages" as an optional inflectional feature that users can turn on and use for describing affixes or as a condition in morphological processes. The morphology sketch output by versions of FLEx before version 3.0 generated a list of inflectional features but only showed the features and not which morphemes have these features, which was a real disadvantage. But, as a result of input from this thesis, version 3.0 now shows both the features and the affixes with which the features are associated.

A weakness of the morphology sketch is that it does not structure pronoun information in a meaningful way, or even show all the pronouns. It only shows the first ten pronouns. Ayta Abellen has 32 pronouns that can be organized into four grammatical and eight semantic categories. It would be helpful to be able to show this in the sketch.

In summary, the morphology sketch that FLEx outputs is a helpful starting point as a description of the morphemes and word formation processes. Inclusion of more detailed information about the verbal affixes with the option of selecting examples from the database to be included was needed to make this document a more significant description of a Philippine language. The morphology sketch does not answer all the questions that linguists

specializing in Philippine-type languages have about a particular language but it does give

a good foundational description.

Chapter 7

SYNTAX RESULTS

This chapter discusses the analysis of the syntax of Ayta Abellen. In 7.1, I summarize work done by others in the area of formal syntax for Philippine languages and show the types of trees that the PC-PATR parser is generating for Ayta Abellen. In 7.2, I look at some of the challenges for syntax parsing in Ayta Abellen, namely case and voice in 7.2.1, pronouns and second position clitics in 7.2.2, and topicalization in 7.2.3.

In 7.3, I evaluate the results for the Ayta Abellen syntactic parser on a training set of sentences and in 7.5, I evaluate the results on two different sets of natural texts. In 7.4, I describe the structures not included in the original PAWS phrase structure rules but which were added as custom rules. In 7.6, I briefly evaluate the syntax sketch automatically generated by PAWS.

7.1 Ayta Abellen syntactical parsing using PAWS

Syntax parsing in general is a vast field of study with special complications for Philippine-type languages. Most of the recent research on formal syntax has been done on Tagalog as seen in Kroeger (1993), Maclachlan (1996), Rackowski (2002), Aldridge (2002),

and Sabbagh (2005) although work has also been done on Cebuano (Bell 1976) and

Kapampangan (Mirikitani 1972).

In order to see a cross-section of work done in Philippine languages related to Ayta

Abellen, I briefly survey the conclusions of Kroeger, Bell, and Mirikitani for Tagalog,

Cebuano, and Kapampangan. Each considered the respective Philippine languages to be

non-configurational and lacking a traditional VP. Kroeger (1993) described the basic Tagalog

sentence with this tree:

(66)

```
            IP
            |
    SPEC         I'
            I        S
                  V   NP  NP
```
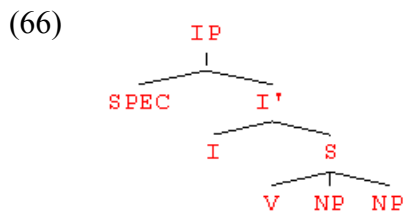
Bell (1976) described the syntactic tree structure of Cebuano in this way:

(67)

```
          S
          |
          VP
     V   NP₁  NP₂  NP₃
```
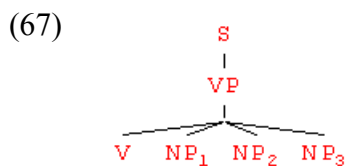
Mirikitani (1972) also proposed a flat tree structure for Kapampangan,[1] as in,

(68)

```
            S
            |
          PredP
     VP      PP      PP      PP
```

---
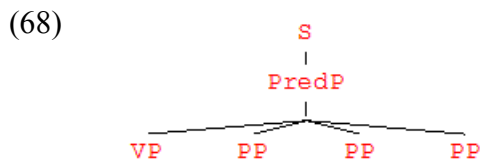
[1]Kapampangan and Ayta Abellen speakers have contact with each other because of geographical proximity. While Kapampangan is not a member of the Sambal subgroup, its relatively high lexical similarity with Ayta languages suggests macrogrouping (Stone 2009).

For each of these Philippine languages something other than a traditional VP is proposed. In a traditional VP there is a close relationship between the verb and the object. But Philippine language syntax is different as even the identification of what constitutes the subject is problematic and has been debated by many linguists. What is clear is that the verb normally precedes one or more noun phrases and hence the notion that Philippine languages are VSO. With this background we are now in a position to examine the tree parses that PC-PATR generates for Ayta Abellen sentences.

In looking at the three models above, the analysis generated by PAWS for a sample Ayta Abellen sentence with three NPs (69) shows similarities to the flat structures previously proposed for Cebuano and Kapampangan.

(69) *In-i-byay   na   ye   haa   kanan Pabling*
      PFV-CV-give   3SG.GEN   NOM   banana   DAT   Pabling
      'He gave the banana to Pabling.'

(70)

```
                              S_27
                               |
                             IP_28
                               |
                             I'_29
                               |
                             VP_30
        _____|_____
       |              |              |                |
     V_5+          DP_6           DP_10+           DP_20+
       |              |          ____|____        ____|____
   Inibiyay        D'_7        |         |       |         |
   PFV CV give       |      Case_11+  DP_12+  Case_21+  DP_22+
                   NP_8         |       |       |         |
                     |          ye    D'_13+  kanan    D'_23+
                   Pron_9      NOM      |      DAT        |
                     |               NP_14+            NP_24+
                    na                 |                 |
                 3SgGenPoss          N'_15+            N'_25+
                                       |                 |
                                     N_16+             N_26+
                                       |                 |
                                      haa             Pabling
                                     banana            pabling
```
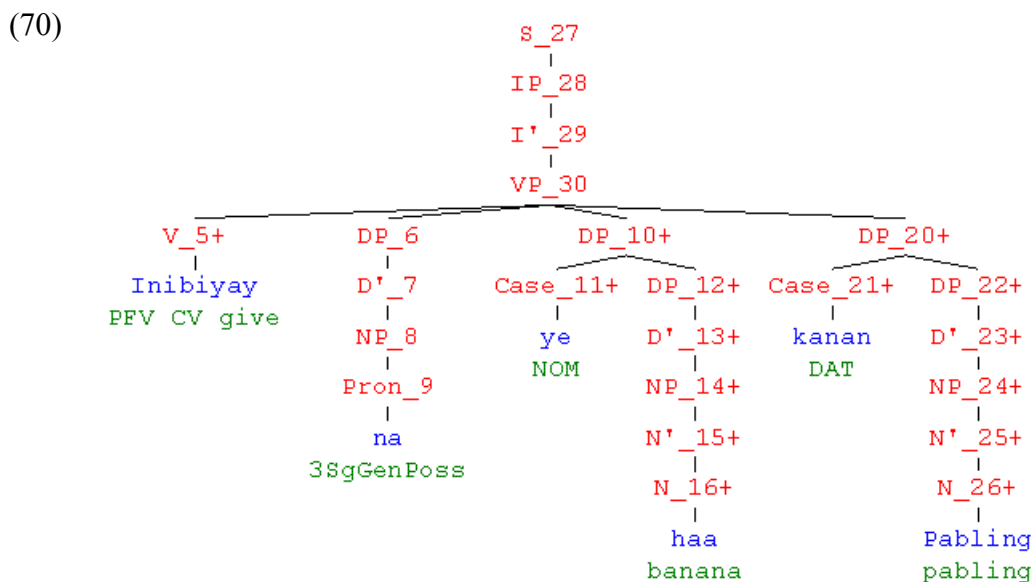
Adverb phrases can often occur in the first position, before the verb phrase as in (72).

(71) *Naapon   ha   mahilem, h<in><om>akit ye   bitoka   ko.*
yesterday DAT afternoon PFV-PV-sick   NOM stomach 1SG.GEN
'Yesterday afternoon my stomach became sick.'

(72)
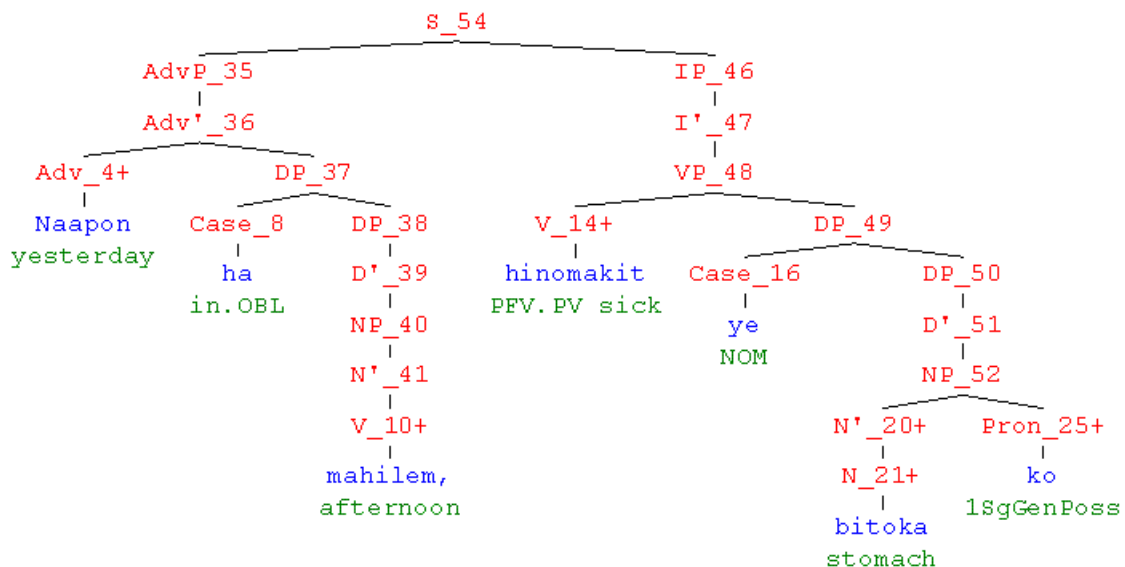
When standalone adverbs are in the first position, pronouns precede the verb and are adjoined to the adverb as in (74).

(73) *Tampol na   ya=n   in-i-byay   kanan Carmelita.*
quickly 3SG.GEN 3SG.NOM=COMP PFV-CV-give DAT   Carmelita.
'Quickly he gave it to Carmelita.'

(74)

```
                                          S_1
                                           |
                                          IP_2
                                           |
                                          I'_3
                    _____
                   I_4+                                      VP_12
          _____               _____
         I_5+              2P_10+         V_13            DP_14              IP_18
      _____              |              |               |                  |
    I_6+    2P_8+        Pron_11+       inibiyay          D'_15             I'_19
      |       |             |          PFV-CV-give          |                 |
    Adv_7+  Pron_9+        yan                            NP_16             VP_20
      |       |      3SgNom GEN.Lnk                         |                 |
   Tampol    na                                          Pron_17           DP_21+
   quickly  3SgGen                                         |                 |
                                                         kanan             D'_22+
                                                   3SgObl GEN.Lnk            |
                                                                          NP_23+
                                                                            |
                                                                          N'_24+
                                                                            |
                                                                          N_25+
                                                                            |
                                                                         Carmelita
                                                                         Carmelita
```
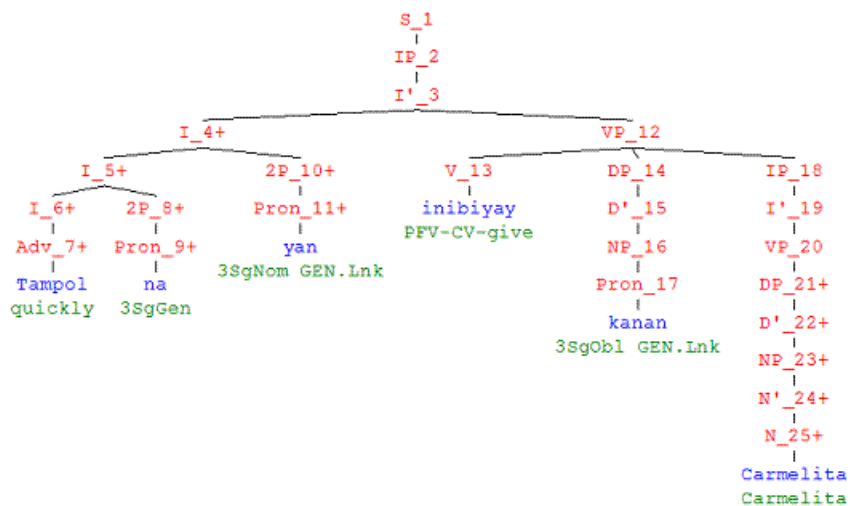
7.2 Ayta Abellen syntactical parsing challenges

This section examines some of the syntax characteristics of Ayta Abellen and other Philippine languages that need special consideration for parsing. I examine case and grammatical relations in 7.2.1, second position pronouns and clitics in 7.2.2, and topicalization in 7.2.3.

7.2.1 Case and grammatical relations

PAWS queries the user about the case system of the language. Linguists have come to varying conclusions about whether Philippine-type languages are more nominative-accusative or ergative-absolutive with regard to case system and so this question is not simple to answer for Ayta Abellen.

PAWS asks the user to select between four options regarding case: (1) no case marking, (2) nominative-accusative, (3) ergative-absolutive, or (4) split ergativity. Ergativity

has been a hot topic in Philippine linguistics in recent years, as has been discussion over what should really be labeled the subject. Some, including Brainard (1994), Gault (1999), and Payne (1982) have called for an ergative analysis for Philippine languages while others including Kroeger (1993) and Barlaan (1999) have concluded that nominative-accusative is more appropriate and that the nominative marked argument is the subject of the sentence.

PAWS also uses features to identify the grammatical relations of the sentence. The analysis of grammatical relations in Philippine-type languages is also debated, specifically, the identity of the subject. Kroeger (1993) has summarized the proposed answers as:

(75)   1. Nominative case marks the grammatical subject as in Bloomfield (1917), Blake (1925), and Bell (1976).
       2. Tagalog has no subject; nominative case marks the Topic as in Schachter (1976).
       3. The actor is always the subject and nominative case marks either the topic as in Carrier-Duncan (1985) or the absolutive argument as in Gerdts (1988) and Payne (1982).

The default grammatical relations values in PAWS are subject (SUBJ), direct object (DO) and indirect object (IO) with the actor assumed to be the subject.[2]

One option was to select nominative-accusative and use the default grammatical relations labels in PAWS, assuming the semantic role of actor to be the subject. This, however, does not enable the parser to correctly identify the grammatical relations for each NP because the case markers don't have a one-to-one correspondence with these grammatical

---

[2]It is beyond the scope of this thesis to define what is the subject for Ayta Abellen syntax. It is also beyond the scope of this thesis to define the function of the grammatical case markers. I have chosen to use nominative, genitive, and dative labels for the glosses because of the necessity for consistency throughout this document.

relations. For example, in (76) *ye* marks what PAWS would call the subject, in (77) it marks

the direct object and in (78) it marks the indirect object.

(76)   *M-am-Ø-yay*   *ye*   *laki* *nin* *haa*   *kangko*
     CTPLT-PL-AV-give NOM man GEN banana DAT.1SG
     'The man will give a banana to me.'

(77)   *K<ing>wa-Ø* *nin*   *bake*   *ye*   *haa.*
     PFV-get-OV   GEN monkey NOM banana
     'The monkey took the banana.'

(78)   *Ø-Biy-an*   *na*   *kawo=n*   *katanaan.*
     CTPLT-give-GV 3SG.GEN 2PL.NOM=GEN peace
     'You will be given peace by him.'

The default grammatical relations defined in PAWS and the notion that the actor is

subject was, however, modeled by marking voice as a feature on the verbal affixes in the

lexicon and by developer Cheryl Black making adjustments to the default rules to account

for voice. The original PAWS rule for a transitive verb sentence in a VSO language is seen

in (79) where the subject (DP) is marked with nominative case and the object (DP_1) is

marked with accusative case. In the three adjusted rules as seen in part in (80)-(82), the type

of voice marked on the verb is used to determine the grammatical relation of DP, DP_1,

and DP_2, whether subject, object, or indirect object.

(79)   *rule {VP option 5e - VSO order, transitive (accusative case object)}*
     VP = V DP DP_1
     <VP head> = <V head>
     <V head subject> = <DP>
     <V head object> = <DP_1>
     <DP head case> = nominative
     <DP_1 head case> = accusative

(80) *rule {VP option 5eAV}*
VP = V DP DP_1
<VP head> = <V head>
<V head subject> = <DP>
<V head object> = <DP_1>
<V head type voice> = actor
<DP head case> = nominative
<DP_1 head case> = genitive

(81) *rule {VP option 7ePronOV}*
VP = V DP DP_1 DP_2
<VP head> = <V head>
<V head subject> = <DP>
<V head object> = <DP_1>
<V head indirectobject> = <DP_2>
<V head type voice> = object
<DP head case> = genitive
<DP_1 head case> = nominative
<DP_2 head case> = dative

(82) *rule {VP option 7eIOGV}*
VP = V DP DP_1 DP_2
<VP head> = <V head>
<V head subject> = <DP>
<V head indirectobject> = <DP_1>
<V head object> = <DP_2>
<V head type voice> = goal
<DP head case> = genitive
<DP_1 head case> = nominative
<DP_2 head case> = genitive

A second option is to choose the ergative-absolutive option in PAWS and mark the lexicon accordingly. This option works well for non-actor voice transitive sentences as in (83) where the ergative marked NP is the subject and the absolutive marked NP is the object. It does not work, however, for antipassive actor voice sentences as in (84) where the ergative marked NP should be oblique and the absolutive marked NP should be the subject. With a

customized rule written by the PAWS developer to handle this special case as partially

shown in (85), the original grammatical relations labels of PAWS (SUBJ, DO, IO) can be

retained. A set of phrase structure rules and adjusted lexicon has also been developed

following this model.

(83) *H<in>abi-Ø nin laki ye balita kanan Emilio.*
PFV-said-OV ERG man ABS news OBL Emilio
'The man told the news to Emilio.'

(84) *M-am-Ø-yay ya=n haa kangko.*
CTPLT-PL-AV-give 3SG.ABS=ERG banana OBL.1SG
'He will give a banana to me.'

(85) *rule {VP option 7f - VSO order, ditransitive with DP (absolutive object), IO,DO order}*
VP = V DP DP_1 DP_2
<V head subject> = <DP>
<V head object> = <DP_1>
<V head indirectobject> = <DP_2>
<DP head case> = absolutive
<DP_1 head case> = ergative
<DP_2 head case> = oblique

A third option is to select nominative-accusative in PAWS and adjust the grammatical

relations feature labels to follow LFG (Kroeger 2004) by allowing the nominative marked

argument to always be the subject and labeling other arguments OBJ, OBL, and OBJ2 (when

there is more than one object NP). Here again, there is not a one-to-one correspondence of

case markers to grammatical relations as can be seen below where the genitive marked

argument *laki* 'man' should be labeled OBJ in (86) but in (87) there are two genitive marked

arguments, the second of which should be labeled OBJ2.

(86)   *Ø-I-byay*   *nin*   *laki* *ye*   *holat* *kanan* *Emilio.*
     CTPLT-CV-give GEN man NOM letter DAT   Emilio
     'The man will give the letter to Emilio.'

(87)   *Ø-Biy-an*   *na*   *kawo=n*   *katanaan.*
     CTPLT-give-GV 3SG.GEN 2PL.NOM=GEN peace
     'He will give you peace.'

So, this model also requires customized rules to be able to correctly apply the grammatical relations labels as features to the constituents. Using the same grammar file generated for solution 1 above, I was easily able to adjust it to use these grammatical relations labels. Here again the rules enabled the parser to use a voice marking feature on the verb to identify the grammatical relations of the NP constituents of the sentence.

In summary, the variety of opinions regarding case and grammatical relations provided motivation to try three different ways of modeling this: (1) nominative-accustive with PAWS default grammatical relations labels (2) ergative-absolutive with PAWS default grammatical relations labels, and (3) nominative-accusative with LFG labels. The first approach was the default used for testing the parser but the system can be switched to use either of the other two grammar models. The issues discussed here are only related to labels and features; the results of the research discussed below is not dependent on any of these three models. Identical parsing percentages should result from using each of the grammar files.

7.2.2 Pronouns and other second position clitics

Ayta Abellen, like Tagalog, has pronoun and particle clitics that occur in what is
called the second position of the sentence, meaning that they usually attach to a sentence
initial verb or adverb. In reference to this phenomenon in Tagalog, Anderson (2008) has
said "Tagalog clitics pose problems of varying severity for an account based on purely
syntactic mechanisms." The second position is problematic to almost all theories of syntax
as the positions of the arguments can move based on whether they are full noun phrases or
pronouns. This can be seen in (88) and (89) where the direct object occurs after the actor
when the direct object is a full noun phrase but occurs before the actor when referenced
with a pronoun.

(88) *Ø-I-byay   na=n   Pabling ye   lanom kanan Emilio*
     CTPLT-CV-give 3SG.GEN=GEN Pabling NOM water DAT Emilio
     'Pabling will give the water to Emilio.'

(89) *Ø-I-byay   na   ya=n   Pabling kanan Emilio.*
     CTPLT-Ø-give 3SG.GEN 3SG.NOM=GEN Pabling DAT Emilio
     'Pabling will give it to Emilio.'

Normally, noun phrase arguments occur after the verb, but with a sentence-initial
adverb, pronominal arguments occur in the second position, in front of the verb as in (90).
If there are two pronouns with a sentence-initial adverb, both occur before the verb as in
(91).

(90) *Tampol na=n   Ø-i-byay   ye   lanom kanan Emilio.*
     quickly 3SG.GEN=COMP CTPLT-CV-give NOM water DAT Emilio
     'Quickly he will give the water to Emilio.'

(91) *Tampol na   ya=n   Ø-i-byay   kanan Emilio*
     quickly 3SG.GEN 3SG.NOM=COMP CTPLT-CV-give DAT   Emilio
     'Quickly he will give it to Emilio.'

This use of the second position in the sentence is not limited to pronouns. Some

clitics or particles also occur in second position. When both particles and a pronoun are

present in the same sentence, the pronoun always precedes the particle as in (92).[3] In the

case of multiple pronouns, both will precede the particle as in (93).

(92) *Tampol na   dayi=n   Ø-i-byay   ye   lanom kanan Emilio.*
     quickly 3SG-GEN OPT=COMP CTPLT-CV-give NOM water DAT   Emilio
     'Quickly he should give the water to Emilio.'

(93) *Tampol na   ya   dayi=n   Ø-i-byay   kanan Emilio.*
     quickly 3SG-GEN 3SG-NOM OPT=COMP CTPLT-CV-give DAT   Emilio
     'Quickly he should give it to Emilio.'

This second position is problematic for syntax parsing and many just choose to call

it a special case not handled by current syntactic theories. Since the goal of my research is

not to make general statements about the underlying syntax structures but rather to determine

what parts of the language could be parsed, custom rules were created (see 7.4) to handle

these preverbal possibilities. Because of the frequency of pronouns and particles in the data

this seemed a better solution than regularizing all the data to full noun phrases and eliminating

all particles.

---

[3]This is in contrast to Tagalog where some particles precede the pronoun and some particles follow pronouns.

7.2.3 Topicalization

Left dislocated topic phrases are marked with *hiyay* or demonstratives. Topicalization

occurs frequently as can be seen in Stone (2008) where a twenty-sentence natural text had

seven instances of topicalization. Because the topic is marked explictly with a separate word

and the position of the topic phrase is always before the verb and bounded by a comma,

this phenomena in Ayta Abellen is relatively easy to model using custom rules that I describe

in section 7.4. A sample topicalized sentence with its parse is shown in (94).

(94)   *Hiyay   Pabling, n-Ø-akew   ya   ha   banwa.*
         TopicM Pabling   PFV-AV-go 3SG.NOM DAT town
         'As for Pabling, he went to town.'
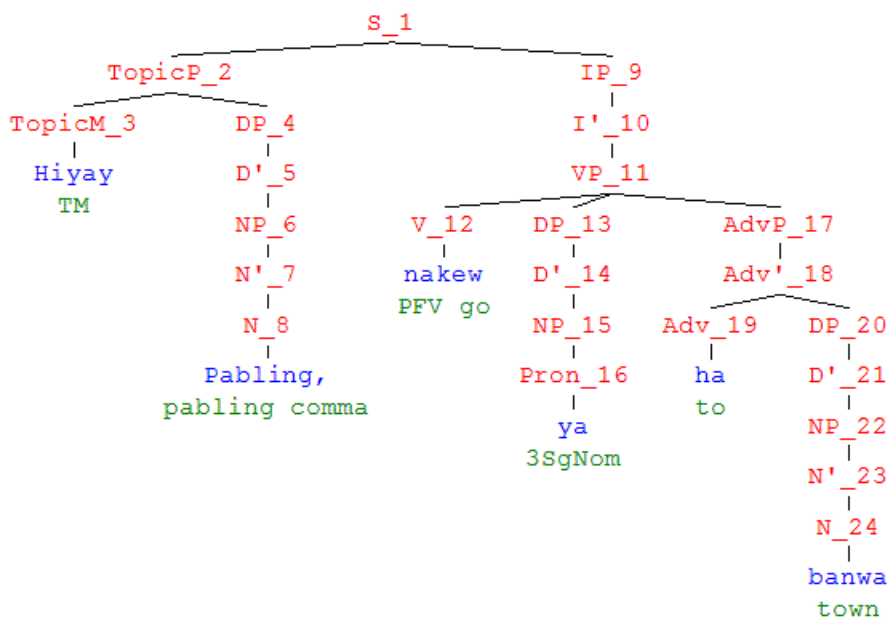
7.3 Evaluating the parser on a training set

One of PAWS' outputs is phrase structure rules that can be used with the PC-PATR syntactic parser. The PAWS Starter Kit says that "those characteristics of the language which are general, usually about 3/4 of the rules, can be modeled via this starter kit." As confirmed in personal communication with the developer, the intent of this statement is that 3/4 of the rules needed to model the syntax of a given language will be automatically generated by the PAWS Starter Kit when the user has finished providing input and verifying it to be correct through testing on training data. To operationalize a test of this claim, I first used a training set of data and counted how many rules were generated automatically versus how many custom rules were needed to make 100% of the training data parse correctly. In order to test how well the parser really works, I then ran it on natural texts to find what percentage of the sentences were parsed (see 7.5).

Since the original development of the PAWS Starter Kit was done without access to Philippine or Austronesian language data, it would be surprising if the original PAWS Starter Kit would be able to generate three-fourths of all needed rules. The voice system of Philippine languages is an integral part of not only the verb, but also the syntax of the whole clause. Through interaction with the developer, the case constraints based on voice were substituted for the normal nominative-accusative constraints as mentioned in section 7.2.1. The test results that I present are not based on the original version of PAWS that was

available at the beginning period of my research, but rather on a newer version that includes

some bug fixes and a few general rule enhancements that are believed to apply to all

languages which the Ayta Abellen data highlighted.

It is hard to accurately test the claim that 75 percent of the needed rules for a given

language are automatically generated by PAWS without actually having already written

rules to cover every potential structure in the language. A method of approximating an

answer to this question, however, is to take the total number of rules generated by PAWS

and divide this by the number of PAWS generated rules plus the number of custom rules.

These custom rules were written by Cheri Black as she worked with me to account for all

the structures contained in a forty-sentence training set. This training set included all the

example data requested by PAWS from the user and a few other example sentences that

were added to make the training set more comprehensive.

PAWS automatically generated 121 rules for Ayta Abellen and it was necessary to

add 28 custom rules[4] in order to get a correct parse for 100% of the sentences in the training

set. This means that 81% of the rules were auto-generated by PAWS. A possible inflation

to this percentage results from my not having manually tested to see if every one of the 121

rules generated by PAWS is used in the parsing of the training set. Without extensive

research into which rules are used for each parse, it is hard to know if there are any extraneous

---

[4]This number refers only to phrase structure rules and not to the features contained in the rules. Thus, the customizations made for Philippine voice were not included in the 28 custom rules because the customizations were necessary to get the grammatical relations features right in the parse tree and were not needed in order to make the sentence parse.

rules in the 121 rule set generated by PAWS. But still it does seem significant that for the 40 sentences, only 28 custom rules were needed. One bit of preprocessing that I did for this training set and the test set was to verify the accuracy of the morphological parser for each sentence to make sure that accurate word parses were being supplied to PC-PATR so that only syntax parsing errors would show up in the results data.

7.4 Structures covered in custom rules

In 7.3 it was noted that 28 custom rules were created to handle language specific structures that were not handled with the original rules generated by PAWS. Of those 28 rules, 11 were related to accounting for linkers, 9 for second position clitics, 6 for topic constructions, 1 for adjoining the case markers to DP, and 1 for handling passive/stative constructions.

One of the first discoveries of a feature not covered in the original set of rules generated by PAWS was the *a* linker which connects a variety of constructions such as modifiers to nouns, degree words to modifiers, modifiers to modifiers, etc. This linker is not optional and the lack of an account for it in the phrase structure rules blocked all adjective phrases from being parsed. In all, eleven custom rules were created and appended to the end of the grammar file to account for the linker. Two of the rules used for linking adjectives to nouns can be seen in (95).

(95) a. *N'_1 = AdjP Linker N'_2*
     b. *N'_1 = N'_2 Linker AdjP*

Custom rules were added to handle the second position pronouns and clitics described previously in section 7.2.2. The basic approach taken in these nine rules is to model manner adverbs and negative auxiliaries as filling the head position of the inflection phrase with the pronouns and clitics being right adjoined to this head. When there is no adverb or negative auxiliary present, the pronouns and clitics adjoin to the verb.

Five custom rules were added to handle various topic constructions. The first rule added the specification that a demonstrative with a linker could also be a topic marker as in (96a). Various nominalized elements that could occur inside a topic phrase were added such as a VP in (96b) or an IP as in (96c) or conjoined topics as in (96d). Equative sentences with a topic marker were handled with a rule to allow the topic phrase to be equated with the VP as in (96e).

(96)   a.   *TopicM = Dem Linker*
       b.   *TopicP = (Deg) TopicM VP*
       c.   *TopicP = TopicM IP*
       d.   *TopicP = TopicP_1 (DP) Conj TopicP_2*
       e.   *VP = TopicP*

Custom rules were also written for case markers, as in (97). The case markers could have been analyzed as prepositions resulting in PP arguments. But since the case markers are really proclitics, it is more optimal to create a custom rule that will adjoin the case marker to DPs rather than the default PP.[5] Another rule as in (98) was added to accomplish the same

---

[5]This custom rule will be added to a future version of PAWS.

for nominalized VPs with a case marker, that they would be parsed as VPs also. This enables

the parser to handle sentences such as in (99).

(97) *rule {DP option Case - case marker on DP}*
DP = Case DP_1
<DP head> = <DP_1 head>
<DP option> = Case

(98) *rule {DP option Case - case marker on VP}*
VP = Case VP_1
<VP head> = <VP_1 head>
<VP_1 head type verbheaded> = +
<VP head type case-marked> < = +
<VP option> = Case

(99)

```
                        S_1
            _____|_____
        TopicP_2                  IP_8
           |                       |
        DP_3+                    I'_9
       ____|____                   |
    D'_4+      Deg_7+            VP_10
      |          |          _____|_____
    NP_5+      bengat    Case_11        VP_12
      |         just        |        ____|____
    Dem_6+                   ye     V_13     DP_14
      |                     NOM      |        |
     Yati                          tanda    D'_15
     this                          know      |
                                            NP_16
                                             |
                                           Pron_17
                                             |
                                             ko
                                           1SgGen
```
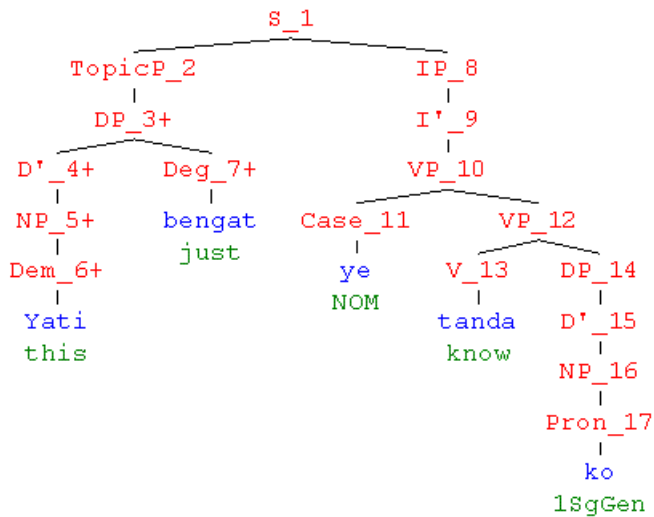
(100) *Yati bengat ye   tanda ko.*
      this   only   NOM   know   1SG.GEN
      'This only is what I know.'

A custom rule was added for stative verbs so that they will not be parsed as actor

voice. Since stative verbs can take an argument structure similar to that of object voice, it

was easiest for the custom rule to link these sentences to the object voice rule. Thus, the

custom rule in (101) defines the VP_1 voice type as object.

(101) *rule {VP option 3cEXTVP - VSO order, stative(=object_voice) VP comp}*
        VP = V VP_1
        <VP head> = <V head>
        <V head type transitive> = -
        <V head type copular> = +
        <V head type motion> = -
        <VP head type pro-drop> = +
        <V head type passive> = -
        <VP_1 head type voice> = object
        <VP_1 head infl aspect stative> =+

As was mentioned previously, the Ayta Abellen language uses pauses in natural

speech to make preverbal phrase boundaries explicit. This was symbolized in the input text

with commas so a feature "comma" was added to the word that comes before the comma.

The comma feature was then used to constrain other rules in the grammar so that the syntactic

parser could recognize phrase boundaries. So, while this was not done with a separate custom

phrase structure rule, it was added as a custom constraint to several existing rules.

7.5 Evaluating the parser on natural text

Testing on the training set substantiates the claim that PAWS generates 75 percent

of the needed rules. However, the validity of the claim is limited to the specific set of

sentences chosen for the training set. Do we still have 75 percent of the needed rules if we

test the parser against any set of sentences?

To investigate this question, I chose two native authored texts which were different than the forty-sentence training set. One was a 21 sentence narrative about a hunting experience told by a man from the northern part of the language area. The other was a 28 sentence narrative told by a man in the central part of the language area where he recounts the traditional story of the Tower of Babel as he heard it from his father. So, a total of 49 sentences were used.

As I began testing the syntactic parser with natural texts, I found that there were phrase structure rules generated to address certain constructions, but the sentences containing those constructions were not parsing for reasons other than failings in the basic syntactic description. Two kinds of adjustments needed to be made to regularize the data: (1) changes related to improving the orthography of the transcribed text and (2) changes related to regularizing forms that would otherwise be extremely difficult for any syntactic parser to account for.

An example of the first kind of adjustment was the adding of commas at phrase boundaries marked phonologically with a pause (see 7.2.3). This improved the writing style of the original document. Because of the significance to the grammar of these pauses (commas), a special custom rule was also written by the developer of PAWS to enable the parser to use this very important clue about phrase boundaries. This notion of writing grammar rules that take note of commas is not original, as Butt and others (1999) mention this as being part of the ParGram parsing system.

Spelling changes were a second kind of adjustment that helped improve the text. Originally the word *kananyati* 'this' and similar demonstratives had been written as two separate words *kanan* 'DAT' *yati* 'this' because in many contexts each are standalone words. But looking at syntactic structure revealed that when together, they really have one grammatical function and one semantic meaning. Phonologically they are also pronounced as one word, so this adjustment is a refinement to the orthography that came as a result of syntax study.

There was one place in the natural texts where the speaker added material to explain the meaning of a word for the sake of the audience. This addition was semantically and syntactically redundant. Removing the phrase *o kaginta a maambal* 'or big python' in (102) did not affect the grammaticality or meaning of the sentence and made the sentence better conform to a writtten style.

(102) *Ket yabayin n-a-patey ya=y maambal*
    And this   PFV-STA.OV-kill 3SG.NOM=NOM python

    a malake o kaginta a maambal.
    LNK big   or big   LNK python
    'The big python was killed.'

Other adjustments were made in order to regularize forms for syntactic parsing. For example, fronted temporal phrases are normally marked with *ha* 'DAT'. At times native speakers leave this out for naturalness reasons even though the full form is grammatically correct. The parser needed to see the full form in order to be able to correctly identify the

phrase type. So, there were instances where *ha* was added to the beginning of the phrase to make the phrase complete.

The quote margin *wana* 'he said' can occur both quote initial and quote medial as in (103). Ayta Abellen speakers frequently use the quote margin for the discourse function of reminding the listener who is talking and quote medial quote margins are not predictable. Because of its random position, several rules would be needed to try to account for its location inside a quotation. For this reason quote medial margins were removed and the quotation was split off into another sentence as in (104).

(103) a. *Hiyay   habi nin   tatang ko,   wana, "Ha   hatew,*
           TopicM word GEN father 1SG.GEN spoke DAT past

           anak ko,"   wana, "ha   onan panaon,
           child 1SG.GEN spoke DAT first   time

           labay la=n   Ø-abot-en   ti   langit nin   tatao."
           want   3PL.GEN=COMP CTPLT-reach-OV NOM heaven GEN people
           'The word of my father, he spoke, "In time past, my child," he said, "at the earliest time, the people wanted to reach heaven.'

(104) a. *Hiyay   habi nin   tatang ko,   wana. "Ha   hatew,*
           TopicM word GEN father 1SG.GEN spoke DAT past

           ha   onan panaon, abay la=n   Ø-abot-en
           DAT first   time   want   3PL.GEN=COMP CTPLT-reach-OV

           ti   langit nin   tatao."
           NOM heaven GEN people
           'The word of my father, he spoke. "In time past, my child, at the earliest time, the people wanted to reach heaven."'

Ayta Abellen has a particle *kano* that can roughly be translated 'it is said'. It normally occurs in the second position of the clause but like Walrod (1979) noted about the use of

kano in Ga'dang, its distribution can best be analyzed from a discourse perspective rather than syntax. The frequency of its use can be related to peak development in both Gadang and Ayta Abellen (Stone 2007). For this reason, it can sometimes occur at the end of a sentence. Because of this irregular positioning, the particle *kano* was deleted from the source texts.

Even with the adjustments described above, there was still one phrase in (103) that would be difficult to parse and needs further adjustment, namely the vocative *anak ko* 'my child'. Because a vocative like this could be placed almost anywhere in the sentence, it was decided to not attempt coverage for vocatives in the syntactic parser and it was removed from the source text.

I ran four different tests on the PAWS generated rules. The first test used the original rules from PAWS with the original, unadjusted source text. The second test used the original source text but used the 28 custom rules added by the developer. The third and fourth tests used the adjusted source text. Predictably the test run with the adjusted source text and including the custom rules produced the highest percentage of sentences parsed, 81%. This result shows that the rules originally generated by PAWS were able to parse 65% of the sentences in a text sample that was independent of the original training examples. All the results can be seen in (105).

(105)

	Original source text	Adjusted source text
Original Rules	38%	65%
Custom Rules (28)	60%	81%
Custom Rules (28+6)	69%	100

After documenting these results, I analyzed the reasons 19% of the sentences still were not parsing. I submitted this analysis to the developer of PAWS who wrote 6 more rules[6] to get all of the remaining sentences to parse. So, while the 34 custom rules raised the percentage with adjusted source text to 100% the addition of the rules raised the percentage with the unadjusted source text from 60 to 69%.

In order to further test the parser, I used a 64 sentence section of a narrative of a man describing how he met his wife. This was a different genre of text, being mostly dialogue in contrast to the monologue in the previous test set. The parsing results can be seen in (106), where all 34 custom rules were used. In spite of being a different genre of narrative, the parser still accounted for 76% of the sentences with the adjusted source text.[7]

(106)

	Original source text	Adjusted source text
Original Rules	24%	41%
Custom Rules (34)	61%	76%

Thus, it can be said that for oral narrative texts which have been adjusted in the ways listed above, the parser with the customized rules currently is able to parse around 75-80%

---

[6]These rules covered the following: (1) Verb functioning as adjective, preceding noun. (2) Verb functioning as adjective, following noun. (3) Verb functioning as noun. (4) Stative verb with genitive marked actor. (5) Time phrase following the noun of the main clause. (6) Copular sentence with multiple arguments.

[7]There were two sentence patterns which occurred repeatedly in the dialogue which were new to the parser. If there had been rules for these two sentence patterns, the total would have been 85%.

of sentences in narrative texts. These percentages can be expected to improve as further adjustments to the grammar are made and more testing is done.

It must be mentioned that these kinds of results are significant for Philippine languages, since previous research by others using a rule-based approach for Philippine languages have not been tested on natural text. In Giganto (2004), while PinoyMMT had a 96% accuracy rate for Cebuano sentences, this was achieved on a dataset of 16 selected base sentences from which 39 additional sentences were derived by reordering NPs of the original 16 sentences. Manguilimotan (2007) developed an LFG model for Tausug but the dataset "covered only simple Tausug verbal sentences." Borra and others (2007) also used LFG in designing "hand-crafted rules" for 12 Tagalog sentences. In each case, testing was done on selected syntactic structures and rules were written specifically to match those structures. While the results with PAWS for Ayta Abellen are not as high, they were done on natural texts. The custom rules were developed to handle the data in the training set, without looking at the kinds of structures present in the natural texts of the test set.

7.6 Evaluating the sketch

The syntax sketch generated by PAWS is shown in Appendix E. Like the process for generating a morphology sketch in FLEx, PAWS generates a syntax sketch in XML format that can be further edited and rendered into other formats such as PDF. The XML sketch is an example of actionable data that comes from a structured knowledge source.

The XML document is automatically created based on an expert system template. The document itself can also be edited using an XML editor and easily output to cross-platform formats like html and pdf. For these reasons, the syntax sketch is an example of the Third Wave Publishing paradigm (see 6.5).

The sketch covers the basic elements of syntax, being the most extensive for areas essential to writing phrase structure rules. Vernacular examples were automatically included in the sketch in their appropriate sections which made for simple final editing of the document.

The original sketch was not very detailed regarding case marking systems and the original had no mention of voice. I added an extra section to the Ayta Abellen syntax sketch to cover more completely these features which are so integral in Philippine-type languages. The original list of verbal inflection features was also incomplete for Ayta Abellen. This was another area where I edited the final document. The developer is planning to incorporate features from the Ayta Abellen sketch into the next version of PAWS.

By virtue of being a standardized outline, the automatically generated sketch contained sections about features not found in Ayta Abellen. The sketch correctly said that the features do not exist but in the interest of making the sketch better fit the language being described, some of those sections were simply removed from the final sketch.

The explanations in PAWS are given in English and all the examples are in English. Analysis of Ayta Abellen structures at times was difficult with only English examples to

consider. This was a factor in several of my initial analyses being wrong. For this reason, in addition to the Ayta Abellen examples, I have added examples from Tagalog for each section in order to make the sketch and the PAWS tool more helpful for linguists already familiar with Tagalog.

Overall, the automated sketch is a helpful starting point for a description of the syntax of Ayta Abellen. The use of this structured information template enabled me to spend the bulk of my time on analyzing the data and very little time on formatting issues. The ability to test the analysis on actual data (by using PC-PATR with the automatically generated PSRs) made the resulting sketch more accurate.

Chapter 8

THE STATUS OF LANGUAGE DOCUMENTATION FOR THE AYTA ABELLEN LANGUAGE

The Ayta Abellen language is an endangered and underdocumented language. This current research will add to the various forms of language documentation already in existence. It would be helpful to examine the complete list of materials produced for this language through the language development efforts of SIL before considering what forms of documentation are still lacking. The first three items in (107) are what Himmelmann (1998) would call "documentary linguistics" while the rest would be in the category of "descriptive linguistics".

(107)

Num	Category	Description	Extent	Media	Format	Creator
26	Stories	Recording	174 minutes	audio	WAV	Nitsch, Stone
26	Story Transcripts	From Recordings	7531 words	data	FLEx/SFM	Nitsch, Stone
17	Stories	Written	2032 words	data	Word	Espinosa
1	Ethnography	Documentary	22 min	video	DVD	Curtis, Stone
10	Lessons	Language Learning	38 pages	data	Word	Nitsch

Num	Category	Description	Extent	Media	Format	Creator
1	Lexicon	Dictionary	3,548 roots	data	FLEx	Nitsch, Stone
1	Parser	Morphology	3,548 roots	data	XAMPLE	Stone
1	Parser	Syntax	156 rules	data	PC-PATR	Stone
1	Linguistic Sketch	Phonology	28 pages	data	Word	Nitsch
1	Linguistic Sketch	Morphology	29 pages	data	XML, HTML, PDF	Stone
1	Linguistic Sketch	Syntax	77 pages	data	XML, HTML, PDF	Stone
1	Linguistic Sketch	Grammar	29 pages	data	Word	Nitsch
1	Linguistic Sketch	Discourse	20 pages	data	Word	Stone
1	Linguistic Sketch	Morphophonemics	4 pages	data	Word	Nitsch
10	Interlinear	Texts	505 sentences	data	XML, HTML, PDF	Stone

In thinking of the various domains of linguistics, there are unpublished manuscripts on phonology, morphophonemics, and discourse. While no published papers can be found on the list for phonetics or semantics, it can be argued that there is actionable data[1] for each of those domains. For example, while no phonetics paper has been published, primary language data in the form of audio recordings exist that researchers could use with tools

---

[1]While the current corpus of primary data is limited, a documentation project will begin in the second half of 2010 that will aim to record oral stories and produce a corpus of transcribed text containing over 300,000 words.

such as Speech Analyzer[2] to study various sounds. No semantics paper is available, but information on semantics can be found in the lexicon and interlinearized texts.

So while the lack of published books describing the grammar of Ayta Abellen would seem to indicate a significant need for language documentation, it can be argued that increasing availability of primary language data in combination with an automated morphology and syntax parsing system is equally as significant because the field of potential researchers is much wider than just the field linguist who gathered the data. Others can take existing text data and supply it as input into the parsers or gather their own language data to use with the parsing tools. New gathered texts can be automatically annotated for part of speech and the syntax of simple sentences can be studied with the PC-PATR parser.

---

[2]Available online at: http://www.sil.org/computing/sa/sa_download.htm.

Chapter 9

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

A part of this research was developing a morphological parsing system that can parse virtually all the words currently occurring in the written text corpus of the Ayta Abellen minority language. This morphological parsing system currently parses 99.8% of wordforms in Ayta Abellen committee-reviewed written texts.

The second part of this research has been the development of a syntax parsing system for Ayta Abellen. This system can be applied to new texts. Initially, unadjusted narrative texts have a parsing rate of approximately 60%. When certain surface adjustments are made to the input text, we would expect the parser to generate a reasonable parse for 75-80% of the sentences in a test set.

The analytical choices made in PAWS for Ayta Abellen could have benefit for other linguists working to document or develop syntax parsing systems for other Philippine languages. The syntax sketch output by PAWS in Appendix E has been edited to include Tagalog examples which reveal very significant similarities between Tagalog and Ayta Abellen syntax. Because the phrase structure rules generated by PAWS are a direct reflection

of the information in the syntax sketch, it can be hypothesized that the Ayta Abellen phrase structure rules have much in common with those that would describe Tagalog and many other Philippine language.

FLEx and PAWS are similar in that they both produce multiple kinds of data that can be used in various ways. FLEx's information store is at the morpheme and word level while PAWS is at the sentence level. Both use templates to automatically generate structured sketches of linguistic data while also generating data necessary for automated parsing.

This type of linguistic documentation is different than other methods not only in the way documents are created but also in the way the information is intended to be used. Because of these differences, the pros and cons of this new kind of approach must be evaluated. What does this type of approach do better than a traditional grammar paper and where does it fall short?

First, a primary advantage of the automated template style of documentation is that the grammatical information is actionable. The formal parsing rules are actionable in that they can be used for analyzing new text data. The linguistic sketches are also actionable by virtue of being expressed in an XML format which can be processed in various ways to be rendered in multiple presentation formats.

A second advantage for researchers is that this style of documentation is verifiable. Researchers can actually check how words are formed by using the XAMPLE parser themselves. Researchers can check how sentences are structured by using the syntactic

parser. Traditional grammar papers are presented as a full accounting of what is happening with the language but since all grammars leak, the problem for the reader is in determining how much of the language is covered with the description in the paper. In contrast, any statement made in the FLEx morphology sketch or the PAWS syntax sketch can be tested empirically with the parser using primary source language data.

A third advantage is that this type of documentation is updatable. A traditional published grammar paper is very rarely updated. A linguist might change his or her analysis and publish it in a different article but the original document is not updated. But this type of documentation, especially when used in conjunction with the internet, can be easily updated. For example, the discovery of a new derivational affix would be a simple matter to update in FLEx with the automatic regeneration of the morphology sketch. In the same way, the discovery of sentences that disprove a previous conclusion about the syntax can lead to changing answers in the PAWS questionnaire which will in turn automatically regenerate the phrase structure rules.

A fourth advantage is that it is interactive. Because the data can be used in conjunction with parsers, research can be interactive. A researcher studying Ayta Abellen can make hypotheses about the grammar and then test them on a data corpus using existing tools. With a traditional grammar sketch, the researcher is left to use the data the writer of the grammar paper chose to include as examples, rather than interacting with the whole text corpus.

A fifth advantage is that it is measurable. With this kind of approach it is possible to run tests and determine what percentage of the data are covered with the existing parsers by running measured quantities of data through the parser and counting the rate of success. A researcher reading a traditional grammar paper has no way to determine what percentage of the language is covered by the description put forth by the author of the grammar paper.

A sixth advantage is that this can be an annotation tool. The XAMPLE and PC-PATR parsers can function as automated annotaters of new texts. When the two parsers are used together they show information from the morpheme level up to the sentence level. A grammar paper can also discuss phenomena on these levels but the paper cannot be used to annotate new language data.

As far as disadvantages of this approach, the first is that it focuses more on surface structure than on semantic structure. Most traditional grammar papers for Philippine languages focus on the semantics of affixes and the syntax issues related to case and voice. These topics are only lightly touched on in the current approach.

A second disadvantage, seen in the FLEx parser, is the difficulty in handling assimilation and deletion at a morpheme boundary, a phenomena common in Philippine languages. At present, FLEx supports only an item-and-arrangement approach to morphology. Thus, an allomorph must be added for each verb stem where this occurs. Fortunately, a rule-based item-and-process component has been released in FLEx version 3.0.

A third disadvantage is in the underlying theory used for the PAWS parser. While the tool is not totally tied to Government and Binding and X-Bar theory, it is heavily influenced by those theories. Some have said that Government and Binding is not optimal for Philippine-type languages. So, while the underlying theory chosen has not surfaced as a significant problem in this research, the unfamiliarity of many Philippine linguists with its use could be viewed as a negative.

The work on this thesis has revealed a number of avenues for further research and development. First, further development is needed for the morphological parser. The analysis using null derivation and three sets of affixes with two being inflectional and one being derivational should be checked with other languages in the Sambal subgroup. It may be that the use of this analysis could lead to the rapid development of morphological parsers for the other seven languages in the subgroup.

Second, it would be very helpful for the FLEx program to export the lexicon in a format that could be used by the XAMPLE parsing engine in a stand-alone mode. This would enable the parser to be used as a stand-alone parser for those without access to FLEx. I was able to make this work by finding the lexicon files stored in temporary files and modifying those files for the purpose. The process would ideally be more accessible to the average user.

Third, the morphological analysis needs to be extended to include an analysis of verb classes. A full analysis of verb classes will improve the quality of the lexicon and further refine the morphological analysis presented here.

Fourth, research is also needed with regard to the syntactic parser. The custom rules should be expanded to cover all the structures of the language so that any adjusted sentence could be parsed. Determining the total number of rules needed for Ayta Abellen would enable more accurate testing of the developer's claim that 75% of the rules needed for parsing are automatically generated by PAWS.

Fifth, it is hypothesized that most of the PAWS answers for Ayta Abellen would be the same for Tagalog and many other Philippine languages. This hypothesis should be tested through the use of the PAWS Starter Kit with another Philippine language such as Tagalog or Cebuano. Comparing the resulting phrase structure rules with those for Ayta Abellen would show the level of similarity between these languages.

Sixth, it would also be helpful to have a customized version of PAWS for Philippine languages. Since the underlying template files that drive PAWS are in XML, it would be possible to modify them and make a version specifically designed for Philippine or Austronesian languages where default values are set to match proto-Austronesian syntax.

Finally, the language data mentioned in chapter 8 should be organized and archived as a data corpus that could be made available online. The various linguistic sketches could be combined into one XML document to be rendered in multiple formats. The audio files

could be made available to go with their transcriptions while the parsers could be used with

any of the primary source language material.

Appendix A

ABBREVIATIONS

(108)

Abbreviation	Meaning
ADJ	Adjective
ADV	Adverb
APT	Aptative mode
ASSOC	Associative
AV	Actor voice
CAUS	Causative
COLL	Collective
COMP	Complementizer
CONT	Continuous aspect
CTPLT	Contemplated aspect
CV	Conveyance voice
D	Determiner
D'	Determiner'
DAT	Dative case
DEM	Demonstrative
DET	Detransitive
DIM	Diminuative
DIS	Distributive mode
DP	Determiner phrase
DUP	Reduplication
DUR	Durative mode
EMPH	Emphatic

Abbreviation	Meaning
EXT	Existential
EXCL	Exclusive
FLEx	FieldWorks Language Explorer
GEN	Genitive case
GER	Gerundizer
GV	Goal voice
I'	Inflection
INCL	Inclusive
INST	Instantaneous action
INV	Inversion
IP	Inflectional phrase
LNK	Linker
N	Noun
N'	Noun'
NLP	Natural language processing
NOM	Nominative case
NP	Noun Phrase
OPT	Optative mood
ORD NUM	Ordinal number
OV	Object voice
PA	Prolonged action
PFV	Perfective aspect
PL	Plural action mode
Pron	Pronoun
PV	Patient voice
REC	Reciprocal action mode
REQ	Social request mode
RES	Respect particle

Abbreviation	Meaning
RHET	Rhetorical question marker
SG	Singular
SOC	Social request mode
SPEC	Specifier
STA	Stative mode
SUP	Superlative
TM	Topic marker
VP	Verb phrase
VSO	Verb-Subject-Object

Appendix B

FLEX MORPHOLOGY SKETCH

B.1 Introduction

Ayta Abellen is a language spoken in the Philippines. This paper gives a preliminary sketch of Ayta Abellen morphology following a basic item-and-arrangement model. The sketch covers the following topics:

- Phonemes in section 2.

- Morpheme types in section 3.

- Word categories in section 4.

- Inflection in section 5.

- Derivation in section 6.

- Clitics in section 7.

- Morpho-syntactic feature system in section 8.

- Allomorphy in section 9.

- Natural classes in section 10.

- Residue in section 11.

There also are two appendices. The first, Appendix C, lists morphemes arranged by morpheme type. The second, Appendix D, lists morphemes arranged by lexical category.

B.2 Phonemes

Ayta Abellen has 20 phonemes as shown in the following table (the first column shows the orthographic representations):

Representation	Basic IPA Symbol	Name	Description
'	ʔ	ʔ	glottal stop
-	ʔ	ʔ	glottal stop
a	a	a	low central unrounded vowel
ā	aː	ā	lengthened open front unrounded vowel
b	b	b	voiced bilabial stop
d	d	d	voiced alveolar stop
e	e	e	close central unrounded vowel
g	g	g	voiced velar stop
h	h	h	glottal fricative
i	i	i	high front unrounded vowel
k	k	k	voiceless velar stop
l	l	l	alveolar lateral
m	m	m	bilabial nasal
n	n	n	alveolar nasal
ng	ŋ	ŋ	velar nasal
o	o	o	mid back rounded vowel
p	p	p	voiceless bilabial stop
t	t	t	voiceless alveolar stop
u	u	u	high back rounded vowel

Representation	Basic IPA Symbol	Name	Description
w	w	*w*	voiced labial-velar approximant
y	y	*y*	close front rounded vowel

B.3 Morpheme types

Words in this analysis of Ayta Abellen are formed from morphemes of 8 types. The following table lists the types along with a count of how many instances are in the lexicon. Appendix C lists some or all of these.

Count	Name	Description	Appendix
2	circumfix	A circumfix is an affix made up of two separate parts which surround and attach to a root or stem.	C.1
19	enclitic	An enclitic is a clitic that is phonologically joined at the end of a preceding word to form a single unit. Orthographically, it may attach to the preceding word.	C.2
4	infix	An infix is an affix that is inserted within a root or stem.	C.3
35	phrase	A phrase is a syntactic structure that consists of more than one word but lacks the subject-predicate organization of a clause.	C.4
51	prefix	A prefix is an affix that is joined before a root or stem.	C.5
3548	root	A root is the portion of a word that (i) is common to a set of derived or inflected forms, if any, when all affixes are removed, (ii) is not further analyzable into meaningful elements, being morphologically simple, and, (iii) carries the principle portion of meaning of the words in which it functions.	C.6
547	stem	"A stem is the root or roots of a word, together with any derivational affixes, to which inflectional affixes are added." (LinguaLinks Library). A stem "may consist solely of a single root morpheme (i.e. a 'simple' stem as in *man*), or of two root morphemes (e.g. a 'compound' stem, as in *blackbird*), or of a root morpheme plus a derivational affix (i.e. a 'complex' stem, as in manly, unmanly, manliness). All have in common the notion that it is to the stem that inflectional affixes are attached" (Crystal 1997).	C.7
2	suffix	A suffix is an affix that is attached to the end of a root or stem.	C.8

B.4 Word categories

In this analysis of Ayta Abellen there are 11 major syntactic categories for words. Some of these in turn have subcategories. The following is a complete list of the categories and subcategories that are posited (along with a count of how many instances of each are found in the lexicon; some or all of these are in the appendix).

- Adjective (408)

- Adverb (122)

- Connective (28)

- Demonstrative (27)

- Interjection (35)

- Interrogative pro-form (16)

- Noun (1846)

- Number (0)

  - Cardinal numeral (32)

  - Ordinal numeral (6)

- Preposition (22)

- Pronoun (47)

- Verb (1952)

  - Pseudo verb (14)

The categories are defined as follows (the category's abbreviation is shown within square brackets):

B.4.1 Adjective [Adj]

An adjective is a word whose main syntactic role is to modify a noun or pronoun, giving more information about the noun or pronoun's referent.

(See instances from the lexicon in appendix D.1.)

B.4.2 Adverb [Adv]

An adverb, narrowly defined, is a part of speech whose members modify verbs for such categories as time, manner, place, or direction. An adverb, broadly defined, is a part of speech whose members modify any constituent class of words other than nouns, such as verbs, adjectives, adverbs, phrases, clauses, or sentences. Under this definition, the possible type of modification depends on the class of the constituent being modified.

(See instances from the lexicon in appendix D.2.)

B.4.3 Connective [Conn]

Also known as a conjunction, a connective is a class of parts of speech whose members syntactically link words or larger constituents, and expresses a semantic relationship between them. A conjunction is positionally fixed relative to one or more of the elements related by it, thus distinguishing it from constituents such as English conjunctive adverbs.

(See instances from the lexicon in appendix D.4.)

B.4.4 Demonstrative [Dem]

A demonstrative is a determiner that is used deictically to indicate a referent's spatial, temporal, or discourse location. A demonstrative functions as a modifier of a noun, or a pronoun.

(See instances from the lexicon in appendix D.5.)

B.4.5 Interjection [Interj]

An interjection is a part of speech, typically brief in form, such as one syllable or word, whose members are used most often as exclamations or parts of an exclamation. An interjection, typically expressing an emotional reaction, often with respect to an accompanying sentence, is not syntactically related to other accompanying expressions, and may include a combination of sounds not otherwise found in the language.

(See instances from the lexicon in appendix D.6.)

B.4.6 Interrogative pro-form [Interog Pro-form]

An interrogative pro-form is a pro-form that is used in questions to stand for the item questioned.

(See instances from the lexicon in appendix D.7.)

B.4.7 Noun [N]

"Nouns are items which display certain types of inflection (e.g. of case or number), have a specific distribution (eg. they may follow prepositions but not, say, modals), and perform a specific syntactic function (e.g. as subject or object of a sentence). Nouns are generally subclassified into common and proper types, and analysed in terms of number, gender, case, and countability." (Crystal 2008:333)

The Noun category has 1 inflectional template:

Number

(See instances from the lexicon in appendix D.8.)

B.4.8 Number [Num]

Number

(See instances from the lexicon in appendix D.9.)

B.4.9 Cardinal numeral [CardNum]

A cardinal numeral is a numeral of the class whose members are considered basic in form, are used in counting, and are used in expressing how many objects are referred to.

(See instances from the lexicon in appendix D.3.)

B.4.10 Ordinal numeral [Ordnum]

An ordinal numeral is a numeral belonging to a class whose members designate positions in a sequence.

(See instances from the lexicon in appendix D.10.)

B.4.11 Preposition [Prep]

"Prepositions are the set of items which typically precede noun phrases (often single nouns or pronouns), to form a single constituent of structure." (Crystal 2008:383)

(See instances from the lexicon in appendix D.11.)

B.4.12 Pronoun [Pro]

"Pronouns are the closed set of items which can be used to substitute for a noun phrase (or single noun)." (Crystal 2008:391)

(See instances from the lexicon in appendix D.12.)

B.4.13 Verb [V]

A Verb is a part of speech whose members typically signal events and actions; constitute, singly or in a phrase, a minimal predicate in a clause; govern the number and types of other constituents which may occur in the clause; and, in inflectional languages, may be inflected for tense, aspect, voice, modality, or agreement with other constituents in person, number, or grammatical gender.

The Verb category has 5 inflectional templates:

AV/PV/CV inflection template
OV/GV inflection template
PFV OV inflection template
Prolonged action template
Stative verb template

These templates are valid for not only this category, but also its subcategory: Pseudo

verb.

The Verb category has 3 inflection classes: partial, motion, and state.

The Verb category has 2 inflectable features: voice in Philippine-type languages

and aspect.

(See instances from the lexicon in appendix D.14.)

B.4.14 Pseudo verb [PseudoV]

A partially inflectable verb.

The Pseudo verb category has 1 inflectional template:

Pseudo verb

(See instances from the lexicon in appendix D.13.)

B.5 Inflection

In this analysis of Ayta Abellen the following word categories are inflected:

In the inflectional templates expressed below, parentheses indicate that a slot is optional.

B.5.1 Noun inflection

This section lists all inflectional templates and slots for the Noun category.

B.5.1.1 Noun Templates

The category Noun has the following template.

B.5.1.1.1 Number

These are the morphemes in the noun prefix slot which inflect the stem for number. Singular nouns are not marked. [C^1][V^1] means that the first consonant and vowel are reduplicated from the stem.

(Noun pre)	Stem
[C^1][V^1]- 'PL'	

B.5.1.2 Noun Slots and Fillers

The following is a listing of the fillers of the slot involved in Noun inflection.

B.5.1.2.1 Noun pre

These are the morphemes in the noun prefix slot.

Form	Gloss	Definition
[C^1][V^1]-	'PL'	Plural

B.5.2 Verb inflection

This section lists all inflectional templates and slots for the Verb category and its subcategories.

B.5.2.1 Verb Templates

The category Verb has the following templates.

B.5.2.1.1 AV/PV/CV inflection template

This is the inflection template for AV/PV/CV verbs. This template is valid for not only the Verb category, but also its subcategory: Pseudo verb.

Aspect	Voice prefix	Stem
aN- 'CONT' *-in*- 'PFV' *m*- 'CTPLT' *n* - 'PFV'	*Ø*- 'AV' *i*- 'CV' *-om*- 'AV ' *-om*- 'PV'	

B.5.2.1.2 OV/GV inflection template

This is the inflection template for OV/GV verbs. This template is valid for not only the Verb category, but also its subcategory: Pseudo verb.

Aspect	Stem	(Voice suffix)
aN- 'CONT' *-in*- 'PFV' *m*- 'CTPLT' *n* - 'PFV'		*-an* 'GV' *-en* 'OV'

B.5.2.1.3 PFV OV/CV inflection template

This is the inflection template for aspect-voice portmanteau forms. This template is valid for not only the Verb category, but also its subcategory: Pseudo verb.

Aspect/voice	Stem
-in- 'PFV, OV' *in--* 'PFV, OV' *iN-* 'PFV, CV'	

B.5.2.1.4 Prolonged action template

This is the inflection template for prolonged action stems that are derived into nouns. This template is valid for not only the Verb category, but also its subcategory: Pseudo verb. This template is a non-final template. That is, when it applies, it does not yet make a well-formed word. It requires a derivational affix to change its category and then the resulting category may have an inflectional template to complete it.

Om prefix	Stem
-om- 'AV ' *-om-* 'PV'	

B.5.2.1.5 Stative verb template

This is the inflection template for stative verbs. This template is valid for not only the Verb category, but also its subcategory: Pseudo verb.

Aspect		Stative voice		Stem	(Stative voice suffix)
aN-	'CONT'	*ka*-	'STA OV'		*ka- -an* 'STA GV'
-in-	'PFV'	*ka- -an*	'STA GV'		
m-	'CTPLT'	*kai*-	'STA CV'		
n -	'PFV'	*paka*-	'STA AV'		

B.5.2.2 Verb Slots and Fillers

The following is a listing of the fillers of the slots involved in Verb inflection.

B.5.2.2.1 Aspect

These are the morphemes in the aspect slot.

Form	Gloss	Definition
aN-	'CONT'	Continuous aspect.
-in-	'PFV'	Perfective aspect.
m-	'CTPLT'	Contemplated aspect.
n -	'PFV'	Perfective aspect.

B.5.2.2.2 Aspect/voice

These are the morphemes in the aspect/voice slot.

Form	Gloss	Definition
-in-	'PFV, OV'	Perfective aspect, object voice.
in--	'PFV, OV'	Perfective aspect, object voice.
iN-	'PFV, CV'	Perfective aspect, conveyance voice.

B.5.2.2.3 Om prefix

These are the morphemes in the om prefix slot.

Form	Gloss	Definition
-om-	'AV '	Actor voice.
-om-	'PV'	Patient voice.

B.5.2.2.4 Stative voice

These are the morphemes in the stative voice slot.

Form	Gloss	Definition
ka-	'STA OV'	Stative patient voice.
ka- -an	'STA GV'	Stative goal voice.
kai-	'STA CV'	Stative conveyance voice.
paka-	'STA AV'	Stative actor voice.

B.5.2.2.5 Stative voice suffix

These are the morphemes in the stative voice suffix slot.

Form	Gloss	Definition
ka- -an	'STA GV'	Stative goal voice.

B.5.2.2.6 Voice prefix

These are the morphemes in the voice prefix slot.

Form	Gloss	Definition	Inflection Features
Ø--	'AV'	Actor voice.	[voice in Philippine-type languages:actor voice]
i-	'CV'	Conveyance voice.	[voice in Philippine-type languages:conveyance voice]
-om-	'AV '	Actor voice.	[voice in Philippine-type languages:actor voice]
-om-	'PV'	Patient voice.	[voice in Philippine-type languages:patient voice]

B.5.2.2.7 Voice suffix

These are the morphemes in the voice suffix slot.

Form	Gloss	Definition	Inflection Features
-an	'GV'	Goal voice.	[voice in Philippine-type languages:goal voice]
-en	'OV'	Object voice.	[voice in Philippine-type languages:object voice]

B.5.2.3 Pseudo verb inflection

This section lists all inflectional templates and slots for the Pseudo verb category.

B.5.2.3.1 Pseudo verb Templates

The category Pseudo verb has the following template.

B.5.2.3.1.1 Pseudo verb

This inflectional template for Pseudo verb has the following slot after the stem.

Stem	(Voice suffix)
	-an 'GV' *-en* 'OV'

B.5.2.3.2 Pseudo verb Slots and Fillers

The category Pseudo verb does not define any slots. Its templates, however, may use any of these slots: aspect, aspect/voice, om prefix, stative voice, stative voice suffix, voice prefix, and voice suffix.

B.6 Derivation

The lexicon currently contains 42 derivational affixes. A number in the table below indicates the number of derivational affixes that attach to a stem of the syntactic category named in the row label to the left and produce a stem of the syntactic category named in the column label above it. (Note that it is possible for a derivational affix to have more than one mapping so the sum of the numbers in the table may be greater than the number of derivational affixes in the lexicon.)

	Adj	Adv	CardNum	N	Ordnum	V
Adj	6			1		1
CardNum		3	3		1	1
N	1	1		2		2
Num						1
Ordnum						1
V				10		14

The following are the derivational affixes in this analysis of Ayta Abellen:

B.6.1 From Adj to Adj

ma-	'Adj, SG'	Adjective, singular.
manga-	'Adj, PL'	Plural adjective marker.
ma-	'Adj, SG'	Singular adjective marker.
Dup-	'DIMUN'	Diminuative root reduplication.
pinaka-	'SUP'	Superlative degree marker.
[C^1][V^1]-	'EMPH'	Emphasis.

B.6.2 From Adj to N

ka- -an 'COLL' Collective noun marker.

B.6.3 From Adj to V

ˆ0- 'Adj > V' Adjective to Verb null derivation.

B.6.4 From CardNum to Adv

titi-	'X per'	X per.
mani-	'each'	Each. Denotes distributivity of numerals.
ti-	'each'	Each. Denotes distributivity of numerals.

B.6.5 From CardNum to CardNum

[C^1][V^1]-	'EMPH'	Emphasis.
manga-	'approximately'	Approximately.
labin-	'10 +'	Ten plus X.

B.6.6 From CardNum to Ordnum

ika- 'ORD NUM ' Ordinal number marker.

B.6.7 From CardNum to V

maika- 'ord num verbalizer' Ordinal number verbalizer.

B.6.8 From N to Adj

ˆ0- 'N > Adj' Noun to adjective null derivation.

B.6.9 From N to Adv

mina- 'every' Every.

B.6.10 From N to N

ka- -an 'COLL'   Collective noun marker.
mi-   'SOC REL' Social relationship noun marker.

B.6.11 From N to V

ˆ0- 'N>V' Noun to verb null derivation.

B.6.12 From Num to V

ˆ0- 'Num>V' Number to verb null derivation.

B.6.13 From Ordnum to V

ka- 'STA NUM' Ordinal number to verb derivation.

B.6.14 From V to N

pāg-	'PA.GER'	Prolonged singular action gerundivizer.
pāy-	'REC.PA.GER'	Reciprocal prolonged action gerundivizer.
pangi-	'PL.DETR.GER'	Plural action detransitive gerundivizer.
pāngi-	'PL.PA.DETR.GER'	Plural prolonged action detransitive gerundivizer.
pāN-	'PL.PA.GER'	Plural prolonged action gerundivizer.
māN-	'OCCUP'	Occupational noun marker.
ka-	'ASSOC'	Reciprocal association marker.
paN -	'PL.GER'	Plural action gerundivizer.
pag-	'SG.GER'	Singular action gerundivizer.
māg-	'OCCUP'	Occupational noun marker.

B.6.15 From V to V

Citation form	Gloss	Definition	To inflection class
ka-	'STA'	Stative.	
paki-	'REQ'	Request mode.	
pangi-	'PL.DETR'	Plural, detransitive.	
pāngi-	'PL.DETR'	Plural, detransitive.	
Dup-	'DIMUN'	Diminuative root reduplication.	
[C^1][V^1]-	'EMPH'	Emphasis.	
pai-	'CAUS.2TRANS'	Causative mode, bitransitive.	partial
pa-	'CAUS'	Causative. Denotes that someone permits or causes someone to do something.	partial
pag-	'DUR'	Durative mode. Verbal action spans a duration of time.	partial

Citation form	Gloss	Definition	To inflection class
paka-	'APT'	Aptative mode. Signals that a noun phrase has an actor relationship to a transitive verb, denoting abilitative action.	partial
paki-	'REQ'	Request mode. Signals that an active verb has an actor relationship to a noun phrase, indicating a request for social action.	partial
paN -	'PL'	Plural action mode. Indicates that plurality of events taking place.	
pay-	'REC'	Reciprocal mode. Signals that the actors referred to by the noun phrase are involved in reciprocal action to each other.	partial
pi-	'HAB'	Habitual mode. Signals that the event is occurring habitually.	partial

B.7 Clitics

In this analysis of Ayta Abellen there are 19 clitics.

Form	Gloss	Definition	Category	Attaches to:
=*ana*	'now'	Now, already. Signals that a condition, an action or process has reached a certain actual state.	Adverb	Any category
=*awod*	'so'	So, therefore, then. Expression signals inference relation.	Adverb	Any category
=*bay*	'EMPH'	Forming long or maybe better emphatic forms of demonstrative pronouns, conjunctions and adverbs.	Adverb	Any category

Form	Gloss	Definition	Category	Attaches to:
=*bega*	'absolutely'	Absolutely. Not even a little bit.	Adverb	Any category
=*bengat*	'just'	Just.	Adverb	Any category
=*dayi*	'OPT'	Optative mood. Added to express a wish of the speaker that an action be possible.	Adverb	Any category
=*ingat*	'on other hand'	On other hand, rather.	Adverb	Any category
=*kano*	'it is said'	Supposedly; so they said; so he said; so it is said; according to. (Often an expression of doubt).	Adverb	Any category
=*lagi*	'possibly'	Possibly; perhaps.	Adverb	Any category
=*laweh*	'uncertainty'	Marks questions, often denoting uncertainty.	Adverb	Any category
=*manayti*	'surprise'	Denotes mild surprise at new information, or an unexpected event or situation, or in expressing an afterthought.	Adverb	Any category
=*modin*	'same'	Same as always.	Adverb	Any category
=*n*	'COMP'	Complementizer.	Connective	Any category
	'GEN'	Genitive case marker contraction.	Preposition	Any category
=*nayi*	'QUES'	Marks an utterance as a question, often denoting politeness.	Adverb	Any category
=*pa*	'still'	Still.	Adverb	Any category
=*po*	'yet'	Yet, still.	Adverb	Any category
=*tana*	'just '	Just.	Adverb	Any category
=*teed*	'also, too'	Also, too.	Adverb	Any category
=*y*	'NOM'	Nominative case marker contraction.	Preposition	Any category

B.8 Morpho-syntactic Feature System

Ayta Abellen has a morpho-syntactic feature system with the feature structure types listed in section 8.1 and the features given in section 8.2.

B.8.1 Morpho-syntactic Feature Structure Types

Ayta Abellen has a feature system with the following feature structure types:

B.8.1.1 Inflection

Ayta Abellen verbs have the following features:

Name	Description
voice in Philippine-type languages	Philippine-type languages (e.g. Tagalog and Cebuano as well as some in Malaysia) have a voice or focus system in which the verb selects the semantic role of the grammatical subject. The verb has an affix which indicates the semantic role of the nominative marked argument.
aspect	Aspect is a grammatical category associated with verbs that expresses a temporal view of the event or state expressed by the verb.

B.8.1.2 Pronoun agreement

Features common to agreement on pronouns. It has the following features:

Name	Description
case	Case is a grammatical category determined by the syntactic or semantic function of a noun or pronoun.

B.8.2 Morpho-syntactic Features

Ayta Abellen has a morpho-syntactic feature system with the following features:

B.8.2.1 Aspect

Aspect is a grammatical category associated with verbs that expresses a temporal view of the event or state expressed by the verb. It has the following possible values:

Name	Abbreviation	Description
continuous aspect	cont	Continuous aspect is an imperfective aspect that expresses an ongoing, but not habitual, occurrence of the state or event expressed by the verb.
perfective aspect	pfv	Perfective aspect is an aspect that expresses a temporal view of an event or state as a simple whole, apart from the consideration of the internal structure of the time in which it occurs.
contemplated aspect	ctplt	Contemplated aspect is an aspect that expresses an action that is not started but is anticipated.

B.8.2.2 Case

Case is a grammatical category determined by the syntactic or semantic function of a noun or pronoun. It has the following possible values:

Name	Abbreviation	Description
dative case	dat	Dative case is a case that marks any of the following: (1) Indirect objects (for languages in which they are held to exist); (2) Nouns having the role of (a) recipient (as of things given), (b) beneficiary of an action, or (c) possessor of an item.
nominative case	nom	Nominative case is the case that identifies clause subjects in nominative-accusative languages. Nouns used in isolation have this case.
genitive case	gen	Genitive case is a case in which the referent of the marked noun is the possessor of the referent of another noun. The genitive case can also be used for the actor in non actor voice sentences.

B.8.2.3 Voice in Philippine-type languages

Philippine-type languages (e.g. Tagalog and Cebuano as well as some in Malaysia) have a voice or focus system in which the verb selects the semantic role of the grammatical subject. The verb has an affix which indicates the semantic role of the nominative marked argument. It has the following possible values:

Name	Abbreviation	Description
actor voice	AV	The nominative marked argument has the semantic role of actor.
patient voice	PV	The nominative marked argument has the semantic role of patient.
object voice	OV	The nominative marked argument is the object toward which the verbal action is directed.
conveyance voice	CV	The nominative marked argument has the semantic role of a theme being conveyed in some way by the verbal action.
goal voice/focus	GV	The nominative marked argument has the semantic role of goal or recipient.

B.9 Allomorphy

This analysis of Ayta Abellen has phonological conditioning of allomorphs.

B.9.1 Phonological Environments

The following is a complete list of the phonological environments that condition allomorphs in this analysis:

Representation	Name	Description	Count
/ # _ [Bil]	_ Bil	Stem Initial, Before Bilabial	2 instances
/ # _ [V]	_ V	Stem initial, Before Vowel	8 instances
/ [C] _	[C] _	After Consonant	2 instances
/ _ [C^1][V^1]	CV Redup	Consonant Vowel Reduplication	2 instances
/ _ [V^1]	_ [V^1]	Before reduplicated Vowel	2 instances
/ # _ [Alv]	_ Alv	Stem Initial, Before Alveolar	2 instances
/ # _ [Vel]	_ Vel	Stem Initial, Before Velar	2 instances
/ # [C] _	C _	After stem-initial Consonant	2 instances
/ # [Nas] _	[Nas] _	After Stem-initial Nasal	11 instances
/ # [Nas] _ [C]	# [Nas] _ [C]	After stem initial Nasal and before Consonant	1 instance
/ # [Nas] _ [V]	# [Nas] _ [V]	After stem initial Nasal and before Vowel	1 instance
/ # _	# _	Word initial	7 instances
/ # _ [V^1]	V Redup	Vowel Reduplication	1 instance
/ # _ h	_ h	Before h	1 instance
/ [Nas] _	_ Nas	Following Nasal	7 instances
/ _ [V]	_ [V]	Before Vowel	3 instances

Representation	Name	Description	Count
/ _ a	_ a	Before a	1 instance
/ _ w	_ w	Before w	2 instances
/ y _	y _	After y	1 instance

The following is a complete list of the phonological environments that condition infix positioning in this analysis:

Representation	Name	Description	Count
/ # _ [V]	_ V	Stem initial, Before Vowel	7 instances
/ [C] _	[C] _	After Consonant	2 instances
/ # [C] _	C _	After stem-initial Consonant	4 instances
/ # _	# _	Word initial	2 instances
/ _ w	_ w	Before w	2 instances

B.9.2 Inflection Classes

This analysis of Ayta Abellen also has allomorphy that is lexically conditioned by inflection class.

The category Verb has the inflection classes shown in the following table. There is no default inflection class for this category.

Name	Description	Stem count	Affix count
Motion	Motion verbs	36 stems	3 affixes
Partial	Partially inflectable	14 stems	1 affix
State	Change of state verbs	35 stems	3 affixes

B.10 Natural Classes

The following natural classes have been defined in this analysis of Ayta Abellen.

Class	Phonemes	Class Name
Alv	*d, h, l, n, t*	Alveolar plus
Bil	*b, m, p*	Bilabial
C	*b, d, g, h, k, l, m, n, ŋ, p, t, w, y, ʔ*	Consonant
Glot	*h*	Glottal
Nas	*m, n, ŋ*	Nasal
V	*a, e, i, o, u*	Vowel
Vel	*g, k, ŋ*	Velar

B.11 Residue

B.11.1 Ad hoc constraints

The following sets of morphemes or allomorphs never co-occur in the same wordform, but the morphological description given above does not yet offer an explanation. They are listed as follows:

- Morpheme ad hoc sequences

- Allomorph ad hoc sequences

B.11.1.1 Morpheme ad hoc sequences

The following table delineates the sets of morphemes which may not co-occur:

Key morpheme	Cannot occur	Other morphemes
Ø- 'N>V' N>V	Adjacent before	*ka-* 'ASSOC' V>N
Ø- 'N>V' N>V	Anywhere around	*pag-* 'GER' V>N
Ø- 'N>V' N>V	Adjacent before	*Ø-* 'N>V' N>V
-in- 'PFV' V:Aspect	Adjacent before	*Ø-* 'AV' V:Voice prefix

B.11.1.2 Allomorph ad hoc sequences

The following table delineates the sets of allomorph/morpheme pairs which may not co-occur:

Key allomorph/morpheme	Cannot occur	Other allomorph/morphemes
pa 'CAUS'	Adjacent after	*m* 'CTPLT'
Ø 'AV'	Anywhere around	*Ø* 'CTPLT'
Ø 'AV'	Adjacent before	*a* 'CAUS'

Appendix C

Ayta Abellen morphemes by type

This appendix lists morphemes by morphological type. Only the first ten morphemes will be listed for each morphological type.

- Circumfix (2).

- Enclitic (19).

- Infix (4).

- Phrase (35).

- Prefix (51).

- Root (3548).

- Stem (547).

- Suffix (2).

C.1 Circumfix

This subsection lists all the instances.

ka- -an 'STA GV'
ka- -an 'COLL'

C.2 Enclitic

This subsection lists the first ten instances.

=*ana*	'now'
=*awod*	'so'
=*bay*	'EMPH'
=*bega*	'absolutely'
=*bengat*	'just'
=*dayi*	'OPT'
=*ingat*	'on other hand'
=*kano*	'it is said'
=*lagi*	'possibly'
=*laweh*	'uncertainty'

C.3 Infix

This subsection lists all the instances.

-in-	'PFV, OV'
-in-	'PFV'
-om-	'AV '
-om-	'PV'

C.4 Phrase

This subsection lists the first ten instances.

agyamakanoman	'whenever'
aniton balang	'evil spirit'
Apo Dioh	'God'

balita kothido	'hearsay'
bato balani	'magnet'
bayontao	'bachelor'
habaw nono	'breast milk'
habot maih	'corn silk'
hipilyon ngipen	'tooth brush'
indyan minggo	'Indian mango'

C.5 Prefix

This subsection lists the first ten instances.

[Cˆ1][Vˆ1]-	'PL'
[Cˆ1][Vˆ1]-	'EMPH'
Ø-	'N>V'
Ø-	'AV'
Ø-	'No Aspect'
Ø-	'Adj > V'
Ø-	'Num>V'
aN-	'CONT'
Dup-	'DIMUN'
i-	'CV'

C.6 Root

This subsection lists the first ten instances.

a'	'REL'
a	'LNK'
aba	'delay'
abagat	'rainy season'

abala	'activity'
abala	'bother'
abalayan	'parents by marriage'
abang	'rent'
abano	'cigar'
abanti	'advance'

C.7 Stem

This subsection lists the first ten instances.

adadoen	'plow'
adapan	'front of'
agapayan	'stay beside'
agwayen	'cut vines'
ahonbalang	'wild creature'
ampagkabawan	'being senile'
ampameyeng-peyeng	'shake head'
anak-anak	'doll'
anemagatoh	'600'
anganagen	'attack by termites'

C.8 Suffix

This subsection lists all the instances.

-an	'GV'
-en	'OV'

Appendix D

Ayta Abellen morphemes by category

This appendix lists morphemes by category. Only the first ten morphemes will be listed for each category.

- Adjective (407).

- Adverb (122).

- Cardinal numeral (32).

- Connective (28).

- Demonstrative (27).

- Interjection (35).

- Interrogative pro-form (16).

- Noun (1846).

- Number

- Ordinal numeral (6).

- Preposition (22).

- Pronoun (47).

- Verb (1966).

D.1 Adjective

This subsection lists the first ten instances.

abli	'valuable'
ada	'beautiful'
aep	'considerate'
aha	'sharp'
ahem	'sour'
ak-haw	'strong'
akokoh	'selfish'
aktong	'stiff, rigid'
alah	'o'clock'
alalang	'sharp'

D.2 Adverb

This subsection lists the first ten instances.

=*ana*	'now'
=*awod*	'so'
=*bay*	'EMPH'
=*bega*	'absolutely'
=*bengat*	'just'
=*dayi*	'OPT'
=*ingat*	'on other hand'
=*kano*	'it is said'
=*lagi*	'possibly'
=*laweh*	'uncertainty'

D.3 Cardinal numeral

This subsection lists the first ten instances.

anem	'6'
anemagatoh	'600'
anemapo'	'60'
apat	'4'
apatagatoh	'400'
apatapo'	'40'
gatoh	'100'
hyam	'9'
hyamagatoh	'900'
hyamapo'	'90'

D.4 Connective

This subsection lists the first ten instances.

=n	'COMP'
a'	'REL'
a	'LNK'
angga	'until'
bana	'because'
bayo	'before'
biha	'and then'
bilang	'like'
boy	'and'
dinan	'comparative'

D.5 Demonstrative

This subsection lists the first ten instances.

ati	'this'
ayti	'this'
di	'here'
do	'there'
hilati	'this'
hiyain	'that (near hearer)'
hiyatew	'that (far)'
ihen	'there (near hearer)'
ihtew	'there (far)'
ihti	'here (near speaker)'

D.6 Interjection

This subsection lists the first ten instances.

adey	'don't'
adi	'don't'
agya	'not 3S'
ah	'ahh'
araay	'ouch!'
arooy	'oh my!'
awo	'yes'
awobay	'yes'
ay	'oh!'
bahala	'it's up to'

D.7 Interrogative pro-form

This subsection lists the first ten instances.

ano	'how many'
aya	'what'
aya	'which'
hinya	'what; who'
hinyain	'what is that'
hinyatew	'what is that'
hinyati	'what is this'
makano	'when'
nakano	'when'
pano	'how'

D.8 Noun

This subsection lists the first ten instances.

abagat	'rainy season'
abala	'activity'
abalayan	'parents by marriage'
abang	'rent'
abano	'cigar'
abaw	'beetle'
abay	'lower abdomen'
Abellen	'Abellen'
Abelling	'Abellen'
abeng	'quail'

D.9 Number

D.10 Ordinal numeral

This subsection lists all the instances.

ikalwa   'second'
ikatlo   'third'
kakalwa 'second'
kakatlon 'third'
kalwa   'second'
katlo   'third'

D.11 Preposition

This subsection lists the first ten instances.

=n   'GEN'
=y   'NOM'
ala   'one'
alas   '> two o'clock'
angga 'until'
bahta 'as long as'
ha   'OBL'
hilay   'PropmrkrPL'
hilay   'PLTopicmrkr'
hiyay   'TM'

D.12 Pronoun

This subsection lists the first ten instances.

agyahinyaman	'whatever'
ako	'1SG.NOM'
hika	'2SG.EMPH'
hikawo	'2PL.EMPH'
hikayi	'1PL.EXCL.EMPH'
hikita	'1DU.EMPH'
hikitawo	'1PL.INCL.EMPH'
hiko	'1SG.EMPH'
hila	'3PL.EMPH'
hilabayin	'3PL.EMPH'

D.13 Pseudo verb

This subsection lists the first ten instances.

ba'ko	'I thought'
ba'la	'they thought'
ba'mo	'you thought'
ba'na	'he thought'
bobokod	'do alone'
labay	'like'
main	'EXT'
padah	'speed up'
tanda	'know'
wanla	'they said'

D.14 Verb

This subsection lists the first ten instances.

aba	'delay'
abala	'bother'
abang	'rent'
abanti	'advance'
abaw	'gather beetles'
abholto	'absolve'
abli	'value'
ablo	'out of line'
abo	'heat'
aboh	'pant'

Appendix E

PAWS SYNTAX SKETCH

E.1 Introduction

The Ayta Abellen language is spoken in the province of Tarlac in the Philippines. There are approximately 3,500 speakers. Ayta Abellen is a member of the Sambal language subgroup.

After giving information on the syntactic typology of the language, the grammar begins by describing the smaller phrases which can modify other phrases before moving through several types of nominal phrases and on to adpositional phrases. Basic, main clause sentences are then covered, followed finally by various types of embedded clauses and constructions involving changes in word order.

This is a description of the syntax of Ayta Abellen based on the answers given to the PAWS Starter Kit. For each section, explanations are initially given with English examples. Because of the differences between English and Austronesian language syntax, examples from Tagalog are supplied for each section for comparison.

E.2 Word Order Typology

Tagalog as well as most Philippine languages are Verb-Subject-Object languages.

The basic word order of Ayta Abellen in transitive sentences is also VSO, as shown in the following examples:

(109) a. *N-am-Ø-yay   ya=n   pilak   ye   Pablo kangko.*
PFV-PL-AV-give 3SG.NOM=GEN money NOM Pablo 1SG.DAT
'Pablo gave money to me.'

b. *Bogbog-en   la   ya=n   la-laki*
CTPLT-beat-OV 3PL.GEN 3SG.NOM=GEN PL-man

   ye   Juan ha   loob   baey   na.
   NOM Juan DAT inside house 3SG.GEN
'The men will beat Juan inside his house.'

c. *Am-paolay-an   na   ya=n   Apo   Dioh ye   laki.*
CONT-abandon-GV 3SG.GEN 3SG.NOM=GEN Lord God   NOM man
'The man is being abandoned by God.'

In keeping with the head-initial typology, Ayta Abellen has prepositions (section 9.2).

Possessors occur after the noun being possessed (section 6.4). Adjective phrases occur on either side (but not both sides) of the noun they are modifying (section 6.5). Relative clauses occur after the head noun (section 14).

In Ayta Abellen, pro-drop of the nominative marked argument is allowed for subject, direct object, and indirect object (section 11.2).

Ayta Abellen has auxiliaries which are written as separate words (section 11.3).

In Yes/No questions, there is a question marker which occurs in the second position of the sentence (section 13.1). In content questions, an interrogative phrase moves to the front of the clause (section 13.2).

E.3 Quantifiers and Quantifier Phrases

Quantifiers can show up in various places in a nominal phrase. Most languages have four sets of quantifiers:

- Those meaning *all* or *not*, which modify the whole nominal phrase,

- quantifiers such as *some* or *no,* which do not co-occur with any other modifers except adjectives,

- other quantifiers such as *many* or *few*, and

- numbers.

Each type will be considered in turn.

E.3.1 Quantifiers which modify the whole nominal phrase

Quantifiers meaning *all* or *not* seem to be a category of their own. They usually occur first (or last) in a nominal phrase, in a position distinct from the other quantifiers but in the same place as degree words like *only* and *just* (which can also act as focus markers, to be discussed in section 18) and words that mean the same as *almost* (which can also co-occur with *all*), so we will refer to them as special degree words. These words modify the whole nominal phrase, as in *all five children, only Sue's three children, nearly all John's young children* and *just this many children*. Note that *not* can occur in place of or before these other special degree words, as in *not the children, not all five children, not just this many children* or *not only Sue's three children*.

Tagalog quantifiers which can modify the whole nominal phrase include words like

lahat 'all', *halos* 'almost all', and *lang* 'only'. Examples of nominal phrases are:

(110) *lahat ng   mga anak ni   Rosie*
       all   GEN PL   child GEN Rosie
       'all the children of Rosie'

(111) *ang   mga hayop sa   bukid   lang*
       NOM PL   animal DAT mountain only
       'only the animals in the mountains'

Ayta Abellen uses the following to express these notions:

(112)

Ayta Abellen	meaning
kaganaan pawa	'all'
haloh	'almost all'
bengat	'only or just'

In Ayta Abellen, these are expressed as separate words which modify the whole

nominal phrase. They occur on either side (but not both sides) of the rest of the nominal

phrase. Ayta Abellen positive nominal phrase examples include:

(113) a. *kaganaan a   tanam   nan laki*
          all   LNK feelings GEN man
          'all the man's feelings'

      b. *hilan kaganaan ti   hahapon a   manga-tobag*
          NOM all   NOM Japanese LNK PL-cruel
          'all the cruel Japanese (soldiers)'

      c. *yabayin anan bengat*
          this   now   just
          'just this now'

d. *maghay kolap bengat*
    one   blink only
    'just the blink of an eye'

E.3.2 Quantifiers which do not co-occur with Determiners

The second set includes quantifiers like *some*, *every*, and *no*, which do not allow any other articles, demonstratives or possessors in the nominal phrase. English examples include *some sad children*, *almost every little girl that I see* and *no black dogs*. Because they take the place of articles, demonstratives and possessors, we will treat these quantifiers as determiners with their own phrase structure rule.

Philippine languages like Tagalog have quantifiers which act as the only determiner in the nominal phrase as in (114) and (115).

(114) *bawat tao*
       each   person
       'each person'

(115) *tuwing umaga*
       every   morning
       'every morning'

Ayta Abellen has the following positive quantifiers which act as the only determiner in the nominal phrase:

(116)   **Positive Quantifiers**

balang	'each'
pawa	'all'
kaganaan	'all'

Ayta Abellen examples of positive quantifier determiners in full nominal phrases include:

(117) a. *balang magha kanla*
       each   one   DAT.3PL
       'each one of them'

     b. *kaganaan a   ta-tao*
        all   LNK PL-person
        'all people'

E.3.3 Other Quantifiers and Degree words

Other quantifiers, like *many*, *much*, and *few*, come in a different position in the nominal phrase. In English, this is after an article, demonstrative or possessor and before an adjective and noun, as in *John's many black dogs.* Further, these quantifiers can be modified by degree words such as *very* and *so*, as in *very many dogs* or *so much waste*.

In Tagalog these quantifiers, **marami** 'many' and **kaunti** *'few'*, generally precede the noun with the possessor or demonstrative following the noun, as in,

(118) *sa   mga maraming aso ni   Juan*
      DAT PL   many   dog GEN Juan
      'to the many dogs of Juan'

(119) *ang   kaunting tubig na   ito*
      NOM little   water LNK this
      'this little amount of water'

Ayta Abellen has the following quantifiers of this type:

Quantifiers

malabong	*'many'*
makandi	*'few'*

Ayta Abellen has degree words which can modify the quantifiers. These degree words occur on either side of the quantifier. Ayta Abellen examples include:

(120) a. *malabong a   tobat hila=y   ta-tao*
           many   LNK very NOM=NOM PL-person
           'very many are the people'

      b. *tobat a   malabong ye   n-a-kwa   na*
           very LNK many   NOM PFV-STA.OV-get 3SG.GEN
           'very many is what he got'

      c. *makandi ya=n   tobat*
           few   3SG=LNK very
           'it is very few'

E.3.4 Numbers

Numbers cannot be modified by the degree words like the quantifiers can, though they occur in the same position. Cardinal numbers can, however, be modified by ordinal numbers and by *next* and *last*, as in *the first twenty boys* or *the next one hundred years*. Ordinal numbers and *next* and *last* can also modify nouns directly, as in *the third girl* and *the last boy*. Since these modifiers (including ordinal numbers) also act as regular adjectives in modifying nouns directly, they act syntactically like adjectives.

E.3.4.1 Cardinal Numbers

Tagalog cardinal numbers can be modified by ordinal numbers and by 'next' and

'last', as in,

(121) *yong unang isang daan*
     those first   one   hundred
     'those first one hundred'

(122) *ang   susunod na   baente na   taon*
     NOM next   LNK twenty LNK year
     'the next twenty years'

The basic forms for Ayta Abellen cardinal numbers are given in the following chart:

(123)   **Cardinal Numbers**

1 through 10	*magha* 'one' *loa* 'two' *tatlo* 'three' *apat* 'four' *lima* 'five' *anim* 'six' *pito* 'seven' *walo* 'eight' *hyam* 'nine' *mapo* 'ten'
11 through 20	*mapo boy magha* 'eleven' *mapo boy loa* 'twelve' *mapo boy tatlo* 'thirteen' *mapo boy apat* 'fourteen' *mapo boy lima* 'fifteen' *mapo boy anim* 'sixteen' *mapo boy pito* 'seventeen' *mapo boy walo* 'eighteen' *mapo boy hyam* 'nineteen' *loampo* 'twenty'
30 through 100 (by tens)	*tatlompo'* 'thirty' *apatapo'* 'forty' *limampo'* 'fifty' *anemapo'* 'sixty' *pitompo'* 'seventy' *walompo'* 'eighty' *hyamapo'* 'ninety' *maghay gatoh* 'one hundred'

Larger cardinal numbers can be compounds which may be written as separate words, such as *one hundred thirty three, four thousand and three* or *six thousand one hundred and thirty three*. Examples of compound numbers in Ayta Abellen include:

(124) a. *malibo   boy hiyam a   gatoh   boy apatapo boy lima.*
thousand and nine   LNK hundred and forty   and five
'1945'

E.3.4.2 Ordinal Numbers

Tagalog ordinal numbers, with the exception of *una* first, are formed by adding the prefix *ika-*, as in *ikalawa* second, *ikatlo* third, *ikaapat* fourth. Ordinal numbers can also modify nouns directly, as in,

(125) *ikalawang buwan*
second   month
'second month'

In Ayta Abellen, ordinal numbers also may be formed by adding a prefix to the cardinal number. The first ten ordinal numbers are:

(126)   **Ordinal Numbers**

first through tenth	*ona*	'first'
	ikalwa	'second'
	ikatlo	'third'
	ikapat	'fourth'
	ikalima	'fifth'
	ikanem	'sixth'
	ikapito	'seventh'
	ikawalo	'eighth'
	ikahyam	'ninth'
	ikapo'	'tenth'

When the ordinal numbers in Ayta Abellen modify a cardinal number, as in *the first twenty boys*, they occur before the cardinal number.

Ayta Abellen examples of the use of ordinal numbers and the modifiers meaning "last" or "next" within a nominal phrase include:

(127) a. *ikaloa=n   baey   ha   wanan*
        second=LNK house DAT right
        'second house on the right'

   b. *hoyot a   mangaamot*
       last   LNK day
       'last day'

E.4 Adverbs and Adverb Phrases

Adverbs express four basic types of information:

- time,

- location,

- manner, and

- reason or purpose.

These types have different characteristics as to which positions they may occupy and how they may be modified. Each type will be covered in turn.

Languages normally have degree words which can modify at least some types of adverbs. Ayta Abellen has such degree words. These degree words occur after the adverb. The degree words can modify temporals, locatives, and manner adverbs, but reason or purpose adverbs are not modified.

E.4.1 Temporal Adverbs

Temporal adverbs include the English words: *already*, *then*, *soon*, *now*, *later*, and *always*. Examples including degree word modifiers are: *very soon*, *much later* and *only now*. Many languages also have a negative temporal adverb meaning 'no when', such as *never* in English.

Tagalog temporal adverbs include: *madalas* 'often', *minsan-minsan* 'sometimes', *mamaya* 'sometimes', *ngayon* 'now', *noong* 'back then', *nang* 'when', *buhat* 'from', *mula* 'from', *palagi* 'always', and *hanggang* 'until'. Degree for temporal adverbs is communicated in a variety of ways like using a past particle in *mula pa noon* 'from back then', or adding a prefix as in *napakadalas* 'very often'. The Tagalog negative temporal adverb phrase *kahit kailan* literally means 'even when'.

Examples of Ayta Abellen positive temporal adverbs and any degree words which can modify them include:

(128) a. *ha   hatew*
      DAT   time-past
      'in time past'

    b. *haanin*
      now
      'now'

    c. *lano*
      later
      'later'

d. *lanang*
   always
   'always'

e. *popoh*
   always
   'always'

f. *papainghan*
   later
   'later (in the day)'

Examples of these positive temporal adverbs in sentences include:

(129) a. *Main ya=n   hakit paibat ha hatew.*
   EXT 3SG.NOM=GEN sickness from when time-past
   'He has been sick from earlier.'

 b. *Haanin lalo ya=n   Ø-om-way   ye ginit na.*
   now more 3SG.NOM=COMP CTPLT-AV-widen NOM tear 3SG.GEN
   'Now, the tear will get even wider.'

 c. *Lano madeglem, m-Ø-akew kitawo ha baey na.*
   later tonight CTPLT-AV-go 1PL.INCL DAT house 3SG.GEN
   'Later tonight we will go to his house.'

 d. *Lanang ya=n   am-pag-Ø-habi=n*
   always 3SG.NOM=GEN CONT-PL-AV-speaking=GEN

  laban ha gobyerno.
  against DAT government
 'Always he is speaking against the government.'

 e. *Popoh hila=n   am-pay-Ø-tipon ha baey la.*
   always 3PL.NOM=LNK CONT-REC-AV-gather DAT house 3PL.GEN
   'Always they are gathering together at their house.'

 f. *"Papainghan ana," wana.*
   later   now 3SG.said
   "Later in the day," he said.

The time element of a sentence can also be expressed by other types of words or phrases, such as: nominal phrases *yesterday*, *last night*, *next year*, *tomorrow*, *a long time ago*; prepositional phrases *for a long time*, *in a little while*, *at 10 o'clock*; and adverbial clauses *while you are working*, *when Sue arrived*. These phrases occur in the same positions as the temporal adverbs, normally either initially or finally in the clause. See section 6 for the structure of nominal phrases and section 9 for information on pre/post-positional phrases. See also section 15.1 for examples of temporal adverbial clauses.

E.4.2 Locative Adverbs

Locative adverbs include the English words: *here*, *there*, *everywhere*. Locative adverbs can be modified by *only* and *right*, as in: *only here*, *right here* and *right there*.

Locative adverbs in Tagalog include: *dito* 'here', *diyan* 'there (near hearer)', and *doon* 'there (far)'. Locative adverbs can be modified by *lang* 'only', as in *diyan lang* 'there only'.

Examples of Ayta Abellen positive locative adverbs and any degree words which can modify them include:

(130) a. *ihti*
           here
           'here (near speaker)'

      b. *ihen*
           there
           'there (near hearer)'

    c. *ihtew*
       there
       'there (far)'

    d. *ihti   bengat*
       here only
       'here only'

    e. *ihen bengat*
       there only
       'there only'

    f. *ihtew bengat*
       there only
       'there only'

Examples of positive locative adverbs and degree words in complete sentences include:

(131) a. *Ø-I-lakew*   *yo*   *ya*   *ihti* *kangko.*
          CTPLT-CV-bring 2PL.GEN 3SG.NOM here 3SG.DAT
          'Bring him here to me.'

      b. *Ano=y*   *tao*   *a*   *anti*   *ihen?'*
          how.many=NOM people LNK present there
          'How many people are present there?'

      c. *Malabong hila=y*   *tatao ihtew a*   *n-ang-Ø-an.*
          many   3PL=NOM people there LNK PFV-PL-AV-eat
          'Many are the people there who ate.'

Location is often also expressed by a prepositional phrase, such as: *in the mountains, at San Jose, over there*. Many languages allow just a nominal phrase, consisting of a proper noun or description of the place. Locative expressions may normally occur either initially

or finally in the clause, or as complements to motion verbs. See section 11.4 for examples with motion verbs.

E.4.3 Manner Adverbs

Manner adverbs directly modify the verb and usually end in *-ly* in English. These manner adverbs can be modified by *very* and *so*, as in: *very quickly*, *so slowly* and *very well*.

In Tagalog manner adverbs tend to occur clause final and are introduced with the word **nang** or the linker **na/-ng**, as in,

(132) *Kailangan natin=g   k<um>ain   na=ng   mabilis.*
must   1PL.GEN=LNK <PFV.AV>eat GEN=LNK quickly
'We must eat quickly.'

(133) *Bigla   siya=ng   <um>upo.*
suddenly 3SG.NOM=LNK <PFV.AV>sit
'Suddenly she sat down.'

(First example from Schachter and Otanes 1972.)

The position of manner adverbs is similar for Ayta Abellen with the adverb being marked with the enclitic *=n* on the previous word. Examples include:

(134) a. *Ø-Bilew-en   mo=n   manged.*
CTPLT-look-OV 2SG=LNK well
'You look at it well.'

    b. *Ø-I-galang   mo=n   loboh.*
CTPLT-CV-honor 2SG=GEN wholeheartedly
'You wholeheartedly honor him.'

E.5 Adjectives and Adjective Phrases

In many grammars, numbers, quantifiers, articles, and demonstratives are considered adjectives, but each of them has a different distribution from the qualitative adjectives, so they will each have distinct syntactic categories. This section is limited to considering qualitative adjectives that express size, color, age, texture, form, or quality. One characteristic of this type of adjective is that more than one can modify a noun and in general the order of the adjectives can change (with a slight change in meaning due to the scope of the modification). For example, *hard dull brass rods* versus *dull hard brass rods* versus *brass hard dull rods*, etc. The position of adjectives within the nominal phrase will be dealt with in section 6.5.

Qualitative adjectives themselves may be modified. Some English examples are: *very big*, *so blue*, *extremely large*, *very well built* and *very easily forgotten*. Not all combinations of words work due to semantics and other considerations, but in general it seems that certain degree words and manner adverbs or manner adverb phrases can modify adjectives.

Tagalog has qualitative adjectives, such as,

(135) *B<in>igay   niya   ang   magandang bulaklak sa   akin*
        <PFV.OV>give 3SG.GEN NOM beautiful   flower   DAT 1SG.DAT
        'He gave the beautiful flower to me.'

(136) *Meron sila=ng   masipag   na   anak*
        EXT   3PL.NOM=LNK industrious LNK child
        'They have an industrious child.'

(Examples taken from Schachter and Otanes (1972).)

These adjectives can be either intensified or moderated. Intensification is done through reduplication as in *pagod na pagod* 'very tired' or *butas-butas* 'full of holes' or by using the *napaka-* prefix as in *napakaganda* 'very beautiful'. Moderation is expressed by using *medyo* 'rather' before the adjective as in *medyo gutom* 'rather hungry' or with *nang kaunti* after the adjective as in *magugulo ng kaunti* 'somewhat troublesome'.

Ayta Abellen has qualitative adjectives as in (137a). Intensification of adjectives is done with the degree marker *tobat* which can occur either after the adjective (137b) or before the adjective as in (137c). Intensification is also expressed through stem reduplication as in (137d). Moderation is expressed with the word *makandi* as in (137e).

(137) a. *maalale a   ihip*
           deep   LNK thought
           'deep thought'

       b. *mayadet a   tobat a   baey*
           big   LNK very LNK house
           'very big house'

       c. *tobat a   mayadet a   baey*
           very LNK big   LNK house
           'very big house'

       d. *matagay a   matagay a   matondol*
           high   LNK high   LNK mountain
           'very high mountain'

       e. *makandi=n kataangan*
           small=LNK   distance
           'small distance'

E.6 Nominal Phrases

Nominal phrases can be very complex. This section concentrates on those with common nouns as the lexical head. Modification by articles, demonstratives, possessors, quantifiers and degree words, adjective phrases, pre/post-positional phrases and noun compounds is considered. Participles are also covered, but nominal phrases headed by proper names are handled in section 7 and those headed by pronouns, demonstratives, and quantifiers are handled in section 8. Relative clauses are covered in section 14.

The basic order of elements allowed in nominal phrases in Ayta Abellen is:

(138) (Dem) (QP) (AdjP) ((N) N) N (AdjP) (QP) (Poss) (Deg)

Each of these phrases will be discussed and exemplified in sections E.6.3-6.6, after first dealing with agreement and case marking.

E.6.1 Agreement

Many languages require agreement between at least some of the modifiers and the head noun. While English only has number agreement, in other languages this agreement may also include animacy, class, or gender.

Tagalog has no agreement features between the noun and modifiers for number, animacy, or gender.

In Ayta Abellen, there is also no number agreement. Instead, number marking comes from the number or quantifier. Further, there is no agreement in animacy, class number, nor gender between the modifiers and the head noun.

E.6.2 Case

Nouns and their modifiers may also be marked for case. The most common case system is nominative-accusative. This is the system English has, though only pronouns are marked for case in English. In a nominative-accusative system, subjects of both transitive and intransitive verbs are in nominative case, while direct objects are accusative case. Possessors are genitive case. The nominative pronouns in English are: *I, we, you, he, she, they*. The accusative pronouns are: *me, us, you, him, her, them*. The genitive pronouns include: *my, our, your, his, her, their*. Therefore, we say *They hit him* and not *Them hit he*.

In an ergative-absolutive case system, subjects of transitive verbs are marked with the ergative case, while objects of transitive verbs and subjects of intransitive verbs are marked with the absolutive case. The case of possessors is not set cross-linguistically. Some languages with an ergative-absolutive system simply have different sets of verbal agreement markers, but no marking on nouns.

Some languages are even more complex in that they normally have a nominative-accusative system but under some conditions the case system switches to ergative-absolutive. This is called "split ergativity."

The case system of Philippine languages has been a topic of debate among linguists studying Philippine languages. Some have argued for an ergative analysis (Brainard, Gault) while others (Kroeger, etc.) have argued for a nominative-accusative analysis of the case system. The existence of different focuses or voices is a primary complicating factor. Just looking at actor focus verbal sentences, Tagalog (and other Philippine languages) have a nominative-accusative system as the actor (subject) is marked with the *ang* 'NOM' marker in both transitive and intransitive sentences. Non actor focus sentences are transitive but the difficulty in determining what the true subject is in these constructions makes it also hard to determine whether Tagalog is truly ergative. It is helpful just to say that Tagalog is nominative-accustive for actor voice verbal predications.

Ayta Abellen has a nominative-accusative case system in the same sense as Tagalog. Subjects of both transitive and intransitive verbs are marked in the same way for actor voice transitive verbal sentences and use the same set of pronouns, but objects of transitive verbs are marked differently.

The pronoun sets will be given in section 8.

E.6.3 Articles and Demonstratives

In English, the articles are *a*, *an*, and *the*. English demonstratives include *this*, *that*, *these*, and *those*. Some simple examples are: *a book, an apple, the book, the books, this apple* and *those books.*

In English, articles and demonstratives do not co-occur in the same nominal phrase, and both occur before the noun. Some languages allow both to co-occur in the same phrase, and they may occur in different positions with respect to the noun. Therefore, they need distinct syntactic categories. Some languages only have one set, either articles or demonstratives. A few languages may not have either type expressed as separate words.

Tagalog does not have articles but does have optional demonstratives that precede the noun, including *ito* 'this', *iyan* 'that (near)', and *iyon* 'that (far)'.

Ayta Abellen has only demonstratives that are realized as separate words, no articles. The demonstratives are not marked for case. The demonstrative occurs before the noun. The presence of a demonstrative is always optional. Examples of nominal phrases including articles and/or demonstratives as allowed in Ayta Abellen include:

(139) a. *yati＝n   pag-katongtong tawo*
        this=LNK   GER-discuss   1PL.INCL.GEN
        'this discussion of ours'

    b. *yain a   mangga*
        this   LNK mango
        'this mango'

    c. *yatew a   kayo*
        that   LNK tree
        'that tree'

E.6.4 Possessors

Possession can normally be expressed by a possessive pronoun (to be addressed in section 8) or by a full nominal phrase. Many languages also add some kind of marking,

such as the *'s* in English, as an indication of possession. These markings are either affixes

(or clitics) on the head noun or phrase-level clitics which attach to one end of the whole

phrase. The English marking is this second type, because alongside phrases where the *'s*

appears to attach to the head noun, as in *the boy's mother*, there are phrases which clearly

show that the clitic attaches to the end of the whole nominal phrase, such as *the boy that I*

just talked to's mother or *the girl in green's wonderful speech*. Possessors can also be

embedded in one another, as in *the boy's sister's dog*.

In Tagalog, possession can be shown with an existential possessive clause, as in,

(140) *May lapis ang   bata.*
    EXT pencil NOM child
    'The child has a pencil.'

or with the possessor modifier following the head noun, as in,

(141) *lapis   ng   bata*
    pencil GEN child
    'child's pencil'

But the possessor does not always follow the head noun as there are constructions

where both a modifier and a possessor can precede the head noun, as in,

(142) *maliit kong   anak*
    small   1SG.GEN child
    'my small child'

(Examples taken from Schachter and Otanes (1972).)

In Ayta Abellen, possession can also be shown with existential possessive clauses. In non existential sentences the possessor modifiers follow the head noun but unlike Tagalog there are no forms where the possessor can precede the head noun. Examples of possessed nominal phrases with simple and embedded possessors in Ayta Abellen include:

(143) a. *Main ya＝n   damwag   ye   Totoy.*
     EXT   3SG.NOM=GEN water.buffalo NOM Totoy
     'Totoy has a water buffalo.'

   b. *damwag   nan   Totoy*
     water.buffalo GEN Totoy
     'Totoy's water buffalo'

   c. *damwag   nin   katongno nan kapitan*
     water.buffalo GEN sibling   GEN captain
     'captain's brother's water buffalo'

   d. *baey nin ali   nan Vangie'*
     house GEN younger.sibling GEN Vangie
     'Vangie's younger sister's house'

Examples with relative clauses within the possessor here, included in simple full sentences are:

(144) a. *In-om-alih   ya＝yna＝y   indo   nin   laki*
     PFV-AV-left 3SG.NOM=now=NOM mother GEN man

      a   k＜in＞atongtong ko.
      REL <PFV>talk   1SG.GEN
     'The mother of the man that I was talking to left already.'

As seen in the examples above, Ayta Abellen does not have any special marking to distinguish possessors from any other nominal phrase. The possessors take genitive case

marking but these same surface forms can function as subjects or indirect objects in other sentences. The possessor occurs after the noun being possessed.

E.6.4.1 Possessors and Articles Or Demonstratives Occurring Together

In English, possessors and articles or demonstratives do not co-occur in the same nominal phrase, unless the possessor is expressed in a prepositional phrase. For example, *those [the boy's] books* is ungrammatical; instead one would use *those books [of his]* to express the same thought. Some other languages allow both possessors and articles or demonstratives to occur in the same phrase, so the first example above would be grammatical.

In Tagalog, nominal possessors can occur in the same phrase as demonstratives where the demonstrative precedes the noun and the possessor follows the noun, as in,

(145) *ito=ng   lapis   ko*
     this=LNK pencil 1SG.GEN
     'this pencil of mine'

In Ayta Abellen, like Tagalog, nominal possessors may occur in the same phrase as demonstratives, as in,

(146) *yati=n   lapih   ko*
     this=LNK pencil 1SG.GEN
     'this pencil of mine'

E.6.4.2 Possessor Case

With the normal nominative-accusative case system followed by Ayta Abellen, possessors are marked with the genitive case.

In some languages, the possessor agrees with the case of the head noun. In other languages, the possessor is marked with its own case, say genitive, while the head noun is marked with case depending on its position relative to the verb: nominative or accusative, ergative or absolutive depending on the case system.

In Tagalog and Ayta Abellen, there is no case agreement for possessed NPs.

E.6.5 Adjective Phrase Modifiers

Adjective phrases (AdjP) were discussed earlier in section 5. Adjective phrases which modify a noun occur close to the noun, yielding nominal phrases such as: *just those very few old, dirty books*, *all the boy's many black dogs* and *only the first one hundred very eager young boys*. As the English examples show, adjectives all occur before the noun, and more than one adjective phrase is allowed. This is handled by a recursive rule.

In Tagalog, adjective phrases tend to precede the noun they modify, as in (147).

(147) *lahat ng   mga hinog na   mangga*
     PL   GEN PL   ripe   LNK mango
     'all ripe mangoes'

But adjective phrases can also follow the noun they modify, as in (148).

(148) *mangga na   nasa   bahay*
     mango   LNK present house
     'mangoes in the house'

In Ayta Abellen, adjective phrases also precede the noun they modify, as in (149).

(149) *kaganaan   a   manga-naom a   mangga*
       all   LNK PL-ripe   LNK mango
       'all the ripe mangoes'

But adjective phrases can also follow the noun they modify, as in (150).

(150) *kaganaan a   ta-tao   a   anti   ihen ha   baey ko*
       all   LNK PL-person LNK present there DAT house 1SG.GEN
       'all the people there at my house'

E.6.6 Noun Compounds

Many languages are like English, in that they have noun modifiers. It may seem like all nouns can also be used as adjectives, but a closer look reveals that true adjectives come before the noun modifier, as in *large, round [syrup bottle]* but not **round, syrup large bottle*. Therefore, noun modifiers will be treated as compound nouns which are written as separate words.

In the English examples above, *bottle* is clearly the head noun rather than *syrup*, because *syrup* is restricting the type of *bottle*, not the other way around. In all compound nouns and noun modifier constructions in English, the head is the right-most noun in the series.

In Tagalog, the left-most noun is the head of the compound, as in,

(151) *tubig ulan*
      water rain
      'rainwater'

(152) *tanod bayan*
      guard town
      'policeman'

Ayta Abellen also has compound nouns that are written as separate words. The left-most noun is the head in these constructions. Examples include:

(153) a. *tambal   dagih*
      medicine rat
      'rat poison'

    b. *polot   panilan*
      honey honeybee
      'honey'

    c. *labok maih*
      hair   corn
      'corn silk'

E.7 Proper Names

Proper names are a special kind of noun that act syntactically just like other nouns. In English, some proper names can stand alone, like *Sue* and *Russia*. Others are like compound nouns, such as *George Washington*. Proper names may also contain articles and/or pre/post-positional phrases, as in *the United States of America*. Many languages require an article before a name. Appositives also occur, possibly including a possessor, as in *my son, David*. In addition, proper names can be modified by adjectives and relative clauses, as in *the good, old USA* or *the Susan Welch that I know*.

In Tagalog, personal proper names are required to be marked for case with *si* (NOM), *ni* (GEN), or *kay* (DAT). Non personal proper names such as place names are marked with *sa* (DAT). Proper names can be modified by demonstratives, as in,

(154) *si   Juan*
      NOM Juan
      'Juan'

(155) *ito＝ng   Juan*
      this=LNK Juan
      'this Juan'

Ayta Abellen has a pronominal agreement system. When there are proper noun NPs, coreferential pronouns agreeing in case, person, and number occur in their usual positions in the clause while the full NPs follow later in the sentence. When there are no other intervening particles or arguments it appears that the personal proper name argument is *yay* as in (156).

(156) *N-ag-Ø-habi   ya＝y   Pabling kanna.*
      PFV-DUR-AV-speak 3SG.NOM=NOM Pabling 3SG.DAT
      'Pabling spoke to him.'

But when there is an intervening particle, like *met* 'also' in (157), it can be seen that there is both a pronoun and a case marker.

(157) *N-ag-Ø-habi   ya   met ye   Pabling kanna.*
      PFV-DUR-AV-speak 3SG.NOM also NOM Pabling 3SG.DAT
      'Pabling also spoke to him.'

This leads to the conclusion that the correct analysis of *yay* in (156) should be *ya＝y* as in (158).

(158) *N-ag-Ø-habi   ya＝y   Pabling kanna.*
      PFV-DUR-AV-speak 3SG.NOM=NOM Pabling 3SG.DAT
      'Pabling spoke to him.'

The same is true for proper nouns marked as genitive (159a) and dative (159b). Non

personal proper names are marked with the dative *ha* as in (159c). Proper names can also

be marked with a demonstrative as in (159d).

(159) a. *baey   na=n   Pabling.*
       house   3SG.GEN=GEN   Pabling
       'Pabling's house'

   b. *kana=n   Apo   Dioh.*
       3SG.DAT=GEN   Lord   God
       'to God'

   c. *ha   Tarlac.*
       DAT   Tarlac
       'to Tarlac'

   d. *yati=n   Pidigo.*
       this=LNK   Pidigo
       'this Pidigo'

Examples of proper names with prepositional phrases in Ayta Abellen include:

(160) a. *ibat   kana=n   Apo   Dioh.*
       from   3SG.DAT=LNK   Lord   God
       'from God'

   b. *palakew ha   Tarlac.*
       headed   DAT   Tarlac
       'headed toward Tarlac'

Examples of proper names modified by relative clauses included in simple full

sentences in Ayta Abellen include:

(161) a. *Am-Ø-pahalamat* *ako* *kana=n* *Apo* *Dioh*
      CONT-AV-thank 1SG.NOM 3SG.DAT=LNK Lord God

      a *n-am-Ø-yay* *kangko* *nin* *kaganaan.*
      REL PFV-PL-AV-give 1SG.DAT GEN all
      'I am thanking God who gave everything to me.'

    b. *Hiyay* *Pidigo* *a* *matoa,* *main* *ya=n* *hakit.*
      TM   Pidigo REL old   EXT 3SG.NOM=GEN sickness
      'Pidigo who is old is sick.'

E.8 Pronouns

Pronouns come in various types, including personal pronouns, possessives, reflexives, reciprocals, and indefinites. These pronouns normally stand alone in the place of a nominal phrase, but in some languages they may be modified by determiners and quantifiers. Each of these types and issues will be looked at in turn, after covering the agreement features on all types of pronouns.

E.8.1 Agreement Features on Pronouns

Pronouns normally have either more than or the same number of agreement features as nouns have.

Tagalog and Ayta Abellen pronouns have agreement features for person and number. The following chart shows the features used for pronouns in Ayta Abellen.

(162)

Type of feature	Feature
person	first exclusive
person	first inclusive
person	second
person	third
number	singular
number	dual
number	plural

E.8.2 Personal Pronouns

Personal pronouns normally have either more than or the same number of case features as nouns have.

In Tagalog, personal pronouns have the same number of case features as nouns. Tagalog can be viewed as having a nominative-accusative case system. Pronouns and nouns are marked for nominative, genitive, and dative case.

Ayta Abellen has a pronominal system similar to Tagalog. There are also three distinct sets of pronouns that correspond to nominative, genitive, and dative case. The personal pronouns which correlate to nominative marked nominal phrases are shown here, beginning with first person, then second person, then the third person pronouns.

(163) **Nominative marked personal pronouns**

Pronoun	Feature	Gloss
ako	1SG	'I'
kita	1PL.DU	'we'
kayi	1PL.EXCL	'we'
kitawo	1PL.INCL	'we'
ka	2SG	'you'
kawo	2PL	'you'
ya	3SG	'he/she/it'
hila	3PL	'they'

Genitive marked personal pronouns

Pronoun	Feature	Gloss
=*ko*	1SG	'my'
=*ta*	1PL.DU	'our'
=*mi*	1PL.EXCL	'our'
=*tawo*	1PL.INCL	'our'
=*mo*	2SG	'your'
=*yo*	2PL	'your'
=*na*	3SG	'his/her/its'
=*la*	3PL	'their'

Dative marked personal pronouns

Pronoun	Feature	Gloss
kangko	1SG	'to me'
kanta	1PL.DU	'to us'
kammi	1PL.EXCL	'to us'
kantawo	1PL.INCL	'to us'
kammo	2SG	'to you'
kanyo	2PL	'to you'
kanna	3SG	'to him'
kanla	3PL	'to them'

Emphatic personal pronouns

Pronoun	Feature	Gloss
hiko	1SG	'I'
hikita	1PL.DU	'we'
hikayi	1PL.EXCL	'we'
hikitawo	1PL.INCL	'we'
hika	2SG	'you'
hikawo	2PL	'you'
hiya	3SG	'he/she/it'
hila	3PL	'they'

E.8.3 Possessive Pronouns

Many languages also have possessive pronouns which can stand alone, such as *yours*, *mine*, *theirs* and *hers* in English. These are distinct from those pronouns which can occur in possessor position, such as *your*, *my*, *their* and *her*, which were exemplified under personal pronouns above.

Tagalog does not have any possessive pronouns that stand alone as a nominal phrase.

Ayta Abellen also does not have any possessive pronouns that stand alone as a nominal phrase.

E.8.4 Reflexives

Reflexives are another type of pronouns that most languages have. They occur only in the same case as direct objects and normally must agree with the subject in person, gender, and number features. Examples in English are: *herself*, *himself*, *themselves* and *myself*. Some languages have a completely separate set of pronoun forms for reflexives, others use a phrase which consists of a noun (usually meaning 'self') that is possessed by the pronoun that refers back to the subject, some languages have a verbal affix which indicates reflexive action, and some languages do not distinguish reflexive constructions at all.

Reflexives in Tagalog are expressed by the noun *sarili* 'self' modified by a pronoun possessor as in *sarili ko* 'my self' or *sa aking sarili* 'to my self'.

Ayta Abellen also has reflexives which are phrases consisting of a certain noun with a pronoun as the possessor. The forms used for reflexives are given here, beginning with first person, then second person, then third person:

Reflexive pronouns

Reflexive	Gloss
hadili ko	'my self'
hadili yo	'your self'
hadili na	'his self'

Some examples with a reflexive used in oblique and object positions of a simple sentence are:

(164) a. *Ahe na   h<in>abi   ye   tongkol ha   hadili na.*
   not 3SG.GEN <PFV.OV>speak NOM about DAT self 3SG.GEN
   'He did not speak about his self.'

  b. *H<in>abi   na   ha   hadili na,*
   <PFV.OV>speak 3SG.GEN DAT self 3SG.GEN

   wana, "Peteg awod yain."
   he-said true then that
   'He said to himself, "So that then is true."'

  c. *Ø-I-ligtah   mo=y   hadili mo.*
   CTPLT-CV-save 2SG.GEN=NOM own 2SG.GEN
   'You save yourself.'

  d. *Ø-Biha-en   yo=y   hadili yo.*
   CTPLT-careful.OV 2PL.GEN=NOM self 2PL.GEN
   'Take care of yourself.'

E.8.5 Reciprocals

Reciprocals are similar to reflexives except that they only occur with plural subjects. They may be formed in the same way reflexives are, or they may be distinct. For example, in English, the reciprocal forms are phrases: *each other* and *one another*. Some languages use only a verbal affix to indicate reciprocal action of the plural subject, others have separate pronouns, others use a possessed noun phrase or a quantified noun phrase (as in English), and some do not mark reciprocal action at all.

Tagalog uses the ***mag- -an*** and ***nag- -an*** circumfixes attached to the verb to indicate reciprocal action, as in,

(165) *Mag-tulung-an   tayo.*
     CTPLT.REC-help-REC   1PL.INCL.NOM
     'We will help each other.'

(166) *Nag-halik-an   sila.*
     PFV.REC-kiss-REC   3PL.NOM
     'They kissed each other.'

Ayta Abellen also uses a verbal prefix *pay-* to indicate reciprocal action.

(167) a. *n-a-pay-Ø-tipon*
       PFV-STA-REC-AV-gather
       'gathered together'

     b. *n-a-pay-Ø-tongtong*
       PFV-STA-REC-AV-talk
       'talked together'

     c. *n-a-pay-Ø-laban*
       PFV-STA.OV-REC-AV-fight
       'fought each other'

Some examples of reciprocals in sentences are:

(168) a. *Hilay Ayta, n-a-pay-Ø-tipon   hila.*
       TM   Ayta   PFV-STA-REC-AV-gather   3PL.NOM
       'The Aytas, they gathered together.'

     b. *Nangon, n-a-pay-Ø-tongtong*
       Earlier,   PFV-STA-REC-AV-discuss

       kayi   ha   baey ko
       1PL.EXCL.NOM DAT house 1SG.GEN
       'Earlier, we discussed together at my house.'

     c. *Hilay Ayta boy Hapon,*
       TM   Ayta and Japanese

       n-a-pay-Ø-laban   hila=yna.
       PFV-STA-REC-AV-fight   3PL.NOM=now
       'The Ayta and Japanese fought each other.'

E.8.6 Indefinites

In English, indefinites include *everybody*, *everything*, *everyone*, *anybody*, *something*, etc. Also, *nobody*, *nothing*, *none* are negative versions of the indefinite pronouns. Note that these are made up of a quantifier (positive or negative) plus an indefinite generic noun. Some languages may not merge them into a single word, as in *no one* for English.

Haspelmath (2000) says that "In Tagalog, existential sentences are used where other languages use specific indefinite pronouns." The following examples are given for positive and negative existential sentences:

(169) *May d<um>ating   kahapon.*
     EXT <PFV.AV>arrive yesterday
     'Someone came yesterday.'

(170) *Walang d<um>ating   kahapon.*
     none   <PFV.AV>arrive yesterday
     'No one arrived yesterday.'

In spite of the preference for existential sentences for indefinites there are indefinite pronouns *sinuman* 'whoever' and *kailanman* 'whenever' which can occur in "a restricted range of functions" Haspelmath (2000).

Ayta Abellen also has a preference for expressing indefinites with existentials but does allow a limited number of indefinite pronouns. In Ayta Abellen, some or all of the positive indefinites are merged into single-word pronouns. Examples of the positive indefinite pronouns include:

Positive Indefinite Pronouns

hinyaman	'whoever, whatever'
makanoman	'whenever'
wayihtewman	'wherever'

An example of a single word indefinite pronoun nominal phrase can be seen in (171a).

(171) a. *Manged ya=n*   *pag-ameyan nin   agya hinyaman.*
           good   3SG.NOM=LNK GER-shade   GEN   even whatever
           'It's good at shading even whatever.'

E.8.7 Pronouns as the Head of a Nominal Phrase

Though pronouns and also demonstratives, quantifiers, and numbers can stand alone as a nominal phrase in the right context, sometimes modifiers are also allowed. In most languages, these elements cannot be modified by possessors, adjectives, or other quantifiers, although they can be modified by a degree word, an article and/or a demonstrative. English examples of this type include the following: *all twelve*, *only he*, *just two* and *just those few*.

In Tagalog, pronouns can be modified by a degree word, as in,

(172) *siya   lang*
        3SG.NOM only
        'only he'

(173) *ito   lang*
        this   only
        'only this'

Examples of this type in Ayta Abellen include:

(174) a. *hiya   bengat*
      3SG.NOM only
      'him only'

    b. *yati bengat*
      this only
      'this only'

In addition, it is possible to have a pronoun or quantifier modified by a relative

clause, as in *all those who will go with us* and *only these two which I saw*. Examples of this

type in Ayta Abellen, included in full sentences are:

(175) a. *Ma-ligha   hila   a*
      Adj-happy 3PL.NOM REL

      n-aki-Ø-lamo   kammi.
      PFV-REQ-AV-companion 1PL.EXCL.DAT
      'They who joined with us were happy.'

    b. *N-a-baha   yati=n   loa a   n-a-kit   ko.*
      PFV-STA.OV-wet these=LNK two REL PFV-STA.OV-see 1SG.GEN
      'These two which I saw were wet.'

This distribution discussed for English can be accounted for by classifying pronouns,

demonstratives, quantifiers, and numbers as NPs (rather than DPs) in the expanded DP tree

structure for nominal phrases. However, some languages allow pronouns to be modified by

possessors and adjectives as well, so the correct category in that case is N. A few languages

may require pronouns to stand alone without any modification at all, making them only

DPs. (Possessive, reflexive, and reciprocal pronouns are always DPs, since they must stand

alone.)

For Tagalog and Ayta Abellen, the correct category or level in the tree structure for personal pronouns and other elements which can occur without a head noun is NP, like in English, because they can be modified by degree words, demonstratives, relative clauses and PPs, but not by possessors or adjectives. The optional PP occurs after the element acting as the head of the nominal phrase.

E.9 Pre/Post-Positional Phrases

This section covers the internal structure of PPs, which can be either prepositional or postpositional phrases. Possible modifiers are considered first, then the type of PP, and finally the complements allowed.

Philippine languages like Tagalog do not have an easily identifiable set of prepositions. The case markers (*ang*, *ng*, *sa*) have many overlapping characteristics with prepositions but they are best labeled "proclitics" (Kroeger 1993). One of those proclitics, *sa*, is sometimes labeled as the only true preposition in Tagalog. But because it also marks dative case, it is probably best not labeled as a preposition. There are a few other words that can be safely called prepositions such as *taga* as in *taga Maynila* 'from Manila', *para* as in *para kay Jose* 'for Jose' or *para sa mga tao* 'for the people', and *galing* as in *galing sa Maynila* 'coming from Manila.'

Ayta Abellen also has case marking proclitics (*ye*, *nin*, *ha*) that are not labeled prepositions in this grammar description. Like Tagalog, there is a small set of words that

can more easily be labeled prepositions such as *taga* 'from', *ibat* 'from', *palakew* 'going to'. These, with the exception of *taga*, are used with the dative case marker *ha* (see 9.3).

E.9.1 Modifiers

There are normally a few degree words that can modify a whole PP, such as *just* and *right* in *just around the corner* and *right to the last second*. Ordinal numbers and adjectives can also modify PPs, as in *next in line* and *first out the door*. In addition, certain manner adverbs can modify PPs, as in *almost in the hole*, *mostly up the tree*, *squarely on the table*, and *nearly upon the enemy*.

Tagalog does not regularly allow modifiers to prepositional phrases although in some special cases this can occur as in *unang galing sa Maynila* 'first one from Manila'. But it should be mentioned that this is really an elided form where a noun like *tao* 'person' has been left out.

Ayta Abellen also does not regularly have any modifiers which occur either before or after PPs syntactically. Like Tagalog, a contracted form of *onan taon ibat ha Maynila* 'first person from Manila' could result in *onan ibat ha Maynila* 'first from Manila', appearing to be a modfier of the PP. But this is a contracted form and neither degree words nor ordinals nor adverbials can modify PPs as separate words.

E.9.2 Head Type

Prepositions come before their complements, while postpositions come afterward. English has prepositions, as shown in *for Joe* and *by the river*. The prepositions or postpositions may be separate words, or they may be clitics which are attached to the edge of the complement or to its head.

Tagalog and Ayta Abellen have prepositions only. They are all separate syntactic words. See the examples in the next section.

E.9.3 Complements within PP

Prepositions or postpositions normally take a nominal phrase for their complements, as in *to [Bill's house]* and *under [the giant Sycamore tree]*.

Tagalog prepositions can take a nominal phrase for the complement of a PP, as in (176).

(176) *para sa   mga importanteng tao   sa   Maynila*
     for   DAT PL   important   people DAT Manila
     'for the important people in Manila'

Ayta Abellen prepositions can also take a nominal phrase complement of a PP, as as in (177).

(177) *palakew ha   baey na=n   Pabling*
     toward   DAT house 3SG.GEN=GEN Pabling
     'toward the house of Pabling'

E.9.3.1 Locative Adverb Complements

Most languages will allow locative adverbs as complements of prepositions or postpositions, such as *over there* and *in here*.

Tagalog allows locative adverbs as complements of prepositions as in:

(178) *Mula roon, p<um>unta sila   sa   Bagiuo.*
      from   there   <PFV.AV>go 3PL.NOM DAT Bagiuo
      'From there, they went to Bagiuo.'

Ayta Abellen also allows locative adverbs as complements of prepositions as in:

(179) *Paibat ihtew, n-Ø-akew   hila   ha   banwa=n Angeles.*
      from   there, PFV-AV-go 3PL.NOM DAT city=LNK Angeles
      'From there, they went to the town of Angeles.'

(180) *Maligha hila=y   tatao a   taga ihtew*
      happy   3PL=NOM people LNK from there
      'The people who are from there are happy.'

E.9.3.2 Locative PP Complements

Locative prepositions and postpositions are a special group in that they can take another locative PP as their complement. English examples of this type include: *up [to the top]*, *around [behind the shed]* and *over [by the table]*.

Tagalog and Ayta Abellen have locative prepositions which take a locative PP complement, as seen in (181) and (182).

(181) *galing sa   ilalim   ng   bahay niya*
      from   DAT underneath GEN house 3SG.GEN
      'from underneath his house'

(182) *ibat   ha   hilong nin   mangga*
     from DAT under GEN mangga
     'from under the mango tree'

E.10 Austronesian Voice and Case patterns

Tagalog and other Austronesian languages have a rich voice and case marking system. Kroeger (1993) says, "Each verbal clause must contain one and only one nominative argument. A characteristic property of all Philippine-type languages is that the thematic role of the nominative element is reflected in a verbal affix which I will call the voice marker."

(183) *B<um>ili   ang   lalake ng   isda sa   tindahan.*
     <PFV.AV>buy NOM man   GEN fish DAT store
     'The man bought fish at the store.'

(184) *B<in>ili   ng   lalake ang   isda sa   tindahan.*
     <PFV.OV>buy GEN man   NOM fish DAT store
     'A man bought the fish at the store.'

(185) *B<in>il-han ng   lalake ng   isda ang   tindahan.*
     <PFV>buy-GV GEN man   GEN fish NOM store
     'At the store the man bought fish.'

Kroeger (1993) further says, "The voice marker in each example reflects the role of the nominative argument in that clause: AV for 'Active Voice' (indicating that the nominative argument is the Actor); OV for 'Objective voice'; DV for 'Dative/Locative Voice.'"

Ayta Abellen has a voice system similar to Tagalog. Voice is marked on the verb which indicates the semantic role of the nominative argument. While different terms are used here for the different voices, they are very similar to the three kinds of voice in the

Tagalog examples above. Here I use AV for Actor voice, CV for Conveyance voice, and

GV for Goal voice.

(186) *M-am-Ø-yay   ya=n   haa   kammo.*
CTPLT-AV-give   3SG.NOM=GEN   banana 2SG.DAT
'He will give a banana to you.'

(187) *Ø-I-byay   na   ye   haa   kammo.*
CTPLT-CV-give 3SG.GEN   NOM   banana 2SG.DAT
'He will give the banana to you.'

(188) *Ø-Bi-an   na   ka=n   haa.*
CTPLT-give-GV 3SG.GEN   2SG.NOM=GEN   banana
'He will give you a banana.'

Each argument is marked for case. For a nominative analysis there are three cases:

nominative, genitive, and dative. Some view the nominative marked argument as the

grammatical subject of the sentence. Others regard the actor or agent of the sentence to be

the subject. With this analysis the case marking of the subject is not the same for all voices

with the subject being marked with nominative case for actor voice and genitive case for

all other voices.

E.11 Basic, Single-Clause Sentences

This section covers intransitive, copular, transitive, ditransitive and passive sentences

in their basic surface word order (i.e. not changed by focus, topic, interrogative, or negation

constructions). As noted in section 1, the typology of Ayta Abellen is VSO, so the basic

word order is the verb in initial position, then the subject, then any complements. The order

with any auxiliaries present will be discussed in section 11.3.

Before discussing each of the sentence types in turn, general issues of inflection and

agreement features, pro-drop, and auxiliaries will be covered.

E.11.1 Inflection Features

Verbs usually carry inflection features, such as tense, aspect, and/or mood. In most

cases, these features are added by affixes, but may also be part of irregular verb forms.

Tagalog verbs are inflected for aspect. Schachter and Otanes (1972) define three

aspects.

(189)

Type of feature	Feature
aspect	continuative
aspect	imperfective
aspect	perfective

Ayta Abellen verbs are also inflected for aspect.

(190)

Type of feature	Feature
aspect	continuative
aspect	contemplated
aspect	perfective

E.11.2 Pro-Drop

Some languages are pro-drop languages which allow a subject and/or object to be missing. For example, in Spanish the normal way to say "I buy a book" is *Compro un libro* rather than with the overt subject pronoun included: *Yo compro un libro*. This is not allowed in English, except for imperatives. (The missing subject of infinitives in embedded complements will be covered in section 12.)

Tagalog is a pro-drop language. Kroeger (1998) says "Virtually any argument can be referred to with zero anaphora, including oblique recipients."

In Ayta Abellen, pro-drop is also allowed in any type of sentence and for any type of argument as evidenced by the deletion of the subject in (191a), the direct object in (191b), and the indirect object in (191c).

(191) a. *Ket ahe ana n-an-Ø-dawa.*
           and not now PFV-PL-AV-fruit
           'And [it] did not now bear fruit.'

       b. *K<ing>wa   ko=yna.*
           <PFV.OV>-take 1SG.GEN=already
           'I took [it] already.'

       c. *Haanin, Ø-i-byay   mo=yna=y   haa.*
           now   CTPLT-CV-give 2SG.GEN=now=NOM banana
           'Now, you give the bananas [to him].'

E.11.3 Auxiliaries

Some languages have auxiliary elements that are separate words, and some can have more than one auxiliary element before the main verb. Examples for English transitive single-clause sentences include: *The boy will hit the ball*, *The boy might have hit the ball* and *The boy might have been hitting the ball*.

In some languages verbal negation is an auxiliary element, while in others it is simply a verbal affix. English has both the negative auxiliary *not*, which is a separate word that must always follow another auxiliary, and the contraction form *-n't*, which is written as a suffix on the auxiliary. Examples of English negative transitive single-clause sentences include:*The boy did not hit the ball*, *The boy won't hit the ball* and *The boy might not have hit the ball*.

Tagalog has a negative auxilary *hindi* 'not'. This is the only type of verbal negation in the language.

Ayta Abellen has a negative auxiliary *ahe* 'not' which is written as a separate word, and that is the only type of verbal negation in the language. There is no negative verbal affix nor negative auxiliary affix. The negative auxiliary may be the only auxiliary in the phrase; no other auxiliary is required.

See section 16.2 for examples of these negative elements used in sentences and further discussion of the negation constructions in Ayta Abellen.

E.11.4 Intransitives and Motion Verbs

Simple intransitives do not have any complements after the verb, such as in *The girl swims*. Motion verbs include a location in the form of either a PP or adverb. English examples of motion verbs include *The girl ran around the track*, *My mother went to the store* and *My teacher came here*.

Examples of Tagalog intransitive verb sentences are:

(192) *Nag-la~laro   ang   mga bata.*
      AV-CONT~play NOM PL   child
      'The children are playing.'

(193) *Um-upo   siya.*
      PFV.AV-sit   3SG.NOM
      'He sat down.'

Examples of simple intransitive sentences in Ayta Abellen include:

(194) a. *T<in><om>angoy ya.*
        <PFV><AV>-swim   3SG.NOM
        'He swam.'

     b. *Am-pang-Ø-an   hila   po.*
        CONT-PL-AV-eat 3PL.NOM still
        'They are still eating.'

An example of a Tagalog motion verb sentence is:

(195) *P<um>unta sila   sa   dagat.*
      <PFV.AV>-go 3PL.NOM DAT ocean
      'They went to the ocean.'

Ayta Abellen examples with motion verbs include:

(196) a. *N-Ø-ayew   ya=y   babayi ha   gilid dagat.*
         PFV-AV-run 3SG.NOM=NOM woman DAT   side   ocean
         'The woman ran to the seashore.'

     b. *N-Ø-akew   ya   ha   banwa.*
         PFV-AV-go 3SG.NOM DAT   city
         'He went to the city.'

     c. *Hiyay maihtodo ko,   n-Ø-akew ya   ihti.*
         TM   teacher   1SG.GEN PFV-Ø-go 3SG.NOM here
         'My teacher came here.'

E.11.5 Copular Constructions

Copular sentences can have an adjective, PP, or nominal phrase following the copular verb. Some languages do not require a copular verb or auxiliary to be present. English examples include *The girl is pretty*, *She seems nice*, *The girl is in charge* and *He is the leader*.

Tagalog copular sentences do not require a copular verb Kroeger (1998) although it has been proposed that Tagalog has a null copula which sometimes is overt in the form of *maging* 'become' Richards (2009).

Ayta Abellen copular sentences also do not require a copular verb as seen in (197) and (198).

(197) *Yabayin ye   baey   ko.*
     that   NOM house 1SG.GEN
     'That   is   my   house.'

(198) *Manged ye   pakibat mo.*
     good   NOM   answer 2SG.GEN
     'Your   answer is   good.'

E.11.6 Transitives and Ditransitives

Examples of English transitive single-clause sentences include *The batter will hit*

the ball, *My baby wants that toy* and *The dog might have buried his bone*.

Examples of Tagalog transitive sentences include:

(199) *B<in>ili   niya   ang   isda.*
      \<PFV.OV\>buy 3SG.GEN NOM fish
      'He bought the fish.'

Ayta Abellen examples of transitive sentences include:

(200) a. *H<in>aliw   na   ya=n   anak ko*
        \<PFV.OV\>buy 3SG.GEN 3SG.NOM=GEN child 1SG.POSS

        ye   digalo.
         NOM gift
        'My child purchased the gift.'

    b. *Ø-Bogbog-en   na   ya=n   Totoy*
        CTPLT-beat-OV 3SG.GEN 3SG.NOM=GEN Totoy

        ye   kaaway na.
         NOM enemy 3SG.GEN
        'Totoy will beat up his enemy.'

Ditransitives are verbs where a subject can take two objects. The second object can

be expressed as a PP or as a nominal phrase. Some languages, like English, allow both types

via dative movement, as in *The girl gave the gift to her mother* and *The girl gave her mother*

the gift.

For Tagalog ditransitive verbs both the theme and the recipient are nominal phrases. In conveyance voice constructions like (201) the theme is a nominative marked nominal phrase while the recipient is a dative marked nominal phrase.

(201) *I-bi~bigay   niya   ang   sapatos sa   nanay   niya.*
      CV-CTPLT~give 3SG.GEN NOM shoes   DAT mother 3SG.GEN
      'He will give the shoes to his mother.'

But for goal voice, the first object is a nominative marked nominal phrase with the semantic role of recipient while the second object is a genitive marked nominal phrase with the semantic role of theme, as in,

(202) *Bi~bigy-an   niya   ang   nanay niya=ng   sapatos.*
      CTPLT~give-GV 3SG.GEN NOM mother 3SG=GEN shoes
      'He will give his mother shoes.'

In Ayta Abellen, like Tagalog, both objects in a ditransitive clause are expressed as nominal phrases. When all three objects are full nominal phrases, the argument with the semantic role of actor is typically marked as topic and left dislocated as in (203a). When pronouns are used for both the actor and the recipient, the order of the arguments remains the same as in (203b). The order of arguments for ditransitive clauses does change, however, for different voices. With the goal voice example in (203a) the order is actor (genitive), recipient (nominative), theme (genitive). For object voice in (203c), the order is actor (genitive), theme (nominative), recipient (dative). This same order (with different case markings) is true for actor voice (203d).

(203) a. *Hiyay Pabling, Ø-biy-an   na   ya＝n   tapih*
        TM   Pabling CTPLT-give-GV 3SG.GEN 3SG.NOM=GEN cloth

        ye   indo   na.
        NOM mother 3SG.GEN
        'Pabling will give some cloth to his mother.'

    b. *Ø-Biy-an   na   ya＝n   tapih.*
        CTPLT-give-GV 3SG.GEN 3SG.NOM＝GEN cloth
        He   will   give   some cloth to his mother.

    c. *Ø-I-byay   mo＝y   kaganaan a*
        CTPLT-CV-give 2SG.GEN=NOM all   LNK

        bandi mo   kanlan   mangaidap.
        wealth 2SG.GEN DAT.PL poor
        'You give all your wealth to the poor.'

    d. *M-an-Ø-aliw   ya＝n   beyah*
        CTPLT-PL-AV-buy 3SG.NOM=GEN rice

        kanan ali   na.
        DAT   younger-sibling 3SG.GEN
        'He will buy rice from his younger sibling.'

E.12 Complement Clauses

There are various types of embedded clauses which are complements of main verbs. Complement clauses function as a argument of a verb. There is some degree of cross-linguistic similarity in the types of clausal complements that verbs subcategorize for. Adjectives acting as the predicate of the sentence may also subcategorize for the various types of clausal complements. Examples of each type of clausal complement will be given in the next section, followed by a discussion of the complementizers and their position in

declarative embedded clauses. See section 13 for similar discussion for questions and section 14 for relative clauses.

E.12.1 Types of Complement Clauses

Each of the following subsections exemplifies a particular subcategorization requirement for a verb (or other category of head - see section 15 for examples of Adverbial Clauses).

E.12.1.1 A Finite Complement

Verbs which take a finite complement, with or without a complementizer, include the following examples in English: *I believe Jill to be my friend*, *Jill insists (that) she is my friend* and *I know (that) he likes me*.

Examples of Tagalog verbs which take a finite complement include (204) and (205) where *na* is the complementizer.

(204) *Alam ko   na   da~Ø-rating   din ang   araw.*
      know 1SG.GEN COMP CTPLT~AV-arrive also NOM day
      'I know that the day will arrive.'

(205) *Na-ni~niwala   ako   na   hindi*
      STA-CONT~trust 1SG.NOM COMP NEG

       na-tu~tulog   ang   Diyos.
       STA-CONT-sleep NOM God
      'I believe that God does not sleep.'

Examples of this type in Ayta Abellen include (206a), (206b), and (206c) where *a* is the complementizer.

(206) a. *M-atanda-an*   *la*   *a*   *n-Ø-akew*   *ya*   *ihtew.*
         CTPLT-know-GV   3PL.GEN   REL   PFV-AV-go   3SG.NOM   there
         'They will know that he went there.'

     b. *Ket an-habi-en*   *la*   *a*   *an-topad-en*
         and   CONT-say-OV   3PL.GEN   REL   CONT-fulfil-OV

        la=y   *kaotohan.*
        3PL.GEN=NOM   law
        'And they are saying that they are fulfilling the law.'

     c. *Am-pag-Ø-habi*   *hila*   *a*   *m-ag-Ø-ilyadi*
         CONT-DUR-AV-say   3PL.NOM   REL   CTPLT-DUR-AV-become

        ka=n   *kapitan.*
        2SG.NOM=GEN   captain
        'They are saying that you will become captain.'

E.12.1.2 A Nonfinite Complement

Verbs which take only a nonfinite complement without a complementizer, but the subject is dropped when it is coreferent with the main clause subject, including the following examples in English: *I want to come to the party* and *I want Joe to come to the party*.

Examples in Tagalog include (207) where the subject is overt and (208) where the subject is coreferent with the main clause subject. The complement is connected to the main clause with the *=ng* linker which Richards (1999) labels an allomorph of the *na* complementizer.

(207) *B<in>a~balak*   *ni*   *Maria=ng*   *m-aka-Ø-pagaral*
        <OV>CONT~plan   GEN   Maria=LNK   CTPLT-APT-AV-study

       ang   *anak niya*   *sa*   *UP.*
      NOM   child   3SG.GEN   DAT   UP
      'Maria is planning for her child to be able to study at UP.'

(Example taken from Aldridge (2006).)

(208) *Gusto ni   Maria=ng b<um>ili   ng   libro.*
      want GEN Maria=LNK <PFV.AV>buy GEN book
      'Maria wants to buy a book.'

(Example taken from Aldridge (2005).)

Examples in Ayta Abellen do not have an overt subject but rather the subject as

coreferent with the main clause subject. The complements are linked to the main clause

using the forms $=y$ (209a) and $=n$ (209b).

(209) a. *Tanda ko=y   m-ag-Ø-habi   nin   Ayta.*
             know 1SG.GEN=NOM CTPLT-DUR-AV-speak GEN Ayta
             'I know how to speak Ayta.'

      b. *Ang-ihip-en   la   ya=n   Ø-paty-en.*
             CONT-think-OV 3PL.GEN 3SG.NOM=LNK CTPLT-die-OV
             'They are thinking of killing him.'

E.12.1.3 A Nonfinite Complement with a Coreferent Subject

Verbs which take only a nonfinite complement without a complementizer, and the

subject must be coreferent with the main clause subject and dropped, include the following

examples in English: *Joe tried to come to the party* and *I tried to think of more examples*.

Tagalog allows a nonfinite complement with a coreferent subject but requires a

complementizer as in (210). In this situation it can be said that there is obligatory control,

that the actor of the complement clause is obligated to be coreferential with the subject of

the matrix clause.

(210) *N-a-pilit-an*   *si*   *Charlie＝ng*   *hiram-in*
        PFV-STA-force-GV   NOM Charlie=COMP borrow-OV

        ang *pera* *sa*   *bangko.*
        NOM money DAT bank
       'Charlie was forced to borrow money from the bank.'

(Example taken from Kroeger (1993).)

The same is true for Ayta Abellen with a complementizer being required and the

actor of the complement clause being obligated to be coreferential with the subject of the

matrix clause, as in (211).

(211) *N-a-pilit*   *ya＝n*   *in-om-alih*   *ihtew.*
        PFV-STA.OV-force 3SG.NOM=COMP PFV-AV-leave there
       'He was forced to leave there.'

E.12.1.4 An Object plus a Complement Clause

Verbs which take a direct object as well as either a finite or nonfinite clause

complement include the following examples in English: *I will persuade Joe that he should*

come to the party and *I will persuade Joe to come to the party*.

Examples of this in Tagalog are:

(212) *S＜in＞abih-an* *ako*   *ni*   *Luz* *na*
        ＜PFV＞say-GV   1SG.NOM GEN Luz COMP

        i-b＜in＞igay *mo*   *na*   *ang* *pera* *kay* *Charlie.*
        CV＜PFV＞give 2SG.GEN already NOM money DAT Charlie
       'I was told by Luz that you already gave the money to Charlie.'

(Example taken from Kroeger (1998).)

(213) *H＜in＞imok*   *ni*   *Maria* *si*   *Juan＝g*   *suriin*
        ＜PFV.OV＞persuade GEN Maria NOM Juan=COMP examine

ng   bago=ng   doktor.
GEN new=LNK doctor
'Maria persuaded Juan to be examined by the new doctor.'

(Example taken from Miller (1988).)

Examples of this type in Ayta Abellen include:

(214) a. *An-habi-en   ko   kanan Alma a*
           CONT-say-OV 1SG.GEN DAT   Alma COMP

   m-Ø-akew   ya   ha   banhal   nan   katongno na.
   CTPLT-AV-go 3SG.NOM OBL wedding GEN sibling   3SG.GEN
   'I am telling Alma that she will go to her brother's wedding.'

   b. *Pilit-en   mo   hila=y   ta-tao*
       force-OV 2SG.GEN 3PL.NOM=NOM PL-person

   a   m-Ø-akew   ihti.
   COMP CTPLT-AV-go   here
   'You force the people to come here.'

E.12.2 Complementizer Position

English has two complementizers which mark declarative embedded clauses, *that* for finite clauses and *for* for nonfinite clauses. In other languages the complementizers for embedded declarative clauses may or may not be overt, and in some languages they are clitics which are written attached to another word.

In Tagalog there are two complementizer forms for declarative embedded clauses: 1) *na* 'that' and 2) *=ng* which Richards (1999) says "corresponds to the affix -*Ø*". Ayta Abellen similarly has two complementizer forms: 1) *a* 'that' 2) *=n* which is a contracted form of *nin*. It also does not have an explicit gloss in English. The complementizer occurs before the embedded clause.

E.13 Questions

This section considers the main clause and embedded clause varieties of Yes/No questions and content questions.

E.13.1 Yes/No Questions

In English, main clause yes/no questions are formed by moving an auxiliary verb in front of the subject. Other languages use a Yes/No question marker or special complementizer instead of the fronted auxiliary. This is the case in embedded Yes/No questions in English; no auxiliary fronting is allowed but *whether* or *if* fills the complementizer position. Since languages may act differently in main clause and embedded clause Yes/No questions, each will be covered separately.

E.13.1.1 Main Clause Yes/No Questions

English examples of main clause Yes/No questions, showing the required fronting of an auxiliary, include: *Are you going to the store?*, *Will Sue buy the present for her mother?* and *Could the boy have hit the ball?*

Tagalog Yes/No questions are formed by inserting the *ba* marker in the 2nd position as in,

(215) *Ikaw   ba   ang   bago=ng   dating?*
        2SG.NOM   QUES   NOM   new=LNK   arrival
        'Are you the new arrival?'

Questions can also be marked with a rising pitch on the last word even without the *ba* marker, as in,

(216) *N-aka-Ø-punta   kayo?*
     PFV-APT-AV-go 2PL.NOM
     'Were you able to go?'

In Ayta Abellen main clause yes/no questions, there are two question markers *nayi* and *laweh*. These question markers are written as separate words and occur in the second position of the sentence (or third if there is a preceding pronoun). In Ayta Abellen also a yes/no question can be marked with rising pitch on the last word, even without an overt question marker in the sentence.

Ayta Abellen examples of main clause yes/no questions include:

(217) a. *M-Ø-akew   ka   laweh ha   banwa?*
       CTPLT-AV-go 2SG.NOM QUES DAT city
       'Will you go to the city?'

    b. *N-Ø-akew   ka   nayi   ihti a   m-ang-Ø-an?*
       PFV-AV-go 2SG.NOM QUES here LNK CTPLT-PL-AV-eat
       'Did you come here to eat?'

E.13.1.2 Embedded Clause Yes/No Questions

English examples of embedded clause yes/no questions, showing the presence of the question markers or complementizers, include: *I wonder whether she will come*, *I will know if she goes* and *Sue wonders whether the boy could have hit the ball*.

Tagalog has a complementizer *kung* 'if, when' for embedded yes/no questions. An example is:

(218) *Hindi ko   alam kung   b<um>ili   ng   isda ang babae.*
    NEG   1SG.GEN know whether <PFV.AV>buy GEN fish NOM woman
    'I don't know whether the woman bought fish.'

(Example taken from Law and Gärtner (2005).)

In Ayta Abellen embedded clause yes/no questions, there is a complementizer *no*

for embedded question clauses. The complementizer is written as a separate word and occurs

before the rest of the embedded question.

Ayta Abellen examples of embedded clause yes/no questions include:

(219) a. *Tawan   no   m-Ø-akew   ya   ihtibay.*
        don't-know COMP CTPLT-AV-go 3SG.NOM here
        'I don't know if he will come here.'

   b. *Tanda ko   no   peteg ye   h<in>abi   na.*
        know   1SG.GEN whether true   NOM <PFV.OV>say 3SG.GEN
        'I know whether what he said is true.'

E.13.2 Content Questions

Content questions differ from yes/no questions in that they contain an interrogative

word or phrase, known as wh-phrases in English because they almost all begin with *wh*.

Tagalog has separate word interrogatives:

(220)   **Interrogatives used in content questions**

alin	'which'
ano	'what'
sino	'who'
kailan	'when'
saan	'where'
bakit	'why'
paano	'how'

The corresponding interrogative words in Ayta Abellen are shown in the following chart.

(221)   **Interrogatives used in content questions**

aya	'what'
hinya	'who'
makano, nakano	'when'
wayihtew	'where'
taket	'why'
wayemen	'how'

Of the examples of interrogative words above, most simply stand alone as pronouns or adverbs, but *how* can either be a manner adverb by itself or a degree word modifying an adjective, as in *how big* or a degree word modifying a quantifier, as in *how many children*; *what* can either be a pronoun by itself or a demonstrative modifying a noun, as in *what books*; *which* never stands alone but always modifies a noun as a demonstrative, as in *which boy*; and *whose* can be either a stand alone possessive pronoun, or it can fill the possessor position in a nominal phrase, as in *whose toys*.

In Tagalog, *paano* 'how' is used as an interrogative pronoun while *gaano* 'how' is used as a degree word modifying a quantity as in,

(222) *Paano ang   lakad niya?*
how   NOM walk   3SG.GEN
'How does he walk?'

(223) *Gaano kalayo ang   ilog?*
how   far   NOM river
'How far away is the river?'

(Examples taken from Schachter and Otanes (1972).)

Tagalog has interrogative words that can modify a noun, as in,

(224) *Alin   libro?*
which book
'Which book?'

The word *kanino* 'whose' fills the possessor position in a nominal phrase, as in,

(225) *Kanino = ng bahay ito?*
whose=LNK house   this
'Whose house is this?'

In Ayta Abellen the word *way-emen* 'how' is used both as an interrogative pronoun (226) and as a degree word modifying a quantity. When modifying a quantity, the word *wayemen* 'how' occurs only before the adjective as in (227),

(226) *Wayemen ye   pag-biay mo?*
how   NOM GER-live 2SG.GEN
'How is your living?'

(227) *Wayemen kakadang ye   labok na?*
how   long   NOM hair   3SG.GEN
'How long is his hair?'

Ayta Abellen has interrogative words that can modify a noun, as in,

(228) *Hinya=y   liblo yati?*
     whose=NOM book this
     'Whose book is this?'

The word *hinya* 'who, whose' fills the possessor position in a nominal phrase, as in,

(229) *Hinya=y   ngalan a   n-aka-Ø-holat   ihti?*
     whose=NOM name   LNK PFV-APT-AV-write here
     'Whose name is written here?'

Depending on the basic word order of the language, content questions may be formed with or without fronting an interrogative phrase. Head-final languages (SOV, OVS, and OSV) usually do not have movement of an interrogative phrase in content questions, while head-initial languages (SVO, VOS, and VSO) usually require fronting of an interrogative phrase. In languages without fronting for interrogatives, the phrase structure of a content question will either be the same as a Yes/No question with a question marker or auxiliary in the complementizer position or just the same as a normal sentence. In either case, the interrogative phrase appears in its normal argument positon. Many SOV languages optionally allow movement of an interrogative phrase to the position immediately before the verb.

In both main clause and embedded content questions in Tagalog and Ayta Abellen, an interrogative phrase moves to the front of the clause.

E.13.2.1 Main Clause Content Questions

English examples of main clause content questions, showing the required fronting

of an interrogative phrase, include: *Which boy hit the ball?*, *Who might have hit the ball?*,

What did the boy hit?, *When will she come?* and *Where is the leader?*

In Tagalog content questions there is fronting of the interrogative pronouns as in

(230). For cleft equative sentences like (231) the noun phrase is marked with nominative

case. When the noun phrase is a headless relative clause as in (232), the clause is also marked

with nominative case.

(230) *Kailan siya   da-Ø-rating?*
      when   3SG.NOM CTPLT-AV-arrive
      'When will he arrive?'

(231) *Nasaan ang   bangko?*
      where   NOM bank
      'Where is the bank?'

(232) *Ano   ang   n-a-kita   niya   doon?*
      what   NOM PFV-STA.OV-see 3SG.GEN there
      'What did   he   see   there?'

In Ayta Abellen main clause content questions there is also fronting of the

interrogative pronouns with no marker or auxiliary fronting as in (233). Interrogative cleft

sentences have similar case marking patterns to Tagalog as seen in (234) and (235).

(233) *Makano ya=n   m-Ø-akew   ihti?*
      when   3SG=COMP CTPLT-AV-go here
      'When will he come here?'

(234) *Wayihtew ya=y   anak ko?*
      where   NOM=NOM child 1SG.GEN
      'Where is my child?'

(235) *Hinya=y   n-a-kit   yo?*
      what=NOM PFV-STA.OV-see 2PL.GEN
      'What did you see?'

E.13.2.2 Embedded Clause Content Questions

English examples of embedded clause content questions, showing the required fronting of an interrogative phrase, include: *I know who hit the ball*, *I wonder when she will come* and *I will ask what the boy hit*.

Tagalog has embedded clause content questions, as in,

(236) *Magta-tanong   ako   kung   kanino=ng lapis   ito.*
      CTPLT.AV-ask 1SG.NOM COMP who=LNK   pencil this
      'I will ask whose pencil this is.'

The complementizer *kung* occurs before the interrogative pronoun.

In Ayta Abellen embedded clause content questions, there is a complementizer *no* 'if, when' written as a separate word. The complementizer occurs immediately before a fronted interrogative pronoun. A complementizer and fronted interrogative phrase are required in every embedded content question.

Ayta Abellen examples of embedded clause content questions include:

(237) a. *Tanda ko   no   hinya ka.*
        know 1SG.GEN COMP who 2SG.NOM
        'I know who you are.'

     b. *M-an-Ø-epet   ako   no   hinya=y*
        CTPLT-PL-AV-ask 1SG.NOM COMP who=NOM

> *n-am-Ø-yay   nin   pilak.*
> PFV-PL-AV-give   GEN   money
> 'I will ask who gave money.'

E.14 Relative Clauses

Relative clauses normally modify a head noun and have one of four structures:

- like a complement clause headed by 'that', as in *the man that came to the meeting late*;

- like a content question with 'who' or 'which' fronted, as in *the man who came to the meeting late*;

- (or for some languages other than English) like a complement clause but with a special relative marker occurring in the complementizer position, or possibly with two markers which occur on each sides of the relative clause;

- (or for some languages other than English) like a regular sentence with a relative marker attached to the verb.

Some languages also allow one of the above structures to occur by itself as a nominal phrase without a head noun to modify, usually called "headless" relative clauses.

"Tagalog and most Philippine-type languages are examples of languages which can relativize only on subjects" (Kroeger 2004:181). Tagalog has a relativizer ligature $=ng$ which can also occur as a separate word *na* depending on phonological environment of the preceding phoneme. Tagalog is an example of the third structure listed above. An example of a relative clause in Tagalog is:

(238) *Matalino ang lalaki=ng b<um>asa ng diyario.*
intelligent NOM man=REL <PFV.AV>read GEN newspaper
'The man who read a newspaper is intelligent.'

Tagalog does allow headless relative clauses, as in,

(239) *ang n-DUR-Ø-nakaw ng kotse mo*
NOM PFV-DUR-AV-steal GEN car 2SG.GEN
'the one who stole your car'

(Examples taken from Sabbagh (2008).)

Ayta Abellen also is an example of the third structure listed above. Only subjects

can be relativized. Relative clauses occur after the head noun. The relative marker *a* occurs

before the sentence part of the relative clause.

Examples of relative clauses in Ayta Abellen include:

(240) a. *kagalingan a d<in>yag la*
skillfulness REL <PFV.OV>do 3PL.GEN
'skillfulness that they did'

b. *laki a ibat ha lale*
man REL from DAT forest
'man from the forest'

Full sentences containing the above relative clauses are as follows:

(241) a. *Hiyay kagalingan a d<in>yag la,*
TM skillfulness REL <PFV.OV>do 3PL.GEN

 n-an-Ø-yag hila=n baey.
PFV-PL-AV-do 3PL.NOM=GEN house
'They skillfully made a house.'

b. *L<in><om>ateng ya=ynay laki a ibat ha lale.*
<PFV><AV>arrive 3SG.NOM=now man REL from DAT forest
'The man from the forest arrived.'

Ayta Abellen does allow headless relative clauses, as in (242).

(242) *Hila=y*   *n-an-Ø-yag*   *nin*   *tori,*
     3PL.NOM=NOM PFV-PL-AV-make GEN tower

     n-ang-a-Ø-igat   *hila.*
     PFV-PL-STA-AV-surprise 3PL.NOM
     'The ones who made the tower, they were surprised.'

E.15 Adverbial Clauses

This section considers the complements that can follow temporal and reason adverbs, which were introduced in section 4.1. All of these adverbial clauses (or phrases) act the same as simple adverbs, so they will be considered AdvPs with the appropriate type of complement. Adverbial clauses of these types normally may occur either sentence-initially or sentence-finally.

E.15.1 Adjuncts of Temporal Adverbs

English examples of temporal adverbs with adjuncts used in full sentences include:

1. [Since then], John has been moody.

2. John has been home [since last week].

3. [Since John has been home from the hospital] he is doing better.

4. John is feeling fine [now that he is home from the hospital].

5. The dog followed Sue everywhere [when she returned from college].

6. [When she returned from college], the dog followed Sue everywhere.

These examples show that particular temporal adverbs subcategorize for certain types of complements. In (1) *since* is followed by another temporal adverb, in (2) by a temporal nominal phrase, and in (3) it has a sentential adjunct. Examples (5) and (6) show that *when* also takes a finite sentential adjunct, whereas *now* in (4) has a finite clausal adjunct with a complementizer.

Tagalog temporal adverbs can occur sentence initial (243), sentence final (244), or sentence medial (245). Adverbial sentential adjuncts are common as in (246).

(243) *Mula noon, hindi siya   n-aka-pag-Ø-trabaho.*
      from then   not   3SG.NOM PFV-APT-DUR-AV-work
      'Since then, he has not been able to work.'

(244) *T<um>ira   ako   roon hanggang 1950.*
      <PFV.AV>live   1SG.NOM   there until   1950
      'I lived there until 1950.'

(245) *N-ag-Ø-bakasyon   si   Pedrong minsan   sa   Bagiuo.*
      PFV-DUR-AV-vacations   NOM Pedro   sometimes DAT Bagiuo
      'Pedro vacations sometimes in Bagiuo.'

(246) *Nang m-aka-Ø-lampas   ng   ala   una, pu-Ø-punta   na   tayo.*
      when CTPLT-APT-AV-pass GEN o'clock one CTPLT-AV-go   now 1PL.INCL
      'After one o'clock we will go.'

In Ayta Abellen, the preference is for temporal adverbial phrases to occur sentence initial (247a). Both a temporal adverb and an adverb sentential adjunct can occur before the main clause of the sentence (247b). The adverbial sentential adjuncts occur after the adverb but before the verb. Temporal adverbial phrases can also occur at the end of the sentence

(247c) but not sentence medial. Examples of temporal adverbial clauses in Ayta Abellen, included in full sentences are:

(247) a. *Ha   anti   ko   ha   lale,   am-pam-Ø-ahaka*
            DAT present 1SG.GEN DAT forest CONT-PL-AV-climbing

           ko   ha   matondol.
            1SG.NOM   DAT   mountain-peak
           'When I am in the forest, I am climbing to the mountain peak.'

       b. *Hatew, ha   n-an-Ø-ambot   hila=y   hahapon,*
            past   DAT   PFV-PL-AV-win   NOM.PL=NOM   Japanese

           in-om-alih   ya=y   General Mcarthur.
            PFV-AV-left 3SG=NOM General McArthur
           'Before, when the Japanese won, General McArthur left.'

       c. *Am-pag-Ø-painawa   kawo   po   nayi   angga haanin?*
            CONT-DUR-AV-rest 2PL.NOM still RHET until   now
           'Are you still resting until now?'

E.15.2 Complements of Reason Adverbs

Turning now to adverbial phrases or clauses which express reason or purpose, we find adverbs which stand alone, such as *therefore* and *so*, along with those that take various types of complement clauses. English examples of reason or purpose adverbs which take clausal complements include:

1. John rushed to the hospital [because Sue was in an accident].

2. [Since Jill said that she likes him], Joe smiles all the time.

3. Bill rushed to the airport [so that he could see Kay before she left].

4. The troops threw grenades at the enemy [in order to escape].

5. The troops threw grenades at the enemy [in order for the prisoners to escape].

In the examples above, *because* and *since* take only finite sentential complements and *so* takes a finite clausal complement with a complementizer. In contrast, *in order* takes either a nonfinite sentential complement with a pro-dropped subject or a nonfinite clausal complement with a complementizer.

Examples of Tagalog reason or purpose adverbial clauses are:

(248)   *Kaya, b<in>ili   niya   ang   lahat ng   sapatos.*
        so   PFV-buy-OV 3SG.GEN NOM all   GEN shoes
        'So she bought all the shoes.'

(249)   *Dahil   sa magandang pananalita niya,*
        Because of good   speaking   3SG.GEN,

         n-an-Ø-iwala   sila   sa   kaniya.
         PFV-PL-AV-believe 3SG.NOM DAT 3SG.DAT
         'Because of his beautiful words, they believed in him.'

(250)   *Ø-Puntah-an   niya   ang   barangay nila*
        CTPLT-go-GV 3SG.GEN NOM barangay 3PL.GEN

         upang   Ø-ipaliwanag   ang   g<in>awa   niya.
         in.order.to CTPLT-CV-explain NOM <PFV.OV>do 3SG.GEN
         'He will go to their barangay in order to explain what he did.'

Examples of reason or purpose adverbial clauses in Ayta Abellen, included in full sentences are:

(251) a. *Kayabay, n-apatey la   ya=y   Amerikano.*
        so   PFV-kill 3PL.GEN 3SG.NOM=NOM American
        'So, they killed the American.'

     b. *Oli   ha   kabiangan nin   nakem   na,*
        because DAT hardness   GEN inner.being 3SG.GEN

         n-a-poot-Ø   ya=y   indo   na.
         PFV-STA.OV-anger 3SG.NOM=NOM mother 3SG.GEN
         'Because of the hardness of his inner being, his mother became angry.'

c. *M-ag-Ø-pahyal   ya   ha   baey nan balatang*
   CTPLT-DUR-AV-visit 3SG.NOM DAT house GEN woman

   taomon na   m-ahadt-an   ti   labay na.
   so.that 3SG.GEN CTPLT-say-GV NOM want   3SG.GEN
   'He will visit the house of the young woman so that he will say what he wants.'

E.16 Negation Constructions

All languages have the ability to express negation. This is usually achieved through the use of:

- special negative degree words, such as *not* (section 3.1), or negative quantifier determiners, such as *no* (section 3.2), which negate a nominal phrase,

- negative indefinite pronouns, such as *noboby* or *nothing* (section 8.6), which express a negative nominal argument by themselves,

- negative adverbs, such as *never* (section 4.1) or *nowhere* (section 4.2),

- negative auxiliaries, such as *not* or the contraction affixal form *-n't* (section 11.3), and/or

- negative verbal affixes (section 11.3).

All of the above have been covered as individual items in previous sections (as noted). After discussing the type of negation system Ayta Abellen uses, the data given will be reviewed here by category or phrase type, as well as adding examples used in complete sentences and discussing any co-occurrence restrictions between the various negative elements.

E.16.1 Type of Negation System

Languages fall into two major classifications with respect to negation: Polarity

Changing languages versus Negative Concord languages, exemplified by standard English

and Spanish or Black English, respectively.

Polarity Changing languages normally allow only a single negative element in the

sentence for negation to be expressed (though technically any odd number of negatives

expresses negative polarity) because the presence of two (or any even number of) negative

elements changes the polarity back to positive. For example, *John saw nothing* is negative,

but *John didn't see nothing* is positive because it means the same as *John saw something*.

The version with a negative auxiliary is expressed as *John didn't see anything*, using what

is known as a negative polarity item in object position. Such negative polarity items may

occur as subjects in some contexts without any other negation present, such as *Anyone may*

come, but they may only occur in object position when some higher (technically

c-commanding) element is negative. For example, one can say *Nobody saw anyone* and

Nowhere will they find anyone but without the negative subject or fronted negative adverb

the negative auxiliary must be used, as in *We did not see anyone* and *They will not find*

anyone anywhere.

In Negative Concord languages, any number of negative elements expresses negative

polarity, so in a Negative Concord version of English (such as Black English) *John didn't*

see nothing is the correct way to express a negative sentence with a negative object. In these

languages, similarly to the restriction on negative polarity items in Polarity Changing languages, a negative element may normally occur in subject position or a fronted position without any other negative element present, but a negative object may only occur when some higher (technically c-commanding) element is negative. Some Negative Concord languages have the stronger restriction that any negative argument must co-occur with verbal negation and still others require fronting of the negative argument, similarly to the fronting of question words or phrases in content questions.

Tagalog and Ayta Abellen are Polarity Changing languages, since the polarity changes between positive and negative with each instance of negation. Tagalog has negative polarity items in: *kahit anuman* 'whatever', *kahit sinuman* 'whoever', *kahit kailanman* 'whenever', and *kahit saanman* 'wherever'. The equivalents in Ayta Abellen are: *agya hinyaman* 'whatever, whoever', *agya makanoman* 'whenever', and *agya wayihtewman* 'wherever'. The negative pronouns, determiners, and adverbs may occur in object position or the normal adverb position even when there is another negative element in the sentence.

E.16.2 Auxiliary and Verbal Negation

Tagalog has the negative auxiliary *hindi* which is written as a separate word. Examples in full sentences are:

(252) *Hindi pwede yan.*
      not   possible that
      'That's not   possible.'

(253) *Hindi siya   masipag.*
     not   3SG.NOM industrious
    'He   is   not   industrious.'

We saw in section 11.3 that Ayta Abellen has one negative auxiliary which is written as a separate word, and that is the only type of verbal negation in the language. There is no negative verbal affix nor negative auxiliary affix. The negative auxiliary may be the only auxiliary in the phrase; no other auxiliary is required. The negative auxiliary word in Ayta Abellen is *ahe* 'not'.

Examples of the negative auxiliary word used in a complete sentence include:

(254) a. *Ahe malyadi yain.*
      not   possible that
      'That is not possible.'

   b. *Ahe ya   n-ang-Ø-an   nangon.*
      not   3SG.NOM PFV-PL-AV-eat earlier
      'He did not eat earlier today.'

E.17 Coordination Constructions

This section considers basic coordination constructions where there is an overt conjunction between the conjuncts, including coordination at the sentence level and at the verb phrase, nominal phrase, and adjective level. More complex types of coordination constructions, such as lists without overt conjunctions between each conjunct, gapping, appositives and comparatives, will need to be dealt with later.

In English, the conjunctions which can be used between two conjuncts are*: and*, *or*, *but* and *and/or*. In addition, there are coordination constructions which require an initial

conjunction, such as *either...or* and *neither...nor* constructions. Some languages also use a word meaning *with* as a conjunction between nominal phrases.

Ayta Abellen uses the following conjunctions between conjuncts:

(255)   **Coordinating conjunctions**

Ayta Abellen	meaning
boy *ket* *ta*	'and'
o	'or'
noba	'but'

E.17.1 Sentence-level coordination

Sentence-level coordination includes most types of sentences, including declaratives and questions. Some English examples are:

1. John went to the office this morning but he did not return.

2. When did the mail come and where is my package?

3. Either I will come to the party or I will send my sister.

Tagalog has sentence-level coordination constructions using conjunctions such as:

at (a), *o* (b), and *pero* (c).

(256) *T<um>ahol   ang   aso at n-a-takot-Ø   ang   pusa.*
      <PFV.AV>bark NOM dog and PFV-STA.OV-afraid NOM cat
      'The dog barked and the cat got scared.'

(257) *M-Ø-agaaral   ka   ba   o m-a-tutulog.*
      CTPLT-AV-study 2SG.NOM QUES or CTPLT-STA.OV-sleep
      'Are you going to study or sleep?'

(258) *T<um>awag ako   kaninang umaga pero wala   ka.*
       <PFV.AV>call 1SG.NOM earlier   morning but   not.exist 2SG.NOM
       'I called this morning, but you were not there.'

(Tagalog examples taken from Gallo-Crail (2010).)

Ayta Abellen examples of sentence-level coordination include:

(259) a. *Ket am-paty-en   la   hila   boy*
          and CONT-kill-OV 3PL.GEN 3PL.NOM and

          antoyhoken   la=n   etak.
          CONT-stab-OV 3PL=GEN machete
          'And they are killing them and stabbing them with machetes.'

    b. *Malabong ana=n   pag-obda noba ahe ko   po*
          many   now=LNK GER-work but   NEG 1SG.GEN yet

          malyadi=n m-ag-Ø-obda.
          can   CTPLT-DUR-AV-work
          'Much now is the working but I still cannot work.'

E.17.2 Verb clause coordination

Verb phrase coordination is exemplified in the following English sentences:

1. Bill ran the race and won the prize.

2. I will come to the party or send my sister.

3. I will not come to the party nor send my sister.

Tagalog allows verb clause coordination where the same subject is used for two

clauses. Examples include:

(260) *P<um>unta na kami   sa   may   tindahan*
       <PFV.AV>go   now 1PL.EXCL.NOM DAT EXIST store

at b<um>ili   ng   inumin.
and <PFV.AV>buy GEN drink
'We went to the store and bought something to drink.'

Ayta Abellen examples with verb clause coordination include:

(261) a. *Hiyay lota ye   am-paka-Ø-pa-tobo*
TM   soil NOM CONT-APT-AV-CAUS-grow

boy am-Ø-pakapadawa nin   bini.
and CONT-AV-yield   GEN seed
'The soil is what makes the seed grow and bear fruit.'

b. *Ahe   m-a-lokoloko*
NEG CTPLT-STA.OV-cheat

boy ahe   m-aki-Ø-halehale   ha   Iloko.
and NEG CTPLT-REQ-AV-mix DAT Ilokano
'They will not be cheated and are not mixing with the Ilokanos.'

E.17.3 Nominal phrase coordination

Nominal phrase coordination is exemplified in the following English sentences:

1. Bill ran in the Boston Marathon and the Olympics.

2. My sister or I will be there.

3. Neither my sister nor I will be there.

4. I will come to the wedding and/or the reception.

Tagalog has nominal phrase coordination as can be seen in this example:

(262) *Hu-hugas-an   ko   ang   mga tenedor at   mga kutsara.*
CTPLT-wash-GV 1SG.GEN NOM PL   forks   and PL   spoons
'I will wash the forks and spoons.'

Ayta Abellen examples with nominal phrase coordination include:

(263) a. *M-ang-Ø-gawa*   *kayi=n*   *Ø-tamn-an*   *mi=n*
CTPLT-PL-AV-work   1PL.EXCL   CTPLT-plant-GV   1PL.EXCL=LNK

  pa-paday boy bo-botey boy nakahinadi.
  PL-rice   and PL-taro   and others
'We will work together to plant rice and taro and other things.'

b. *N-ag-Ø-kamain*   *ana=n*   *Iloko*   *boy Ayta.*
PFV-DUR-AV-have now=GEN Ilokano and Ayta
'There came to be now Ilokanos and Aytas.'

E.17.4 Adjective coordination

Adjective coordination is exemplified in the following English sentences:

1. The black and white dog belongs to my son.

2. The sore on your leg is ugly and painful and dangerous.

3. My sister wants a red or blue candle.

Tagalog adjective coordination is done with the conjunction *at*, as in:

(264) *Maganda at   mayaman si   Rosa.*
  beautiful   and rich   NOM Rosa
  'Rosa is beautiful and rich.'

(Tagalog example from Schachter and Otanes (1972).)

Ayta Abellen adjective coordination is done with the conjunction *boy*, as in:

(265) a. *manga-yadet boy manga-tampa a   baey*
      PL-big   and PL-beautiful   LNK house
      'big and beautiful house'

E.18 Topic and Focus Constructions

This section considers two types of constructions which draw attention to a particular element. Since topics occur outside of focus constructions, they will be considered first. The following example shows that a single sentence may have both a topic and a focused phrase. In this case, the topic *Bill* is introduced by the topic marker *as for*, then the focused phrase *soccer*, which has been moved out of its regular place in the sentence follows.

- As for Bill, soccer he plays __ best.

E.18.1 Topics and Topic Markers

As defined here, topic constructions consist of a topic phrase followed by a complete sentence or question and usually set apart by punctuation. Certain markers may be used to set off topics, as shown in the following English examples:

1. As for John, I think he will make a good doctor.

2. Speaking of Jill, where is she?

Topic in Tagalog is marked syntactically through left dislocation with a resumptive pronoun or morphologically by the word *ay* which occurs after the topic phrase. This understanding of the "ay inversion" is supported by McKaughan and Al-Macaraya (1996), Kroeger (1993), Kaufman (2005), and Hirano (2006). Sentence examples include:

(266) *Si   Juan ay   isang magsasaka.*
      NOM Juan INV   a   farmer
      'Juan is a farmer.'

(267) *Ang   mga tao   rito   ay   mabait.*
      NOM   PL   person here INV   nice
      'The people here are nice.'

Ayta Abellen has topic phrases introduced by *hiyay* which occur before the main sentence and are separated by a pause. Some examples include:

(268) a. *Hiyay Pabling, n-Ø-akew   ya   ha   banwa.*
          TM   Pabling   PFV-AV-go   3SG.NOM DAT   city
          'Pabling went to the city.'

      b. *Hiyay laki, labay na   ya=n*
          TM   man   want   3SG.GEN 3SG.NOM=COMP

          kalogoden ti   babai.
          love-OV   NOM   woman
          'The man wants to be in love with the woman.'

      c. *Hiyay alaga nan   beyah ha   hatew,   baente pisos*
          TM   value GEN   rice   DAT   time.past twenty pesos

          ye   maghay kaban.
          NOM   one   kaban
          'The value of rice before was twenty pesos per kaban.'

In Ayta Abellen, the topic phrase is marked by certain words or phrases. The topic markers occur before the topic phrase itself. For more information about topic marking in Ayta Abellen, see Stone (2008).

(269)   **Topic markers**

Topic Markers	Meaning
hiyay	'SG'
hilay	'PL'

Topic Markers	Meaning
yati	'this'
yain	'that'
yatew	'that (far)'

E.18.2 Focused Phrases and Focus Markers

In contrast to topics, while focus constructions also have an element fronted which may be set off by punctuation, the phrase in focus is moved from its normal position in the sentence, leaving a gap. Certain markers, such as *only* in English, may be used to mark the focused phrase. Focus constructions are not used frequently in SVO languages, since the subject is usually in focus and is already first in the normal order. Other word orders, such as VSO languages, may make extensive use of focus constructions. English examples of focus constructions include:

1. Girls, I like __.

2. Soccer, Bill plays __ best.

3. Only that boy, Jerry hit __ .

Tagalog and Ayta Abellen do not allow a focus phrase to be moved before or after the rest of the sentence out of its normal position. Focus phrase constructions in these languages are usually cleft sentences as in (270) for Tagalog and (271) for Ayta Abellen.

(270) *Isda ang gusto ko.*
fish NOM want 1SG.GEN
'Fish is what I want.'

(271) *Yati ye   labay ko=n   habi-en.*
this NOM want 1SG.GEN=COMP say-OV
'This is what I want to say.'

E.19 Exclamations and Greetings

This section considers various types of common utterances which are not complete sentences. We will deal with greetings, interjections and exclamations in turn.

E.19.1 Greetings

Greetings typically consist of a particular word or phrase, either standing alone or followed by a name or nominal phrase describing the peson or animal being greeted. Some are full sentences or full questions, such as *How do you do? How are you?* and *Have a good day!* which do not need to be treated separately. Some examples of English greetings are:

1. Hi!

2. Howdy!

3. Hello, Jill.

4. Goodby.

5. Good day, sir.

6. Good morning, Bob.

7. Good afternoon.

8. Good evening.

9. Good night, sweet baby.

Some Tagalog examples are:

(272) a. *Kumusta?*
      how
      'How are you?'

   b. *Magandang umaga   po.*
      good   morning RESP
      'Good morning.'

Examples of greetings in Ayta Abellen include:

(273) a. *Komohta?*
      how-are
      'How are you?'

   b. *Koko=yna.*
      1SG=now
      'I will go now.'

E.19.2 Interjections

Interjections typically consist of a particular word or phrase that is used alone, usually with an exclamation point following it. Though many of the words in interjections are not used in any other situation, regular adjectives are also used in this way. Some examples of interjections in English are:

1. Ouch!

2. Gross!

3. Cool!

4. Jumping Jehoshaphat!

5. Jiminy Cricket!

Examples of interjections in Tagalog include:

(274) *Aba!*
     hey
     'Hey!'

(275) *Salamat!*
     thanks
     'Thanks!'

Examples of interjections in Ayta Abellen include:

(276) a. *Hi!*
         expr
         'Hi! (expression denoting disagreement)'

Yes and No answers to questions will be treated similarly to interjections. Ayta

Abellen uses the following words to express these answers:

(277)   **Interjections**

Ayta Abellen	meaning
Awo *Awobay*	'yes'
Ahe *Aliwa*	'no'

Appendix F

PAWS GENERATED PHRASE STRUCTURE RULES

S = (Conj) (PP_1) ((AdvP_2) AdvP_1) (PP_2) {IP / CP}

S = (Conj) (PP_1) ((AdvP_2) AdvP_1) TopicP (PP_2) {IP / CP}

S = S_1 Conj S_2

S = Greet (DP)

S = Intj / Adj

CP = (InitConj) CP_1 Conj CP_2

CP = C'

CP = DP C'

CP = PP C'

CP = AdvP C'

CP = C DP IP

CP = C PP IP

CP = C AdvP IP

CP = IP FocusP

C' = IP

C' = C IP

TopicP = DP

TopicP = TopicM DP

FocusP = FocusM DP

IP = I'

I' = VP

VP = (InitConj) VP_1 Conj VP_2

VP = VP_1 AdvP

VP = VP_1 PP

VP = V DP

VP = V

VP = V DP {PP / AdvP}

VP = V {PP / AdvP}

VP = V {PP / AdvP}

VP = V DP {PP / AdjP / DP_1}

VP = V {PP / AdjP / DP}

VP = V {PP / AdjP / DP}

VP = DP {PP / AdjP / DP_1}

VP = {PP / DP}

VP = AdjP

VP = V {PP / AdjP / DP_1}

VP = {PP / AdjP / DP}

DP = (InitConj) DP_1 Conj DP_2

DP = Pron DP = {D' / D"}

DP = {D' / D"} Deg

D' = Det N'

D' = D'_1 CP

D' = (Dem) NP

NP = {N" / N'}

NP = {N" / N'} {DP / Pron}

NP = {Pron / Dem / Q / Num / Deg} (PP)

N" = QP N'

N" = N' QP

N'_1 = AdjP N'_2

N'_1 = N'_2 AdjP

N'_1 = VP N'_2

N'_1 = N'_2 VP

N' = ((N_3) N_2) N_1

N' = VP

PP = (InitConj) PP_1 Conj PP_2

PP = P'

P' = P DP

P' = P AdvP

AdjP = (InitConj) AdjP_1 Conj AdjP_2

AdjP = Adj'

AdjP = Deg Adj'

AdjP = Adj' Deg

Adj' = Adj

Adj' = Adj {CP / IP}

AdvP = Adv'

AdvP = Adv' Deg

Adv' = Adv

Adv' = Adv Adv_1

Adv' = Adv DP

Adv' = Adv IP

Adv' = Adv CP

QP = Q

QP = Deg Q

QP = Q Deg

QP = (((((Num_5) (Conj_4) Num_4) (Conj_3) Num_3) (Conj_2) Num_2) (Conj_1) Num_1)

(Conj) Num

QP = Adj (((((Num_5) (Conj_4) Num_4) (Conj_3) Num_3) (Conj_2) Num_2) (Conj_1)

Num_1) (Conj) Num

VP = V DP DP_1

VP = V DP_1

VP = V DP

VP = V (DP_1) DP

VP = V DP (DP_1)

VP = V DP_1

VP = V

VP = V DP

VP = V DP DP_1 DP_2

VP = V DP_1 DP_2

VP = V DP DP_1

VP = V DP DP_1

VP = V (DP_2) DP DP_1

VP = V DP (DP_2) DP_1

VP = V DP DP_1

VP = V DP_1

VP = V DP DP_1

VP = V DP {CP / IP}

VP = V {CP / IP}

VP = V DP DP_1 {CP / IP}

VP = V DP_1 {CP / IP}

DP = Case DP_1

VP = Case VP_1

D' = Dem Linker NP

DP = Deg Linker {D'/D"}

D' = D'_1 Linker V

Adv' = Adv'_1 Linker V

D' = Det Linker N'

N" = QP Linker N'

N" = N' Linker QP

N'_1 = AdjP Linker N'_2

N'_1 = N'_2 Linker AdjP

AdjP = Deg Linker Adj'

AdjP = Adj' Linker Deg

I' = I VP

I = Aux

I = Adv

I = I_1 2P

2P = Mood

2P = Pron

Adv = Adv_1 2P

V = V_1 2P

DP = DP_1 2P

VP = TopicP

D' = D'_1 VP

TopicP = (Deg) TopicM VP

TopicP = TopicM IP

TopicP = TopicP_1 (DP) Conj TopicP_2

TopicM = Dem Linker

VP = V VP_1

VP = V DP DP_1 DP_2

VP = V DP DP_1 DP_3 DP_2

References

Alcantara, Danniel, and Allan Borra. 2008a. Constituent structure for Filipino: Induction through probabilistic approaches. Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation (PACLIC22). Cebu, Philippines.

Alcantara, Danniel, and Allan Borra. 2008b. A greedy approach to unsupervised grammar induction for Filipino. Proceedings of the 5th National Natural Language Processing Symposium. Manila: De La Salle University.

Aldridge, Edith. 2002. *Ergativity and word order in Austronesian languages.* New York: Cornell University dissertation.

Aldridge, Edith. 2005. Phase theory account of absolutive extraction in Tagalog. *MIT Working Papers in Linguistics* 49.1-28.

Aldridge, Edith. 2006. Against case agreement in Tagalog. (http://ling.nthu.edu.tw/afla13/ abstracts/SFm-1.pdf). (Accessed Dec. 26, 2009.)

Anderson, Stephen R. 2008. Second position clitics in Tagalog. In S. Inkelas and K. Hanson, eds. *The nature of the word,* 549-66. Cambridge, MA: MIT Press.

Antworth, Evan. 1979. *Grammatical sketch of Botolan Sambal.* Manila: Linguistic Society of the Philippines.

Antworth, Evan. 1990. *PC-KIMMO: A two-level processor for morphological analysis.* Dallas: Summer Institute of Linguistics.

Barlaan, Rodolfo. 1999. *Aspects of focus in Isnag.* Manila: Linguistic Society of the Philippines.

Beesley, Kenneth R., and Lauri Karttunen. 2003. *Finite-state morphology.* Stanford, CA: Center for the Study of Language and Information Publications.

Bell, Sarah J. 1976. *Cebuano subjects in two frameworks.* Vancouver: University of British Columbia dissertation.

Black, Cheryl A., and H. Andrew Black. 2009. PAWS: Parser and writer for syntax: Drafting syntactic grammars in the third wave. SIL Forum for Language Fieldwork. (http://www.sil.org/silepubs/Pubs/51432/SILForum2009-002.pdf). (Accessed May 23, 2010.)

Black, H. Andrew. 2006. A conceptual introduction to FieldWorks Language Explorer Stage 1 Morpho-phonology. (http://www.sil.org/computing/fieldworks/flex/ GrammarConceptualIntro/ConceptualIntroduction.htm). (Accessed January 31, 2010.)

Black, H. Andrew. 2009. Writing Linguistic Papers in the Third Wave. SIL Forum for Language Fieldwork. (http://www.sil.org/silepubs/Pubs/52286/SILForum2009-004.pdf). (Accessed May 23, 2010.)

Black, H. Andrew, and Gary F. Simons. 2006. The SIL FieldWorks Language Explorer approach to morphological parsing. Computational Linguistics for Less-studied Languages: Proceedings of Texas Linguistics Society 10. Austin, TX.

Blake, Frank R. 1925. *A grammar of the Tagalog language.* New Haven: American Oriental Society.

Bloomfield, Leonard. 1917. *Tagalog texts with grammatical analysis.* Volume 3: University of Illinois Studies in Language and Literature.

Bonus, Don Erick J. 2004. The Tagalog stemming algorithm (TagSA). 1st National Natural Language Processing Symposium. Manila: De La Salle University.

Borra, Allan, Erwin O. Andrew, Chris Ian R. Lim, Richard Bryan S. Tan, Marlon Tong, and N. Cromwell. 2007. LFG-Based machine translation engine for English and Filipino. 4th National Natural Language Processing Symposium. Manila: De La Salle University.

Brainard, Sherri. 1994. *Voice and ergativity in Karao.* Eugene: University of Oregon dissertation.

Butt, Miriam, and Tracy Holloway King, Maria-Eugenia Nino, and Frederique Segond. 1999. *A grammar writer's cookbook.* Stanford, CA: Center for the Study of Language and Information Publications.

Carrier-Duncan, Jill. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16.1.1-34.

Ćavar, Damir. 2008. Efficient morphological parsing with a weighted finite state transducer. *Informatica* 33.107-13.

Cena, Resty M. 2006. A not-so-gentle morphological analyzer of Tagalog. International Conference on Austronesian Linguistics. Puerto Princessa, Philippines.

Crystal, David. 1997. *A dictionary of linguistics and phonetics. Fourth edn.* Cambridge, MA: Blackwell.

Crystal, David. 2008. *A dictionary of linguistics and phonetics. Sixth edn.* Cambridge, MA: Blackwell.

Errington, Ross. 1979. A transition network grammar of Cotabato Manobo. *Studies in Philippine Linguistics* 3(2).105-63.

Fortes, Farrah Cherry L. 2004. Optimality theory in morphological analysis. 1st National Natural Language Processing Symposium. Manila: De La Salle University.

Gallo-Crail, Rhodalyne. 2010. Coordinating conjunctions. (http://www.seasite.niu.edu/Tagalog/Grammar%20Activities/Grammar%202/Sentence%20Expansion/coordinating.htm). (Accessed January 28, 2010.)

Gault, JoAnn Marie. 1999. *An ergative description of Sama Bangingi'.* Manila: Philippine Journal of Linguistics. Special Monograph Issue, Number 46.

Gerdts, Donna B. 1988. Antipassives and causatives in Ilokano: Evidence for an ergative analysis. In Richard McGinn, eds. *Studies in Austronesian linguistics,* Ohio University Monographs in International Studies, Southeast Asia Series No. 76.295-321. Athens, OH: Ohio University Press.

Giganto, Reynaldo T. 2004. Exploiting structural similarities of Philippine languages for multilingual machine translation system. 1st National Natural Language Processing Symposium. Manila: De La Salle University.

Grimes, Joseph E. 1975. *Network grammars.* Norman, OK: Summer Institute of Linguistics.

Grimes, Joseph E. 1983. *Affix positions and cooccurrences: The PARADIGM program.* Arlington, TX: The Summer Institute of Linguistics and the University of Texas at Arlington.

Hammarström, Harald. 2009. Poor man's word-segmentation: Unsupervised morphological analysis for Indonesian. Third International Workshop on Malay and Indonesian Language Engineering (MALINDO). Singapore.

Haspelmath, Martin. 2000. *Indefinite pronouns.* Oxford: Oxford University Press.

Himmelmann, Nikolaus P. 1991. *The Philippine challenge to universal grammar.* Arbeitspapier 15. Ko□ln: Institut fu□r Sprachwissenschaft, Universita□t zu Ko□ln.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36.165-91.

Hirano, Takanori. 2006. Subject and topic in Tagalog. 10th International Conference on Austronesian Languages. Puerto Princessa, Philippines.

Kaufman, Daniel. 2005. Aspects of pragmatic focus in Tagalog. In I Wayan Arka and Malcom Ross, eds. *The many faces of Austronesian voice systems: Some new empirical studies,* 571.175-96. Canberra: Pacific Linguistics.

Koskenniemi, Kimmo. 1983. *Two-level morphology: A general computational model for word-form recognition and production.* Helsinki: University of Helsinki Department of General Linguistics.

Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog.* Stanford, California: Center for the Study of Language and Information Publications.

Kroeger, Paul. 1998. Clitics and clause structure in Tagalog. In Ma. Lourdes S. Bautista, ed. *Pagtanaw: Essays on language in honor of Teodoro A. Llamzon,* 53-72. Manila: Linguistic Society of the Philippines.

Kroeger, Paul. 2004. *Analyzing syntax: A lexical functional approach.* Cambridge: Cambridge University Press.

Kroeger, Paul. 2005. *Analyzing grammar: An introduction.* New York: Cambridge University Press.

Law, Paul, and Hans-Martin Gärtner. 2005. Postverbal WH-phrases in Malagasy, Tagalog and Tsou. Heinz, Jeffrey and Dimitris Ntelitheos, eds. *Proceedings of AFLA XII.* 12.211-26. Los Angeles: UCLA Working Papers in Linguistics.

Maclachlan, A. E. 1996. *Aspects of ergativity in Tagalog.* Montreal: McGill University dissertation.

Manguilimotan, Erlyn Q. 2007. Syntactic representation of simple verbal Tausug sentences using lexical-functional grammar formalism. 4th National Natural Language Processing Symposium. Manila: De La Salle University.

McConnel, Stephen. 1995. PC-PATR reference manual: A unification based syntactic parser. (http://www.sil.org/pcpatr/manual/pcpatr.html). (Accessed May 18, 2009.)

McKaughan, Howard P., and Batua Al-Macaraya. 1996. *A Maranao dictionary.* Manila: De La Salle University Press, Inc and Summer Institute of Linguistics.

Miller, Barry W. 1988. *Non-configurationality in Tagalog.* Ann Arbor: University of Michigan dissertation.

Mirikitani, Leatrice T. 1972. *Kapampangan syntax.* Oceanic Linguistics Special Publication 10. Honolulu: University Press of Hawaii.

Nelson, Hans J. 2004. *A two-level engine for Tagalog morphology and a structured XML output for PC-KIMMO.* Provo, UT. Brigham Young University.

Newell, Leonard E., and Francis Bon'og Poligon. 1993. *Batad Ifugao dictionary with ethnographic notes.* Manila: Linguistic Society of the Philippines.

Nitsch, Wilhelm. 2009. Some Ayta Abellen grammar. (http://www.sil.org/silepubs/Pubs/51243/51243_NitschW_Ayta_Grammar.pdf). (Accessed May 15, 2009.)

Nitsch, Wilhelm. 1998a. Ayta Abellen morpophonemic findings. SIL, ms.

Nitsch, Wilhelm. 1998b. Ayta Abellen phonological statement. SIL, ms.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6.75-106.

Porter, M.F. 1980. An algorithm for suffix stripping. *Program* 14.130-37.

Prószéky, Gábor, and Balázs Kis. 1999. A unification-based approach to morpho-syntactic parsing of agglutinative and other (highly) inflectional languages. Robert Dale and Ken Church, eds. *37th Annual Meeting of the Association for Computational Linguistics.* 37.261-68. College Park, MD: Association for Computational Linguistics.

Rackowski, A. 2002. *The structure of Tagalog: Specificity, voice, and the distribution of arguments.* Cambridge, MA: MIT dissertation.

Richards, Norvin. 1999. Complementizer cliticization in Tagalog and English. C. Smallwood and C. Kitto, eds. *Proceedings of AFLA VI (The sixth meeting of the Austronesian Formal Linguistics Association).* 16(2).297-311. Toronto: University of Toronto.

Richards, Norvin. 2009. The Tagalog copula. Sandy Chung, Daniel Finer, Ileana Paul, and Eric Potsdam, eds. *Proceedings of the Sixteenth Meeting of the Austronesian Formal Linguistics Association.* 16.181-95. Santa Cruz: University of California.

Rubino, Carl Ralph Galvez. 1998. *Ilokano-English English-Ilocano dictionary and phrasebook.* New York: Hippocrene Books.

Sabbagh, Joseph. 2005. *Non-verbal argument structure: Evidence from Tagalog.* Cambridge, MA: MIT dissertation.

Sabbagh, Joseph. 2008. Right node raising and extraction in Tagalog. *Linguistic Inquiry* 39.502-11.

Sak, Haşim, Tunga Güngör, and Murat Saraçlar. 2009. A stochastic finite-state morphological parser for Turkish. *ACL-IJCNLP 2009 Conference.* 47.273-76. Suntec, Singapore: The Association for Computational Linguistics and The Asian Federation of Natural Language Processing.

Sapir, Edward. 1921. *Language: An introduction to the study of speech.* New York: Harcourt Brace & Company.

Schachter, Paul, and Fe T. Otanes. 1972. *Tagalog reference grammar.* Berkeley, CA: University of California Press.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In Charles N. Lii, eds. *Subject and Topic,* 493-518. New York: Academic Press.

Shieber, Stuart M. 1986. *An introduction to unification-based approaches to grammar. Volume 4 of CSLI Lecture Notes Series.* Stanford, California: Center for the Study of Language and Information.

Simons, Gary F., and H. Andrew Black. 2009. Third wave writing and publishing. SIL Forum for Language Fieldwork. (http://www.sil.org/silepubs/Pubs/52287/SILForum2009-005.pdf). (Accessed May 23, 2010.)

Starosta, Stanley. 2002. Austronesian 'focus' as derivation: Evidence from nominalization. *Language and Linguistics* 3.2.427-79.

Starosta, Stanley, Andrew K. Pawley, and Lawrence A. Reid. 1982. The evolution of focus in Austronesian. Third International Conference on Austronesian Linguistics. Canberra, Australia.

Stone, Roger. 2004. Exploring Ayta Abellen verbal morphology. SIL, ms.

Stone, Roger. 2007. Analyzing Ayta Abellen narratives for peak, participant reference, information type and fronting. Graduate Institute of Applied Linguistics, ms.

Stone, Roger. 2008. The marked topic slot in Ayta Abellen. (http://www.gial.edu/GIALens/vol2-3/Stone-Topic-Focus-Ayta-Abellen.pdf). (Accessed May 18, 2009.)

Stone, Roger. 2009. The Sambalic languages of central Luzon. *Studies in Philippine Languages & Cultures* 19.158-83.

Uliniansyah, Mohammad Teduh, Shun Ishizaki and Kiyoko Uchiyama. 2002. Indonesian morphological parser with minimum connectivity cost to solve ambiguities. Mitsuru Ishizuka and Abdul Sattar, eds. *7th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence.* 2417/2002.606. London: Springer-Verlag.

Walrod, Michael R. 1979. *Discourse grammar of Ga'dang.* Arlington: The Summer Institute of Linguistics and the University of Texas at Arlington.

Weber, David J., and William C. Mann. 1979. *Prospects for computer-assisted dialect adaption.* Notes on Linguistics Special Publication 1. Dallas: Summer Institute of Linguistics.

Weber, David J., H. Andrew Black, and Stephen R. McConnel. 1988. *AMPLE: A tool for exploring morphology.* Dallas: Summer Institute of Linguistics.

Winograd, Terry. 1971. *Procedures as a representation for data in a computer program for understanding natural language.* Cambridge, Massachusetts: MIT.

Wolfenden, Elmer. 1961. *A re-statement of Tagalog grammar.* Manila: Summer Institute of Linguistics and the Institute of National Language.

Woods, W. A. 1970. Transition network grammars for natural language analysis. *Communications of the ACM* 13(10).591-606. New York.