INFERENCE OF MEANINGS FOR DERIVATIVE WORDS IN MALAYSIAN LANGUAGE

3 R − 6

Siti Noramisah Ramli, Masaya YAMAGUCHI, Nobuo INUI, Yoshiyuki KOTANI, Hirohiko NISIMURA

Dept. of Computer Science, Tokyo University of Agriculture and Technology

1 Introduction

Malaysian (Malay) language has numerous derivative words. Some of them are predictable by rules because the derivational processes in Malay language are highly regular. In the previous attempts at implementation of a Malay-to-Japanese machine translation, all those words were listed in the lexicon. Therefore, many lexical entries were required. In our attempt, we have eliminated the need for many lexical entries by reducing the Malay derivational processes to a set of morphological rules. We have developed a Malay derivative word analysis system that investigates affixes and base words in derivative words, provides some information about them and finally, infers their meanings from the categories of the affixes and the base words. To evaluate the system, we have applied it to our Malay-to-Japanese machine translation.

2 Malay Derivation

In this section we will briefly explain how Malay derivation works. In Malay language, affixes are divided into four categories: prefixes, suffixes, confixes, and infixes. Malay language has a number of rules that change the spellings of prefixes depending on the initial letter of the word to which it is added. For example, when the initial letter of a word is k, the spelling of meN-prefix will be meng- and the letter k will be omitted. For instance, if we prefix meN-prefix to a word *kacau*, it becomes *mengacau*. There are more rules like this in Malay language that are known as pronunciation rules.

3 Inference of Meaning

Malay language makes great use of affixes for grammatical purposes and to form new words from old. For example, the use of meN-prefix is to indicate the active form of a verb. *Membaca* (to read), for instance, is an active form of a verb *baca* (to read) by prefixing meN-prefix to the verb. There is very little difference in English or Japanese between the meanings of *baca* and *membaca*.

Furthermore, we can get a passive form of *membaca*, which is *dibaca*, by replacing the meN-prefix with di-prefix. We conclude that listing every form of this word is obviously inefficient because it needs many lexical entries in the lexicon. Therefore, we have reduced Malay derivation system to a list of rules so called inference rules in developing a Malay-to-Japanese machine translation.

4 System Structure

The system structure of Malay derivative word analysis is shown in Figure 1. It is divided into two stages: specification of affixes and base words categories and inference of meanings for the derivative words.

At the first stage a derivative word is first separated to affixes and a base word by consulting the list of affixes, the pronunciation rules and the Malay lexicon. Then the information such as a grammatical function and a lexical category of the affixes and the base word are given. At the second stage, the meaning of the derivative word is determined by inferring it from the categories of the affixes and the base word, which are obtained from the previous stage.

There are two types of lexicon that are used in this system: Malay lexicon, which consists of Malay words and information about these words and Malay-Japanese lexicon, which consists of Malay and Japanese words that are equivalent to each other.

To translate in Japanese a derivative word such as a di-form that is a passive form of a verb, a set of inflection rules, as shown in Figure 5, is required.

5 Samples of Analysis

Table 1 shows two samples of analysis for Malay derivative words. Figure 2 shows the examples of a Malay-to-Japanese translation that are obtained by using the derivative word analysis in our machine translation.

6 Conclusion

We have applied the Malay derivative word analysis system to our Malay-to-Japanese machine translation, which works on K-Prolog. As a result, the morphological rules certainly reduced many lexical entries but sometimes it took up time to determine the meaning of one derivative word.

An implementation of Malay derivation is a kind of trade-off. Listing every form of every word is apparently inefficient; the morphological rules can certainly save space. Unfortunately, a highly complicated rule system can be inefficient in the opposite way. It can be faster to work with an incomplete set of rules, looking up some forms in the lexicon, than to work with a more complex set of rules that captures everything.

However, some of Malay derivative words are predictable by simple rules and have no big difference in Japanese or English between the meaning of the original word and the new word. From that reason, it is more effective to implement morphological rules than to list all the forms of each word in the lexicon in developing a machine translation system for Malay language.

References

1. Covington,Michael A.:Natural Language Processing For Prolog Programmers, Prentice Hall,1994.
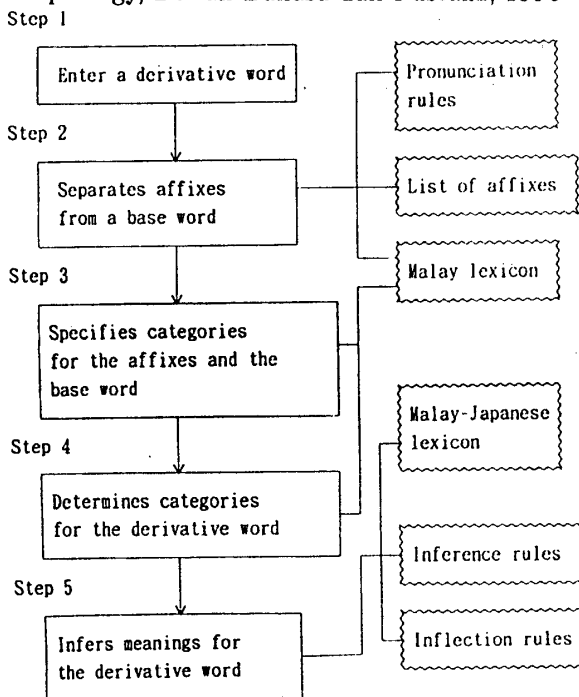2. Nik Safiah Karim:Tatabahasa Dewan, Vol.2: Morphology, Dewan Bahasa dan Pustaka, 1994

Step 1

Figure 1 System Structure

Table1 Samples of Analysis

Step	di-	ter-
1	ditulis	tertinggi
2	di+tulis	ter+tinggi
3	verbal prefix+tverb	adjective prefix+adjective
4	verb	adjective
5	書かれる (be written)	娶も高い (the tallest)

Figure 2 Examples of Malay-to-Japanese Translation

```
lex(baca,primary word,tverb).
lex(tulis,primary word,tverb).
lex(putih,primary word,adjective)
lex(buku,primary word,noun).
lex(oleh,primary word,preposition).
```

Figure 3 Lexicon of Malay words

```
prefix([m,e,m],meN,verbal prefix,verb).
prefix([d,i],di,verbal prefix,verb).
prefix([t,e,r],ter,adjective prefix,adjective).
```

Figure 4 List of Affixes

```
baca(tverb, 読む, 他動詞二段).
tulis(tverb, 書く, 他動詞二段).
putih(adjective, 白い, 形容詞1 ).
buku(noun, 本, 名詞).
oleh(preposition, に, 前置詞).
```

Figure 5 Lexicon of Malay-Japanese

```
他動詞二段 (う, われる).
他動詞二段 (く, かれる).
他動詞二段 (む, まれる).
```

Figure 6 List of Inflection Rules