

DESCRIPTIVE STUDY OF INDONESIAN AND KOREAN TIME EXPRESSION WITH LGGs

Prihantoro*, Sébastien Paumier **

*DICORA, Hankuk University of Foreign Studies, Korea/ ** IGM, University of Paris-Est, France

prihantoro2001@yahoo.com, paumier@univ-mlv.fr

Abstract

This paper aims at considering some problems encountered when we translate time expressions between Korean and Indonesian. We consider in this paper the time expressions corresponding to the questions of type *what time* or *when*, rather than those related to the questions of type *how long* or *how often*. Especially the time expressions referring to calendar, day and clock time in Indonesian and Korean are considered in this study. Our discussion underlines four different aspects in translating the time expressions between these two languages: word order, numeral system, abbreviation, and time segmentation. The LGG(Local Grammar Graph) model will be used in the description of these aspects. The result of our study can be used as a linguistic resource not only for the learning system of Indonesian and Korean, but also for the machine translation system between these two languages.

1. Introduction

Indonesian is the official language of Indonesia and was politically created in October 28th 1928, as an effort of unifying Indonesian people, who was colonized by Netherlands. At that time, Indonesian communicated mostly in their first language only. Indonesian is the development of Malay used in Riau (Sneddon, 2004:14), in order to fulfill the plural characteristic of around 250 million Indonesian people. In other hand, Korean is the official language of Korea, both South and North. There are about 78 million Korean speakers. In the 15th century, a national writing system was commissioned by Sejong the Great. The system is currently called Hangeul. Prior to the development of Hangeul, Korean people had used Hanja (Chinese characters) to write for over a millennium (Sohn, 1999:122). Recently, the bilateral relation between Indonesia and Korea is getting more intense. Those two countries exchange many things such as, trading objects, technology, knowledge, culture, and of course language. For this reason, the need of understanding Korean and Indonesian language is also increasing, as well as the translation between these two languages.

Korean and Indonesian are not genealogically related language. The gaps between these two languages pose challenges for translation, especially when this translation task is performed by machine.

The computer scientists and language experts are working to provide solution for this problem. However, before coming to that part, it is important to understand the different aspects between Indonesian and Korean Languages. That is the purpose of this paper. The paper is organized as follows. Section 1 is a brief introduction to both Korean and Indonesian languages. Section 2 shows the scope of problems that we present in this paper. In Section 3, we discuss four different aspects of time expressions between these two languages. This paper will be ended with a short summary and future perspective in Section 4.

2. Scope of the Problems

We here are based on one of the previous studies related to time expressions in Korean language, which was done by J.-S. Nam(2009a, 2009b). In her papers, she introduces the LGGs(Local Grammar Graphs) to describe and recognize Korean time expressions. From the observation on Nam’s studies, this paper is focused on the comparative studies of ‘date’, ‘day’, and ‘clock’ time expressions between Indonesian and Korean. Our discussion will be focused on exact time expressions rather than duration expressions, which are classified as WHEN class in Nam(2009a:69, 2009b:164). These expressions will not answer to the questions like “*How much time do I need to study programming?*” or “*How many days will you spend in Hawaii?*”. Instead, they correspond to the questions like “*When did you study Korean Linguistics?*” or “*What time will you go to Hawaii?*”.

3. Different aspects between Indonesian and Korean Languages

Translation is not only one-on-one word form conversion from source language to target language. Beyond that, translation conveys message from source language to target language. Winter(1961:98) believed, in order to achieve ideal translation, a translator must be able to handle the gaps, either linguistic or non-linguistic gaps, between source and target languages. Consider the following example, where *Google translator* failed to achieve a qualified translation.

Illustration 1. Google Translation



In the above example, source language is Indonesian, and target language is Korean. The Indonesian

phrase in the input box means “Monday, January 29th, 9 o’clock in the morning”. Korean speakers might be able to guess what the translated result signifies in the output box, but they cannot obtain an appropriate translation. It should be translated into “2009 nyon 1 wol 29 il wolyoil achim 9 si”.

In this paper, we underline four problems in translating ‘calendar’, ‘day’ and ‘clock’ expressions from Korean to Indonesian and vice-versa. These problems are “word order, numeral system, use of abbreviation, and time segmentation”. Our description is done as a DAG(Directed Acyclic Graph) we can build by using *UNITEX GraphEditor*¹.

3.1 Word Order

The order of Korean time expressions is usually like <Year, Month, Date, Day, Hour, Minute and Second>. We can see that the expressions are structured from the widest to the narrowest time entities. In Indonesian, there are no rigorous rules in ordering time expressions, but common order is like <Day, Date, Month, Year, Hour, Minute and Second>. The table 1 summarizes the order of time expressions in both languages. We here notice the gap of word orders between Indonesian and Korean languages in general.

Table 1. Word Order in Time Expressions in Indonesian and Korean

Korean	Year	Month	Date	Day	Hour	Minute	Second
Indonesian	Day	Date	Month	Year	Hour	Minute	Second

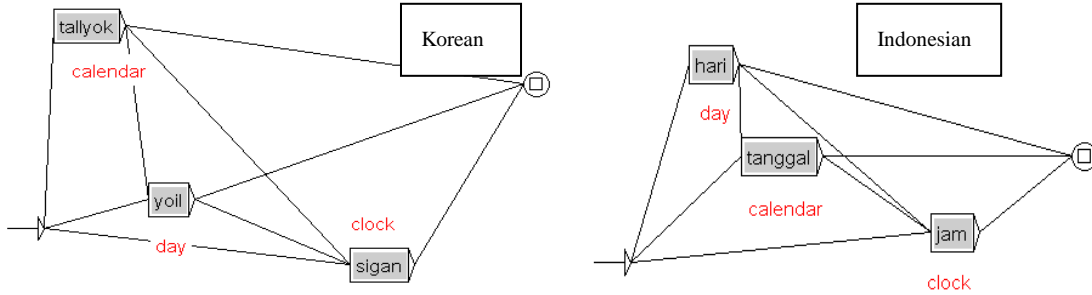
The following example shows the translation of the English phrase “Monday, January the 2nd 2009, nine forty five and twenty second” into both Korean and Indonesian.

- (1) Korean : 2009 nyon 1 wol 2 il Wolyoil 9 si 45bun 20 cho
 2009-year 1-month 2-date Monday 9-hour 45-minute 20-second
- (2) Indonesian : Hari senin, tanggal 2 bulan Januari tahun 2009, jam 9 45 menit 20 detik
 Day-Monday date-2 month-January year-2009 hour-9 45 minute 20-second

The following Graphs represent the order of time expressions in both languages:

¹ These graphs built in Unitex are precisely Local Grammar Graphs(LGGs) which are suggested by French computational linguist Maurice Gross. The system UNITEX is a corpus-based processor, using an automata technology. (Paumier, 2003)

Graph 1. Korean and Indonesian Time Expression from Calendar to Clock Time Expression



We observe that the order of Korean time expressions is structured from the widest range (Calendar) to the narrowest one (clock). Whereas in Indonesian, the order starts from day, calendar and ends by clock time. These graphs show that it is possible to put only one time slot as well as to combine them.

When we discuss time expressions, there are mainly two points: numbers and time slot units (i.e. minute, hour, year etc). In Korean, all expressions indicating time slots are located after numeral expressions, which is not the case in Indonesian. Some of them can be located after numeral expressions and some of them can be before them. In Indonesian, Arabic numbers cannot be used to indicate *Days* and *Months*, while they are forbidden in the sequences of *Days* in Korean. This is a complete list in Table 2:

Table 2. Order of Numeral-Classifier in Time Expressions in Indonesian and Korean

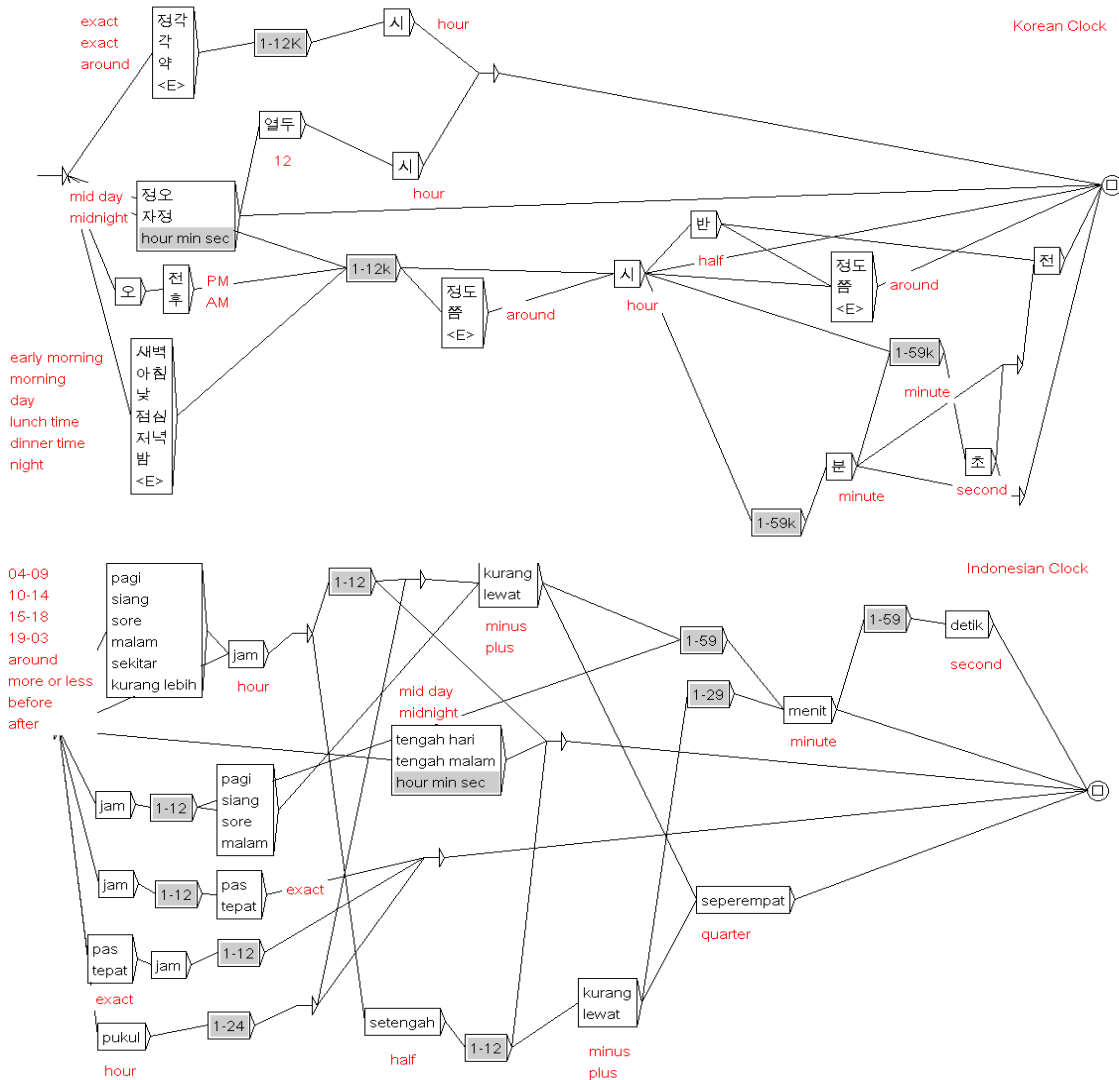
	Indonesian			Korean		
	Classifier	Numeral	Classifier	Classifier	Numeral	Classifier
Second		20	<i>detik</i>		20	<i>cho</i>
Minute		45	<i>menit</i>		45	<i>bun</i>
Hour	<i>Jam</i>	09			09	<i>si</i>
Day	<i>Hari</i>	<i>(senin+*Num)</i>			<i>(wol+*Num)</i>	<i>yoil</i>
Date	<i>Tanggal</i>	02			2	<i>il</i>
Month	<i>Bulan</i>	<i>(Januari+*Num)</i>			1	<i>wol</i>
Year	<i>Tahun</i>	2009			2009	<i>nyon</i>

In English clock time expressions, we recognize some units like *half* or *quarter*. Such expressions are also available in Indonesian and in Korean, but the order of these units are different. In Indonesian, there are some expressions like *setengah*(=half) and *seperempat*(=quarter), while in Korean, there is only *ban*(=half). Let us consider the following examples for Indonesian and Korean clock expressions.

- | | | | | | |
|-----|-------------|--------------------------|--|----------------------------|--|
| | | 1: 30 | | 1: 15 | |
| (3) | Korean | : <i>Han si ban</i> | | \emptyset | |
| | | One hour half | | \emptyset | |
| (4) | Indonesian: | <i>Jam setengah satu</i> | | <i>Jam satu seperempat</i> | |
| | | Hour half one | | Hour one quarter | |

The following LGGs are representing some of these time expressions:

Graph 2. Korean and Indonesian Clock Expression



3.2 Numeral System

In Indonesian, there is only one numeral system. This makes the translation from Korean to Indonesian easier. Different from Indonesian, Korean has two numeral systems. One is native Korean numeral system (limited to 1~99), and the other is a system adopted from Chinese numeral system (Hobin, 2001: 144, Lee, 2004: 77). These two numeral systems are used in different contexts. Regarding time expressions, the challenging part is not when numerals are represented in Arabic numbers, but when they are in word forms. In written language, Arabic numbers could be used frequently, which do not make difference in both native Korean and Chinese numeral systems. However, in spoken language, even though Arabic numbers are used in the text, they have to be pronounced in a different way according to

their correspondence with native Korean or Chinese numeral system. The following table shows the correspondence of Arabic numbers with these 2 systems in a given context, i.e. the nature of time slot units:

Table 3. Correspondence of Numerals with Time slot units in Korean

Time slot units	Numeral System
Year	Chinese
Month	Chinese
Date	Chinese
Day	Chinese
Hour	Korean
Minute	Chinese
Second	Chinese

In the above table, *hour* is the only part requiring native Korean numerals. The problem does not stop here. In Indonesian, numerals are not distinguished between noun and determiner positions, differently from Korean. Let us consider the following table that sums up the different forms between noun and determiner positions in Korean:

Table 4. Nouns and Determiners in Native Korean Numeral System

Arabic Number	Noun	Determiner
1	hana	han
2	tu	tu
3	ses	se
4	nes	ne
20	seumul	seumu

<01:30 >

- (5) Korean : *Han* *si* *samsip* *bun*
 One (Korean Determiner) hour thirty (Chinese Determiner) minutes
 **il* *si* *samsip* *bun*
 *One (Chinese Determiner) hour thirty (Chinese Determiner) minutes

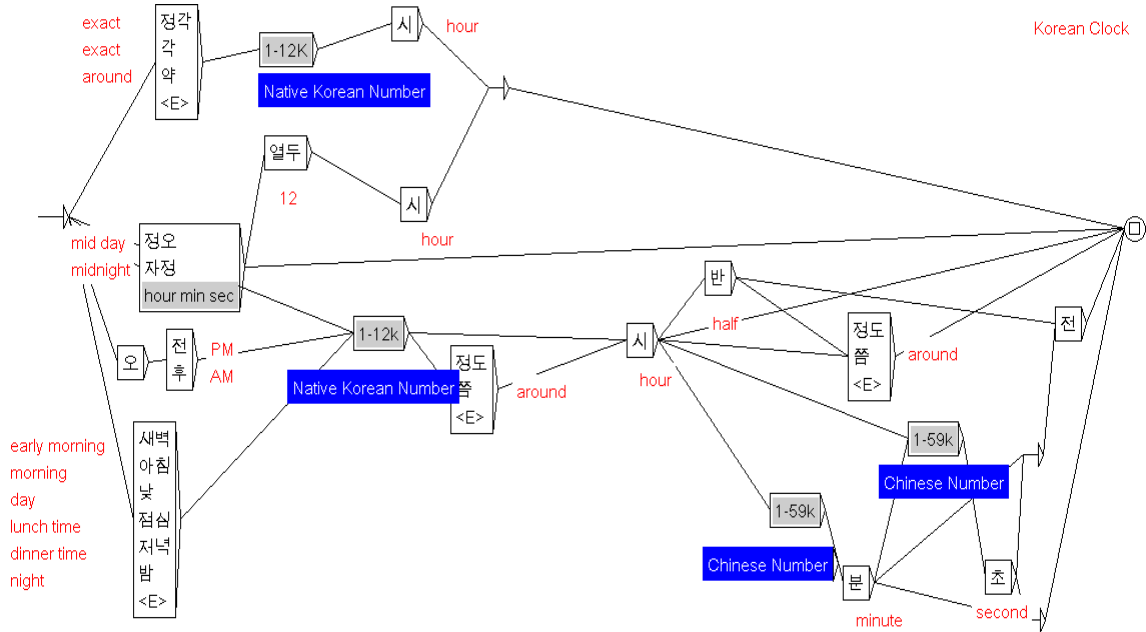
- (6) Indonesian : *Jam satu lewat tiga puluh menit*
 Hour one plus thirty minutes

<02:15>

- (7) Korean : *tu* *si* *samsip* *bun*
 two (Korean Determiner) hour 30(Chinese Determiner) minutes
 **tu* *si* *samsip* *bun*
 *Two (Korean Noun) hour 30(Chinese Determiner) minutes

- (8) Indonesian: *Jam dua lewat lima belas menit*
 Hour two plus fifteen minutes

Graph 3. Korean Clock Time Expression

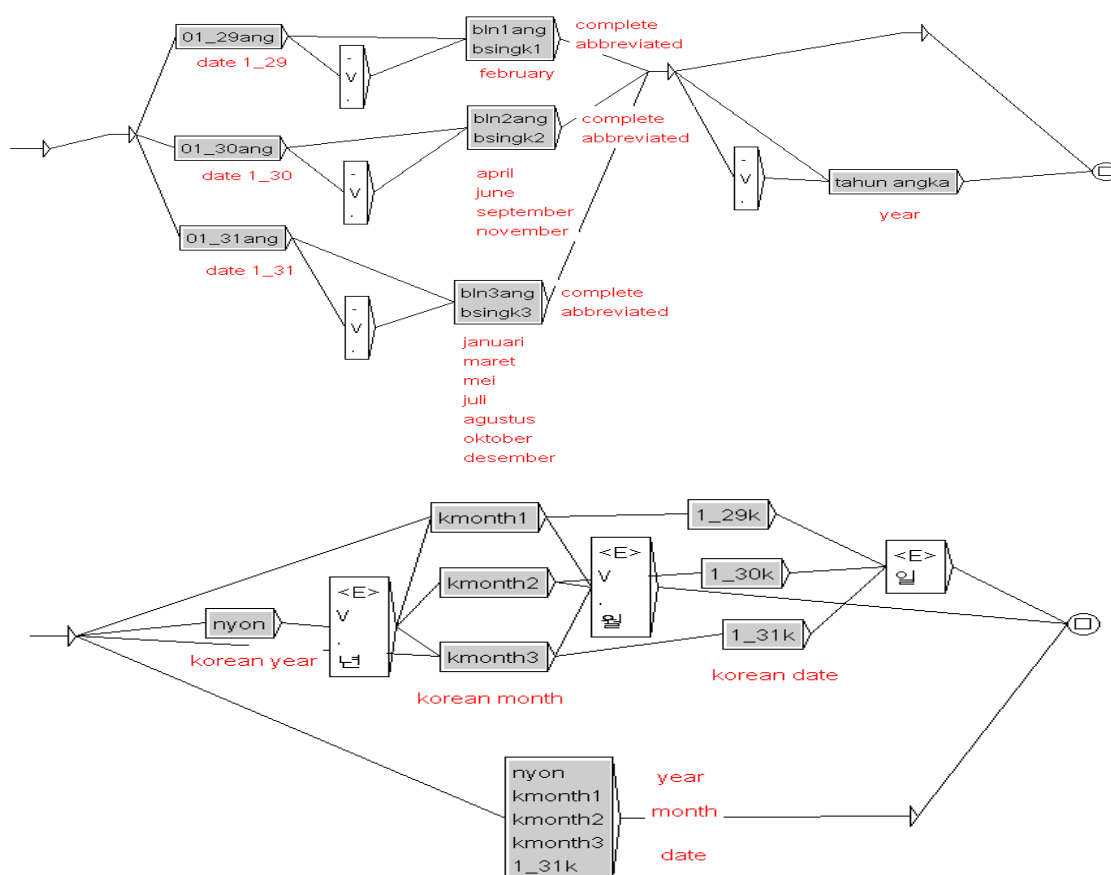


3.3 Use of Abbreviation

This part describes the types of separators and abbreviations used in Indonesian and Korean time expressions. The main reason in using separators in time expressions is for the space economization. In newspapers or memos, where the number of words is an important factor, separators and abbreviations are frequently used. In both Korean and Indonesian, separators such as “.” and “/” are used to separate date, month and year. However, there is one separator that is used quite frequently in Indonesian, differently from in Korean: the separator “-(=hyphen)”. Consider the following examples:

- (9) Indonesian : <02.05.1998>, < 02/05>, <02-05-1998>
- (10) Korean : <1998.05.02>, < 05/02>, <??1998-05-02>

Graph 4. Indonesian and Korean Calendar Time Expression



In Korean, for calendar time, the expressions like “*nyon*(year) *wol*(month) *il*(date)” are used, and there are some equivalent expressions in Indonesian like “*tahun*(year) *bulan*(month) *tanggal*(date)”, but the latter is not used frequently in current usage. Shortly say, these expressions are highly optional in Indonesian. The most frequently used one is *tanggal*(date), as it refers to a calendar as a whole. Consider the following examples:

- <27 January 2003>
- (11) Indonesian : *Tanggal 27 Januari 2003*
 Date 27 January 2003
- (12) Korean : *2003 nyon 1 wol 27 il*
 2003 year 1 month 27 day

In Indonesian, the months are usually expressed as an abbreviated form with first three letters like in English. For example, *Nopember*(=November) is abbreviated into *Nop*. However, because of the influence of English, it can be wrongly abbreviated into *Nov*. It is the same in the case of *Desember* (=December): it can be abbreviated into either *Des*(Indonesian) or *Dec*(like English). For the other nouns

of month, we cannot distinguish whether they follow English style or Indonesian style, because they are abbreviated in the same way. For example, for the noun *March*, the abbreviation in Indonesian will be *Mar* like in English. Consider the following table which shows the system of abbreviation of Month nouns in Indonesian and that of Korean requiring <Num+Classifier> instead of lexical words like in Indonesian:

Table 5. System of Abbreviation of Month in Indonesian compared to Korean System

Month in English	Full Form in Indonesian	Abbreviation in Indonesian	Month in Korean <Chinese Numeral>
January	<i>Januari</i>	<i>Jan</i>	<i>il(1) wol</i>
February	<i>Pebruari</i>	<i>Feb / Peb</i>	<i>i(2) wol</i>
March	<i>Maret</i>	<i>Mar</i>	<i>sam(3) wol</i>
April	<i>April</i>	<i>Apr</i>	<i>sa(4) wol</i>
May	<i>Mei</i>	<i>Mei</i>	<i>o(5) wol</i>
June	<i>Jun</i>	<i>Jun</i>	<i>yu(6) wol</i>
July	<i>Juli</i>	<i>Jul</i>	<i>chil(7) wol</i>
August	<i>Agustus</i>	<i>Ags / Agu</i>	<i>phal(8) wol</i>
September	<i>September</i>	<i>Sep</i>	<i>gu(9) wol</i>
October	<i>Oktober</i>	<i>Oct/Okt</i>	<i>si(10) wol</i>
November	<i>November</i>	<i>Nov / Nop</i>	<i>Sib-il(11) wol</i>
December	<i>December</i>	<i>Des</i>	<i>Sib-i(12) wol</i>

Abbreviation will be found also in the expressions of Days. In Korean, full names like *wol-yoil*(Monday), *hwa-yoil*(Tuesday), or *su-yoil*(Wednesday) can be abbreviated into *wol*, *hwa*, or *su*. In Indonesian, *senin*(Monday), *selasa*(Tuesday), or *rabu*(Wednesday) are abbreviated into *Sen*, *Sel*, or *Rab*. In computer systems, as the abbreviations appear with the first letter of day names, they can make confusion with ‘S’ among *senin*(Monday) or *selasa*(Tuesday), or even *sabtu*(Saturday). Let us consider:

Table 6 . Abbreviation of Day Names

	Indonesian			Korean	
	Full Form	Abbreviation		Full Form	abbreviation
Monday	<i>Senin</i>	<i>Sen</i>	<i>S</i>	<i>wolyoil</i>	<i>Wol</i>
Tuesday	<i>Selasa</i>	<i>Sel</i>	<i>S</i>	<i>hwayoil</i>	<i>Hwa</i>
Wednesday	<i>Rabu</i>	<i>Rab</i>	<i>R</i>	<i>suyoil</i>	<i>Su</i>
Thursday	<i>Kamis</i>	<i>Kam</i>	<i>K</i>	<i>mokyoil</i>	<i>Mok</i>
Friday	<i>Jumat</i>	<i>Jum</i>	<i>J</i>	<i>kemyoil</i>	<i>Kem</i>
Saturday	<i>Sabtu</i>	<i>Sab</i>	<i>S</i>	<i>toyoil</i>	<i>to</i>
Sunday	<i>Minggu</i>	<i>Ming</i>	<i>M</i>	<i>ilyoil</i>	<i>il</i>

In Indonesian, there is a rule stating that the beginning of a sentence must be started by an uppercase letter (Waridah, 2008:6). This implies that a sentence must begin with a word, not with an Arabic number. Therefore, if a sentence starts by an Arabic number, the first sequence must be converted

into word forms. Compare the following examples in Indonesian:

<1945 is the independence year>

(13) *Seribu sembilan ratus empat puluh lima* adalah tahun kemerdekaan
1945 is year independence

(14) *Tahun 1945* adalah tahun kemerdekaan
Year 1945 is year independence

(15) *1945 adalah tahun kemerdekaan
1945 is year independence

The availability of different separators and abbreviations in both Indonesian and Korean make the calendar time expressions be represented in various ways. The following tables show possible representations of calendar time expressions in both Indonesian and Korean:

Table 7. Calendar Time Expressions in Indonesian

Date	Month		Year		Sequence Type	
25		01		2003	25 01 2003	
25-		01-		2003	25-01-2003	
25/		01/		03	25/01/03	
25		<i>Jan</i>		2003	25 Jan 2003	
25		<i>Januari</i>		2003	25 Januari 2003	
25/		01			25/01	
<i>Tanggal</i>	25	<i>Bulan</i>	<i>Januari</i>	<i>Tahun</i>	2003	<i>Tanggal 25 bulan Januari tahun 1945</i>
<i>Tanggal</i>	25		<i>Januari</i>		2003	<i>Tanggal 25 Januari 2003</i>

Table 8. Calendar Time Expressions in Korean

Year		Month		date		Sequence Type
2003	<i>nyon</i>	01	<i>wol</i>	20	<i>il</i>	2003nyon 01wol 20il
2003	.	01	.	20	<i>il</i>	2003.01.20
03	.	01		20		03.01.20
2003	.	01				2003.01
		01	.	20		20.01
03	/	01				03/01

3.4 Time Segmentation

Linguistic aspects are not the only factor that should be noticed when we discuss the difference of time expressions in Indonesian and Korean. There are also some extra-linguistic aspects such as a geographical difference between Indonesia and Korea. Indonesia is a very large country, and indeed, Indonesia is the largest archipelago country in the world. Therefore, when we refer to one time in a part of Indonesia, it can be different in another part of Indonesia. For example, Sunday 22:30 in *Jakarta*(capital of Indonesia) will be Sunday 23:30 in *Bali*(middle part of Indonesia). Sunday 23:30 in *Bali* will be

Monday 00:20 in *Papua*(eastern island of Indonesia). Basically, there are three time segmentations in Indonesian region, which are west, central, and east sides with one-hour difference in each area. On the contrary, Korean does not have time difference problems. There is no significant time gap between *Jeju*(southern part) and *Seoul*(northern part). Sunday 22:30 in Seoul will also be Sunday 22:30 in *Jeju*, *Busan*, *Daegon* and other cities in Korea.

Moreover, in Indonesia, sunrise time is always around 6 pm ~ 6: 30 p.m., since Indonesia is geographically located right on the equator line. In the case of Korea, there are 4 seasons: spring, summer, fall, and winter. Therefore, sunset times in winter and in summer can be different in Korea. In Indonesian, *Malam* or *pam* refers to night time, and around 6 p.m. can be called as evening or night. But in Korea, or in other countries where there are four seasons, the time referring to the beginning of the evening can be different in each season.

Another problem is related to cultural aspects. Eating seems to be an inseparable part from Korean life. The expressions such as “*pab mog-ess-ni?*(Have you had your meal?)” are used frequently among Korean people in order to replace formal greetings. As a result, some expressions related to Meal such as *cemsim*(lunch) or *cenyok*(dinner) can be used to refer to some particular time. This does not happen in Indonesian, while some expressions familiar to Moslem prayers are frequently used to refer to some time. The expressions such as *magrib*(sunset time, an hour length) or *subuh*(from dawn to sunrise) are the examples related to Moslem culture, which are now used by any Indonesian people.

4. Summary and Perspectives

In this paper, we discussed the different aspects of time expressions between Indonesian and Korean languages. We here considered four differences: word order, numeral expression, use of abbreviation, and Time recognition. The last difference reveals a geographical and cultural gap between Indonesia and Korea. In order to perform a reliable translation between these two languages, we need to understand these differences observed in both languages. In this paper, we shortly introduced the LGG-based description of time expressions in Korean and Indonesian, focused on the expressions corresponding to the WHEN-question type.

In our future works, we will present in detail the LGGs of time expressions in both languages so that we can use them in automatic translation of time expressions. The UNITEX system(Paumier 2003) is an appropriate framework to construct the LGGs which are transformed into finite-state automata in text processing.

Acknowledgements

We are grateful to Jee-Sun Nam for her insightful comments and suggestions on this paper.

References

- Alwi, Hasan, et al. 1997. *Tata Bahasa Baku Bahasa Indonesia*. Balai Pustaka. Jakarta.
- Ihm Hobin, Kyungpyo Hyong and Sukin Chang. 2001. *Korean Grammar for International Learners*. Yonsei University Press. Seoul.
- Lee, Iksop. 2004. *Korean Grammar*. Seoul National University Press, Seoul.
- Nam, Jee-Sun. 2009a. Description of Korean Time Expression including Digits for Automatic Extraction of Time Information, *Journal of Language Sciences*, 16-1, 59~87.
- Nam, Jee-Sun. 2009b. Finite Local Grammars of Korean Time Expressions, *Journal of Korean Linguistics* 42, 155-191.
- Paumier, Sebastien. 2003. *Unitex Manual*. Université Paris-Est Marne-la-Vallée. Paris.
- Sohn, Homin. 1999. *The Korean Language*. Cambridge University Press. Cambridge.
- Sneddon, James. 2004. *The Indonesian Language: Its History and Role in Modern Society*. UNSW Press. Sydney.
- Waridah, Ernawati. 2008. *EYD dan Seputar Kebahasa-Indonesiaan*. Kawah Media. Jakarta.
- Winter, W. 1964. Impossibilities of translation, in W. Arrowsmith and R. Shattuck (eds.) *The Craft and Context of Translation*, New York: Anchor.