
Creating a Komering Dictionary based on the Semantic Domain Approach

Dita Dewi Palupi
Deny Arnos Kwary

English Department, Universitas Airlangga

Abstract

There are many endangered languages in the world, especially in Indonesia which has 719 individual languages. One of the endangered languages in Indonesia is Komering language, a language that exists in South Sumatera. This language is labeled 6b or Threatened by Ethnologue. This study aims to propose a concept for creating a dictionary of Komering language based on the semantic domain of daily life and using the software called WeSay. The theory of semantic domain is based on Moe (2003). By conducting an interview with four native Komering people, it is discovered that there are 114 words in the “food” subdomain. Thirteen words of this subdomain are found to have been excluded from the dictionary due to five possible causes. These causes are expected to be corrected in the future when making the dictionary. The dictionary database which has been constructed using WeSay is expected to add the resource for creating a better Komering – Indonesia language dictionary in the future.

Keywords: daily life domain, endangered language, food subdomain, Komering, semantic domains

Introduction

As the fourth biggest country in the world, Indonesia has 719 individual languages that are united by the Indonesian language as the lingua franca. From the 719 individual languages, 365 of them are living, 341 of them are endangered, and 13 of them are extinct (Ethnologue). One of the endangered languages is Komering language, an ethnic language that exists in South Sumatera. According to Ethnologue, a language is considered as endangered based on two characteristics; the number of speakers and the usage of the language by a particular speech community. When a language has fewer and fewer speakers and it is no longer serving its function as a mean of communication for the speech community that is used from one generation to another, so the language is endangered. Following the characteristics given by Ethnologue, thus Komering language is considered as endangered. If this language were not maintained, the language would be led to death.

There are many factors that can lead a language into death. Crystal (2000) argues that the death of a language can occur due to speakers’ reduction by catastrophic natural causes, disease, war, and conflict (p. 71). It also can happen because of the change of the culture by assimilation or globalization (p. 77). When there is nobody speak a language, thus the language is dead. Even if the speaker is only one person left, the language is still considered dead (Crystal, 2000, p.1).

In order to prevent language loss or language death, the language needs to be preserved. One way to preserve a language is to document it by making a dictionary of the language. According to Atkins and Rundell (2008), dictionary is “a description of the vocabulary used by members of a speech community” (p. 2). The maker of a dictionary is called lexicographer. In preserving endangered languages, a dictionary has a significant role. Dictionary assists in maintaining the existence of endangered languages since it serves a function as language documentation.

There are two ways in gathering the data for the dictionary making, i.e. using corpora and conducting an interview. Corpora are used when a lexicographer wants to make a dictionary for big languages, i.e. English. In the other hand, interviews are conducted when a lexicographer wants to make a dictionary for endangered languages, i.e., Komering. In conducting an interview, a lexicographer can use semantic domains for constructing the questionnaire. After doing the data gathering, a lexicographer can use software that can assist them in building a word database.

There are some previous researches which focus is creating a lexical resource. Warner, Butler, and Luna-Costillas (2006) attempt to revitalize a dormant language called Mutsun. They collect the data from recordings and written documentation since the last speaker of the language died in 1930. By using the data collected, they create an English-Mutsun Mutsun- English dictionary to help revitalizing the language. Then, Noor, Sapuan, and Bond (2011) try to create WordNet in the Indonesian language that also works for Malay language. They do it by combining information from some lexical resources, i.e. French-English-Malay dictionary, Kamus Melayu-Inggeris, and some WordNets in English, French, and Chinese versions. Then, bin Mohd Rosman, Bond, and Kratochvil (2014) try to create an open-source mapping between the SIL's semantic domains and WordNet. They do this by mapping the semantic domain files to WordNet files in both English and Indonesian versions. From these previous researches, it can be seen that there has not been any research on an endangered language in Indonesia that use semantic domains.

This paper aims at creating a dictionary of Komering language based on Moe's (2003) theory of the semantic domain. An open lexical resource software called WeSay is utilized in conducting this study as the database for the data acquired from interview. The importance of this paper is making an endangered ethnic language's dictionary in order to conserve the language. Thus, by focusing on the making of Komering-Indonesian dictionary, the author attempts to use semantic domain of daily life in constructing the questionnaire and build a word database by using a software called WeSay by SIL.

Method of the study

This study uses the qualitative approach since the author develops questions and procedures for the study, gathers data in the location where the informants stay, analyzes the data, and interprets the meaning of the data (Creswell, 2009: 4). Following this requirements, thus this study utilizes qualitative approach.

The data required is from Komering language, the endangered language used by Komering tribe in Ogan Komering Ulu, one of the regencies in South Sumatera. The data is taken in Palembang, South Sumatera, instead of in the rural area since the informants already move to Palembang. In order to gain the data required for this study, the author should select some informants to be interviewed.

In choosing the informants, there are some criteria that are included by the author. The informants should be the people whose childhood was spent in Komering area or Komering is their first language besides the Indonesian language. They also should be able to use Komering language actively. Then, they have to be old enough. The selected informants should be in the range age of 35-90 years old. Based on these criteria, there are four informants selected for this study since they fulfill the requirements.

In collecting the data, there are three steps taken by the author. The first step done by the author is constructing the questionnaire. By following Moe's (2003) theory of semantic domain, the questionnaire is composed based on an open source software called WeSay, a lexical resource constructed by SIL. The questionnaire is based on the subdomain of "food" in the daily life domain. The second step in doing the data collection is traveling to the location of the informants to conduct the interview. The author travels to Palembang, South Sumatera, two times to gather the data. Then, the third step that is conducted by the author is interviewing the informants by asking them the questionnaire constructed from WeSay.

After conducting the data collection process, the author conducts some steps of analyzing the data gathered. The first step is counting the number of words obtained from the interview. Then, the author attempts to discover whether the words are already included in Kamus Bahasa Daerah: Indonesia – Komering dan Komering – Indonesia (2011). After that, the author attempts to figure out the possible cause of the exclusion of the words from the dictionary. Then, the author creates a tabulation of the words and their possible causes of exclusion to sum them up.

Results and Discussion

From the interview, there are 114 words obtained for "food" subdomain. From these 114 words, there are 13 words that are missing from the dictionary. These 13 words are divided into five possible causes of the exclusion of the words from Kamus Bahasa Daerah: Indonesia – Komering dan Komering – Indonesia (2011).

The first possible cause of the exclusion is the failure of inserting the derived words from the lemma. There are three words that are excluded due to this reason, i.e., *ambawak* (to peel), *ngadun* (to knead), and *tialusko* (mashed). Yet, the root words of these words, i.e., *bawak* (skin), *adun* (knead), and *alus* (mash), are available in the dictionary. This may happen because the drafting team of the dictionary employs glossary-like format for the dictionary. Since the format is glossary-like, the drafting team may not consider the derived words of some root words to be inserted into the dictionary. Even if there are some derived words included into the dictionary, but those words are placed in their own entries and they stand as the lemmas. Whereas, the derived words should be placed under a root word so that the users of the dictionary are able to refer to the root words in finding the meaning of the derived words. Thus, some of the derived words are not included in the dictionary.

The second possible cause of the exclusion is the inaccuracy of the format of the dictionary. The word *kucik* (to peel) is the word in which this reason occurs. The word *kucik* is actually included in the dictionary, yet it is not found in its own entry. It is placed in the entry of the word *kubak* which also has the same meaning as the word *kucik* and the word *kucik* is only separated with *kubak* by a slash. Even the word *kubak* is mentioned twice in the different entries in this dictionary. If the drafting team of the dictionary attempts to employ the glossary-like format, they should put the word *kucik* as its own entry and not put it in an entry of another word with similar meaning. The drafting team should also recheck the dictionary before eventually publishing it so that this kind of error may not happen to the dictionary.

The third possible cause of the exclusion is the influence of the Indonesian language to Komering language. There are three words excluded due to this reason, i.e., *golpung* (flour), *nyamcami* (to taste), and *sodop* (delicious). These words are excluded from the dictionary because the drafting team selects other words with similar meanings to be inserted into the dictionary. Those other words apparently have the similar forms with the words in the Indonesian language. The replacement of these three words may occur since the drafting team are more influenced by the Indonesian language. Because the drafting team are influenced more by the Indonesian language, they tend to include the words of Komering language which have the similar forms with the ones in the Indonesian language. They may employ those words more than these three words so that they insert those words into the dictionary instead of placing these three words into the dictionary.

The fourth possible cause of the exclusion is the old words which are not recognized nowadays. There are three words that are excluded due to this reason, i.e., *pangut* (a knife for weeding fish), *panyuik* (a utensil for moving rice from a pot into a rice container), and *timpung* (a container made of white pumpkin). These words are traditional kitchen utensils which exist in Komering traditional kitchen. It means that these utensils are really old. Even not all the informants of this study has these utensils in their kitchen anymore. Since the utensils exist in Komering traditional kitchen, these utensils may not be available in modern kitchen. Even if they exist in modern kitchen, not all Komering people utilize these utensils nowadays. Because of the rareness of these utensils in Komering modern kitchen, the drafting team may not know the existence of these utensils. Thus, the drafting team do not include these utensils into the dictionary due to the rareness of the use of these utensils.

The last possible cause of the exclusion is the different dialects employed by the drafting team and the informants. There are three words excluded from the dictionary due to this reason, i.e., *nyadiakok* (to prepare), *nyotarko* (to serve), and *sorkit* (spatula). These words may be excluded from the dictionary because the dialect employed by the drafting team of the dictionary do not include these words and the dialect may have other terms for these words. However, these words are obtained from the interview. It means that the informants of the interview may employ this words in their daily activity, which means that these words are available in the dialect chosen by the informants. Because the drafting team of the dictionary does not employ the same dialect as the informants, the drafting team may not know these words and they may utilize other words with similar meanings to be included in the dictionary instead. Thus, these three words are failed to be included in the dictionary.

These possible causes of the exclusion of words in “food” subdomain are summarized in Table 1, along with the excluded words.

Table 1: The possible causes of the exclusion of words in “food” subdomain

No.	Possible Causes of the Exclusion	Words
1	Derivation	Ambawak Ngadun Tialusko
2	Inaccurate formatting	Kucik
3	The influence of the Indonesian language	Golpung Nyamcami Sodop
4	Obsolete words	Pangut Panyuik Timpung
5	Different dialects	Nyadiakok Nyotarko Sorkit

Conclusion and Suggestion

Komering is one of the endangered languages that exists in Indonesia. The existence of Komering language should be maintained so that this language will not die in the future. In order to maintain this language, the documentation such as creating a dictionary for it is required. In order to create a good ethnic language dictionary, the use of semantic domain is required. By utilizing WeSay as the database for the data obtained from conducting interview, the questionnaire is constructed by focusing on the “food” subdomain contained in the domain of daily life. From conducting an interview, there are 114 words acquired. By comparing the data obtained from the interview with Kamus Bahasa Daerah: Indonesia – Komering dan Komering – Indonesia (2011), it is discovered that there are 13 words missing. These words are excluded from the dictionary due to five possible causes.

It is suggested for further research to employ other eight subdomains of daily life domain for Komering language along with the culture. Any research on Komering language is also allowed to be conducted based on other domains in WeSay other than daily life domain. Any further research is also allowed to conduct a similar research by employing other lexical resource software besides WeSay, such as FLEx, for instance. There may also be similar research by comparing another endangered language with its local dictionary. This may assist the cultures and the languages to be maintained so that both of the core elements of an ethnic group will be preserved.

References

- Atkins, B. T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford, UK: Oxford University Press.
- Badan Pusat Statistik. (n.d.) Kebutuhan Data Ketenagakerjaan untuk Pembangunan Berkelanjutan. Retrieved from http://www.ilo.org/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-jakarta/documents/presentation/wcms_346599.pdf. (Accessed 12 September 2015).
- Bin Mohd Rosman, M. Z., Bond, F., & Frantisek Kratochvil. (2014). Bringing together over- and under-represented languages: Linking Wordnet to the SIL Semantic Domains. *Volume editors*, 40.
- Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative. and Mixed Methods Approaches* (3rd ed.). California, USA: Sage Publications, Inc.
- Crystal, D. (2000). *Language Death*. Ernst Klett Sprachen.
- Ethnologue. (n.d.). Endangered Languages. Retrieved from <http://www.ethnologue.com/endangered-languages>. (Accessed 12 September 2015).
- _____. (n.d.). Indonesia. Retrieved from <https://www.ethnologue.com/country/ID>. (Accessed 12 September 2015).
- _____. (n.d.). Komering. Retrieved from <https://www.ethnologue.com/language/kge>. (Accessed 12 September 2015).
- _____. (n.d.). Language status. Retrieved from <http://www.ethnologue.com/about/language-status>. (Accessed 12 September 2015).
- Kamus Bahasa Daerah: Indonesia – Komering dan Komering – Indonesia, 1st ed. 2011. FOKKU Sumatera Selatan.
- Moe, R. (2003). Compiling dictionaries using semantic domains. *Lexikos*, 13(1).

- Noor, N. H. M., Sapuan, S., and Francis Bond. (2011). Creating the Open Wordnet Bahasa. In *PACLIC* (pp. 255-264).
- Warner, N., Butler, L., & Luna-Costillas, Q. (2006). Making a dictionary for community use in language revitalization: The case of Mutsun. *International Journal of Lexicography*, 19(3), 257-285