Resolving Structural Conflicts in English-Korean MT System[1]
– Pronouns and Passives as examples –

Jeong-Ryeol Kim[*]

I. Introduction

There are at least two distinct problems that concern translators in the translation between English and Korean. One is the use of overt pronouns between two languages, and the other is the different grammatical devices for focusing on the theme of a sentence.

Consider the following contrast between the two languages in the use of pronouns:

(1) Overt pronoun to reflexive pronoun
   he_i washed his_i car.
   ku-ka caki-uy cha-lul ssis-ess -ta
   he-NOM self-GEN car-ACC wash-PAST-DECL

(2) Overt pronoun to zero
   he_i thought that he_i washed the car.

---

*한국해양대학교 교양과정부 전임강사

ku-ka ku cha-lul ssis-ess -ta -ko sayngkakhay-ss-ta
-NOM the car-ACC wash-PAST-DECL-COMP think-PAST-DECL

(3) Awkwardness of repeated pronouns (Hong 1987)
he$_i$ does his$_j$ work$_j$ when he$_i$ feels like doing it$_j$.
?ku-ka ku-uy il -ul ku-ka kukes-ul ha-ko
he-NOM he-GEN work-ACC he-NOM it -ACC do-COMP
siph-ul   ttay ha-n -ta
want-COMP when do-PRES-DECL

The next issue is that the use of passive voice is very common in English text especially when it comes to technical manuals. Consider the following text (Wada 1990):

Before the engine is started, the gear lever should be shifted to the neutral position to disconnect the engine from the drive shaft. Now the ignition key is inserted into the ignition switch. When the key is turned to the right, the steering-wheel lock is first released, so that the wheel can be freely rotated.

One of the natural Korean translations of this text would be:

eyncin-ul sicakha-ki-cen-ey ku kia leypa-nun eyncin-ul traipa syaputhu-eyse celtanha-ki-wihay cwunglip wichi-lo itongha-yeya ha-n-ta. kuliko, cemhwa khi-nun cemhwa suwichi-ey sapipha-n-ta. khi-lul olunccok-ulo tolli-myen, wuncentay camwulsoy ka mence phwulli-ko pakhwi-nun cayuloi tol swu iss-ta.

Three types of translation of English passives are found in Korean text:

(4) passive to active with the topic marker *-nun*
    ...the gear lever should be shifted to the neutral position...

...ku kia leypa-nun ...cwunglip wichi-lo itongha-yeya ha-n-ta.
the ignition key is inserted into the ignition switch.
cemhwa khi-nun cemhwa ssuwichi-ey sapipha-n-ta.
...the wheel can be freely rotated.
...pakhwi-nun cayuloi tol swu iss-ta.

(5) passive to active without topic
   Before the engine is started ...
   eyncin-ul sicakha-ki-cen-ey ...
   When the key is turned to the right ...
   khi-lul olunccok-ulo tolli-myen ...

(6) passive to morphological passive
   ...the steering-wheel lock is first released, ...
   ...wuncentay camwulsoy-ka mence phwulli-ko, ...

The data show that all English passives in the text but one sentence are translated into active voice in Korean. There are three types of correspondences depending on grammatical environments. Passive to active with topic is preferred when a passive construction occurs in a matrix sentence. Also, passive to active without topic is used when topic construction is not available in Korean. Note that topic in subordinate clause often results in an awkward translation when another topic already exists in a matrix clause as follows:

(7)
   ?? eyncin-un sicakha-ki-cen-ey ku kia leypa-nun ...
      cwunglip wichi-lo itongha-yeya ha-n-ta.
   ?? khi-nun olunccok-ulo tolli-myen, ... pakhwi-nun cayuloi
      tol swu iss-ta.

The data finally show that in limited environments where a morphological passive (as opposed to syntactic passive *-ye-ci*)

exists as the translation of a particular English verb and the pre-existing topic blocks the repetition of another topic, then it appears that the passive construction is used in Korean as well.

2. Resolution of Structural Differences

2.1 Pronouns

The ECS MT System employs Lexical-Functional Grammar (LFG: Bresnan 1982a, 1982b, 1982c, 1987, Bresnan and Mchombo 1986, Bresnan and Kanerva 1989, Kaplan 1989, J. Kim 1990, 1991, Tomii 1990, Wescoat 1987) as its linguistic framework.[2] Thus, the formalism used in the ECS MT System is based on LFG with some modification to improve the expressive power of the grammar. Thus, it is necessary to state sentence structures in LFG terms. In doing so, I will provide a short definition of those terms.

The repeated use of overt pronouns in a single sentence often results in awkward/unacceptable translation in Korean, when those pronouns coreference each other.[3] The conditions to block such awkward translation of pronouns can be found in the study of distribution of Korean anaphoric pronouns (Y. Cho 1985, Kaplan and Bresnan 1985, Pentheroudakis 1990).

There are three factors involved in causing an awkward occurrence of pronouns when English text is translated pronoun by pronoun. When an overt pronoun exists in the f-command domain of what the pronoun refers to, the use of an overt pronoun is resisted in Korean. F-command is defined as

2. Also, Refer to K. Hong 1987, Sells 1985, Kuroda 1965, Nirenburg 1987 for the motivation behind using LFG as the linguistic framework for the ECS MT System.
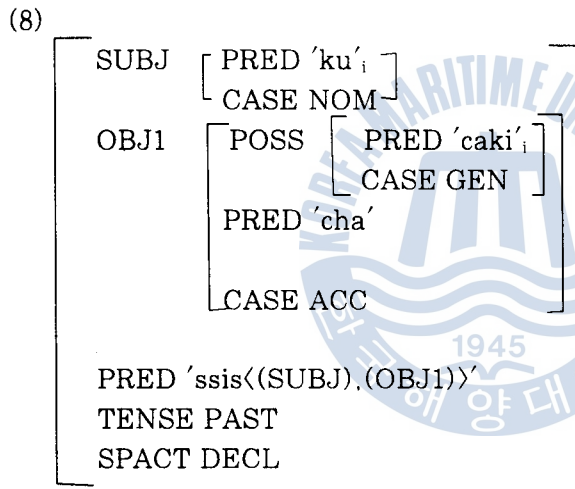
3. Refer to o. Her 1990 for the use of anaphoric pronouns in intersentential levels.

follows (Tomii 1990):

An antecedent α f-commands a pronominal β iff

a) α does not contain β, and

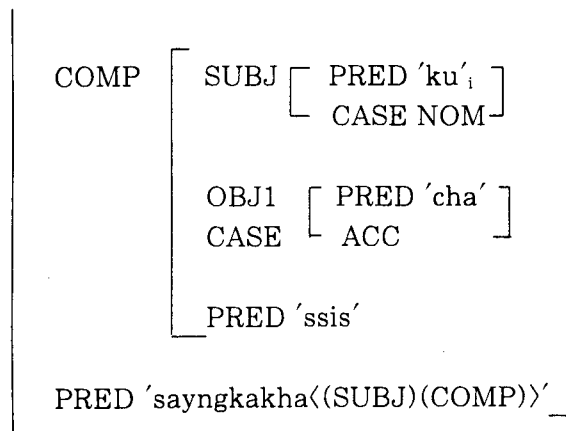b) every nucleus that contains α contains β.

A nucleus is a structure in which PRED and its argument functions such as SUBJ, POSS and/or OBJs exist. Consider the examples of pronouns in the previous section and their representation of f-structures:

(8)

$$
\begin{bmatrix}
\text{SUBJ} & \begin{bmatrix} \text{PRED 'ku'}_i \\ \text{CASE NOM} \end{bmatrix} \\
\text{OBJ1} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED 'caki'}_i \\ \text{CASE GEN} \end{bmatrix} \\ \text{PRED 'cha'} \\ \text{CASE ACC} \end{bmatrix} \\
\text{PRED 'ssis} \langle \text{(SUBJ),(OBJ1)} \rangle ' \\
\text{TENSE PAST} \\
\text{SPACT DECL}
\end{bmatrix}
$$

The SUBJ 'he' f-commands the POSS of OBJ1 'caki', but not vice versa. Also, they are coindexed. Following the definition of f-command, the nucleus containing the SUBJ is the whole sentence, while the nucleus containing the POSS is the OBJ1, and the sentence contains the POSS of OBJ1, and OBJ1 does not contain SUBJ.

The next example also shows that the zero-pronoun position is f-commanded by what it refers to as follows:

(9)

$$
\begin{bmatrix}
\text{SUBJ} & \begin{bmatrix} \text{PRED 'ku'}_i \\ \text{CASE NOM} \end{bmatrix}
\end{bmatrix}
$$

$$
\left[
\begin{array}{l}
\text{COMP} \left[
\begin{array}{l}
\text{SUBJ} \left[
\begin{array}{l}
\text{PRED 'ku'}_i \\
\text{CASE NOM}
\end{array}
\right] \\[2ex]
\text{OBJ1} \left[
\begin{array}{l}
\text{PRED 'cha'} \\
\text{ACC}
\end{array}
\right] \\
\text{CASE} \\[2ex]
\text{PRED 'ssis'}
\end{array}
\right] \\[4ex]
\text{PRED 'sayngkakha}\langle(\text{SUBJ})(\text{COMP})\rangle'
\end{array}
\right]
$$

There are typologically two different types of grammatical devices to express coreferent pronouns: one is Binding and the other is Control. For example, in English a verb *think* uses Binding and another verb *want* uses Control. That is, *He*_i *thinks that he*_i *washes his car* but not *He thinks to wash his car* and in contrast, *He wants to wash his car* but not *He*_i *wants that he*_i *washes his car.* However, Control is widely used in Korean wherever a sentence allows the use of Control. The environments of Control is that the controller and controllee are unifiable with each other in addition to coreference. In English the use of Control and Binding is determined as a side effect of verb subcategorization. In contrast, in Korean it is determined by the ability of unification between controlling nouns and controlled nouns. When a controller and a controllee cannot be unifiable with each other, a reflexive pronoun is preferably used in Korean. Note that the distribution of reflexive pronouns is widely different from that in English (Y. Cho 1985, Kaplan and Bresnan 1985, Pentheroudakis 1990).
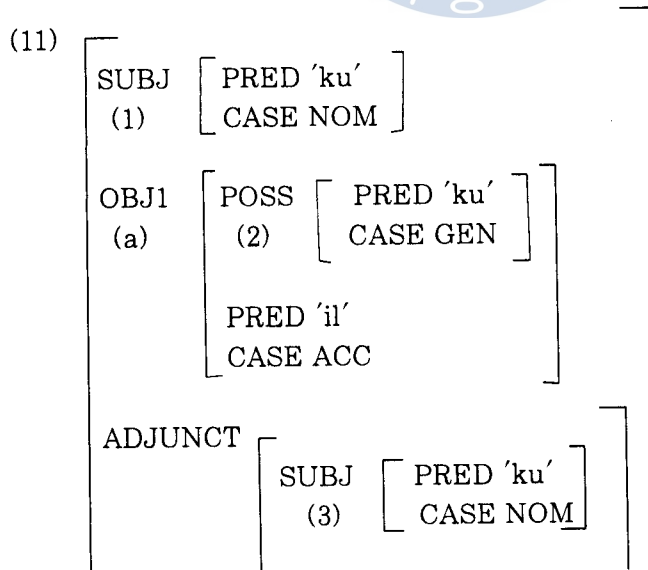
Controlled zero pronoun is preferred where the SUBJ in a subordinate clause can be unified with the SUBJ in a upper nucleus and thus, the missing pronoun is recoverable. A reflexive pronoun *caki* is realized where the SUBJ cannot be unified with the POSS due to the different grammatical
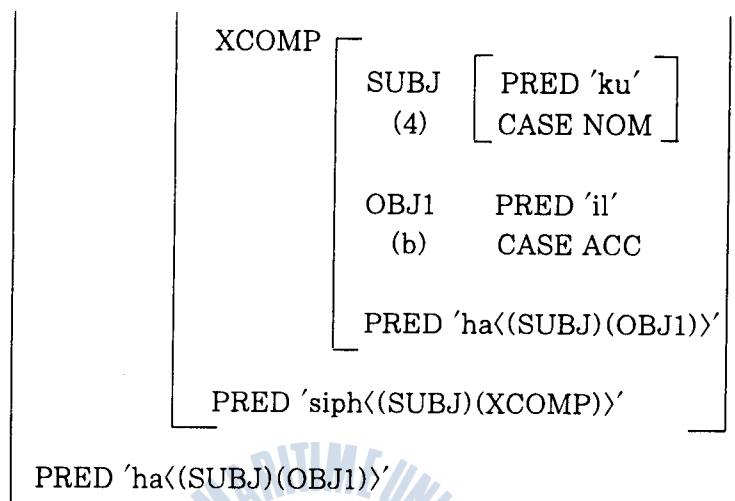
function expressed by different cases as in (8). Thus, we can formulate a functional rule to generate an appropriate form of pronouns based on the observation as follows:

(10)

$\langle \hat{} / \{ GF1\ GF2\ ...GFn \}^* \backslash GF\ \alpha \rangle = \langle \hat{}\ GF\beta \rangle$

$\langle \hat{}\ GF\alpha\ PRED \rangle = 'caki'$

iff $\langle \hat{} / \{ GF1\ GF2\ ...GFn \}^* \backslash GF\alpha\ PRED_i \rangle =$

$\langle \hat{}\ GF\beta\ PRED_i \rangle$

(10) states the fact that Control is used where a current grammatical function is unifiable with the same grammatical function within f-command domain of a controlling noun in a matrix sentence. Otherwise, a reflexive pronoun *caki* is used where the controlling noun cannot be unified with the controllee.

The unacceptable translation we saw in the previous section violates rules of pronominalization in Korean three places. Consider the f-structure of this sentence:

(11)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED 'ku'} \\ \text{CASE NOM} \end{bmatrix} \\ (1) & \\ & \\ \text{OBJ1} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED 'ku'} \\ \text{CASE GEN} \end{bmatrix} \\ (2) & \\ & \\ \text{PRED 'il'} \\ \text{CASE ACC} \end{bmatrix} \\ (a) & \\ & \\ \text{ADJUNCT} & \begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED 'ku'} \\ \text{CASE NOM} \end{bmatrix} \\ (3) & \end{bmatrix} \end{bmatrix}$$

$$
\left[
\begin{array}{l}
\text{XCOMP} \left[
\begin{array}{ll}
\text{SUBJ} & \left[\begin{array}{l} \text{PRED 'ku'} \\ \text{CASE NOM} \end{array} \right] \\
(4) & \\
\text{OBJ1} & \text{PRED 'il'} \\
(b) & \text{CASE ACC} \\
\text{PRED 'ha}\langle(\text{SUBJ})(\text{OBJ1})\rangle'
\end{array}
\right] \\
\text{PRED 'siph}\langle(\text{SUBJ})(\text{XCOMP})\rangle' \\
\text{PRED 'ha}\langle(\text{SUBJ})(\text{OBJ1})\rangle'
\end{array}
\right]
$$

All the pronominal positions in the structure are numbered in the parentheses: the third person masculine pronouns in (1-4) and the third person neuter pronouns (a-b). According to our rules in (10), since all these pronominal positions satisfy the first unification condition except (1-2) and (a), they should be realized as zero pronouns and (2) should be realized as the reflexive pronoun *caki* as follows:

(12)

ku-ka caki-uy il -ul ha-ko siph-ul ttay-man
he-NOM self-GEN work-ACC do-COMP want-COMP when-only.

ha-n-ta.
do-PRES-DECL
'He does his work when he feels like doing it.'

As predicted by the rules in (10), the sentence in (12) is perfectly natural in Korean as opposed to the unacceptable sentence in (3).

2.2 Passives

Passive sentences in English, as demonstrated earlier, correspond to active sentences in Korean except in some very restricted environments. Two basic mechanisms are employed in Korean to reflect the focus on the SUBJ of passive sentences in English: one is in matrix sentences where English SUBJs are translated into TOPICs in Korean and the other is in subordinate clauses that English SUBJs are translated into OBJs in Korean.

Thus, the following changes are in order from English passives to Korean actives:

(13)

The English SUBJs become Korean OBJs.
If those OBJs are in matrix sentences, they are unified with TOPICs.
English agentive phrases become Korean SUBJs, if any.
Otherwise, dummy SUBJs will be created.

3. ECS MT System

The ECS English-Korean Bidirectional Machine Translation System (ECS MT System) uses an indirect transfer method (Nirenberg 1987, Pentheroudakis 1990) and thus, it consists of three distinctive components: analysis, transfer and generation. Since the paper is mainly concerned with the resolution of the differences between two languages, the transfer component will be of our major concern.

The transfer component is a bridge between two independently motivated linguistic analyses within each individual language. It is useful at this point to introduce a schematic flow of the ECS MT System as follows:

| Input Text | → | Word Analysis | → | Sentence Analysis |

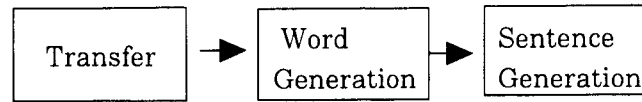| Transfer | → | Word Generation | → | Sentence Generation |
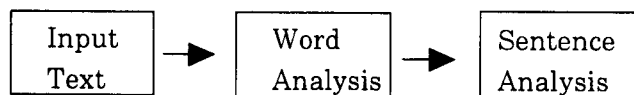
fig. 1

When an input sentence is introduced to the system, the system applies the word grammar of a given language and analyses each word in the sentence. When a word is recognized by the dictionary, the system loads the lexical information of the word from the dictionary. Having completed the analyses of all the words appeared in a sentence, the system further applies a set of sentential rules to analyze a given sentence on the basis of the lexical information provided by the dictionary. As the analysis of an input sentence is concluded, a set of transfer rules are introduced by means of a source-target language dictionary and a set of feature inheritance framewords in the lexicon. That is, when a source word is translated into a target word, the entry will discharge a set of transfer information which will be executed upon the satisfaction of conditions.

The relevant parts of the translation system to this paper can be represented as in the fig. 2.

TRANSFER

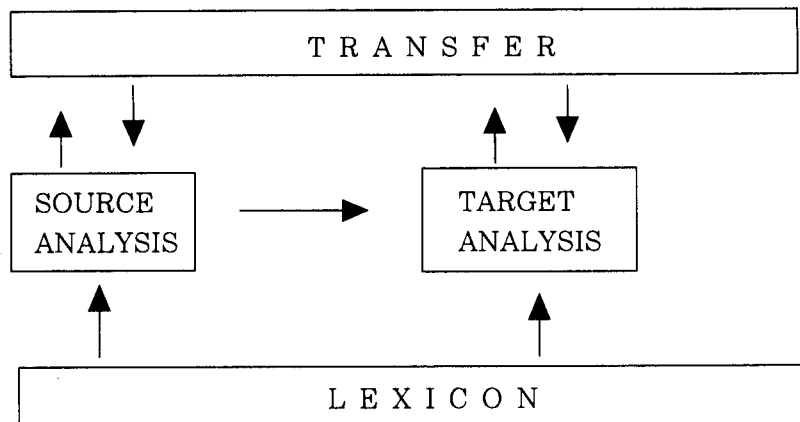SOURCE ANALYSIS → TARGET ANALYSIS

LEXICON

fig. 2

A set of applicable transfer rules interacts with an already analyzed source text through the parsing of the source text to convert into the appropriate counterpart in the target language. These transfer rules will be effective only when a source-target bilingual dictionary entry satisfies a specified condition and discharges these rules from its entry.
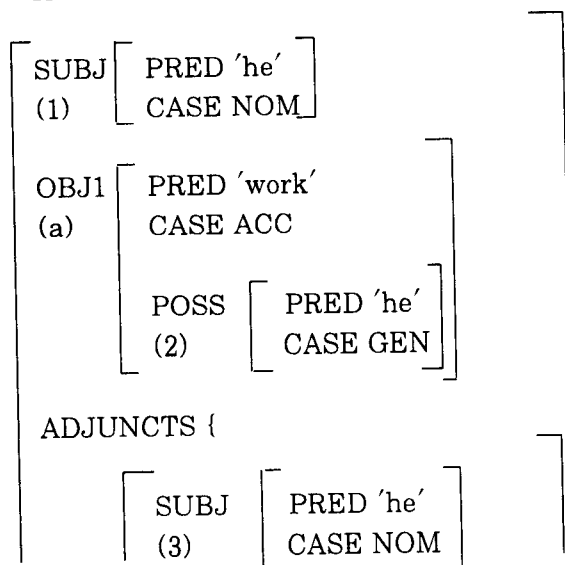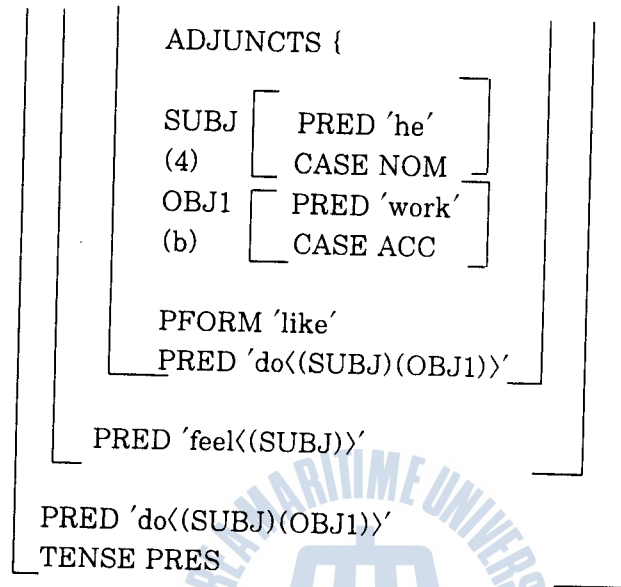
4. Implementation

4.1 Pronouns

The suppression of pronouns is a distinctive property we have to take care of from English to Korean, while the recovery of pronouns is the issue we have to look into from Korean to English.

Consider the following analysed English sentence in f-structure:

(14)
He does his work when he feels like doing it.

$$
\begin{bmatrix}
\text{SUBJ} \\ (1) \end{bmatrix}
\begin{bmatrix} \text{PRED 'he'} \\ \text{CASE NOM} \end{bmatrix}
$$

$$
\begin{bmatrix}
\text{OBJ1} \\ (a) \end{bmatrix}
\begin{bmatrix} \text{PRED 'work'} \\ \text{CASE ACC} \\ \begin{bmatrix} \text{POSS} \\ (2) \end{bmatrix} \begin{bmatrix} \text{PRED 'he'} \\ \text{CASE GEN} \end{bmatrix} \end{bmatrix}
$$

$$
\text{ADJUNCTS} \{
$$

$$
\begin{bmatrix} \text{SUBJ} \\ (3) \end{bmatrix} \begin{bmatrix} \text{PRED 'he'} \\ \text{CASE NOM} \end{bmatrix}
$$

$$
\begin{bmatrix}
\begin{bmatrix}
\begin{bmatrix}
\text{ADJUNCTS }\{ \\
\\
\begin{array}{ll}
\text{SUBJ} \\
\text{(4)}
\end{array}
\begin{bmatrix}
\text{PRED 'he'} \\
\text{CASE NOM}
\end{bmatrix} \\
\begin{array}{ll}
\text{OBJ1} \\
\text{(b)}
\end{array}
\begin{bmatrix}
\text{PRED 'work'} \\
\text{CASE ACC}
\end{bmatrix} \\
\\
\text{PFORM 'like'} \\
\text{PRED 'do}\langle(\text{SUBJ})(\text{OBJ1})\rangle' \\
\end{bmatrix} \\
\text{PRED 'feel}\langle(\text{SUBJ})\rangle'
\end{bmatrix} \\
\text{PRED 'do}\langle(\text{SUBJ})(\text{OBJ1})\rangle' \\
\text{TENSE PRES}
\end{bmatrix}
$$

If we apply the rules described in (10), all occurrences of the third person pronoun **he** (1-4) and the noun **work** (a-b) are related with each other by either Control or Binding. That is, the pronoun **he** in (1) controls pronouns in (3) and (4), and binds the pronoun in (2). The noun (a) controls the pronoun **it** in (b). Thus, the f-structure must undergo the changes compatible with the rules in (10) in addition to the obvious functional label change of the innermost ADJUNCTS to XCOMP.

First of all, the pronoun (2) will be translated into a reflexive pronoun in Korean since the pronoun is coreferent but not unifiable within the f-domain of SUBJ **he** of (1). The others must be realized as zero-pronouns in Korean since they are coreferent with SUBJ **he** of (1) and unifiable in the f-domain of the pronoun.

Secondly, the noun **work** in (b) is coreferent with that of (a) within (a)'s f-domain and is unifiable. Thus, the realization of the pronoun must be zero.

After the application of these rules, the output translation

will be as follows:

(15)

    ku-ka caki-uy il -ul ha-ko   siph-ul ttay-man
    he-NOM self-GEN work-ACC do-COMP want-COMP when-only

    ha-n-ta.
    do-PRES-DECL
    'He does his work when he feels like doing it.'

4.2 Passives

The verbs used in the text can be classified into three different groups:

(16)

    Class 1: Transitive to Transitive
        start -> sicakha
        shift -> itongha
        insert -> sapipha
        turn -> tolli
    Class 2: Transitive to Intransitive
        rotate -> tol
    Class 3: Transitive to Morphological Passive
        release -> phwulli

The implementation of transfer rules for those in Class 1 consists of switching SUBJ to OBJ and unifying OBJ with TOPIC, and thus the focus is placed on the OBJ with a topic marker. The SUBJ to OBJ conversion will take place only if the SUBJ of source language is in a matrix clause. However, in a position where TOPIC is suppressed such as in a subordinate clause, the conversion takes place without the unification with TOPIC. Thus, the transfer process discussed

so far can be stated as follows:

(17)

CHANGE ⟨ ˆ SUBJ ⟩ TO ⟨ˆ OBJ ⟩

&

CHANGE ⟨ ˆ VOICE PASSIVE ⟩ TO ⟨ ˆ VOICE ACTIVE ⟩

　　　 iff ⟨ ˆ VOICE ⟩ = PASSIVE

⟨ ˆ OBJ ⟩ = ⟨ ˆ TOPIC ⟩

　　　 iff ~ (⟨ ˆ SPACT ⟩ = MINUS)

CHANGE ⟨ ˆ ADJUNCT ⟩ TO ⟨ ˆ SUBJ ⟩

　　　 iff ⟨ ˆ ADJUNCT ⟩ :- [PCASE AGENT]

otherwise:

⟨ ˆ SUBJ PRED ⟩ = DUMMY

The rules in (17) state that the grammatical function SUBJ changes into OBJ if and only if the voice of currently analysed sentence is passive. The OBJ will be unified with the TOPIC if and only if the matrix sentence feature SP(eech)ACT has a value other than MINUS. The ADJUNCT turns into the SUBJ if and only if it is subsumed by an agentive pcase, otherwise the SUBJ will be formed with a dummy form.

The remaining two verbs, **rotate** and **release** Class 2 and Class 3 respectively, act differently from the rest of verbs in the text in one important aspect which no such conversion of grammatical function noticed in Class 1 verbs will take place.

The transfer rule for Class 2 verbs are stated as follows:

(18)

⟨ˆSUBJ⟩ = ⟨ˆTOPIC⟩

& CHANGE ⟨ˆVOICE PASSIVE⟩ TO ⟨ˆVOICE ACTIVE⟩

　 iff⟨ˆVOICE⟩ = PASSIVE

　　　 && ~ (⟨ˆADJUNCT⟩ :- [PCASE AGENT])

However, the pair, **rotate->tol,** has to meet one condition in

order to maintain the SUBJ. That is, the ADJUNCT in the same nucleus should not be subsumed by agentive pcase. In other words, the sentence must be an agentless passive, otherwise, the pair will undergo the same change as the rest of verbs did.

The other pair, *release->phwulli,* will select the alternative form to the base form *phwul* by not changing the feature voice.

5. Conclusion

The paper discusses two structural differences between English and Korean, and its resolution based on LFG framework. Though LFG is originated and developed for a purely syntactic analysis, the expressive power of grammar utilized widely in the field of natural language processing is equally applicable to express the resolution of structural differences between languages.

The study illustrates examples demonstrating such differences and describes the linguistic environments. The linguistic environments are stated in LFG functional expressions, which are in turn implemented in the ECS system.

References

Bresnan, J. "Control and Complementation," J. Bresnan (ed.). *The Mental Representation of Grammatical Relations.* MIT Press, Cambridge, Massachusetts. 1982.

Bresnan, J. "Polyadicity," J. Bresnan (ed.). *The Mental Representation of Grammatical Relations:* MIT Press, Cambridge, Massachusetts. 1982.

Bresnan, J. "The Passive in Lexical Theory," J. Bresnan (ed.). *The Mental Representation of Grammatical Relations.* MIT Press, Cambridge,

Massachusetts. 1982.

Bresnan J. and S. A. Mchombo "Grammatical and Anaphoric Agreement," A. M. Farley, P. T. Farley, and K.-E. McCullough, (eds.). *CLS 22* Part 2: Papers from the Parasession on Pragmatics and Grammatical Theory at the Twenty-Second Regional Meeting. Chicago Linguistic Society: Chicago, Illinois. 1986.

Bresnan, J. *Lexical-Functional Grammar.* Duplicated Lecture Notes, Linguistic Institute, Stanford University. 1987.

Bresnan, J. and J. Kanerva. "Locative Inversion in Chichewa: A Case Study of Factorization in Grammar." *Linguistic Inquiry 20.* 1989.

Cho, Young-mee. "An LFG Analysis of the Korean Reflexive *Caki*," Susumu Kuno et al. (eds.). *Harvard Studies in Korean Linguistics I.* 1985

Hong, Ki-Sun. "Discourse Binding of the Korean Reflexive *Caki*," Susumu Kuno et al. (eds.). *Harvard Studies in Korean Linguistics II.* 1987.

Her, O. *Grammatical Functions and Verb Subcategorization in Mandarin Chinese.* The Crane Publishing Co., Taipei, Taiwan. 1990.

Kim, J. *A Lexical-Functional Grammar Account of Light Verbs.* Ph.D. dissertation. University of Hawaii at Manoa. 1991

Kaplan Ronald and Joan Bresnan. "Lexical-Functional Grammar: A Formal System for Grammatical Representation," Joan Bresnan (ed.). *The Mental Representation of Grammatical Representation.* 1985.

Kaplan, R. "The Formal Architecture of Lexical-Functional Grammar," *Proceedings of ROCLING II.* 1989.

Kuroda, S-Y. *Generative Grammatical Studies in Japanese Language.* Ph.D. dissertation, MIT. 1965

Kim, J. "Bracketing Paradoxes in Natural Language Processing." *Proceedings of Seoul International Conference on Natural Language Processing I.* 1990.

Masterson, Deborah. "An LFG solution of the MT problem of category switching."Paper presented at the 7th International Conference on Korean Linguistics, Osaka, Japan. 1990.

Nirenberg, S. (ed.). *Machine Translation: Theoretical and Methodological Issues.* Cambridge University Press. 1987.

O'Grady, William. "Anaphora and Discontinuous Constituents in Korean."Susumu Kuno et al. (eds.). *Harvard Studies in Korean Linguistics I.* 1985

Pentheroudakis, Joseph. "You can get there from here: design and implementation issues in machine translation systems."Paper presented at 1990 annual meeting of the American Nuclear Society, Nashville, Tennessee. 1990.

Sells, Peter. *Lectures on Contemporary Syntactic Theories.* CSLI, Stanford University. 1985.

Tomii, Atsushi. "Tips on English-Japanese Translation." *Proceedings of the 31st Annual Conference of the American Translators Association.* 1990

Wada, Hajime. "Discourse Processing in MT: Problems in Pronominal Translation," Paper presented at COLING-90 in Helsinki, Finland. 1990

Wescoat, M. *Lexical-Functional Grammar.* Duplicated Lecture Notes, Linguistic Institute, Stanford University. 1987.