Complex Bidirectional Transfer in the ECS English–Korean System

Joseph E. Pentheroudakis

This paper discusses some aspects of the transfer component in the ECS machine translation system. We present the structure of the transfer grammar and the transfer entries by using examples from our English–Korean/Korean–English (EK/KE) system; we then discuss some examples of compositional and non-compositional translation and we show how these are treated in our system.

1. System Overview

The ECS system is an implementation of the Lexical Functional Grammar theory and formalism (Bresnan & Kaplan (eds.) (1982)). The system is stratificational in its organization: for a given sentence, analysis rules generate a constituent structure (c-structure), which is mapped to the functional structure (f-structure) by means of functional equations annotated to the lexicon and to the phrase structure rules. Transfer then produces an f-structure appropriate for the target language, which is submitted to the generation grammar for a more complete target-language specification (for example, to introduce target-language function words such as auxiliaries, specifiers etc., and to generate inflected target forms), leading to the generation of the target sentence. Transfer may be viewed as the first stage in generation, since it bears the initial responsibility for generating a well-formed, albeit abbreviated, structure with respect to the target language. This structure will be augmented by the generation grammar: the output of transfer must subsume the final f-structure produced during generation. Transfer will be discussed in more detail later in the paper.
The dictionary design follows the conventional organization for transfer-based systems: there are monolingual lexical entries, containing information used during analysis or generation; and bilingual, or transfer, entries specifying lexical and/or structural transfer conditions and actions.

The description of each language is independently motivated and is in principle not affected by the description of possible target languages. This allows the modular development of grammars: for instance, the English grammar and lexicon developed over the last five years have been used in the analysis portion of our English-Chinese, English-Japanese, and English-Korean systems and in the generation portion of our Korean-English system. The same grammar was also used in the development of English-French/French-English, English-Spanish/Spanish-English, and English-German/German-English prototypes.

The transfer grammar and dictionaries, on the other hand, are language-pair specific. The transfer grammars currently consist of rules organized in explicitly ordered groups associated with each translation, specifying contrastive and/or structural correspondences (mappings) of feature sets in the functional structure.

In a transfer entry, a source word may specify several translations. In this case, each translation specification may be annotated with appropriate selection conditions and/or actions. This allows us to handle cases of polysemy as well as to select one among several disjunctive target structures configurations corresponding to different source configurations for a given source word. Selection of the appropriate translation involves the evaluation of information from several levels, namely:

a) the technical domain(s) of the text (currently specified by the user when the parser is invoked): a given translation can specify a disjunction of appropriate technical fields;

b) conditions over the feature set configuration in the f-structure: these conditions can be expressed as constraints, or in terms of subsumption or unification operations. For example, the English verb 'eat' must translate in Korean as 'mek-' if it is used transitively, as in 'they always eat kimchee at picnics', but as 'pap-ul mek-' 'rice+ACC eat' when used intransitively, with a generic object interpretation, as in the sentence 'they always eat at picnics'. In the first case, the input structure must be sub-
sumed by the control feature structure

\[
\begin{array}{l}
[ \text{SUBJ ANY} \\
\text{OBJ ANY} \\
]\end{array}
\]

where the special attribute value ANY will subsume any instantiated value of the functions SUBJ and OBJ. The second case, which will be listed as the next translation, will be evaluated next since the subsumption above will fail (there will be no OBJ in the intransitive sentence) and will need to be subsumed by the control feature structure.

\[
\begin{array}{l}
[ \text{SUBJ ANY} \\
\text{VOICE ACTIVE} \\
]\end{array}
\]

c) semantic preference: a semantic preference network is also implemented to help in the resolution of polysemy. Each translation in the bilingual entry can select a node on the network, and it can also express a preference for the semantic marking of the translation of one or more of its arguments or modifiers or for the head of the construction in which it appears. For example, the English verb 'draft' in a sentence such as 'they drafted a plan' vs. the sentence 'John was drafted' can be resolved if the translation 'cingcipha-' (conscript) in the EK transfer lexicon states a preference for a PATIENT role marked as HUMAN or any node subordinate to it, while the translation 'koanha-' (draw up) states a preference for a PATIENT role marked as, say, ABSTRACT, or any node subordinate to it. Grammatical functions are mapped to semantic roles, thus allowing the statement of preferences without direct reference to the functional structure.

The system can also be operated interactively during transfer, to request the user's assistance in resolving structural or translational ambiguities. Note that this feature was incorporated to conform to the needs of a specific installation; the system as originally designed did not require user interaction.

The system also includes a discourse analysis function for the resolution of pronominal anaphora to ensure the correct translation of pronouns (Wada (1990)).
Feature sets in the ECS system are uniformly expressed as directed acyclic graphs (DAGs). The grammar development environment includes a formal language, LECS, and offers the linguist extensive rule tracing and debugging tools; it also includes a complete set of utilities to maintain the linguistic database (rules, lexical entries and various control tables).

Finally, the system includes a user-friendly dictionary maintenance interface, allowing the end-user to update the information in the dictionaries. For a good overview of the ECS system, see Her (to appear); for a description of the ECS grammar development environment, see Pentheroudakis (1990).

2. General Approach

We share the view that a translation system and the linguistic descriptions that comprise it should be extensible both vertically within a given language pair and horizontally, across to other language pairs.

The first requirement means that, as in any natural language processing system, system improvements should not require grammar and lexicon rewrites, and, conversely, that grammar and lexicon extensions should not require arbitrary system redesigns. Linguistic information should be represented in a principled and independently motivated manner.

The second requirement is peculiar to MT, and it means that the representation of linguistic phenomena and the formalism used to express them ought to be such that porting a grammar and lexicon to another language pair would in principle (and, one hopes, in practice) not require rewriting any already defined portions of the system.

The components of a translation system should, therefore, be independently extensible and portable. Our decision to implement our system within the LFG formalism was largely motivated by the need to satisfy both these requirements. In the remainder of this paper we will discuss our approach to some difficult translation problems in our English-Korean and Korean-English system, which will illustrate our approach to transfer.

The formalism used is a modified version of the actual formalism as currently implemented, and reflects research in progress. This does not affect the validity of our remarks or the content of our output.
3. Non-Compositional Translation

A translation correspondence is said to be compositional if ‘the translation of the whole is some reasonably straightforward function of the parts taking account of the way they are combined together’ (Arnold et al. (1988)) (our italics). Consider the English sentence ‘Bush was elected in 1988’ translates in Korean as ‘Bush-nun 1988 nyen-ey senchwultoy-ess-ta’. Regardless of what structure(s) one uses to represent the linguistic information in this sentence, the source and target representations will be isomorphic and the translation will be compositional: the translation of the sentence is the translation of each of its parts, taking into account the information conveyed by the passive voice in the source. The command relation between the head and its arguments and modifiers in both sentences will match in a straightforward way. Thus, in our LFG framework, the f-structures of the two will be as shown below (in abbreviated form):²

English f-structure:

```
[ SUBJ [ FORM “Bush” ]
  FORM “elect”
  PRED # SUBJ #
  VOICE PASSIVE
  TENSE PAST
  ADJUNCTS { [ FORM “1988”
    OBL IN
    ]
    }
  ]
<⋯> ]
```

Korean f-structure:

```
[ SUBJ [ FORM “Bush”
  TOPIC PLUS]
```

² In our implementation of LFG, the value of Pred is the list of a verb’s governable (and locally governed) functions; the string form of the word appears as the value of the attribute FORM.
There are many cases, however, where the source and the target representations are severely non-isomorphic, their translation being characterized as irregular or non-compositional. The most notorious such case, discussed amply in the literature (Arnold et al. (1988), Kaplan & Netter (1989), Zajac (1990)), is the case of modifier-head switching, illustrated by the English-French pair 'they just arrived'/‘ils viennent d’arriver’, or the German-English pair ‘ich schwimme gern’/‘I like to swim’. In these examples the meaning contributed by the adverbial in the source sentence (‘just’ and ‘gern’) is expressed in the target by a higher verb, and the proposition the adverb modifies corresponds to the untensed complement of that higher verb.  

Two approaches have been suggested to handle these cases: (a) representing the two sentences as generally isomorphic on some projection or level (Kaplan & Netter (1989), Zajac (1990)), thus reducing the complexity and the representational requirements of the transfer relation, but at the cost of introducing information in the source language analysis that is inspired (if not motivated) by the structure of the target sentence; and (b) defining the contrastive correspondence between the two structures during

---

3 The underlying assumption here is that it is desirable to define a non-compositional translation relation between two sentences if that translation is the most natural one, even if there may exist a compositional but more awkward translation. This is partly due to the difficulty of identifying “acceptable awkwardness”: a word-for-word translation is, of course, the limiting case. There is a stronger reason, however: current MT systems normally use generation grammars that impose target-language structural and lexical constraints. It is much more difficult to produce a compositional, but awkward translation, since that translation may be ungrammatical with respect to the target grammar.
transfer, as in the EUROTRA system, the system at PAHO and many other transfer-based systems. We subscribe to the view that, given the requirement for modular and independent grammar development, as well as the fact that the translation correspondence can often be structurally arbitrary, any attempt to resolve any structural mismatches belongs in transfer and not in the source language analysis. Undeniably, this reflects our decision to represent linguistic information on the functional structure level alone.

4. The Problem: Adjunct Mapping

We examine some examples where in a well-formed description of the target sentence a modifier (adjunct) of a word in the source cannot modify the translation of that word in the target; rather, in a target sentence with the same meaning, the counterpart of this modifier is realized as an adjunct of a different level in the functional structure. The ill-formedness of a compositional translation is often due to idiosyncratic structural and lexical differences between the two languages.

We will discuss two types of such problematic correspondences:

a) scope-sensitive mapping, where an adjunct may appear on a different level of the target functional structure to reflect a difference in the scope of the adjunct in question;

b) headless adjunct mapping, where the head of an adjunct forms part of a discontinuous lexical item, no mapping for which may be defined compositionally.

a) Scope-sensitive Mapping

Consider the following examples and their relevant abbreviated functional schemata (each example is followed by a literal translation; ‘(n)e’ is the English sentence, and ‘(n)k’ its desired Korean translation):

John’s failure to win surprised us.

[John-uy iki-nun silphay-ka wuli-lul nollakey ha-ess-ta]

[ SUB [ POSS [ FORM “John”] ] FORM “failure” POLAR NEG ]

(1e)
The fact that John was/is not able to win made us surprised.

John's repeated failure to win surprised us.

John-i iki-ci mosha-n/nun kes-i wuli-lul nollakey ha-ess-ta

[The fact that John was/is not able to win made us surprised]
The fact that John was/is repeatedly unable to win made us surprised.

[The fact that John was/is repeatedly unable to win made us surprised] (note ADJUNCTS placement:)

John's unexpected failure to win surprised us.

[John-uy iki-nun ttuspakkuy silphay-ka wuli-lul nollakey ha-ess-ta]
The unexpected fact that John was/is not able to win made us surprised.

In examples (1e/1k) the correspondence is straightforward: the English-Korean transfer entry for failure will specify the following operations:

a) translate 'failure' to 'kes';
b) map its XCOMP to an SCOMP;
c) unify POLAR NEG into the target’s SCOMP; and

---

We will not be concerned here with the projection of the source verb to lower levels of functional structure (inside a modal or a caustive): these operations are straightforward and non-problematic.
d) unify into the target structure TENSE, VOICE and any other attributes necessary to constrain the information in the target sentence.

The transfer entry for 'failure' that would perform these operations would look like this:

\[
\text{ek_failure} :: \\
[T \{\text{FORM "kes"} \}
\{[\text{MAP (E(<^\text{XCOMP}>, K(<^\text{SCOMP}>)))} \\
\quad <K^\text{SCOMP POLAR}> = \text{NEG} \\
\quad \{[<E^\text{/}{\text{SUBJ OBJ ADJUNCTS}}\text{\TENSE}> \\
\quad \quad = \text{c PAST} \\
\quad <K^\text{SCOMP TENSE}> = \text{PAST}] \\
\quad [<K^\text{SCOMP TENSE}> = \text{PRES}] \}
\}
\]

The curly brackets {} enclose disjunctions of rule or attribute-value sets, each disjunct being enclosed in square brackets [ ]. The caret (^) represents LFG's up-arrow metavariable.

The MAP function maps the English XCOMP to the Korean SCOMP; the metavariable ^ can be modified by a source or target language operator (E and K in this case). The / and \ symbols are the implementation of upward-looking functional uncertainty, an extension to the functional uncertainty algorithm introduced by ECS: the / (up) operator searches for the listed sequence of mothers from the level specified (here, the up-arrow ^), while the \ (down) operator will search for the specified sequence of daughters down from a given level, which will f-command the source function level. Thus, in the example above the path

\[
<^\text{/}{\text{SUBJ OBJ ADJUNCTS}}\text{\TENSE}> 
\]

will have a solution if the level pointed to by the up-arrow (^) is the value of the function SUBJ, or OBJ or is in the head's ADJUNCTS set, and furthermore if the path can end in an f-commanding attribute TENSE. This notation allows us to handle such structurally divergent sentences as 'I
was surprised by his failure to win' or 'I didn't anticipate his failure to win', where 'failure' may be projected to ADJUNCTS and OBJ, respectively.

Let us now look at the slightly more complex transfer operations in (2e/2k) and (3e/3k). In (2k), the noun 'kes' is not (and cannot be) modified by an adjunct; rather, the schema of the sentence that represents the most natural translation of (2e) includes the adjunct 'panpokhayse' ('repeatedly') modifying the verb 'mosha-' in the SCOMP. To handle this, we will need to augment the transfer rules for 'ek _ failure' so as to allow:

a) the correct translation of 'failure' ('kes' of 'sasil') depending on the presence or not of sentence adverbials modifying the source noun; and
b) the correct structural mapping of any modifiers of the noun 'failure'.

To ensure the correct translation define an operation that would successfully translate (2e) to (2k), for example, and (3e) to (3k), we need to ensure that the translations of 'repeated' and 'unexpected' are the appropriate ones, given their new placement in the functional structure. In our system, this is accomplished by the generation grammar; the generation rules will generate the appropriate form given the environment of the word in the functional structure.

Let us now revise the EK transfer entry for 'failure'. First, we define a transfer rule that will be shared by both translations, namely, a rule specifying the mapping operation and the assignment of tense, polarity, etc.:

Rule1 ::

[MAP (E(<^XCOMP>), K(<^SCOMP>))
  <K^SCOMP POLAR> = NEG
  [[<E/{SUBJ OBJ ADJUNCTS}\TENSE> = c PAST
  <K^SCOMP TENSE> = PAST]
  [<K^SCOMP TENSE> = PRES]
]

We then define the EK entry for 'failure':

ek _ failure ::

[T { [ FORM "sasil"
The symbol :- is the subsumption operator; the expression “A :- B” is interpreted as “A is subsumed by B”. The purpose of the operation

\[
<&^\text{ADJUNCTS}>&:-([\text{TYPE SENT}])
\]

is to inspect the members of the ADJUNCTS set; the ECS formalism allows inspection and other operations on individual, several or all of the elements in such sets. The distinguished metavariable !^ is used to access those elements of such sets that satisfy the subsumption relation (in our example, that are subsumed by the attribute-value pair \{ [ TYPE SENT ] \} in the first case and \{ [ TYPE FREQ ] \} in the second, thus identifying different adjectives modifying ‘failure’). If the subsumption relation is satisfied, RULE1 will fire and will carry out the mapping specified by MAP, resulting in the correct target structure.

b) “Headless” Adjunct Mapping
This type of adjunct mapping involves cases where the head of the construction combines with another word to map to a single structure or a multiple word expression of a different structure in the target language. Since the head of the construction has no counterpart in the translation, there is no syntactic environment for the modifier in the target. Consider the following example:

\(\text{He took terrible advantage of his relatives.}\)  
\([\text{ku-ka ku-uy chinche-k-tul-ul caninha-n iyongha-ess-ta}]\)
In this example, the meaning contributed to the English sentence by the adjective ‘terrible’ as the modifier of ‘advantage’ is conveyed in the Korean sentence by the adverb ‘caninhakey’ modifying the verb ‘iyongha’. The only translation in Korean of the discontinuous item ‘take advantage of NP’ is the verb ‘NP+ACC iyongha’, which lacks an appropriate compositionally defined candidate that can be modified by the adjective ‘terrible’ in the example sentence.

Once again, if we wish to define a correspondence between the disparate English and Korean structures, we need to determine where the correspondence will be stated and how it will be expressed. We note, first, that this mapping is not a property of the adjective ‘terrible’; its translation in examples such as the following is simple and compositional:
They saw a terrible man. (5e)
Kutul-i chaninha-n namca-lul po-ass-ta

The most general environment where this correspondence can be staged is the transfer specification for 'take advantage of'. In the ECS EK transfer grammar, translation of verbal idioms is specified under the head verb, in this case the verb 'take'. One of the translations of 'take' will be 'iyongha' iff the functional structure of the sentence is subsumed by the following

[VOICE ACTIVE
  OBJ [form "advantage"
       number sg]
 ]

or, in the event the sentence is something like 'terrible advantage was taken of NP',

[VOICE PASSIVE
  SUBJ [form "advantage"
       number sg]
 ]

This is specified under ek_take as follows:

ek_take ::
[T { [FORM "iyongha"
    {[^ :- [VOICE ACTIVE
      OBJ [FORM "advantage"
      number sg
    ]
     OBL-OF ANY
    ]
    MAP (E(<OBL-OF>), K(<OBJ>))
    <other mapping operations>
  ]
  [ ^ :- [VOICE PASSIVE
     SUBJ [FORM "advantage"
     number sg
  ]]
OBL-OF ANY
]
MAP (E(<^OBL-OF>), K(<^SUBJ>))
<other mapping operations>
]
]
<more translations>  ] ]

These mappings will generate the correct target argument structure. We still need, however, to accommodate any modifiers of 'advantage', which has not been mapped to (that is, has no counterpart in) the target structure. This is done by a rule mapping the OBJ's (or SUBJ's) ADJUNCTS to the sentence level:

ek _ take ::
[T  { [ FORM "iyongha"
    {[^ :- [VOICE ACTIVE
       OBJ [FORM "advantage"
            NUMBER SG
    ]
       OBL-OF ANY
    ]
    MAP (E(<^OBL-OF>), K(<^OBJ>))
    MAP (E(<^OBJ ADJUNCTS>),
         K(<^ADJUNCTS>))
    <other mapping operations>
    ]
[^ :-[VOICE PASSIVE
       SUBJ [FORM "advantage"
            NUMBER SG
    ]
    ]
    OBL-OF ANY
    ]
    MAP(E<^OBL-OF>, K(<^SUBJ>))
    MAP(E<^SUBJADJUNCTS>,
         K(<^ADJUNCTS>))}]}
<other mapping operations>
]
}
]
<more translations>
}
]

It may appear unnecessary at first to have two separate transfer specifications, one to handle the active form of the sentence and one to handle its passive counterpart; thus, it might be argued that a semantic representation using roles like AGENT and PATIENT would allow us to express the transfer correspondence in one rule. There are well-known collocations, however, whose translation is sensitive to the configuration of the sentence: thus, 'John finally kicked the bucket' may be ambiguous between the idiomatic and the literal interpretation, but 'the bucket was finally kicked by John' is not.

The rules shown above will map the adjunct to the sentence level, thus providing a target placement for the translation of 'terrible'. Nothing more needs to be specified in the transfer entry for 'terrible'; generation grammar rules will ensure the adverbial realization of Korean, by taking account of its functional projection in the verb's ADJUNCTS.

5. Discussion

The transfer operations described above are motivated by the need to generate an accurate, acceptable and natural translation for the sentences in the examples. Our approach is unabashedly contrastive yet principled: transfer rules state specific operations to resolve any differences in the source and target f-structure configurations, but they do not introduce information which should be, and normally is, described in the target language grammar. Rather, they ensure that the organization of the target functional structure will be correct with respect to the target language. As mentioned earlier, the output of transfer is a structure that will subsume the f-structure produced by the generation grammar. Given a source language f-structure, transfer rules generate a coherent
(but not necessarily complete) target f-structure in the most efficient and properly constrained way. For example, as mentioned earlier, the Korean translation of the intransitive use of the verb ‘eat’ requires a generic object, ‘pap’ (‘rice’); this will be introduced by the generation grammar provided the appropriate conditions are satisfied.

Finally, we would like to touch briefly on the issue of the bidirectionality, or reversibility, of transfer grammars. The formalism of the ECS transfer grammar is not bidirectional in its current implementation: separate transfer grammars had to be written for our English-Korean and Korean-English systems. Several bidirectional transfer grammars have been described in the literature (Estival et al. (1990)), but none have been implemented in full-scale, production-oriented systems. Although the idea is appealing, since it suggests a certain parsimony of representation and implementation, the constraints that would have to be introduced to prevent overgeneration (for example, to ensure that the Korean sentence ‘kuka kuuy chinchektulul caninhakey iyonghaessta’ ‘he used his relatives terribly’ or ‘he took terrible advantage of his relatives’ will generate only one English translation) will, in effect, introduce an additional parameter, thus making the two approaches not directly comparable.

6. Conclusion

We have discussed some aspects of a transfer-based approach to some non-compositional problems in English-Korean translation. Specifically, we addressed the problems posed by the non-compositional mapping of two classes of adjuncts, involving target structures which do not normally allow the source modifier to appear attached to the translation of its source head. These contrasts are due to arbitrary configurational differences, which we resolve during the transfer portion of our grammar.

7. Acknowledgment

I am indebted to my colleagues Dan Higginbotham, Jay Jeong-Ryeol Kim, One-Soon Her and Hajime Wada for their comments and support.
References


Executive Communication System
455 North University Avenue
Suite 202, Provo, Utah 84601
U. S. A.