Analyzing and classifying the Yami emotion phrases using ontological structure and computation

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

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

Yi-Hsin Wu  
Institute of Linguistics  
National Chung Cheng University  
Chiayi, Taiwan, ROC  
yscatherine@gmail.com

Abstract—this paper describes an approach for using an ontological structure to analyze Yami ika- emotion phrases. The study of the emotion phrases of the Austronesian language is still in a very early stage. In this study, we designed an approach of using an ontology to simulate the previous study on emotion and cognition of another Philippine language. Based on the implementation and experimental results, the proposed method was found to be able to obtain similar semantic functions by clustering the Yami emotion phrases into twelve conceptual domains for representing the Yami emotion. These results can help us find the strategy for analyzing and computing other similarly under-researched or endangered Austronesian languages.

Keywords: Austronesian language, natural language processing, ontology, emotion analysis

I. INTRODUCTION

Numerous endangered languages are on the verge of extinction. The scarce web language resources for the endangered languages not only make these language communities totally invisible in the web community. The unbalanced distribution of resources can also worsen the situation by making the Internet virtually inaccessible for people using the endangered languages. Currently, many endangered languages are still understudied. Most research projects related to endangered languages focused on small-scale language documentation. However, the lack of large corpus presents a challenge for researchers doing semantics and ontological studies for this endangered languages[1].

In this paper, we report our progress in studying the emotion phrases comprising the morpheme ika- “the reason for feeling such and such” in the Yami language. In the past, the research team has studied Yami for a decade[13][14]. 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. The first version of a Yami online corpus, Yami online dictionary, and the Yami eLearning system have been developed[15].

This paper aims to describe one project in our recent 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[4], a lexical database for the English language produced by Princeton University. This fundamental lexical database can led to many important research project. 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[9]. 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].

To achieve this goal, we have developed a framework by using ontologies to support our study in analyzing the semantics in the Yami language[18]. In this study, we focus on building an ontological structure for the emotion phrases. Emotional concepts in languages and the phrases with emotion semantics have become the new focus in natural language processing. To do the natural language processing and to find the emotion phrases in the language is a very challenging task.

In this study, we describe our approach for studying the ika- prefix in Yami emotion phrases. The study follows Church et al’s cluster analysis for grouping emotion phrases[2]. In addition to using a corpus linguistic approach, we also developed an ontological framework for identifying the concept domains for the Yami emotion phrases. The results showed that we can find the diversity and categories of Yami emotion by using this integrated ontological approach. In Section II, we describe the whole framework of our study of the Yami ika- emotion phrases.

Section III describes the approach of using the ontological structure for measuring and classifying the ika-emotion phrases. The proposed method can help us bring the ideas of natural language processing into investigation of under-researched Asian language. The results of the clustering and discussion are shown in Section IV. Conclusion and directions for future works are in Section V.

II. STUDYING YAMI IKA- EMOTION PHRASES

Yami is a Philippine language with similar cognates as the Filipino. In a study of the emotion phrases of Filipino, Church et al. [2] found that the most important or
“hypercognized” [8] emotions in Filipino are anger, anxiety/fear, happiness, contentment, sadness, and arousal, whereas the relatively “hyponcognized” or minimally lexicalized emotion domains include feeling tired, guilty, surprised, contemptuous, and aspiring. They also recommended that the terms in all three subcategories of affective-cognitive states in Clore et al.’s study[3] taxonomy of emotion terms be viewed as referring to emotions. Although Church et al.[2] used a quantitative experimental method to validate Clore et al.’s emotion model, they relied on a translation of emotion terms from English rather than searching directly in Filipino.

Our study aims to analyze the Yami emotions encoded by the prefix ika- “the reason to be in such and such a state,” following Huang’s[6] study on emotion in Tsou by focusing on a grammatical model to conceptualize emotion concepts, complemented by an ontological and anthropological approach to categorizing Yami emotions. As Yami is the only Philippine language in Taiwan, it is important to compare the results of Yami emotions from a corpus linguistic approach with the results obtained by Church et al.[2].

With the guidance of Clore et al.’s [3], we coded our 258 tokens of emotion expressions from the 20 texts into 10 categories (7 categories of Clore et al. and two categories of interjections and metaphors/metonyms). The coding was jointly decided by the second and the third authors of this paper. The seven categories are presented in Table 1. The distinction was first made between internal and external conditions. Clore et al.’s [3] external conditions refer to subjective evaluations of character or stable characteristics (e.g., attractive, trustworthy) and objective conditions, such as things done to a person (e.g., abandoned, insulted), but we did not find any token in our data that would fit nicely in the category of external conditions. The internal conditions were further divided between mental and nonmental states. The internal nonmental states refer to physical and bodily states (e.g., sleepy, seasick). The internal mental states consist of affective conditions and cognitive conditions. Under the category of affective conditions, we further separated pure affective states (e.g., afraid, angry, happy) from affective-behavioral states and affective-cognitive states, depending on whether the affective emotion is followed by an action (e.g., scare away, fight) or a cognitive consequence of the emotion (e.g., impatient, sorrowful). The category of cognitive conditions, on the other hand, was further distinguished between pure cognitive states and cognitive-behavioral states with the former referring to the internal mental states in which cognition is dominant (e.g., smelly, stuffy) and the latter being followed by an action (e.g., picky, discreet). The very last two categories of emotion contain interjections/curse (e.g., ouch, damn) and metaphors/metonyms (e.g., the body is as healthy as light feathers or someone is being despised as goats’).

<table>
<thead>
<tr>
<th>Category</th>
<th>Yami example</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Internal nonmental states</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Internal mental states</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 1. affective conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.1 Pure affective states</td>
<td>ikaniahey</td>
<td>afraid</td>
</tr>
<tr>
<td>2.1.2 Affective-behavioral states</td>
<td>ikavozoaaw</td>
<td>scare away</td>
</tr>
<tr>
<td>2.1.3 Affective-cognitive states</td>
<td>ikaotok</td>
<td>impatient</td>
</tr>
<tr>
<td>2.2 Cognitive conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.1 Pure cognitive states</td>
<td>ikangot</td>
<td>smelly</td>
</tr>
<tr>
<td>2.2.2 Cognitive-behavioral states</td>
<td>ikapili</td>
<td>picky</td>
</tr>
<tr>
<td>3. Interjections</td>
<td>ananay</td>
<td>ouch</td>
</tr>
<tr>
<td>4. Metaphors/metonyms</td>
<td>nimanant so</td>
<td>terrified (in the</td>
</tr>
<tr>
<td></td>
<td>velek</td>
<td>stomach)</td>
</tr>
</tbody>
</table>

The discovery of most important emotions and the minimally lexicalized emotion domains in Yami generally matches Church et al.’s [2] findings in Filipino data except that there are no external conditions in Yami. As shown in Figure 1, Yami emotion constitutes three internal conditions: affective, cognitive, and physical with the three affective conditions and the two cognitive conditions “hyponcognized”. This finding supports Church et al.’s recommendation that the terms in all three subcategories of affective-cognitive states in Clore et al.’s [3] taxonomy of emotion terms be viewed as referring to emotions.

Our study is based on a cognitive linguistic exploration of the Yami emotion phrases by extracting data from Yami corpora, while Church et al.’s work gathered data from questionnaires filled out by Filipino college students. Their results of emotion phrases were mapped into the knowledge concepts. Our study includes two stages.

A preliminary classification of the Yami emotion terms found in the 20 texts and further examination of the other 146 narratives from the three Yami websites led us to find all relevant expressions of emotion. The list of lexical items extracted from events invoking emotion consisted of emotion related words, metaphors, and interjections. The final list of emotion terms was streamlined and yielded 126 emotion terms by excluding all words without the ika-prefix. This decision is to ensure an emic categorization of Yami emotion based purely on linguistic construction.

Secondly, a hierarchical cluster analysis was conducted to produce distinct domains of Yami emotion based on our classification of emotion terms. Following Church et al.’s procedure [2], a between-clusters linkage algorithm was calculated to produce comparable dendograms for comparison with Filipino emotions. The cross-relationship between the emotion terms was calculated to build the hierarchical structure by using the knowledge extracted from our proposed ontological computation procedure. In contrast to Church et al.’s questionnaire methods for data collection, our study used a bottom-up corpus approach to create and grow the ontologies of the emotion concepts manually. The process of the proposed approach will be described in Section III.

Table 1: Coding Categories for Yami Emotion
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 [14] [19]. Therefore, ontology is an important tool for our study. The construction of these ontologies follows the detailed step-by-step procedure in Noy and McGuinness [12]. In addition, our approach follows the idea of upper-level ontology[11] for sharing the meaning different languages. In one of the previous studies, we showed that meaning in Yami can be connected to meaning in Chinese and English [18]. We will describe the ontologies for the emotion phrases analysis in the following.

A. building the ontologies for emotion phrases

The ontologies for emotion concepts in our study include two major ontologies that created for this study and other ontologies from our previous projects. These two ontologies are used to create the influence rules for deriving the possible emotional concepts linked to these ika- emotion phrases of Yami. The first ontology is a semantics representation for the ika- emotion phrases. The ontology includes two parts, the indigenous knowledge, and the metaphors from the Yami contexts. Each Yami ika- emotion phrase is put into the leaf node following the hierarchy of the emotional structure shown in Figure 1. In addition, the emotion factors by the metaphor[9] are used for creating the second parts of the ontologies. The bottom node represents our findings of the categories of the Yami emotion terms (as shown in Figure 1). The top node represents the metaphorical structures of human and life proposed by Lakoff & Johnson’s [9] Conceptual Metaphor Theory. In our ontology, all the ika- emotion phrases were classified by using Lakoff and Johnson’s root analogies into four classes: (1) human = action, (2) human = expression, (3) human= relations, and (4) human = tools. We called this ontology the Yami emotion domain ontology (Figure 2). The hierarchical structure describes the classifications in our study and the emotion factors by the metaphor nodes [9].

The metaphorical structure put into the ontology is used as the assisted categories for the ika- emotion terms. All the ika- emotion terms were put into the sub-node of each class in the ontology, so that the ika- emotion terms with similar metaphors can be grouped together and put into a semantic set. This information can help us cluster these ika- emotional terms into different emotion conceptual domains.

In summary, our cluster analysis of Yami emotion identified the same 11 emotion domains as in Church et al.’s (1998) without the guilty domain. The most important emotions in Yami are anxiety/fear, arousal, contentment, anger, happiness, and sadness, whereas the relatively minimally lexicalized emotion domains include feeling tired, quiet/shy, contemptuous, aspiring. Our findings also confirm that the terms in all three subcategories of affective-cognitive states in Clore et al.’s[3] taxonomy of emotion terms be viewed as referring to emotions.

B. Emotion Concept Clustering By the Ontological Structure

We then proposed an approach of using the ontological structure to compute the judgment factors, so that we could use these factors for studying and calculating the conceptual structure of the Yami emotion terms. In the Yami emotion domain ontology, each ika- word was put in the leaf node with the upper-level nodes shown in Figure 1. Two weight factors were produced by calculating the length from the root (thing) of the ontology to each node and the correlational links, using the induction rules in the ontology to find the links between two nodes not in the same upper-level semantic node. Afterwards, the weight factors for each node in the English semantics ontology were calculated following the Church et al.’s categories and the leaves with the words adopted from Wordnet and the English translations of these ika- words. Later, the ontology merge process proposed in [12] was used for connecting the Yami emotion ontology and the English semantic ontology and creating the fifth factor, i.e., the merged influence factor. These five factors were used for the clustering process to put each ika- word into the category in the set of Church et al.’s emotion categories. In addition, the k-NNR clustering algorithm was used to find the subset in each category.

These factors were used to calculate the judgment values for evaluating whether an ika- emotion term could be put into a specific English emotion cluster.

IV. IMPLEMENTATION AND RESULTS

A. Implementation of Clustering

We implemented the ontologies and rules of inductions by using the protégé software and the clustering using the SPSS software.

The results show that the ika- emotion phrases can be classified by the proposed method.

B. Typological Representation of the Emotion Concepts

Figures 2 shows one of the summarized dendograms of our ontology experiments. In figure 2, the dendograms of the clusters of the happy domain, aroused domain, and contented domain are illustrated. The clusters in each domain are represented by using the key words and the numbers of Yami ika- words with the similar semantics.
shown in parentheses after the *ika*-word. The relative distance from each *ika*-emotion term with the English emotion terms was calculated by using these factors. The Yami emotion terms yielded 11 distinctive domains, almost identical to the Filipino emotion terms, except that the guilty domain is lacking in Yami.

The dendograms of this figure show the selected clusters in each domain. The scale shown is a simulation calculation derived by the weight function of semantic distance between each cluster in the emotional ontology. This simulation is used for emulating the judgment process in Church et al.’s study. If the weight of the lexical word is close to 0.9, it indicates the word fits better in its semantic domain. On the other hand, if the weight is close to 0.1, the word does not fit the domain as well.

**Figure 2. Dendogram of Happy, Aroused and Contented Concepts**

V. CONCLUSION

This study has classified Yami emotion into 6 internal conditions. Similar to Filipino, the most important emotions in Yami are anxiety/fear, arousal, contentment, anger, happiness, and sadness, whereas the relatively minimally lexicalized emotion domains include feeling tired, quiet/shy, contempluous, aspiring. In general, the Yami language contains more negative emotion terms than positive ones. Except for the guilty domain, the Yami emotions share the same 11 distinct domains as Filipino emotions.

The proposed approach by using the ontological structure to find the emotional concepts for the *ika*-emotion phrases shows an alternative approach for natural language processing for these languages without the large language resource. The same methods can be used for future development in conducting the semantic study of other 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 101-2631-H-126-001.

REFERENCES


