A Constraint on Lexical Transfer: Implications for Computer-Assisted Translation (CAT)

Kabyong Park*

Abstract

The central goal of the current paper is to investigate lexical transfer between Korean and English and to identify rule-governed behavior and to provide implications for development of computer-assisted translation (CAT) software for the two languages. It will be shown that Sankoff and Poplack’s Free Morpheme Constraint can not account for all the range of data. A constraint is proposed that a set of case-assingers such as verbs, INFL, prepositions, and the possessive marker may not undergo lexical transfer. The translation software is also expected to be equipped with the proposed claim that English verbs are actually borrowed as nouns or as defective verbs to escape from the direct attachment of inflectional morphemes.

Keywords: lexical transfer, language mixing, computer-assisted translation, code-switching, constraints.

I. Introduction

A computer is a system in the sense that it is governed by a definite set of rules. It might be safe to say that a human language is also a system in the same sense. One of the eventual practical goals of linguistic inquiry would be to reach a stage in which the findings of universal properties and language-particular features would enable a translation software to correctly translate a language to another. The aim of the current paper is to discuss the data of lexical transfer or language mixing between Korean and English and to provide implications for a translation software. Machine translation or automated translation refers to translation of one language to a target language by a computer. There seem to be two types of machine translation: rules-based and statistical[14][15]. The former type should contain a combination of rules and the lexicon, whereas the latter is mainly based on big data. As Delpech (2014) points out, an accumulated set of previous translations might constitute a big data[2]. A rule-based translation machine, in turn, heavily depends on an accurate understanding of the differences between the two languages as well as the rules of the two individual languages involved. However, machine translation is far from being perfect. Computer-assisted translation (CAT) might help make appropriate corrections.

Bilingual speakers often show alternation of two languages in their actual speech behavior for some reason. Kusati (2014), for example, argued that teachers may deliberately use language mixing to make a certain point clear[7]. The study of language mixing has been one of the hottest and most interesting areas in the field of bilingualism. Recent studies generally agree that ‘language mixing’ of bilingual speakers is not a random phenomenon but rather a rule-governed behavior. Thus, the findings of this area would provide meaningful implications for machine translation or computer-assisted translation.

*First Author: Kabyong Park, Corresponding Author: Kabyong Park

*Kabyong Park(kpark@nsu.ac.kr) Dept of Global Studies, Namseoul University

*Received 2016 11 01, Revised 2016 11 15, Accepted 2016 11 26

*This work is supported by a Research Grant of Namseoul University
I will use the term 'language mixing' as a sort of general cover term for all kinds of language contact phenomena including code-switching, lexical transfer and language borrowing. Most studies of language mixing, however, have dealt with code-switching of a pair of languages which share basic grammatical structures such as phrase structure rules. Thus, various kinds of linguistic constraints governing code-switching have been proposed in the literature. Zainuddin (2016), for example, reported that code-switching may occur at every level of syntactic categories: word, phrase or sentence[16]. As far as I know, there is not much research reported in the literature on lexical transfer. There is not much research done on any kind of language mixing for a pair of structurally quite different languages, either.

Therefore, the study of lexical transfer data from a pair of quite different languages might be expected to identify systematic patterns or generalizations, which can be utilized in an attempt to design a translation software or computer-assisted translation. I believe that lexical transfer is as much a principled linguistic behavior as code-switching, as Pfaff (1979) assumes[10]. The purpose of this paper is to examine Korean/English mixed data and to explore the question how lexical transfer (from English to Korean) is constrained.

The content of the present paper can be summarized as follows. In section 2, I will briefly address the issue of the distinction between code-switching and lexical transfer, and also the contrast between lexical transfer and language borrowing. Though it is difficult in a strict sense to determine how to distinguish these kinds of language contact phenomena, I suppose that some kind of guideline as to the concept of lexical transfer is necessary for the present discussion. Section 3 will be devoted to the examination of actual Korean/English mixed data. I will propose a constraint that a set of case-assigners such as verbs, INFL, prepositions, and the possessive marker may not undergo lexical transfer. I do not claim that this is a general principle for lexical transfer of any pair of languages. A linguistic theory of any subfield within the generative framework is expected to be universal, since linguistics is not a science of any particular language but of human language faculty. Therefore, this constraint must be tested empirically (and conceptually, of course) for universality, if it is to be a valid linguistic constraint. I assume that language assignment is possible for all the mixed sentences, whether code-switched or lexically transferred, and that the matrix language of all of my examples is Korean.

II. Concept of Lexical Transfer

Most researchers agree that lexical transfer is different from code-switching, though there is still controversy as to how to distinguish one from the other. Pfaff (1979), for example, proposed a speakers’ competence-based hypothesis: ‘borrowing’ (‘lexical transfer’ in my terminology) may occur in the speech of those with only monolingual competence, while ‘code-switching’ implies some degree of competence in two languages[10]. I think, however, there is no reason to believe that only monolingual speakers but not bilingual speakers employ lexical transfer. Lexical transfer may also occur in the speech of bilinguals, if necessary situations arise.

Another distinction between code-switching and lexical transfer is found in Clyne (1987)[11]. Code-switching refers to the cases when “the speaker stops using Language A and uses Language B, so that syntactic connections are now with items from the speaker’s Language B system.” In the case of lexical transfer, on the other hand, a single lexical item is transferred from Language A to B (or vise versa), whether integrated into grammatical and/or phonological system of the recipient language or not. She is not clear, however, what she means by ‘a single lexical item’. Since there are many instances of lexical transfer of compound nouns and verbs, I consider the term ‘lexical’ as the counterpart opposed to ‘phrasal’ in terms of X-bar theory.

I do not agree to Clyne’s (1987) assumption that grammatical adaption is not important[11]. My definition is as follows: Rules -- phonological, morphological, and/or syntactic--that already exist in Language A determine the grammatical features of a transferred word from Language B and regulate the accomplishment of mixing it into Language A. In other words, a transferred word is subject to all rules of the recipient language, though it is not fully incorporated into the lexicon of the recipient language. From the speakers’ point of view, when a bilingual utters a code-switched sentence, (s)he is aware that (s)he starts speaking Language A, stops speaking Language A, and switches to Language B. In the case of lexical transfer, on the other hand, the speaker thinks that (s)he is speaking the same language L1, though (s)he is aware that (s)he employs some linguistic entities, whether they are just single lexical
entries—as assumed in the standard definition—or phrases, from L2.

This is similar to Grosjean’s (1982) distinction: a code-switch can be of any length (a word, a phrase, a sentence) and is a complete shift to the other language, whereas a borrowing (my ‘lexical transfer’) is a word or a short expression that is adapted phonologically and morphologically to the language being spoken[4].

I am not saying, however, that there must be always a clear-cut distinction, and that a given mixed sentence, therefore, ought to belong to one of the two categories. As indicated in Grosjean’s definitions, there seems to be some kind of overlap between code-switching and lexical transfer. Since the exact distinction between these two kinds of language mixing is not my main concern here, I will simply assume, according to the above definitions, that the Korean/English data in this paper are instances of lexical transfer.

Let us discuss, in turn, the relationship between lexical transfer and language borrowing. While ‘lexical transfer’ refers to the alternation of two languages at the level of individual speech — given that elements from L2 are adjusting to the structure of L1—, ‘language borrowing’ is considered to have taken place at the level of language and is considered part of L1. From the speakers’ point of view, when they utter a borrowed word at the level of language borrowing, they are not conscious that they are borrowing it from L2. However, most linguists seem to agree that a borrowed word, though not always, transforms gradually in most cases from the level of lexical transfer (‘speech borrowing’ in Grosjean’s (1982) terminology) to the level of language borrowing[4]. In other words, a borrowed word is synchronically at some point of lexical transfer—language borrowing continuum. Therefore, the distinction between lexical transfer and language borrowing is, in fact, as difficult as the distinction between lexical transfer and code-switching.

III. A Constraint on Lexical Transfer from English to Korean

Consider the following examples. The following acronyms are used throughout the paper:

(a) a. DOOR-({k, i} yol- li-ass-ta.
NOM open-PASS-PAST
‘The door was opened.’
b. mun-({k, i} yol- li-ass-ta.
NOM opened
‘The door was opened.’
c. kake-({k, i} yol- li-ass-ta.
store NOM opened
‘The store was opened.’

‘ka’ and ‘i’ are the two allomorphs of the same morpheme, the Nominative Marker. As illustrated in the contrast between (1b) and (1c) above, the allomorphy is determined by the value of the final sound of the subject NP: ‘ka’ after a consonant, and ‘i’ after a vowel. However, the Korean–English mixed sentence (1a) is not consistent with this allomorphy. Notice that the English word door ending with a consonant takes the allomorph ‘ka’ instead of ‘i’. This is a clear instance of lexical transfer. What it implies is that once an English word is lexically transferred into Korean, it is subject to the rules that already exist in Korea. A morpheme structure constraint in Korean does not allow [r] to occur at the word-final position. Thus, Korean speakers borrowing the English word door would reconstruct the structure of syllables and consider it having two syllables [do-er]. Now, the reason for the contrast between (1a) and (1b) is clear.

Now let us discuss the question of what kinds of linguistic element may not be transferred. First, let us look at the verb phrases.

(2) Mary-ka Room-({ lul }) yeyak- ha-ass-ta.
NOM ACC reserve-PAST-DEC
‘Mary reserved a room.’
(3) *Mary-ka ROOM-({ lul }) RESERVED.
NOM ACC
‘Mary reserved a room.’
(4) *Mary-ka bang-({ lul }) yeyak- ha-ass-ta.
NOM room ACC reserve-PAST-DEC
‘Mary reserved a room.’
(5) *Mary-ka bang-({ lul }) RESERVED-ASS-ta.
NOM room ACC reserved-PAST-DEC
‘Mary reserved a room.’

The example in (2) shows that nouns inside VP as well
as the external argument, as shown in (1), can be transferred. The parenthesis indicates that the accusative marker is optional. Given that nouns may be transferred, the ungrammaticality—actually the impossibility of transfer of (3)—must be due to the transfer of the verb reserve. This assumption is again supported by the starred example (4), where only the verb is transferred. The example in (5), in turn, illustrates that the past morpheme may not attach to the English verb reserve. This might be explained by Sankoff and Poplack’s Free Morpheme Constraint which can be stated as follows[11]:

(6) A switch may not occur at the boundary of a free morpheme and a bound morpheme.

In fact, there are some instances that the condition (6) can account for. Not only inflectional affixes such as the past morpheme as but also some derivational suffixes may not attach to borrowed English roots, as shown below.

(7) a. arumtap-ke 'beautifully'
   beautiful-ADVERBIAL
   b. * BEAUTIFUL-ke
(8) a. ppop-ki 'selection'
   select-NOMINAL
   b. * SELECT-ki

However, the constraint in (6) can not account for all the range of data to prove that it is valid. For example, the plural affix 'tul' and some derivational suffixes such as the verbalizing affix 'hwaha' may attach to a borrowed word, as illustrated in (9).

(9) a. iron-till 'theories'
   theory-PL.
   b. THEORY-til
   c. iron-hwaha 'to theorize'
   theory-ize
   d. THEORY-hwaha

Now, one might ask why any korean affix may not attach to a borrowed English verb. Put in a different way, why nouns but not verbs can escape from the effect of the Free Morpheme Constraint? One could simply stipulate that verbs may not be transferred from English to Korean. There is some reason to believe that a Korean speaker considers the borrowed English verb in (5) and (8b) as an occurrence of a noun instead of a verb, as will be shown later. I will suggest later that a more general constraint will exclude the transfer of verbs. One might also ask whether there is any way to incorporate an English verb to a Korean-English mixed sentence. The answer is “Yes”. Consider the following examples.

   NOM room-ACC do-PAST-DEC
   'Mary reserved a room.'
   b. ? Mary-ka bang-lul
   RESERVATION-ha-ass-ta.
   'Mary reserved a room.'

The only difference between (10) and (4-5) is the presence of the operator ‘ha’ in the former. Similar type of lexical transfer is found in Japanese and Panjabi.

Romaine’s (1986) study of Panjabi-English mixing shows that the process of compound mixture using the operator 'kerna' (to do) is highly productive (For example, phonam kerna ‘to telephone’). Nishimura (1986) points out that Japanese also exhibits the same kind of phenomenon[8].

(11) a. RIGHT IN THE CENTER GROW-shitara
   'IF (they) grow (it) in the center of…'
   b. * RIGHT IN THE CENTER GROWTH-shitara
   'IF (they) grow (it) in the center of…'

The contrast in (10) and (11) is interesting, since the Japanese operator suru and the corresponding Korean one ha attach to only nouns in Japanese and Korean, respectively. I will turn to the discussion of the so-called Sino-Korean verbs to account for the contrast between (5) and (10) and the contrast in (10–11). It is well known that a large portion of Korean and Japanese vocabulary is borrowed from the Chinese language. A substantial class of compound verbs in Korean, that is, Sino-Korean verbs can be traced back to Chinese origin: a Sino-Korean verb can be characterized as compound verbs consisting of a noun of chinese origin and the general verb ‘ha’. Japanese also abound with borrowings from Chinese and a Sino-Japanese verb shows the same morphology as a Sino-Korean verb: a borrowed noun plus ‘suru’. Some examples are given below in (12).
Park (1987) shows that the class of Sino-Korean verbs constitutes an exceptional class on the basis of their exceptional phonological and syntactic behavior[9]. For example, the causative construction in (13) and the passive construction in (14) are confined to only this group of verbs.

population problem-ACC study-do-PAST  
‘John studied the problem.’

him-DAT study-CAUS-PAST  
‘Sue made him study the problem.’

(14) a. ku-ka ku pemin-lul chepho-ha-ass-ta.  
He=NOM the criminal-ACC arrest-do-PAST  
‘He arrested the criminal.’

the criminal-NOM he-by arrest-PASS-PAST  
‘The criminal was arrested by him.’

The following examples show that a transferred English verb falls under the same mechanism as a Sino-Korean verb by analogy. Notice that a Korean native verb 'camku' (to lock) and the borrowed English verb exhibit different passive morphology, as in (15-16) and, again, the same causative morphology confined to the class of Sino-Korean verbs is employed in (17).

(15) a. mun-i camku-ki-ass-ta.  
door-NOM lock-PASS-PAST  
‘The door was locked.’

door-NOM lock-PASS-PAST  
‘The door was locked.’

door-NOM PASS-PAST  
‘The door was locked.’

door-NOM PASS-PAST  
‘The door was locked.’

The assumption that borrowed English verbs in the above examples should be considered categorically as nouns can be supported by the fact that some grammatical function markers may appear between the borrowed English verbs and the operator ha, as in the case of Sino-Korean verbs. Park (1987) shows that some grammatical function markers can appear inside the compound verbs: such as the object marker lul 'the plural marker til and the so-called delimiters (e.g. 'to' also)[9]. Delimiters are sort of particles such as 'nin' (only) and 'to' (also), which are attached to nouns and thereby semantically delimit the scope of meaning. In English, adverbs (often called 'quasi-quantifiers') like 'only' and 'also' are independent words, whereas, in Korean, the corresponding forms mentioned above act like affixes. What is important here is that these grammatical function markers may attach to only nouns, not to verbs. Park (1987) shows that these delimiters may attach to only nouns by showing that they can also modify verbs if the affix 'ka' is attached to verbs and nominalizes the verbs, as illustrated in (21)[9].

(17) Mary-eke bang-lul RESERVE=sikhi-ass-ta.  
DAT room-ACC CAUS-PAST  
‘(Someone) made Mary reserve a room.’

(18) * mun-i LOCKED-ass-ta.  
door-NOM PAST  
‘The door was locked.’

(19) mun-i LOCKED-day-ass-ta.  
door-NOM PASS-PAST  
‘The door was locked.’

(20) nay-ka CARD-lul LOST-ha-ass-ta.  
I=NOM ACC do-PAST  
‘I lost the card.’

NOM apple-ACC eat-NOMI-DLT  

NOM apple-ACC eat-DLT-PAST
The examples in (22) show that delimiters can appear inside the compound verbs such as 'sijak-ha' (to begin) and the English-Korean mixed verb 'PUBLISH-ha' (to publish).

   N homework-ACC beginning-DEL do-PAST
   "John began the homework, but...."

   N the paper-ACC also do-PAST
   "John also published the paper (he edited it)."

The plural affix 'tul' offers another evidence in favor of the assumption that the Korean speakers consider the borrowed English verbs as actually nouns if they appear in a mixed sentence. The plural affix in the following examples indicates the plurality of the subject NP rather than the root noun of the derived verb.

(23) a. kutul-in tu sikan chwichim-tull ha-ass-ta.
   they-TOP two hour sleep-PL do-PAST
   "They slept for two hours."

b. salam-tul-i chayk-tul RECOMMEND-tul
   people-PL-NOM book ACC PL
   "People recommended this book."

More evidence in favor of the analogy-hypothesis comes from the negation constructions: a Sino-Korean verb and a English-borrowed verb show the same behavior with respect to negative constructions. Therefore these two must have the same internal structure. It follows that the borrowed English verb should be considered as a Noun instead of a Verb, or perhaps as a defective verb. Consider the examples in (24–26). Notice that a negative formative an 'not' appears in front of a native Korean verb, as shown in (24), but may not appear before a borrowed compound verb, whether the verb is borrowed from Chinese or English, as shown in (25–26).

   NOM party-to not go-PAST
   "John did not go to the party."

   *NOM-ACC the borrowed-PL ACC
   "It does not make sense to borrow a noun from English."

b. ku-ka ku saken-lul josa an ha-ass-ta.
   NOM-ACC the borrowed-PL ACC
   "He did not investigate the incident."

(26) a. *John-i ku nommun-lul an
   NOM the paper-ACC not
   PUBLISH-ha-ass-ta.
   do-PAST
   "John did not publish the paper."

b. John-i ku nommun-lul PUBLISH an
   NOM the paper-ACC not
   ha-ass-ta.
   do-PAST
   "John did not publish the paper."

Park (1987) claims that in the internal structure of a Sino-Korean verb X-ha, the noun X carries argument structure of the derived verb and determines theta role assignment[9]. The Korean-English mixed data given here, however, seem to lead us to the assumption that only verbs carry argument structure and that the internal structure of the borrowed English verb in (10a), for example, would look like (27). Sino-Korean verbs would, of course, have the same structure.

```
V
|   \ N V
|    | V
|    |  RESERVE
```

Figure 1

Now, I can account for the contrast between (5) and (10a) and the contrast between (10a) and (10b). The examples are repeated here for the sake of convenience. The syntactic node N in (27) is responsible for the ungrammaticality of (5), since the verbal affix may not attach to a noun.

   NOM room ACC reserve-PAST-DEC
   'Mary reserved a room'

Let us turn to another issue of this paper. We have seen so far that English verbs may not be transferred into Korean. Let us look at the prepositional phrase. Consider the following examples.

(27) a. Mary-'ka LIBRARY-ese kongbu-ha-n-ta. NOM in studies 'Mary is studying in the library.'

b. Mary-'ka tosekwon-IN kongbu-ha-n-ta. NOM library studies 'Mary is studying in the library.'

The contrast in (27) shows that a preposition may not be borrowed. Thus, one might propose the following constraint, since both verbs and prepositions are heads of the phrases VP and PP, respectively.

(28) A head of a phrase may not undergo the process of lexical transfer.

However, a head noun of a Noun Phrase may be transferred, as illustrated in (29). Thus, we need to modify the condition (28).

(29) a. arumtaun House 'beautiful house'

b. BEAUTIFUL cip 'beautiful house'

c. BEAUTIFUL-ha-n cip 'beautiful house'

The contrast between (29b) and (29c) can be accounted if we adopt (i) Park's (1987) claim that the second group of Sino-Korean verbs, whose example is given in (30a), has the same internal structure as those we have discussed and (ii) that the internal structure looks like Fig. 1[9]. The contrast between (30b) and (30c) might be explained by the same account of the contrast (10a) and (10b).

(30) a. Mary-nin hayngpok-ha-ta. TOP happiness-be 'Mary is happy.'

b. * Mary-nin HAPPINESS-ha-ta. TOP happiness-be 'Mary is happy.'

c. Mary-nin HAPPY-ha-ta. TOP happiness-be 'Mary is happy.'

Now, I will propose a condition which can explain all the data discussed in this paper.

(31) A set of case-assigners may not undergo the process of lexical transfer.

This constraint explains why (i) English verbs (For example, (10)), (ii) inflectional elements (for example, (5)), (iii) prepositions (for example, (27)) may not be transferred into Korean. It also accounts for the following examples, assuming that the possessive affix assigns a Case to the preceding noun.

(32) a. emeni-iy nal mother-POSS day 'Mother’s day'

b. MOTHER-iy nal
c. MOTHER-iy DAY

d. emeni’s nal

Pfaff (1979), in the study of Spanish-English mixed speech, offers a quantitative analysis of instances of single lexical item transfer[10]. The result of his study in Table 1 seems to be consistent with the constraint in (31). Compare the number of transferred nouns to that of verbs.

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>S – E</td>
<td>818</td>
<td>71</td>
<td>43</td>
</tr>
<tr>
<td>E – S</td>
<td>14</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Pfaff (1979) also observed that morphological adaptation of verbs has been dominant in many language-contact situations among Indo-European languages, though it is not productive at all[10]. The constraint (30), on the other hand, would predict at least that a verb will not be easily transferred from another languages. This contradiction is probably because (i) the
situation depends on the degree of differences between languages, or (ii) the predominance of morphological adaptation in verbs is true of only code-switching but not of lexical transfer, or (iii) the constraint in (31) is not universal at all.

IV. Concluding Remarks

Believing that lexical transfer is as much a principled linguistic behavior as code-switching, I have tried to find how lexical transfer is constrained in the Korean–English mixed data. The constraint proposed excluding lexical transfer of a set of case-assigners such as verbs, INFL, prepositions, and Possessive Marker should be tested cross-linguistically to be a valid constraint. One of the implications for computer-assisted translation would be that the software must know that English is very rich in conversion: a word can be used as more than one syntactic category and that English verbs are borrowed as nominal elements into Korean, perhaps because they may not directly attach to inflectional morphemes. Otherwise, the most frequent mixed expression like ‘PUBLISH-ha-ass-ta’ would be translated as ‘did publish’, which means that all the verbs like that would be incorrectly considered having an emphatic meaning.

REFERENCES


Author

Kabyong Park received his B.A. degree in English from Hankuk University of Foreign Studies in 1989, and M.A. and Ph.D. degrees in Linguistics from University of North Carolina at Chapel Hill in 1986 and 1992, respectively. Dr. Park joined the faculty of Department of English at Namseoul University located in Cheonan, Korea, in 1994. He is currently a Professor in the Department of Global Studies at Namseoul University. He is interested in morphology, syntax, mobile teaching/learning and translation.