Two Ontological Approaches to Building an Integrated Semantic Network for Yami ka-Verbs

Meng-Chien Yang, S-W Huang
Department of Computer Science and Communication Engineering
Providence University Taichung, Taiwan, ROC
mcyang2@pu.edu.tw, g9873006@pu.edu.tw

D. Victoria Rau
Institute of Linguistics
National Chung Cheng University Chiayi, Taiwan, ROC Lngrau@ccu.edu.tw

Abstract—This paper describes a proposed ontological language processing system for integrating two semantic sets for a group of important verbs with the prefix ka- in Yami, an Austronesian language in Taiwan. The two semantic sets represent two different classification approaches. One approach follows the concepts and rules of WordNet and the other uses the metaphors in Yami indigenous knowledge. The ontologies are used for classification and semantic integration. The results of implementation are used for building the Yami lexical database. This paper illustrates how the methodology and framework used in classifying Yami can be applied to Austronesia language processing.

Keywords: Austronesian language, natural language processing, ontology

I. INTRODUCTION

Building on past field studies and work on language documentation, Austronesian researchers have recently been using semantic analysis to develop resources from natural language processing for Austronesian languages. However, unlike natural language processing for major languages, such as English or Chinese, the study of natural language processing for Austronesian languages still faces many uncertainties. One of the major problems is that the necessary corpora for many Austronesian languages are still being developed. Therefore, to develop language resources for Austronesian languages, one must use new approaches and bring in methodology from artificial intelligence, such as ontology.

Our research team has studied the Yami language for a decade. Yami is an endangered Austronesian language spoken on Orchid Island in Taiwan. It is closely related to Itbayaten. Ivatan and other Bashiic languages spoken in the Philippines. We have developed the first version of a Yami online corpus, Yami online dictionary, and the Yami eLearning system [14] [15].

This paper aims to describe the ongoing work toward developing a Yami language lexical database, i.e., Yami WordNet. An online lexical database for a specific language is an invaluable resource for many research projects, web applications and natural language processing applications. For example, WordNet [3], a lexical database for the English language produced by Princeton University, has led to many new applications and research projects [18]. Many researchers are following the WordNet experience to produce online lexical databases for other languages (e.g., EuroNet). As a result, researchers in major languages have access to many archives and documents from the past to support their semantic analyses [8].

Creating an online lexical database for Austronesian languages, on the other hand, is a totally different story. Most Austronesian languages spoken by Asian indigenous people are classified as endangered languages. An endangered indigenous language is used by a small group of people and is likely to be extinct in the next decade or so. It is a very challenging task to collect a comprehensive corpus for the endangered language, as the scarce resources and lack of archives present a problem in establishing a sizable corpus. Therefore, many organizations, such as SIL, DOBES and UNESCO, have sponsored projects to help researchers carry out endangered language documentation [1]. However, few research teams have focused on studying the semantics of these languages. As a result, it is very difficult to link the information in these endangered languages to major languages. In our previous studies, we have proposed a framework for building the lexical database for Austronesian languages. This paper describes a semantic analysis of the ka- verbs in Yami.

Analyzing and classifying the verbs is an important task for building the lexical database for a language. In this paper we chose the most polysemous verbal prefix ka- in Yami, and seek to group them as verb sets with different semantics. The related studies of Yami verbs are described in Section II.

In this study, we used two approaches to classify the verb sets. The first approach follows the same approach used in WordNet and the second one is a newly proposed method using Yami cultural metaphors. The methods and results are described in Section III.

The results from the two classifications are two groups of semantic sets. These sets were put into a single ontology. In the process, an algorithm for building links between verbs in different groups was also developed. The final results were generated by a process of ontology integration. The details of the implementation are described in Section IV. The results of the semantic integration can be used to find the different meanings of the verbs with ka- prefix in phrases automatically. The discussion and applications are in Section V. Conclusion and directions for future work are in Section VI.
II. RELATED STUDY IN AUSTRONESIAN LANGUAGES

The discussion of Yami verbs should be put in the context of Austronesian verbs in typology. The study of verbs expressing motion events originated with the study of Talmy [20]. He proposed a classification between two major patterns, V-languages, or verb-framed languages and S-languages, or satellite languages. The V-language is a path or frame event language and has more elaborations, e.g., Spanish. In contrast, the S-language conveys more manners and attitudes, e.g., English.

According to the study by Huang and Tanangkinsing [5], examining six Western Austronesian languages, path information is given more attention in these languages. As a result, they claimed that the Proto-Austronesian language was probably a path-salient language. Furthermore, Rau, Wang, and Chang also found that Yami is a path-salient language [17]. The results can be helpful in classifying verbs in Yami.

The study of the semantics in Yami is still at its very early stage. Currently, we are focusing on proposing models to construct Yami WordNet and study Yami ontology [13] [24]. Therefore, ontology is an important tool for our study. A detailed step-by-step procedure of ontology building can be found in Noy and McGuinness [10].

Studies of metaphors have had a long historical link with cultural studies [7]. Some studies on indigenous cultures regarded the metaphors of the indigenous culture as parts of the ecological system [1]. In the next section, we will show how to use the ecological system in Yami culture as a starting point to establish our verbal classification system.

III. KĀ-PREFIX IN YAMI LANGUAGE AND CLASSIFICATION

A. Kā-Prefix in Yami Language

The reasons that we selected verbs with kā-prefix for semantic analysis are as follows. First, the prefix kā-occurs with verbs and nouns most frequently. Second, kā-is the most polysemous morpheme in the Yami language. Verbs with the kā-prefix are varied, with very different meanings and aspects. This analysis will help determine the information connected with kā-. Thirdly, the meanings of verbs with kā- have been archived and documented by our research team. The study of verbs with kā-prefix can be found in Rau and Dong’s unpublished manuscript, entitled “Serves you right: Kā-as an attitudinal marker in Yami.” [16]

B. Classification following the Wordnet approach

The classification of Yami kā- verbs followed the verb categories in WordNet. Using these verb categories, the verbs with kā-prefix can be categorized into 14 classes. These classes are motion, perception, contact, communication, competition, change, cognition, consumption, creation, emotion, possession, bodily care, social behavior and interactions. The verbs with kā-were put into categories following these guidelines:

- Put the verb in the class where the Chinese translation matches the meaning of the class. For example, the verbs put in the “motion” class include: kacinon (then weave), kadasan (have a bowel movement), kahai (then peel skin and pulp to keep fiber), etc.
- Put the verb in several different classes, if the verb matches the Chinese meaning of these classes. For example, kapivatek (then carve) was put in two classes, the “motion” class and the “creation” class.

The relations and categories are shown in Figure 1. Based on the results of the classification, we can find that verbs in Yami have different semantics from verbs of English when these verbs were put in the same classes. This observation is consistent with the findings in Chung [2], which showed that the structure of ontological source domain cannot predict the specific and regional sense differences and construction differences. The classification following the WordNet categories could partially represent the relations and semantics of these Yami verbs. The rest of the information should be represented by a different classificatory approach of analyzing Yami metaphors and traditional knowledge.

C. Classification using Yami metaphors

The metaphors in Yami culture include many different aspects. In our study, we collected all documents describing Yami metaphors and used a quantitative analysis to analyze them. We transformed the frequency of a particular metaphor into a weight count to form a histogram of all counted metaphoric descriptions. After the histogram was produced, the major metaphors with the highest value were represented in the histogram. These metaphors were then validated by our Yami consultants. The process of validating these metaphors was transformed into a knowledge estimator that can evaluate the weight of these metaphors automatically.

The results of the analysis are shown in Figure 2. There is one core metaphor which is the origin of the other metaphors and three major metaphors which are regarded as mega-metaphors or a collection of metaphors. The core metaphor is the metaphor of Orchid Island. The three major metaphors consist of natural metaphors, legend metaphors, and ceremony metaphors. The collection of metaphors is represented as a tree structure. The leaf node is the fundamental metaphor. Each verb with a ka-prefix was classified and put into a leaf node. The verb set linked to a particular leaf node indicates that each verb in the set has a meaning connected to this metaphor.

The metaphors in the leaf node contain time frame motion, changes of nature, material metaphors,
physiological metaphors, food collection metaphors, taboo metaphors, ritual metaphors, hunting metaphors and fishing metaphors.

Figure 2. The Yami metaphors for classifying the verbs

IV. ONTOLOGICAL APPROACH FOR SEMANTIC INTEGRATION

A. Creating Ontologies for verbs with Ka-prefix

The steps for building the ontology of Yami verbs using the categories in WordNet follow the seven-step process proposed in our previous report [24]. The steps for building the ontology are as follows:

- Use the verbs of motion as the domain for creating the ontology;
- Develop connection links with existing domain knowledge ontologies. In our study, the interface connecting the common upper-level ontology, SUMO [9], is created using the approach developed in our previous study [23];
- Use the fourteen classes in the previous classification as the important terms;
- Define the class and class hierarchy; each class is connected with three major semantics, motion-interaction, cognition, and creation-possession;
- Define the properties of each class and define the properties of each verb class in the slot of the ontology;
- Place each verb in the appropriate category.

The results of the ontology building represent the semantics of the fourteen verb classes with ka-prefix. A part of the ontology with three classes is shown in Figure 3.

The ontology for the classification using the Yami metaphors follows the bottom-up approach, proposed in our previous study [24]. The bottom-up approach for creating the ontology starts with organizing and classifying the semantic entries in the database of the upper-level metaphors. The steps for using the bottom-up approach to create the ontology are described as follows:

- Build the small ontology for each metaphor in the leaf node;
- Build the ontology following the hierarchy of the structure of metaphors: build the ontology for the parent node of the leaf node, followed by the ontology for the upper level;
- Use ontology integration to merge the ontologies in the same level; after finishing all the merging operations, a new ontology will be established by merging all the individual ontologies;
- Generate each individual instance in the new ontology.

The final ontology which is created by merging all the small individual ontologies for each metaphor is the result of the whole ontology creation process.

B. Establish connections between the two ontologies

The connections of the two ontologies are regarded as the similarities between these two ontologies. Each possible close pair will be selected as a reference tool for semantic retrieval. The procedure for retrieving the following three sets of pairs as a reference tool is described as follows.

- If a verb is classified in a verb class set and a verb metaphor set then we pair the class and the metaphor.
- If there are several verbs in a verb class set, we will find whether these verbs are in other classes. These classes are grouped in a set and paired with each other class in this set.
- If there are several verbs in a verb metaphor set, we will find whether these verbs are in other metaphor classes. These metaphor classes are grouped in a set and paired with each other class in this set.
The reference tool was used to search for multiple semantics in the Yami texts. The functions of the reference tool are described in the next section.

V. IMPLEMENTATION

The ontologies with ka-prefix verb and the reference tool with the pair information are used for the following purposes:

- Multiple meaning extraction: to find multiple meanings of Yami phrases and sentences. The reference tool and the ontology can help extract different contextual information from the same sentence.
- Interface for connecting WordNet and Chinese BOW: the reference tool can be used as a set of retrieval keys for finding similar meaning verb sets in English and Chinese. It can be used for constructing the interface between Yami and the large corpora of English and Chinese.
- Lexicon development: the ontologies and the reference tool can be used to create lexical entities from the Yami language resources.

Currently, we are implementing the multiple meaning extraction system and conducting module testing. In the system, a Yami phrase with ka- verbs is matched with the ka-verb ontology. Using the induction rules in the ontology, a set of multiple meanings for this phrase is generated. Each meaning is a semantic entry and is used as the search key to search two lexical databases, the English WordNet and the Chinese BOW. In each search task, a set of corresponding Chinese semantic entries and a set of corresponding English entries are retrieved. The framework of the multiple meaning extraction system is shown in Figure 4.

![Image](image-url)

**Figure 4.** Ka-verb multiple meaning extraction system

VI. CONCLUSION AND FUTURE WORK

This paper has addressed the semantic processing of the ka-prefix verbs in Yami. We began with the process of classifying the verbs by two approaches, i.e., the verb categories in WordNet and the Yami cultural metaphors. Later, we showed how to use the classified verb sets to construct semantic ontologies. We also described the usages of the ontologies. This is a step toward building the lexical database for the Yami language. The study has paved the way for further study of lexical semantics and ontologies in Austronesian languages.

ACKNOWLEDGEMENT

This research was supported in part by Taiwan E-learning and Digital Archives Programs (TELDAP) sponsored by the National Science Council of Taiwan under the NSC Grant: NSC 99-2631-H-126-001. The authors would like to express our appreciation for the editorial service of Dr. Gerald Rau.

REFERENCES


