CS-TR-3233 March, 1994
UMIACS-TR-94-26

Development of Cross-Linguistic Syntactic and Semantic Parameters for Parsing and Generation[1]

Bonnie J. Dorr
Institute for Advanced Computer Studies
and Department of Computer Science
A.V. Williams Building, College Park, MD 20742

Jye-hoon Lee
Department of Computer Science
A.V. Williams Building, College Park, MD 20742

Sunki Suh
Department of Linguistics
A.V. Williams Building, College Park, MD 20742

Keywords: Parsing, Generation, Parameters, Machine Translation,
Lexical Semantics, Korean

Abstract

This document reports on research conducted at the University of Maryland for the Korean/English Machine Translation (MT) project. The translation approach adopted here is interlingual i.e., a single underlying representation called Lexical Conceptual Structure (LCS) is used for both Korean and English.

The primary focus of this investigation concerns the notion of 'parameterization' i.e., a mechanism that accounts for both syntactic and lexical-semantic distinctions between Korean and English. We present our assumptions about the syntactic structure of Korean-type languages vs. English-type languages and describe our investigation of syntactic parameterization for distinguishing between these two types of languages. We also present the details of the LCS structure and describe how this representation is parameterized so that it accommodates both languages.

We address critical issues concerning interlingual machine translation such as locative postpositions and the dividing line between the interlingua and the knowledge representation. Difficulties in translation and transliteration of Korean are discussed and complex morphological properties of Korean are presented. Finally, we describe recent work on lexical acquisition and conclude with a discussion about two hypotheses concerning semantic classification that are currently being tested.

[1]This research has been partially supported by the Army Research Office under contract DAAL03-91-C-0034 through Batelle Corporation, by the National Science Foundation under grant IRI-9120788, and by the Army Research Institute under contract MDA-903-92-R-0035 through Microelectronics and Design, Inc.

Contents

1 Introduction

Over the last year, the computational linguistics group at Maryland has been involved in the problem of 'interlingual' machine translation (i.e., translation of the source language into a single language-independent representation from which the target language is then produced) of Korean and English. We have adopted a pivot form that was used in an earlier machine translation system called UNITRAN (Dorr (1992b, 1993a,b) which translates English, Spanish, and German bidirectionally. This form, called Lexical Conceptual Structure (LCS), was developed on the basis of work by Jackendoff (1983, 1990).[2] As part of this investigation, we have developed an algorithm for recursive composition and decomposition of the interlingual representation; this algorithm allows the LCS to be linked systematically to the syntactic structure both during parsing as well as during generation.

The primary focus of this investigation concerns the notion of 'parameterization' i.e., a mechanism that accounts for both syntactic and lexical-semantic distinctions between Korean and English.[3] Syntactic parameterization is based on a current linguistic theory called 'Government-Binding' (GB) developed by Chomsky (1981, 1986a,b) and his followers. Within this theoretical framework, syntactic distinctions between languages are accounted for by settings of parameters associated with the universal principles. For example, there is a universal principle that requires there to be a conceptual subject for each predicate of a sentence. Whether or not this conceptual subject is syntactically realized is determined by a parameter associated with this principle: the 'null subject' parameter. This parameter is set to 'yes' for Spanish but 'no' for English and German.

While most of the work over the last year has concentrated on parameterization of syntactic distinctions between English and Korean (e.g., the parametric distinction that forces verbs to precede their objects in English but follow their objects in Korean), we have also been investigating the problem of lexical-semantic parameterization between Korean and English. The lexical-semantic component has been designed to allow principles of the lexicon to be parameterized. While syntactic parameterization applies during parsing and generation, lexical-semantic parameterization applies during composition of the LCS. Figure 1 illustrates the overall design of interlingual MT system.

Within the lexical-semantic level, the language-independent and language-

[2] Others who have studied this representation are Hale and Laughren (1983) and Hale and Keyser (1986a,b). For alternative (lexical-)semantic and case representations, see, for example, Fillmore (1968), Gruber (1965), Schank (1972), and Wilks (1973).

[3] Although we have focused primarily on Korean and English for this project, we are also investigating the applicability of the parameterization technique to French, German, Spanish, and Arabic.

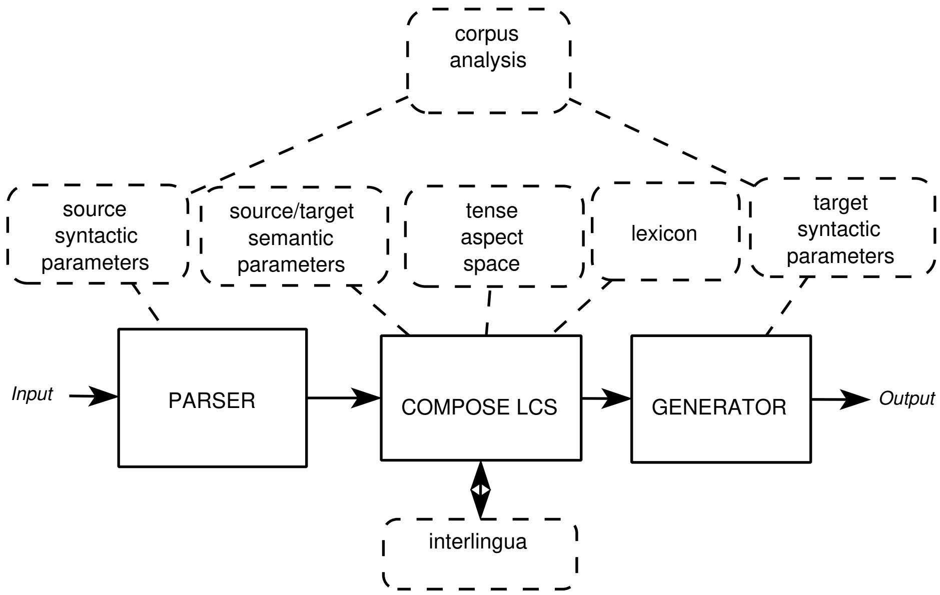

Figure 1: Parametric Approach to Interlingual MT

specific information are supplied by a set of general LCS mappings and the associated parameters for each language, respectively. The interface between the syntactic and semantic levels allows the source-language structure to be mapped systematically to the conceptual form, and it allows the target-language structure to be realized systematically from lexical items derived from the conceptual form. This work represents a shift away from complex, language-specific syntactic processing without entirely abandoning syntax. Furthermore, the work moves toward a model that employs a well-defined lexical conceptual representation without requiring a 'deep' semantic conceptualization.

The next section presents our assumptions about the syntactic structure (and, consequently, syntactic processing) of Korean-type languages vs. English-type languages. Throughout this and later sections, we will use the Yale romanized form as the transliteration for Korean examples. There are also other types of transliteration systems, but Yale romanization is the most commonly used among linguists.

The notion of parameterization is the focus of a large part of this report. The third section describes our investigation of syntactic parameterization for distinguishing between English and Korean. The details of the LCS representation are introduced in the fourth section and the parameterization of the LCS is presented in the fifth section.

The sixth section turns to the definition of the dividing line between the interlingua and a knowledge representation. Locative postpositions are discussed in the seventh section. General issues in translation and transliteration are presented in section eight. Morphological properties of Korean are presented in the ninth section. Finally, the tenth section presents recent work on lexical acquisition and concludes with a discussion about two hypotheses concerning semantic classification that are currently being tested.

2 Grammatical Framework: Korean vs. English

The assumptions underlying our MT research on syntactic processing have originated from Sungki Suh's doctoral dissertation (University of Maryland, forthcoming) which explores parsing phenomena in Korean. Many alternative approaches to sentence comprehension have been made based on the analysis of English data. The exact relation between a grammar and a parser is quite controversial. However, Suh's approach assumes that the human language faculty is innate and universal, and that the parser reflects (at least) some properties of grammar. Thus, a parsing theory based on an analysis of English is expected to be applicable to different type of languages such as Korean, if it is a psychologically plausible theory.

There is, however, an important question to be addressed in applying the previous parsing theories to Korean-type languages as opposed to English-type languages. Provided that parsing is done incrementally from left-to-right and the information from the head is essential in computing constituent structure, we need to make a decision as to how soon we may assign structure upon encountering a Korean word, given that the head occurs at the end of a syntactic phrase. This problem arises precisely because two methodologies that are often adopted in parsing theory — incremental structure building and structural commitment based on the information from the head — are in conflict with the processing of head-final languages. Consider, for example, the problem of computing clausal structure in Korean-type languages. It may seem impossible to compute clausal structure before the end of the clause since the verb occurs at the end of the clause.

It appears that, if the information from the head is the only source for computing constituent structure, structural commitment must be delayed in head-final languages to the extent that receiving the information from the head is delayed. Such a speculation, however, is not compatible with the data obtained from Korean. Suh's experimental results from a rating task strongly suggest that constituent structure is computed prior to the appearance of the head in parsing Korean. Therefore, we assume that some information other than that of the head can function as a major source guiding first-pass processing. This assumption is also motivated by the natural assumption that head-final and head-initial languages are parsed equally efficiently, observing incremental structure building and incremental interpretation.

The fact that, in Korean, constituent structure can be computed in the absence of the information from the head is not surprising in the respect that the grammatical role of argument and adjunct is usually signaled by its morphology such as case markers or adjective/adverb markers. The current project assumes that such morphological information guides the initial structural commitment and the information from the head is used in confirming or revising the initial structural commitment.

The model of grammar assumed for this project is Government-Binding theory (or the Principles-and-Parameters approach).[4] Below, we give a brief overview of the grammar, including phrase structure, levels of representation, each subtheory of the grammar, and an itemization of the syntactic parameter settings for English and Korean.

2.1 Phrase Structure Trees and Dominance/Precedence

We assume that the structural representation of a sentence computed by the parser is equivalent to a phrase structure tree. A phrase structure tree encodes the information about the hierarchical relation and linear order between constituents, as well as on the grammatical type of the constituents. The hierarchical relation between constituents is represented in terms of the predicate 'dominance', which is defined in (1).

(1) For two nodes x and y, x dominates y(= d(x,y)) iff the connection between x and y is composed exclusively of descending branches.

If x dominates y and there is no node between x and y, then x 'immediately' (or 'directly') dominates y.

The linear order between constituents is represented in terms of the predicate 'precedence'. Following the Exclusivity Condition by Wall (1972) in (2), we consider 'precedence' to be more than left-to-right order. That is, two nodes x and y are in a precedence relation (= p(x,y) or p(y,x)) if and only iff they are not in a dominance relation.

(2) In any well-formed tree, either p(x,y) or p(y,x) is true iff neither d(x,y) nor d(y,x) is true.

According to (2), Y in the following diagram precedes not only Z, but also A and B, which are NOT in a dominance relation to Y.

[4] See Chomsky (1981, 1986a) for details. By assuming such a grammatical framework, we do not mean that other grammatical theories are excluded. Rather, we think our approach is compatible with other theories which assume a phrase structure equivalent to the one outlined here.

```
(3)        X                d(X,Y)
         /   \              d(X,Z)   p(Y,Z)
       Y       Z            d(X,A)   p(Y,A)   d(Z,A)
             /   \          d(X,B)   p(Y,B)   d(Z,B)   p(A,B)
           A       B
```

One generalization we can draw from (1) and (2) is that if x precedes y, then all nodes dominated by x precede all nodes dominated by y.

As will be seen below, dominance and precedence are the primary structural relations. Other relations such as government or c-command are defined in terms of these primary relations.

2.2 Levels of Representation

Government-Binding theory assumes the following levels of representation:

```
                 D-structure
                      |
                      |  Move-alpha
                      |
                 S-structure
                    /   \
      Move-alpha   /     \   Move-alpha
                  /       \
       Phonetic Form    Logical Form
            (PF)            (LF)
```

D-structure is a level of representation at which the thematic relations between arguments and their predicates are directly represented. S-structure can be characterized as a level which should be properly related to all of the other structures simultaneously. PF is the interface between the grammar and articulatory and audio-visual systems. LF is the interface between the grammar and other cognitive capacities. The scope of quantifiers/operators is represented at this level.

Note that each of the levels is related by applications of the transformational rule 'Move-alpha', which states 'Move anything anywhere'. The application of Move-alpha and resulting representations are constrained by the principles or conditions of various modules of the grammar.

We assume that the representation computed by the parser amounts to S-structure representation. This assumption is necessary in the respect that the representation computed by the parser should be able to recover immediately both the sound and the meaning, i.e., PF and LF, of input string, and S-structure is the only level directly related to both PF and LF. Note that PF and LF are not directly related in the above diagram. Hence, if the parser computes either LF or PF, it may have difficulty recovering the other.

2.3 X-bar Theory

The gist of X-bar theory is that a phrasal constituent has a layered structure. Every phrasal constituent is considered to have a head, which determines the properties of the phrase containing it. It is generally represented as a zero-level category (X0). It can take a constituent as its complement, resulting in projecting one-bar level (X'). A one-bar level category can project a double-bar level (X") by taking a constituent as its specifier (or modifier). These processes are represented in the following diagram:

(4)
```
              XP(=X'')
            /    \
   Specifier     X'
              /     \
    Complement       X0
```

We consider the double-bar level to be the maximal projection of the category in question. While a projection of a category should have zero-level (= head) and its maximal projection, the intermediate level can be absent (or recursive): If X0 takes neither specifier nor complement, it projects XP without the X' level. Meanwhile, X' can recursively occur if adjuncts (= modifiers) are added to the structure in (4).[5]

The relative order between the head and its complement can vary, depending on whether the language in question is head-initial or head-final. The structure in (4) represents the relative order observed in head-final languages such as Korean.

In the recent syntactic literature, it has been assumed that functional categories, i.e., Comp(lementizer), Infl(ection), and Det(erminer), as well as lexical categories project their maximal projection as shown in (4). Thus, we represent the projection of the functional category 'Comp' or 'Infl' by using this notational system, i.e., we use IP (Infl phrase) and CP (Comp phrase).[6] As an illustration of this notation, consider the following two examples:[7]

[5] We assume that modifiers are distinguished from complements in the respect that the former is adjoined to one-bar level (or double-bar level) while the latter is always attached as a sister to the head. We will, henceforth, write X0 simply as X.

[6] The other functional category D(eterminer) and its projection are not discussed here. Thus, we omit the DP structure from our phrase structure tree.

[7] For the purposes of this report, we will be using the following abbreviations: Nom = nominative marker, Acc = accusative marker, Top = topic marker, Comp = complementizer, Rel = relative clause marker.

```
(5)  (i)   John-ul  Mary-ka  Bill-eykey  sokayhayssta
               -Acc     -Nom     -to     introduced
           'Mary introduced John to Bill'
     (ii)                CP
                       /    \
                     NP      C'
                    /      /   \
               John1    IP      C
                       /    \
                     NP      I'
                    /      /   \
                Mary    VP      I
                       /    \
                     PP       V'
                    /       /   \
                Bill     NP      V
                          |      |
                         t1   introduced
```

```
(6)  (i)   ku   hayk-un    John-i   Mary-eykey cwuessta
           that book-Top       -Nom     -to    gave
           'As for that book, John gave it to Mary'
     (ii)                CP
                       /    \
                     NP      C'
                    /      /   \
            that book1    IP      C
                       /    \
                     NP      I'
                    /      /   \
                John    VP      I
                       /    \
                     PP       V'
                    /       /   \
             to Mary     NP      V
                          |      |
                         t1    gave
```

Note that the relation between CP, IP and VP in (5)(ii) and (6)(ii) is characterized as follows: C takes IP as its complement and IP in turn takes VP as its complement.[8] This notation allows us to represent the hierarchical

[8] Some researchers have argued that Korean-type languages lack a CP (cf. Fukui (1986) and Kim (1989)). Those who assume a CP structure in Korean consider Spec of CP to

relation between the subject and the 'peripheral' phrases (i.e., scrambled elements such as 'John' in (5) or topicalized elements such as 'that book' in (6)).[9] The subject occurs in a position dominated by IP whereas 'peripheral' elements occur in the position directly dominated by CP node. The CP node can be created recursively when more than one 'peripheral' phrase occurs.

2.4 Theta Theory

Theta theory is concerned with how theta roles (thematic roles) are represented grammatically. Theta roles are typically assigned by a lexical head to its complement or specifier. The class of theta roles include Agent, Patient (Theme), Experiencer, Goal, Location, Instrument, etc. Theta role assignment is constrained by the following condition:

(7) Theta Criterion: Each argument bears one and only one theta-role, and each theta-role is assigned to one and only one argument. (Chomsky (1981)).

2.5 Case Theory

Case theory requires that every NP be assigned abstract case. This requirement is formulated as the Case Filter or the Visibility Condition below.

(8) Case Filter: *NP if NP has phonetic content and has no case. (Chomsky (1981))

(9) Visibility Condition: An element is visible for theta-marking only if it is assigned case. (Chomsky (1986a))

These conditions are satisfied by virtue of structural case assignment or through the presence of inherent case. The former includes nominative case and accusative case, which are assigned at S-structure to a syntactic position governed by a case assigner. Structural case-assigning categories include transitive verb and tensed Infl. In contrast, inherent case is assigned to an NP at D-structure and realized at S-structure. Preposition is a typical category assigning inherent case.

---

be the landing site for a topicalized item (cf. Yoon (1990) and Moon (1989)). Those who assume a parallelism between English-type languages and Korean, i.e., who consider the Spec of CP to be a landing site for a Wh phrase at LF, argue either for CP adjunction (or IP adjunction) of topicalized elements or for a separate Topic node above CP.

Also, there have been various proposals on the structure of IP in Korean. The central issues on the IP structure have been: (1) whether IP has more than one specifier position and (2) whether IP-adjoined position is A or A-bar position. See Suh (forthcoming) for relevant discussion.

[9] The category t1 is one of two types of empty categories as discussed below in section 2.7.

2.6 Government

The notion of 'government' is central to the grammatical framework: The relation of 'government' is required in many cases for proper licensing of grammatical structures. For instance, theta-marking and case-marking are possible only under a certain configuration of government. A trace and its antecedent should be in a proper government relation, which is a subcase of government.

The definition of government involves two parts, one pertaining to a structural requirement between the governor and governee, and the other pertaining to a minimality requirement, i.e., a constraint that prohibits a barrier between governor and governee. The former requirement is often characterized as c-command (or m-command), defined as follows:

(10) A c-commands B iff A does not dominate B and every Z that dominates A dominates B. (Chomsky (1986b))

A m-commands B when Z is restricted to maximal projections. For the notion of barrier to government, see Chomsky (1986b).

2.7 Types of Empty Categories and Constraint on Movement

Empty categories (EC) are generally classified as two types, pronominal and non-pronominal EC's. The former includes PRO and pro, and the latter includes anaphoric trace (=NP trace) and variable (=Wh trace etc.).

A trace represents the position from which some element has been extracted. A trace and its antecedent constitute a syntactic chain, which is required to have only one case and one theta role as a whole (cf. Chain condition in Chomsky (1986a)).

The landing site of the extracted element can be either an A(rgument)-position or an A-bar position. Movement to the former is called A-movement and movement to the latter is called A-bar movement. It has often been observed that the distinction between A and A-bar is not so clearcut in a certain syntactic configuration. A construction involving scrambling is such an instance. (See Suh (forthcoming) for additional discussion.)

Syntactic movement is constrained by the Subjacency condition, among others. This condition prohibits a single application of Move-alpha from crossing more than one bounding node. NP and CP nodes are generally considered to be bounding nodes in Korean.

In our syntactic framework, a non-pronominal EC (= trace) will be represented by 't' and a pronominal EC (=PRO or pro) will be represented by 'e'. The former category will occur mostly in relative clauses and constructions involving scrambling. Other types of movement involve 'e', not 't'. Note that, since Korean wh-phrases stay in situ at S-structure, 't' does not represent wh-trace.

2.8 Binding Theory

The relations between anaphoric/pronominal elements and their antecedents are constrained by the Binding theory. The core idea in is that anaphors must have an antecedent within a certain domain, whereas pronominals must not have one. Consider the following conditions:

(11) Binding conditions (Chomsky (1981)):

A. An anaphor is bound in its governing category.

B. A pronominal is free in its governing category.

C. An R-expression is free.

Binding between two constituents, x and y, occurs if and only if (i) x c-commands y and (ii) x and y are coindexed. An element is free if it is not bound. Note that both overt and empty categories are subject to Binding conditions.

The notion of 'Governing Category', i.e, the domain in which Binding Conditions should be satisfied, roughly amounts to the minimal NP or S containing the binder and bindee. 'Governing category' may vary across languages or it may depend on the properties of lexical items within one language (cf. Wexler and Manzini (1987)).

3 Parameterization of English and Korean: Alternative to the Context Free Grammar Approach

An approach to syntactic processing of Korean has been proposed by researchers at Systems Engineering Research Institute at Korea Institute of Science and Technology (Park (1992)). The system built by this research team is called MATES/EK. The approach assumes a BNF grammar for Korean which takes the following form:

(12)
```
S ::= (BCL) (NPP | ADP)* VG
NPP ::= NP PP | NCL PP
NP ::= (ART) ((AP | ACL)*) N (PE)
ADP ::= ADV | VV BE | N BP
AP ::= DET | A AE | NP POP
VG ::= (ADV) VV (AUX)* (TE) (EE)
```

```
NCL ::= S NE
BCL ::= S BE
ACL ::= S AE
CORS ::= S (CONJ | CONE) S
CORNP ::= NPP CONE NPP
CORADP ::= ADP (CONJ | CONE) ADP
CORAP ::= AP CONE ADP
```

This grammar contains most of the phrase structure rules necessary for the analysis of Korean. However, the BNF grammar has both theoretical and descriptive shortcomings:

- NPs accompanied by the case marker (i.e., nominative or accusative marker) as well as by the postposition are considered to be a postpositional phrase (=NPP). This consideration cannot be maintained from the syntactic viewpoint since there are several cases in which the 'NP+Case marker' results in the maximal projection NP (or DP) whereas 'NP+postposition' results in a maximal projection PP.

- In cases where the matrix verb is followed by an auxiliary verb in Korean, the verb potentially takes various types of endings; the form of verbal ending is mainly determined by which auxiliary verb is employed. This cannot be deduced from the BNF given above; whether it would be possible to accommodate such an idiosyncrasy in a phrase structure grammar is an open question.

- In addition to case markers and postpositions, there is another category that attaches to NPs called a 'delimiter' in the literature. For example, the second element of 'John-man-i' (John-only-Nom marker), functions as a delimiter. The BNF grammar does accommodate such constituents. In fact, it may not be possible to differentiate systematically between delimiters and case markers/postpositions in a phrase structure grammar.

In addition to these points, processing of Korean requires more than a CFG-style grammar in order to capture the full extent of parameterization. In particular, it would be impossible to derive the full range of syntactic parameters of GB theory from the BNF above. In fact, the only syntactic parameters we could possibly derive from this grammar would be those pertaining to X-bar theory, e.g., the constituent order parameter.

There are a number of ways that the syntactic properties of Korean are different from those of English. The BNF approach is significantly inadequate for capturing such differences. Our approach is to adopt a framework that uses syntactic parameterization of general principles rather than employing a complex specification of construction-specific rules.

3.1 Parameterization of the Binding Condition

As observed by many linguists, Korean-type languages are different from English-type languages in that Binding Condition A and C (see (11) above) are not applicable to Korean in the same way as to English. In particular, anaphors can be bound from outside of the minimal clause. In English, this would be a violation of Condition A. Furthermore, a name can be bound by another name in Korean; this would be a violation of Condition C in the case of English.

As an illustration of this point, consider the Korean reflexive 'caki'/'casin' (self) which can be bound from both inside and outside of its minimal clause. The result is that sentences such as the following are ambiguous:

```
(13) John1-i [Ray2-ka  caki1/2-lul  piphanhayssta-ko]
        -Nom              -Nom           self-Acc
       malhayssta
       criticized-Comp   said
     'John1 said that Ray criticized John1' or
     'John said that Ray2 criticized Ray2 (himself)'
```

The English counterpart of (13), however, is not ambiguous: The reflexive '(him)self' must be construed with 'Ray', not 'John'. Thus, only the second reading is possible in English. Provided that the Binding Condition in (11) is effective across languages, we may account for the syntactic difference between English anaphors and Korean anaphors by parameterizing the notion of 'governing category':

(14) Governing Category: y is a governing category for x if and only if y is the minimal category which contains x and y (i) has a SUBJECT or (ii) has a root TENSE.

We assume that the setting for English would be (i) and the setting for Korean would be (ii). According to the above definition, sentence (13) is expected to be ambiguous since 'caki' (self) has two antecedents in its governing category, i.e., in the root clause.

3.2 Parameterization of Case Theory

In Korean, so-called multiple subject constructions are available. Such constructions should be ruled out under the assumption that the relation between a case assigner and a case assignee is biunique. In particular, we should not allow Infl(ection) to assign nominative case to more than one subject NP.

Consider the following example:

```
(15) John-i     phal-i   pwureciessta
        -Nom   arm-Nom  was broken
     'John is in the situation that his arm has been broken'
```

The Case Filter states that overt NPs must have case. The existence of two nominative NPs in this example would be a problem if we use the version of Case theory that applies to English-type languages since Infl(ection) assigns nominative case in English and the relation between Case assigner and assignee is supposed to be biunique (i.e., at tensed verb can assign nominative Case to only one NP).

The grammaticality of the above example suggests that the relation between the Case assigner and assignee is not biunique in Korean or that nominative Case in Korean can be assigned by something other than tensed Infl, presumably via the predication relation. Thus, Case theory must be parameterized to allow for this distinction, perhaps as follows:

(16) Case Assignment: Structural case assignment allows each assigner to discharge case (i) to at most one NP or (ii) to more than one NP in multiple-subject constructions.

In a biunique case-assignment language such as English, the setting would be (i); in a Korean-type language, the setting would be (ii).

3.3 Parameterization of Bounding Theory

In Korean, the head noun of a relative clause may be construed with the empty category across more than one intervening CP node, as shown in the following:

```
(17) [cp[cp e1 e2  kyengyengha-ten]  hoysa2-ka  manghayperi-n]
                   managed-Rel      company-Nom is bankrupt-Rel
       Bill1-un  yocum       uykisochimhay issta
            -Top these days  is depressed
     'Bill is such a person that the company which was managed
     by him has been bankrupt, and he is depressed these days'
```

The subject NP 'Bill' is coindexed with the empty category in the more deeply embedded relative clause. (Note that there are two relative clauses and that the first clause is embedded under the second one.) If we assume, following Chomsky (1986a), that relative clause formation involves empty operator movement, then the grammaticality of the above example suggests that the Bounding theory must be parameterized so that crossing more than one bounding node is allowed in Korean (at least for some particular structures).

3.4 Summary of Parameter Settings for English and Korean

We have just seen that syntactic parameterization cannot be captured by a context-free grammar like BNF grammar. In order to account for the phenomena described above, we need to parameterize the submodules of GB theory.

According to Dorr (1993a), the syntactic parameters are set as shown in figure 2 for English. Following this paradigm, our analysis of Korean has revealed the parameter settings shown in figure 3.

During the last two months of 1993, the MT project has focused heavily on the design of a parameterized syntactic processing component that takes these parameter settings into account. Our implementation is based on a GB-based parsing system that uses the message-passing framework proposed by Lin (1993) and Lin and Goebel (1993).

The main innovation of the message-passing paradigm is that it allows linguistic principles to be applied to 'descriptions' of structures instead of to the structures themselves. This leads to a very efficient approach to structure compilation: the final phrase structure representation is built only after all principles are satisfied.

The grammar for each language is stored in what is called a 'grammar network', which has language-specific information compiled into it (e.g., constituent order). The algorithm for the message-passing approach to parsing is the following:

- Look up all words in lexicon and create initial item for each lexical item.
- For each initial lexical item:
  - Send messages induced by this item, creating more items.
  - Call the message-passing procedure recursively on each of these items.
- When all the lexical items are processed, search the CP node for items whose surface string spans on the whole sentence and trace its origins.
- Retrieve structure (by following traces of operations) and return a packed, shared parse forest, from which each parse tree can be retrieved one by one.

The main message-passing operation relies on combining two items in the grammar network. Because of this, the algorithm need not be head-driven. Thus, the model is consistent with Suh's experimental results which suggest that constituent structure is computed prior to the appearance of

Parameter	**English Setting**
Basic Categories	C I V N P A
Pre-terminals	ADV NUM DET
Constituent Order	I: SPEC-INITIAL HEAD-INITIAL N: SPEC-INITIAL HEAD-INITIAL C: SPEC-INITIAL HEAD-INITIAL A: HEAD-INITIAL P: HEAD-INITIAL V: HEAD-INITIAL
Complements	V: NP, PP, CP, AP, ADV, VP P: NP N: PP A: CP I: VP C: IP
Specifiers	I: NP N: DET, NP
Adjunction	IMAX: PP (left), ADV (right) VBAR: PP (right), ADV (left), NP (right) NBAR: NUM (left), A (left), CP (right), PP (right) V: CLITIC (right)
Proper Governors	V, P, AGR
Bounding Node	IP, NP

Figure 2: Syntactic Parameter Settings for English

Parameter	**Korean Setting**
Basic Categories	C I V N P A
Pre-terminals	ADV NUM
Constituent Order	I: SPEC-INITIAL HEAD-FINAL N: SPEC-INITIAL HEAD-FINAL C: HEAD-FINAL A: HEAD-FINAL P: HEAD-FINAL V: HEAD-FINAL
Complements	V: NP, PP, CP, IP P: NP N: NP (+gen), PP (+gen) A: CP I: VP, AP C: IP
Specifiers	I: NP N: NP (+gen), Demonstrative
Adjunction	IBAR: PP (left), ADV (left), NP (left) VBAR: PP (left), ADV (left), NP (left) NBAR: A (left), NP(+gen) (left) A: ADV (left)
Proper Governors	V, P, I
Bounding Node	CP, NP

Figure 3: Syntactic Parameter Settings for Korean

Verb	PRO to do	NP	ECM-ed NP to do	that clause	for NP to do
want	E: yes	E: yes	E: yes	E: no	E: no
	K: yes	K: yes	K: ??	K: ??	K: N/A
think	E: no	E: no	E: no	E: yes	E: no
	K: yes	K: yes	K: yes(?)	K: yes	K: N/A
expect	E: yes	E: yes	E: yes	E: yes	E: no
	K: yes	K: yes	K: ?	K: ?	K: N/A
consider	E: ?	E: yes	E: yes	E: yes	E: no
	K: no	K: no	K: yes	K: yes	K: N/A
prefer	E: yes	E: yes	E: yes	E: no	E: yes
	K: yes	K: yes	K: ??	K: ??	K: N/A
believe	E: no	E: yes	E: yes	E: yes	E: no
	K: ??	K: yes	K: yes(?)	K: yes	K: N/A

Figure 4: Syntactic Distribution of English and Korean Verbs

the head in parsing Korean. The result is that Korean is as efficiently parsed as English.

Over the last two months, we have worked with the research team headed by Dr. Lin at University of Manitoba to modify the message-passing parser called PRINCIPAR so that it handles head-final languages. Our current goal is to design a method for compiling out the X-bar parameter settings into the grammar network automatically so that we are not forced to build a new grammar network for each language being processed. Preliminary results will be reported in Dorr et al. (1994).

We have now begun to build vocabulary for the Korean and English verbs that we plan to test out initially. The PRINCIPAR system requires that the distribution of each verb be spelled out in terms of syntactic features in the dictionary. Figure 4 shows the entries for certain types of English verbs and their Korean counterparts.

4 Lexical Conceptual Structure

In addition to our work on representing Korean syntactic structures, we have also developed a framework for representing the lexical semantics of Korean. In order to adopt an interlingual approach to MT, one must construct a language-independent representation that lends itself readily to the specification of a systematic mapping that operates uniformly across all languages. Recent work (Dorr (1993a)) has adopted the LCS representation as the basis of an interlingual approach to MT. We will now briefly describe the units of meaning underlying this representation.

The field of MT has (almost from the beginning) been concerned with the use of a 'deep semantic representation' and with looking for 'universals' for translation. One of the biggest objections to the use of an interlingual representation is that it relies on defining a set of primitives (to represent the information to be translated) which allow a mapping to be defined among the languages in question. Because it is generally difficult to define such a set, many researchers have abandoned this model. (See, for example, Vauquois and Boitet (1985).) However, recently, there has been a resurgence of interest in the area of lexical representation and organization (with special reference to verbs) that has initiated an ongoing effort to delimit the classes of lexical knowledge required to process natural language. (See, e.g., Grimshaw (1990), Hale and Keyser (1986a,b, 1989), Hale and Laughren (1983), Jackendoff (1983, 1990), Levin and Rappaport (1986), Levin (1985, 1993), Pustejovsky (1988, 1989, 1991), Rappaport et al. (1987), Rappaport and Levin (1988), Olsen (1991), and Zubizarreta (1982, 1987).)

As a result of this effort, it has become increasingly more feasible to isolate the components of meaning common to verbs participating in particular classes. These components of meaning can then be used to determine the lexical representation of verbs across languages. Consequently, the representation adopted for this project (which is by no means exhaustive) is based on an adapted version of the LCS proposed by Jackendoff that takes into account recent theories of the lexicon.

The LCS approach views semantic representation as a subset of conceptual structure, i.e., the language of mental representation. Jackendoff's approach includes 'Types' such as Event and State, which are specialized into 'Primitives' such as GO, STAY, BE, GO-EXT, and ORIENT. As an example of how the primitive GO is used to represent sentence semantics, consider the following sentence:

(18) (i) `The ball rolled toward Beth`

(ii)
```
[Event GO
  ([Thing BALL],
   [Path TOWARD
     ([Position AT ([Thing BALL], [Thing BETH])])])]
```

This representation illustrates one dimension (i.e., the 'Spatial' dimension) of Jackendoff's representation. Another dimension is the 'Causal' dimension, which includes the primitives CAUSE and LET. These primitives take a Thing and an Event as arguments. Thus, we could embed the structure shown in (18)(ii) within a causative construction:

(19) (i) `John rolled the ball toward Beth`

(ii)
```
[Event CAUSE
  ([Thing JOHN],
   [Event GO
     ([Thing BALL],
      [Path TOWARD
         ([Position AT
             ([Thing BALL], [Thing BETH])])])])]
```

Jackendoff includes a third dimension by introducing the notion of 'Field'. This dimension extends the semantic coverage of spatially oriented primitives to other domains such as Possessional, Temporal, Identificational, Circumstantial, and Existential. For example, the primitive `GO_Poss` refers to a GO event in the Possessional field as in the following sentence:

(20) (i) `Beth received the doll`

(ii)
```
[Event GO_Poss
  ([Thing DOLL],
   [Path TO_Poss ([Position AT_Poss
                    ([Thing DOLL], [Thing BETH])])])]
```

To further illustrate the notion of field, the GO primitive can be used in the Temporal and Identificational fields:

(21) (i) `The meeting went from 2:00 to 4:00`

(ii)
```
[Event GO_Temp
   ([Thing MEETING],
    [Path FROM_Temp
       ([Position AT_Temp
           ([Thing MEETING], [Time 2:00])])]
    [Path TO_Temp
       ([Position AT_Temp
           ([Thing MEETING], [Time 4:00])])])]
```

(22) (i) `The frog turned into a prince`

(ii)
```
[Event GO_Ident
   ([Thing FROG],
    [Path TO_Ident
        ([Position AT_Ident
           ([Thing FROG], [Thing PRINCE])])])]
```

As these examples illustrate, there are also other primitives that are included in the LCS framework. In particular, the Position and Path types are used to include primitives such as AT and TO. Furthermore, the Thing, Location, Time, Manner, and Property types are used. Some examples of LCS types and primitives are given in figure 5. For more details, see Dorr (1993a).

Type	Primitives
Event	CAUSE, LET, GO, STAY
State	BE, GO-EXT, ORIENT
Position	AT, IN, ON
Path	TO, FROM, TOWARD, AWAY-FROM, VIA
Thing	BOOK, PERSON, KNIFE-WOUND, KNIFE, SHARP-OBJECT, WOUND, FOOT, CURRENCY, PAINT, FLUID, ROOM, SURFACE, WALL, HOUSE, BALL, DOLL, MEETING, FROG
Property	TIRED, HUNGRY, PLEASED, BROKEN, ASLEEP, DEAD, STRETCHED, HAPPY, RED, HOT, FAR, BIG, EASY, CERTAIN
Location	HERE, THERE, LEFT, RIGHT, UP, DOWN
Time	TODAY, SATURDAY, 2:00, 4:00
Manner	FORCEFULLY, LIKINGLY, WELL, QUICKLY, DANCINGLY, SEEMINGLY, HAPPILY, LOVINGLY, PLEASINGLY, GIFTINGLY, UPWARD, DOWNWARD, WITHIN, HABITUALLY

Figure 5: LCS Primitives and Types

5 Lexical Semantic Parameterization

Just as we have provided a foundation for parameterization of the syntactic processing modules, we have also addressed the parameterization of the lexical-semantic representation in lexical entries. The reason that parameterization is necessary for mapping between lexical items is that processing is that the existence of translation divergences Dorr (1990) makes the straightforward lexical transfer impractical. The English, Spanish, and German examples given in (23)–(29) illustrate the nature of divergence classes in MT. (Literal translations are included for the Spanish and German cases.)

(23) **Conflational Divergence:**
E: I stabbed John
S: Yo le di puñaladas a John
'I (to) him gave knife-wounds to John'
G: Ich erstach John
'I (fatally) stabbed John'

(24) **Structural Divergence:**
E: John entered the house
S: John entró en la casa
'John entered in the house'
G: John trat ins Haus hinein
'John stepped in the house to'

(25) **Thematic Divergence:**
E: I like Mary
S: Mary me gusta a mí
'Mary (to) me pleases to me'
G: Ich habe Mary gern
'I have Mary likingly'

(26) **Categorial Divergence:**
E: I am hungry
S: Yo tengo hambre
'I have hunger'
G: Ich habe Hunger
'I have hunger'

(27) **Promotional Divergence:**
E: John usually goes home
S: John suele ir a casa
'John tends to go to home'
G: John geht gewöhnlich nach Hause
'John goes usually toward home'

(28) **Demotional Divergence:**
E: I like eating
S: Me gusta comer
'(To) me pleases eating'
G: Ich esse gern
'I eat likingly'

(29) **Lexical Divergence:**
E: John broke into the room
S: John forzó la entrada al cuarto
'John forced the entry to the room'
G: John brach ins Zimmer ein
'John broke in the house to'

The divergence 'classes' shown here isolate different types of distinctions across languages. Many sentences may fit into these divergence classes. Furthermore, a single sentence may exhibit any or all of these divergences. These are the systematic types of translation problems that motivate the research for this MT project.

Consider the conflational divergence between Spanish and English:

(30) **Conflational Divergence:**

(i) I stabbed John

(ii) Yo le di puñaladas a John
'I gave knife-wounds to John'

This example illustrates 'conflational' divergence, i.e., a case where the source-language main verb, 'stab', is mapped to more than one target-language word, 'dar puñaladas a'. The approach adopted in the current project handles this case by allowing the 'knife-wound' argument to be implicit (part of the lexical entry) in the source language, but explicit (syntactically represented as a noun phrase) in the target language.

Our research has uncovered a number of similar divergence types in Korean. The examples below illustrate the divergences between Korean and the three languages above. The language-independent LCS representation is included in each case.

(31) **Structural Divergence:**

```
K: John-i     pang-ey/ -ulo   tulekassta
       -Nom   room-Locative     entered
   'John entered the room'
S: John entro' en el cuarto
```

```
   'John entered in the room'
LCS: [Event GO_Loc
      ([Thing John],
        [Path TO_Loc
          ([Place IN_Loc ([Thing John], [Location Room])])])]
```

(32) **Structural Divergence:**

```
K: John-i   Sally-wa    kyelhonhayssta
       -Nom      -with  married
   'John married Sally'
E: John married Sally
LCS: [Event GO_Ident
       ([Thing John],
        [Path TO_Ident
          ([Place CO_Ident ([Thing John], [Thing Sally])])],
        [Manner MARRYINGLY])]
```

(33) **Conflational Divergence:**

```
K: John-i    Bill-eykey    towum-ul   cwuessta
       -Nom      -Dative    help-Acc  gave
   'John helped Bill'
S: John le ayudo' a Bill
LCS: [Event CAUSE
       ([Thing John],
        [GO_Poss
          ([Thing HELP],
           [Path TO_Poss
              ([Position AT_Poss
                ([Thing HELP], [Thing Bill])])])])]
```

(34) **Structural Divergence:**

```
K: John-i    Tom-ul      khal-lo          ccilessta
       -Nom     -Acc    knife-Instrument  poked
   'John stabbed Tom with a knife'
S: John le dio pun~aladas a Tom con un cuchillo
LCS: [Event CAUSE
       ([Thing John],
        [GO_Poss
          ([Thing KNIFE-WOUNDS],
           [Path TO_Poss
```

```
              ([Position AT_Poss
                 ([Thing KNIFE-WOUNDS], [Thing Tom])])])],
          [WITH_Instr ([Event *HEAD*], [Thing KNIFE])])]
```

(35) **Categorial Divergence:**

```
K: John-un      khi-ka    khuta
       -Top  height-Nom   is big/huge
   'John is tall'
E: John is tall
LCS: [Event BE_Ident
       ([Thing John], [Place AT_Ident ([Thing John], [Property Tall])])]
```

(36) **Categorial Divergence:**[10]

```
K: na-nun       pay-ka   kophuta
     I-Top  stomach-Nom   is hungry(?)
   'I am hungry'
G: Ich habe Hunger
LCS: [Event BE_Ident
       ([Thing I], [Place AT_Ident ([Thing I], [Property Hungry])])]
```

6 Dividing Line Between Interlingua and Knowledge Representation

Beyond the lexical-conceptual representation that is used for the interlingua for our MT system, we have studied the issue of how 'deep' the knowledge should be for the capturing cross-linguistic distinctions under investigation. MT theory has done little to elucidate the issues surrounding the relation between the interlingua (IL) and the knowledge representation (KR), either in terms of primitives, structures, or overall MT system computational issues, such as efficiency. In the development of grammatical theory, for example, the 'points of contact' between the syntax and the real world knowledge (e.g., the requirement that the verb 'sleep' has an animate subject explains the anomaly of 'The ideas are sleeping') have been addressed in natural language processing (NLP) systems (e.g., Winograd (1973), and others in Grosz et al. (1987)). However, with respect to a theory of the IL, these issues are more complex because no consensus exists yet on the criteria for evaluating ILs.

[10] The predicate 'kophuta' is used only in circumstances like (36) where it is preceded by the noun 'pay' (= stomach). In this respect, the meaning 'be hungry' in (36) is produced compositionally from 'pay' and 'kophuta'.

Info. Type	IL	KR
Aspect	Process/State/Event	Duration/Iteration/Linearity; Coercion Functions
Primitives and Structure	LCS Primitives; Well-formedness Constraints	Ontological Categories; Hierarchical Structures; Inheritance

Figure 6: Dividing Line between IL and KR modules of an Interlingual MT System

Clare Voss's dissertation work (Voss (forthcoming)) addresses the critical issue of whether the IL should exist as a part of the KR component, or whether it should be viewed as an integral part of the lexical-semantic component. Addressing this issue is an important part of her study of semantic features with respect to spatial relations. This work attempts to identify the problem areas in the spatial domain for English, French, Spanish, and German; we are currently examining spatial phenomena in the context of parsing and generating Korean and English as well. This investigation has led to the definition of the dividing line between the IL and the KR shown in figure 6 (Dorr and Voss (1993)).

The view adopted here is one in which the IL and the KR are considered to be separate components of an MT system. We argue that the 'languages' of the IL and KR system share many of the same predicates, but are not identical. Instead, the IL predicates are a proper subset of those in the KR system because we wish to allow, in principle, for KR concepts that are not needed for language-to-language translations. This avoids the problem of trying to represent a 'full' meaning for each word in a sentence being translated.

Note that Aspect crosses the boundary between the IL and the KR depending on the nature of the concept that is being modeled. For example, the verb 'sit' is a state, and is modeled as such in the IL used at Maryland; however, the sentence 'he sat under the table everyday' provides a repetitive (habitual) reading that would be modeled in the KR component.

This view of an interlingua contrasts with an approach such as that of Nirenburg et al. (1992) (at CMU) which focuses on the development of knowledge intensive morphological, syntactic, and semantic information for the lexicon. Nirenburg's group has developed tools for creating ontologies and has provided a framework for automatic lexical acquisition. While this paradigm falls into the interlingua (IL) category, it differs from our approach in that it is more knowledge intensive. The use of an interlingua text representations (ILT) is characteristic of Nirenburg's approach to Knowledge-Based Machine Translation.

The approach at Maryland complements this framework in that it also aims to provide an interlingual system and focuses heavily on the development of a lexicon. However, the interlingual representation (i.e., the LCS) occurs at a level higher than that of the knowledge-based structures proposed by the CMU research team. That is, the LCSs provide an interface between the knowledge required for deep semantic meaning and inference and the actual structures that appear on the surface in the source and target languages.

Where these two approaches come together is at the linking rules, i.e., the mapping between the semantic representation and the syntactic structure. The framework of Levin (1993) is relevant to the mapping assumed by both research teams. Levin has developed a very comprehensive and systematic classification of the types of verb classes that occur across languages; her classification constitutes a universal language in which we define our linking routines. (See additional discussion in section 10.) The Maryland research team has defined the LCS representation in terms of this taxonomy and has proposed a systematic mapping between this representation and the surface syntactic structure based on an isomorphism that exists between the two. CMU uses this taxonomy as well, but in a more direct way: the subcategorization frames for verbs in the taxonomy are hand-coded directly into the lexical entries that are used to realize knowledge representation (KR) concepts in the surface structure. Because CMU has focused more heavily on the development of the underlying KR concepts, little attention has been devoted to providing a systematic 'linking' of these concepts to the syntax. The Maryland team, on the other hand, has focused heavily on the linking routines and the systematic relation between the syntax and the lexical-semantic representation.

This distinction is further clarified by the following example:[11]

(37) Drop by your old favorite Dunkin' Donuts shop.

For this example, CMU's conceptual representation (ILT) involves a set of frames that look like the following:

(38) (i)
```
(make-frame text_1
    (clauses (value clause_1))
    (relations (value relation_1))
    (attitudes (value attitude_1))
    (producer-intentions (value producer_intention_1)))
```

[11] This is an example from a domain that CMU worked on in the late 80's. Some of the details concerning the knowledge encoded in this representation are not explained here.

(ii)
```
(make-frame clause_1
   (head (value %visit_1))
   (aspect (duration prolonged)
           (phase begin)
           (iteration 1))
   (time (value time_2)))
```

(iii)
```
(make-frame %visit_1
   (is-token-of (value *visit))
   (agent (value *consumer*))
   (destination (value %shop_1)))
```

(iv)
```
(make-frame relation_1
   (type (value domain-conditional))
   (first (value %visit_1))
   (second (value %involuntary-perceptual-event_1)))
```

(v)
```
(make-frame attitude_1
   (type (value &evaluative))
   (attitude-value (value .9))
   (scope (value %shop_1))
   (attributed-to (value *consumer*))
   (time (value (since time_5))))
```

(vi)
```
(make-frame producer-intention_1
   (is-token-of (value *commissive-act))
   (scope (value relation_1)))
```

The CMU approach requires a syntactic frame (in the lexicon) for each conceptual event that occurs in the ILT. The frame that provides the meaning for the 'drop by' action (which is really a form of 'visit') looks like the following:

(39)
```
(drop
     (make-frame
      +drop-v1
      (cat (value v))
      (syn-struc
       ((root $var0)
        (subj ((root $var1) (cat n)))
        (obliques ((root $var2) (prep by)))))
      (sem
       (lex-map
        (%visit
         (agent (value ^$var1))
         (theme (value ^$var2)
```

```
(sem *location)
(relaxable-to *object)))))))
```

Note that the syntactic realization information appears in the hand-coded syn-struc slot. The linking rule that relates this information to the ILT is implicit in the specification that appears in the sem slot. In particular, this information specifies that the agent (var1) maps into syntactic subject position and the theme (var2) maps into a syntactic oblique position.

This encoding of the linking rule identifies the precise syntactic realization for the underlying concept (i.e., the concept is realized as the verb 'drop' with a subject and an oblique argument). The reason such specific syntactic information is required is that the conceptual representation diverges significantly from the surface realization. In other words, there is no single systematic mapping between a concept and its surface realization.

At Maryland, the mapping between the IL representation and its surface realization is systematic and uniform across different concepts, as well as across different languages. In the above example, the LCS corresponding to (39) would be:

(40)
```
[Event GO_Loc
  ([Thing X],
   [Path BY_Loc
     ([Place AT_Loc ([Thing X], [Location Y])])],
   [Manner VISITINGLY])]
```

The linking routine that maps this concept to the syntactic structure is implicit in the well-formedness conditions assigned to the LCS. For example, the logical subject (i.e., the highest, left-most argument in the LCS, which is [Thing X] in the above structure) maps into the syntactic subject position. This linking is applicable to all conceptual structures in all languages; thus, no specific syntactic frames (i.e., slots analogous to syn-struc and sem in (39) above) are needed for individual concepts.

The knowledge-based techniques proposed by the CMU team are clearly necessary for the general solution to the problem of machine translation. While these techniques are not necessary for the solution to a class of problems studied by the Maryland team (namely, translation divergences of a specific type), the use of a lexical-semantic representation does not preclude the possibility of superimposing deeper knowledge onto the LCS framework. On the contrary, a knowledge-based meaning representation such as that of Nirenburg and Levin (1989) and Nirenburg et al. (1992) could significantly enhance the translation mapping, particularly during the processes of lexical selection and generation.

Both Maryland and CMU are forced to address many of the same issues since both are approaching the MT problem within the IL framework. In particular, the debate concerning 'primitives' and 'depth of decomposition' of the IL is discussed frequently by both groups of researchers. This debate still a hot one, as evidenced by the recent MT Journal special issue (7:4, 1993). In that issue, Hutchins (1993) reminds us that 'The aim should be to isolate what is apparently intrinsic to computer-based translation.' and 'What is required in MT is a set of principles of universal validity.' Both research teams have attempted to do this, although at different levels: the Maryland team has focused on defining a universal (reversible) syntax-to-IL mapping, while leaving aside issues of KR; the CMU team has focused on defining the universal units of meaning in the KR, while leaving aside issues of syntax-to-IL linking. In the former, the notion of parameterization is a crucial component of the model since the translation mapping must remain constant across the languages under consideration. In the latter, no such notion is needed since the KR component is truly intended to be a language-independent entity; that is, all parametric distinctions are hand-coded in the frames (as shown in (39) above). While questions concerning the status of the IL still remain, both research teams are on what Hutchins (1993) would consider to be the right track: both are taking on a contrastive analysis of languages to isolate cross-linguistic commonalities prior to defining the components of their respective IL representations.

7 Locative Postpositions

Clare Voss's initial investigation of spatial relations has led to the study of locative postpositions in Korean. Over the last year, Sungki Suh has analyzed the distribution of locative postpositions in Korean. Surprisingly, we have discovered this to be a meeting ground for constraints from various linguistic levels, including syntax, lexical semantics, and aspect. This section addresses a number of issues relevant to locative postpositions.

7.1 Distinction Between Dative and Locative Markers

During our analysis of locative postpositions, we have investigated the distinction between dative and locative markers. Ihm, Hong and Chang (1988) classify 'ey' as a dative marker. This classification is not desirable and, in some sense, incorrect. There are many cases where 'ey' is not used as a dative marker. Consider the following:

```
(41) John-i  uyca-ey       ancassta
         -Nom chair-ey(Loc)   sat
     'John sat on the chair'

(42) John-i  hakkyo-ey      kassta
```

```
        -Nom school-ey(Loc)  went
    'John went to school'
```

```
(43) Bill-i  namwu-ey     olakassta
         -Nom  tree-ey(Loc)  climbed
     'Bill climbed a tree'
```

All of the 'ey'-marked phrases in the above sentences cannot be considered dative. Rather, the 'ey' morpheme is a locative marker. One might claim that there are two kinds of 'ey', one that is dative and the other that is locative, and the above sentences are instances of the latter usage. Such a claim, however, does not seem to be convincing. If we adhere to the standard notion that 'Dative' Case typically expresses an Indirect Object relationship and ANIMATE NPs are usually employed as an indirect object, 'ey' cannot be considered a dative marker since it is not compatible with Animate NPs, as shown in the following:

```
(44) (i)   John-i  Mary-eykey /*-ey   mwul-ul  cwuessta
               -Nom    -eykey(Dat)    water-Acc  gave
           'John gave some water to Mary'
     (ii)  John-i    namwu-ey /*-eykey   mwul-ul  cwuessta
               -Nom  tree-ey(Loc)        water-Acc  gave
           'John sprayed some water to the tree.'
```

Given that animacy is a necessary condition for defining 'dative', it is not correct to consider 'ey' to be a dative marker. Arguably, only the markers that are compatible with animate NPs (i.e., 'eykey', 'hanthey', etc.) should be viewed as dative markers.

The observation made thus far does not necessarily imply that the dative and locative cases are clearly distinct. In fact, rather than distinguishing between these two, we might consider dative to be subsumed by locative. That is, dative can be regarded as locative with a [+animate] feature, as suggested in (44)(i) and (44)(ii) above. Such a consideration is conceptually appealing and also empirically supported: There is no morphological distinction between dative and locative markers in many languages.

To summarize our assumptions, we view 'ey' as a locative marker and we consider the dative markers ('eykey', 'hanthey', 'kkey', etc.) to be subsumed by locative markers.[12]

[12] The same relation holds among the three locative markers cited in Ihm, Hong, and Chang (1988): 'eyse' is an unmarked locative marker and 'eykeyse' and 'hantheyse' are special locative markers which are employed only when the NP they are attached to is [+animate].

7.2 Distribution of 'ey' and 'eyse'

The distribution of Korean locative postpositions 'ey' and 'eyse' has hardly been accounted for in a systematic way. At first glance, it may seem that too many factors are involved in their distribution. However, the fact that the two postpositions are in complementary distribution in many cases suggests that their usage is not random. In the following, we will see that 'ey' and 'eyse' are instances of VP-internal and VP-external locative phrases, respectively, and when 'eyse' functions as a VP-internal locative it should be treated as a compositional postposition. Consequently, two types of 'eyse' are identified, each of which accidently has the same phonetic form.

Consider the following pair:

```
(45) (i)   John-i  cip-ey/*-eyse tulekassta/tochakhayssta/kassta
               -Nom house-Loc      entered /   arrived   / went
           'John entered/ arrived at/ went to his house'
     (ii)  John-i  cip-eyse/*-ey nolayhayssta/wulessta/cassta
               -Nom  house-Loc         sang  /  cried / slept
           'John sang/cried/slept in his house'
```

The above data suggest that an appropriate locative postposition is selected according to the semantic requirements of the verb. The crucial difference between the two groups of verbs in (45) boils down to whether the locative phrase is an argument or a non-argument. The verbs 'enter', 'arrive', and 'go' take a locative phrase as an argument. On the other hand, 'sing', 'cry' and 'sleep' do not.[13] We draw the generalization that verbal argument locatives are headed by 'ey' and non-argument locatives are headed by 'eyse'.

In terms of phrase-structure position, we assume that the 'ey' locative occurs as a sister of the verb due to its argument status whereas the 'eyse' locative is placed outside of VP. Hence, the former may be called 'VP-internal locative' and the latter 'VP-external locative'.

7.3 Two Kinds of 'eyse'

There is a potential problem with the above generalization on the distribution of 'ey'/'eyse': 'eyse' sometimes functions as an argument locative, as in the following:

[13] The term 'argument' here includes both obligatory and optional arguments. Although both types of arguments are distinguished in terms of the obligatoriness of syntactic realization, they are not differentiated in the LCS. That is, not only do verbs such as 'put' take locative PPs as LCS arguments, but verbs such as 'lie' also take locative PPs as LCS arguments, even though the locative argument of the latter may not be realized in syntax.

```
(46) John-i   namwu-eyse  tteleciessta
         -Nom  tree-Loc     fell
     'John fell from the tree'
```

When the 'eyse' locative functions as an argument of the verb, it is always construed as the 'source'. Such a usage is contrasted with that of 'ey', which never represents the 'source' argument. In other words, when the 'eyse' locative functions as an argument it is in complementary distribution with the 'ey' locative.

The observation made thus far is summarized in the following way: 'ey' locatives are arguments while 'eyse' locatives are not. The only exception to this generalization is the usage of the 'eyse' locative as a 'source' argument, as seen in (46). In the following, we will see that 'eyse' in (46) is a totally different postposition from 'eyse' in (45)(ii) although they accidently have the same phonetic form.

First of all, the semantics of the two instances of 'eyse' are entirely different: The argument version of 'eyse' is interpreted as a 'source', whereas the non-argument version of 'eyse' has nothing to do with such an interpretation. Second, we can view the argument version of 'eyse' as a word that is comprised of two individual morphological units, 'ey' and 'se', while the non-argument 'eyse' is considered a single morphological unit. In other words, the former is derived from the locative 'ey' in (45)(i) by adding morpheme 'se', which is responsible for the 'source' interpretation. The derivation of a 'source' meaning by adding 'se' is also observed in other cases such as 'eykey-se' and 'hanthey-se'.[14] Summarizing, 'eyse' in (45)(ii) and (46) should be considered two different postpositions, sharing the same phonetic form by accident.

7.4 Interpretation of 'ey'

It is important to note that 'ey' cannot be properly interpreted independently from the semantics of the verb. Consider the following:

```
(47) John-i namwu-ey   olakassta
```

[14]Both 'eykey' and 'hanthey' (to) are dative postpositions. By adding 'se' to these postpositions (i.e., 'eykeyse' and 'hantheyse' (from)) we derive locative postpositions with the meaning of 'source' (See (iii) below.)

Note that 'ey' is, in a sense, a default locative postposition which can be combined with bound morphemes such as 'key' or 'se'. ('key' is a bound morpheme associated with the [+animate] feature.) The following are possible derivations from 'ey':

(i) ey-key : [+animate] locative (= dative)

(ii) ey-se : [+source] locative

(iii) ey-key-se: [+animate] & [+source] locative

```
     -Nom tree-Loc   climbed
'John climbed a tree'
```

(48)
```
John-i namwu-ey     tol-ul  tenciessta
     -Nom tree-Loc  stone-Acc    threw
'John threw a stone to a tree'
```

(49)
```
John-i namwu-ey       ilum-ul  saykyessta
     -Nom tree-Loc     name-Acc   engraved
'John engraved his name on a tree'
```

The above examples indicate that the same locative 'namwu-ey' (i.e., 'tree-Loc') can be interpreted in various ways. Unlike the English preposition, Korean locative 'ey' is not specified in terms of whether some sort of 'movement' or 'directionality' is involved. Selecting such semantic feature seems to rely on the semantics of the verb. For instance, in (48) 'tenciessta' (threw) selects a locative PP as the target of 'throwing a stone'. Selecting such a target PP implies that some sort of directionality/movement is involved in the action represented by the VP. Hence, 'namwu-ey' is interpreted as 'to a tree' rather than 'on a tree' etc.

Appendix A summarizes the distribution of locatives more generally with additional examples.

8 Issues in Translation and Transliteration

This section describes difficulties in translation and transliteration of Korean. A number of issues have arisen during the translation of the military message corpus shown in Appendix C. The four most difficult problems concerning the translation of sentences in this domain are PPs modifying NPs, the occurrence of acronyms, variations among different phrasal conjunctions, and negation. Examples of each of these are given below.

- PPs modifying NPs

  Korean requires 'attributive phrases' to be realized as full relative clauses, not as simple modifying expressions. This differs from English, which allows either construction to be used.

(50)
```
(i)  *?John-i   pyek-ey     kulim-ul  poassta
           -Nom  wall-Loc  picture-Acc   saw
(ii)  John-i   pyek-ey  iss-nun      kulim-ul  poassta
           -Nom  wall-Loc exist-Rel  picture-Acc   saw
      'John saw the picture on the wall'
```

(51) *?John-eykeyse phyenci
-Loc(from) letter

(i) John-eykeyse-uy phyenci
-Loc(from)-Gen letter
'the letter from John'

- Acronyms

  The composition of acronyms with other morphemes is not a fully productive process: 'han-mi' is Korea-US, but composing 'han' with other morphemes associated with countries (e.g., *han-hwa (Korea-Netherlands)) is not necessarily allowed.

(52) han-mi-yenhap-sa
Korea-US-combination-command

- NP Conjunction vs. VP Conjunction

  Korean allows NP conjunction to be formed in two ways, one with the morpheme 'wa' and the other with the morpheme 'mich'; there is a subtle semantic difference between these. VP conjunction is formed with the morpheme 'ko', which forces a subordination relationship.

(53) (i) senswu-wa khochi
player-AND coach

(ii) senswu mich khochi
player AND coach

(54) John-un cengcikha-ko pwucilenha-ta
-Top be honest-AND be diligent-Ind
'John is honest and diligent'

- Negation and Negative Polarity Item

  Negation is a very complex phenomenon in Korean. In order to employ the equivalent of the word 'any' (amwu-len) Korean requires verbal morphology 'ci'. In addition, scoping problems (e.g., everyone-not(X) vs. no one (X)) make the translation of Korean difficult.

(55) We are receiving no reports....

(i) wuli-nun amwu-len poko-to pat-ci
we-Top any-not report-Delimiter receive-ci

moshayssta
(can/could)not

(56) `Everyone did not leave.`

(i)

```
motun    salam-i      ttena-ci-nun
every/all person-Nom   leave-ci-Delimiter

anassta
(did) not

'Everyone did not leave'
```

(ii)

```
motun      salam-i     ttena-ci    anassta
every/all person-Nom   leave-ci   (did) not
'No one left'
```

In addition to translation issues, there are issues concerning the transliteration program written by Jye-hoon Lee that converts the Yale romanized format into the Hangul character set. (The program also goes in the other direction.) The transliteration system is hooked into 'Hanterm' (a Korean version of Xterm) that displays Korean characters in X windows.

The Korean alphabet has 14 consonants and 10 vowels. In general, Korean characters may be parsed as follows:

```
Vowel
Consonant + Vowel
Consonant + Vowel + Consonant
```

The major problem in converting Hangul into the romanized form is that ambiguity makes it difficult to parse characters. For example, the string C + V + C + V + C is ambiguous; there are two ways of dividing this string into two characters:

```
C + V & C + V + C
C + V + C & V + C.
```

In order to resolve such ambiguities, the current system employs a period (.) as a boundary delimiter. The next version of the system will signal the start of character by capitalizing it.

Another difficulty is that the Yale romanization system is far from what native Korean speakers are likely to understand. However, this is why the transliteration program is an important tool; since the transliteration is bidirectional, the native Korean speaker need not see the romanized form, but can read the Hangul form of the characters from the screen.

9 Morphological Properties of Korean

We are currently implementing a morphological component in the MT project. Our intent is to use PC-KIMMO, a two-level processor for morphological analysis that runs on the PC, Mac, and Sun. This PC version of the program is named after its inventor Kimmo Koskenniemi. (See Karttunen and Wittenburg (1983) and Antworth (1990) for complete details.)

The program is designed to generate (produce) and/or recognize (parse) words using a two-level model of word structure in which a word is represented as a correspondence between its lexical level form and its surface level form. PC-KIMMO is language-independent. For each language description the user prepares two input files: (1) a set of rules that govern phonological/orthographic alternations and (2) a lexicon that lists all words (morphemes) in their lexical form and specifies constraints on their order. The rules and lexicon are implemented computationally using finite state machines.

The remainder of this section outlines morphological properties for Korean. We are in the process of developing the morphology rules corresponding to this analysis; these will then be programmed into the PC-KIMMO system.

9.1 Voice: Active/Passive

The Korean passive voice is indicated by the passive morpheme 'ki'.

(57)
```
kay-ka  koyangi-lul  ccoch-ass-ta
dog-Nom     cat-Acc  chase-Pst-Ind
'A dog chased a cat'
```

(58)
```
koyangi-ka  (kay-eykey)  ccoch-ki-ess-ta
   cat-Nom   dog-by     chase-Pass-Pst-Ind
'A cat was chased (by a dog)'
```

9.2 Mood: Indicative/Interrogative/Imperative/Suggestive

Korean mood is indicated by the morpheme 'ta' for indicative, 'ni' for interrogative, 'ra' for imperative, and 'ca' for suggestive.

(59)
```
John-i   pap-ul   mek-ess-ta
    -Nom meal-Acc  eat-Pst-Ind
'John ate the meal'
```

(60) (i)
```
John-i   pap-ul   mek-ess-ni
    -Nom meal-Acc  eat-Pst-Intr
'Did John eat the meal ?'
```

(ii)
```
John-i  mwues-ul   mek-ess-ni
    -Nom  what-Acc  eat-Pst-Intr
'What did John eat ?'
```

```
(61) (ne)   pap-ul   meke-ra
       you  meal-Acc  eat-Impr
     'Eat the meal !'
```

```
(62) (wuri)  pap-ul   mek-ca
        we   meal-Acc  eat-Suggest
     'Let's eat the meal'
```

9.3 Relative Clauses

Relative clauses are indicated in Korean by the morphemes 'ko' and 'n'.

```
(63) John1-i [e1 Mary-lul manna-ss-ta-ko]    malha-yss-ta
          -Nom        -Acc meet-Pst-Ind-Comp   say-Pst-Ind
     'John said that he met Mary'
```

```
(64) John-i [e1 Mary-lul ttayli-0-n]  namca1-lul  manna-ss-ta
         -Nom        -Acc  hit-Pst-Rel    man-Acc  meet-Pst-Ind
     'John met the man who hit Mary'
```

9.4 Subordinate Clauses

Subordinate clauses are indicated in Korean by a number of different morphemes including 'umyen' (if), 'ciman' (though), 'se' (because), and 'ko' (and).

```
(65) nayil nalssi-ka   coh-umyen  na-nun pichi-ey ka-keyss-ta
     tomorrow weather-Nom be good-if  I-Top beach-to go-Futr-Ind
     'If the weather is good tomorrow, I will go to the beach'
```

```
(66) Mary1-nun [e1 pwucai-ciman] kemsohakey saynghwalha-n-ta
          -Top    be rich-though  frugally      live-Pres-Ind
     'Mary lives a frugal life although she is rich'
```

```
(67) John1-un onul   [e1 apa -se ]     pyengwen-ey ka-ss-ta
          -Top today   be sick-because  hospital-to go-Pst-Ind
     'Today, John went to the hospital because he was sick'
```

```
(68) John-i Mary-lul manna-ss-ko  Jay-ka Sue-lul manna-ss-ta
         -Nom    -Acc meet-Pst-and    -Nom   -Acc  meet-Pst-Ind
     'John met Mary and Jay met Sue'
```

10 Lexical Knowledge Acquisition

The issues raised in section 6 are relevant to another area that lies at the intersection of research by the CMU and Maryland teams, i.e., the problem of automatic lexical acquisition. Nirenburg et al. (1992) remind of us of the importance of this problem: 'Scaling up dictionaries and other knowledge bases of a knowledge-based machine translation system is essential for the overall success of the field of machine translation.' CMU aims to provide automatic acquisition procedures for a feature-based lexicon (as described in Nirenburg, et al. (1992) along with efforts by Wilks, et al. (1990)). This research has involved the development of methods for automatic transformation of information in human-oriented dictionaries into a form suitable for MT. In Wilks's terminology, this is the problem of converting machine readable dictionaries (MRD's) into machine tractable dictionaries (MTD's).

Although Maryland's view of the lexicon does not involve deep KR information of the kind that is required for the ILT in (38) above, the problem of lexical acquisition is still of utmost importance. It is, in fact, the lack of deep KR information that allows us to eliminate much of the complexity that would be involved in general knowledge acquisition that the CMU team is forced to undertake. Instead, we focus on acquiring structures such as (40) above, which isolate properties relevant to the syntax-to-semantics mapping. By restricting the type of knowledge we are attempting to acquire (i.e., by staying to the left of the dividing line in figure 6), we are able to address certain issues currently discussed in the literature such as the two latest volumes of the Computational Linguistics journal (special issue on Using Large Corpora, 19:1-2, 1993). In particular, Pustejovsky, et al. (1993) propose to automate the process of acquiring the information relating to the qualia structure of nouns. The problem is that this information is often quite complex and difficult to discern from corpora. Fortunately, this type of information is generally not relevant to MT (i.e., subtle distinctions such as the difference in semantics for the word 'bake' in 'bake a potato' vs. 'bake a cake'); thus, we are able to simplify the acquisition procedure considerably by viewing this information as part of the KR (i.e., aspectual coercion) and focusing only on inherent features in the IL (i.e., process vs. state vs. event). Although our focus is different than that of CMU, our ultimate goal is the same, i.e., to use MRD's and corpus information to provide MTD's that are usable for MT of different languages.

10.1 Application of Dowty's Tests of Verb Aspect for Automatic Lexical Acquisition

One of our goals for building a framework for automatic lexical acquisition is to develop a program that acquires aspectual representations from corpora

	Test	STA	ACT	ACC	ACH
1.	X-ing is grammatical	no	yes	yes	yes
* 2.	has habitual interpretation in simple present tense	no	yes	yes	yes
3.	spend an hour X-ing, X for an hour	yes	yes	yes	no
4.	take an hour to X, X in an hour	no	no	yes	yes
* 5.	X for an hour entails X at all times in the hour	yes	yes	no	no
* 6.	Y is X-ing entails Y has X-ed	no	yes	no	no
7.	complement of stop	yes	yes	yes	no
8.	complement of finish	no	no	yes	no
* 9.	ambiguity with almost	no	no	yes	no
*10.	Y X-ed in an hour entails Y was X-ing during that hour	no	no	yes	no
11.	occurs with studiously, carefully, *etc.*	no	yes	yes	no

Figure 7: Dowty's Eleven Tests of Verb Aspect in English

by examining the context in which all verbs occur and then dividing them into four groups: state, activity, accomplishment, and achievement. The division of verbs into these four groups is based on several syntactic tests that are well-defined in the linguistic literature such as those by Dowty (1979) shown in figure 7.

As described in Dorr (1992a), it is possible to use these tests for determining the aspectual category of verbs in a corpus.[15] Preliminary results were obtained by running the program on 219 sentences of the Lancaster-Oslo-Bergen corpus as shown in figure 8.[16] Note that the program was not able to pare down the aspectual category to one in every case. We expect to have a significant improvement in the classification results once the sample size is increased. Presumably more tests would be needed for additional improvements in results.

[15]Some tests of verb aspect shown here could not be implemented in the acquisition program because they require human interpretations. These tests are marked by asterisks (*). For example, Test 2 requires human interpretation to determine whether or not a verb has habitual interpretation in simple present tense.

[16]For brevity, only a subset of the verbs are shown here.

Verbs	**Aspectual Category(s)**
doing	(ACC)
facing	(ACC ACT)
asking	(ACC ACT)
made	(ACC)
drove	(ACC ACT)
welcome	(STA ACC ACT ACH)
emphasized	(STA ACC ACT ACH)
thanked	(ACC ACT STA)
staged	(ACC)
make	(ACC)
continue	(ACC ACT)
writes	(ACC)
building	(ACC)
running	(ACC ACT)
paint	(ACC)
finds	(ACC ACT)
arrives	(ACC ACT)
jailed	(ACC ACT STA)
nominating	(ACH ACT ACC)
read	(ACC ACT)
ensure	(STA ACC ACT ACH)
act	(ACT ACC)
carry	(ACC)
exercise	(ACC)
impose	(STA ACC ACT ACH)
contain	(STA ACC ACT ACH)
infuriate	(ACC ACT)

Figure 8: Aspectual Classification Results

	Test	STA	ACT	ACC	ACH
1.	X-'MYENSE' is grammatical	no	yes	yes	yes
* 2.	has habitual interpretation in simple present tense	no	yes	yes	yes
3.	X for an hour 'han sikan tongan'	yes	yes	yes	no
4.	X in an hour 'han sikan maney'	no	no	yes	yes
* 5.	X for an hour entails X at all times in the hour	yes	yes	no	N/A
* 6.	Y is X-ing entails Y has X-ed	N/A	yes	no	no
7.	complement of stop 'kumantwuta'	no	yes	yes	no
8.	complement of finish 'kkutnayta'	no	yes	yes	yes
* 9.	ambiguity with almost 'keuy'	no	no	yes	no
*10.	Y X-ed in an hour entails Y was X-ing during that hour	N/A	N/A	yes	no
11.	occurs with studiously, carefully, *etc.*	no	yes	yes	no

Figure 9: Dowty's Eleven Tests of Verb Aspect in Korean

Over the last year, Sungki Suh has developed an analogous set of tests for Korean as shown in figure 9. Surprisingly, many of the aspectual tests used for English apply in the same way to Korean. We will now look at the analysis of each of these cases.

1. X-'MYENSE' is grammatical: Modified form is analogous to English.

   Because Korean does not have a 'progressive' form, the English test 'X-ing is grammatical' is not applicable. However, the following test seems to be relevant:

   (a)
   ```
   *?John-un [Mary-lul miweha-myense] hakkyo-ey  kassta
         -Top      -Acc   hate-MYENSE  school-to   went
   'John went to school and hated Mary'
   'John went to school, hating Mary'
   ```

   (b)
   ```
   John-un [ppang-ul  meku-myense]  hakkyo-ey  kassta
   ```

```
     -Top  bread-Acc  eat-MYENSE  school-to   went
 'John went to school, eating bread'
```

The English equivalent for 'MYENSE' seems to be 'while' (or 'and'). The construction involving 'MYENSE' is used to describe a situation where the same person does more than one action at the same time. In general, the '-MYENSE' action is semantically subordinate, as seen in (b) above. Using this modified version of test 1 results in a distribution of Korean verbs that is analogous to that of English verbs. However, In order for the test to be effective, the matrix verb should be [+dynamic].

2. Habitual interpretation in simple present tense: Analogous to English.

   This test results in a distribution of Korean verbs that is analogous to that of English verbs.

3. X for an hour: Analogous to English.

   The construction 'X for an hour' results in a distribution of Korean verbs that is analogous to that of English verbs. However, the construction 'Spend an hour X-ing' does not provide the same distribution for 'State' verbs:

```
(c) *?John-i Mary-lul miweha-nun-tey  il nyen-ul ponaysst a
         -Nom     -Acc  hating  in    one year-Acc  spent
    'John spent one year hating Mary'
```

4. X in an hour: Analogous to English.

   The construction 'X in an hour' results in a distribution of Korean verbs that is analogous to that of English verbs. However, the construction 'Take an hour to X' does not provide the same distribution for 'Activity' verbs. This is because the semantics of the verb 'kellyessta' (elapsed/took time) coerces an 'Activity' verb to an 'Accomplishment' verb:

```
(d) John-un swuyengha-nun-tey   han sikan-i  kellyessta
        -Top      swimming in    one hour-Nom  elapsed
    'It took one hour for John to swim'
    <Coercion: [-telic] --> [+telic]>
```

5. X for an hour entails X at all times in the hour: Analogous to English.

   This test results in a distribution of Korean verbs that is analogous to that of English verbs.

6. Y is X-ing entails Y has X-ed: Analogous to English.

   This test results in a distribution of Korean verbs that is analogous to that of English verbs.

7. Complement of stop ('kumantwuta'): Not analogous to English.

   The Korean 'kumantwuta' construction does not have an analogous distribution of 'State' verbs to that of English. The reason seems to be that 'kumantwuta' tends to be associated with a 'volitional' situation; thus, the presence of 'state' verbs such as 'miwehata' (hate) or 'coahata' (like) creates semantic conflict:

```
(e) *?John-un [Mary-lul coaha-/miweha-ki]-lul  kumantwuessta
          -Top     -Acc like-/hate   -Nomlz-Acc stopped
    'John stopped liking/hating Mary'
```

8. Complement of finish ('kkutnayta'): Not analogous to English.

   The Korean 'kumantwuta' construction does not have an analogous distribution of 'Activity' and 'Achievement' verbs to that of English. This is because the semantics of 'kumantwuta' coerces such verbs to be an 'Accomplishment':[17]

```
(f) John-un [swuyengha-ki]-lul  kkutnayessta
        -Top     swim-Nomlz-Acc    finished
    'John finished swimming.'
    <Coercion: [-telic] --> [+telic]>
(g) John-un [cengsang-ey tatalu-ki]-lul   kkutnaysssta
        -Top   summit-at   reach-Nomlz-Acc   finished
    'John finished reaching the summit.'
    <Coercion: [+atomic] --> [-atomic]>
```

   Note that the construction 'cengsang-ey tataluta' (to reach the summit) is originally an 'achievement' verb as in English since it fails in test 3:

```
(h) *John-i   han sikan tongan  cengsang-ey  tatalassta
         -Nom  one hour    for    summit-at     reached
    'John reached the summit for one hour'
```

9. Ambiguity with almost ('keuy'): Analogous to English.

10. Y X-ed in an hour entails Y was X-ing during that hour: Analogous to English.

11. Occurs with studiously, carefully, etc.: Analogous to English.

[17] In English, coercing 'Activity' verbs to 'Accomplishment' verbs is also possible in the construction involving 'finish'.

10.2 Using Syntax to Bootstrap Semantics: Levin's Verb Classification

Our research in lexical knowledge acquisition for the MT project has involved the construction of lexical representations that will be stored in the dictionary for each of the source and target languages. We are currently using an editor that allows us to construct LCS definitions for words (see Appendix B). Jye-hoon's dissertation work involves the investigation of lexical representations based on work by Jackendoff (1983, 1990), Pustejovsky, et al. (1993), Pustejovsky (1991,1994), and also the verb classification hierarchy developed by Levin (1993). This investigation is part of an ongoing effort to develop an automatic lexical acquisition scheme for building lexical representations from online corpora and machine-readable dictionaries. We are investigating the syntactic criteria that are used to delimit the semantic categories of verbs in Korean and we intend to use this to provide a more automatic dimension to the LCS editor.

Results obtained so far, by Mr. Sungki Suh, have shown that the syntactic tests for English do not apply directly to verb classes in Korean since there is not always an exact counterpart of (certain) English verbs in Korean. The remainder of this section will focus on the distinctions between verb classes in Korean and English.

The transitivity alternation, which is an important criterion for verb classification in Levin's theory, is irrelevant (inapplicable) to Korean. This is because it is not possible for a Korean verb to function as both a transitive and an intransitive verb. (Transitivity alternation is not possible without adding a derivational morpheme, i.e., a passive or causative morpheme.) Examples (69)–(71) are relevant to this point:

```
(69) (i)  elum-i    nok-ass-ta                     [base]
           ice-Nom  melt-Pst-Ind
          'The ice melted'

     (ii) John-i   elum-ul     nok-i -ess-ta      [derived]
              -Nom   ice-Acc  melt-Caus-Pst-Ind
          'John melted the ice'
          (John made the ice melt)

(70) (i)  John-i  changmwun-ul  kkay-ss-ta        [base]
              -Nom   window-Acc break-Pst-Ind
          'John broke the window'

     (ii) changmwun-i    kkay-ci-ess-ta           [derived]
             window-Nom  break-Pass-Pst-Ind
          'The window broke'
          (The window was broken)
```

```
(71) (i)   ai-ka    kel-ess-ta                      [base]
           child-Nom  walk-Pst-Ind
           'The child walked'
     (ii)  John-i    ai-lul   kel-li-ess-ta         [derived]
               -Nom child-Acc walk-Caus-Pst-Ind
           'John walked the child'
           (John made the child walk)
```

Another interesting phenomenon related to transitivity alternation is that verbs like 'eat', which can be used both transitively and intransitively in English, are always accompanied by an overt object NP in Korean, even though semantically it is not necessary to specify the object of 'eat'. This is shown in (72)(i) and (72)(ii).[18] (Employing a null object is possible only when the context forces a specific entity to be the object, as in (73).)

```
(72) (i)   ?? John-un   ku ttay      mek-ko    iss-ess-ta
                   -Top  at that time  eat-Conj  Aux-Pst-Ind
              'John was eating at that time'
     (ii)     John-un   ku ttay         mwues-ul    mek-ko
                   -Top  at that time  something-Acc  eat-Conj
              iss-ess-ta
              Aux-Pst-Ind
              'John was eating at that time'

(73) (i)     nwu-ka   ku kheyik-ul  mek-ess-ci ?
             who-Nom  the cake-Acc  eat-Pst-Q
             'Who ate the cake?'
     (ii)    Mary-ka  (mek-ess-e)
                 -Nom  eat-Pst-Ind
             'Mary ate it'
```

The above data demonstrate that there is a strict boundary between transitive and intransitive verbs in Korean; thus, the transitivity alternation is not available in Korean.

In addition to the transitivity alternation, English verbs exhibit a number of other alternations. The range of alternations that a verb may participate in is used as a criterion for the classification of verbs into semantic categories in Levin (1993). The following is an instance of alternation within VP:

[18] Note that if appropriate context is provided, several types of null constituents are allowed in Korean including the object, the subject, the VP, etc.

```
(74) John separated A from B / John separated A and B

(75) John detached A from B / *John detached A and B
```

The alternation observed in (74), which is called the reciprocal alternation, is not possible in (75). Meanwhile, the Korean counterparts of these two examples do not show such a contrast:

```
(76) John-un  A-lul B-lopwuthe  pwunlihayssta
         -Top  -Acc  -from      separated
     'John separated A from B'

     John-un  A-wa  B-lul     pwunlihayssta
         -Top  -and  -Acc     separated
     'John separated A and B'

(77) John-un  A-lul B-lopwuthe  tteyenayssta
         -Top  -Acc  -from      detached
     'John detached A from B'

     John-un  A-wa  B-lul     tteyenayssta
         -Top  -and  -Acc     detached
     'John detached A and B (= A from B)'
```

The fact that the contrast observed in English is not found in Korean does not necessarily mean that the reciprocal alternation is not a good criterion for verb classification. The different result from the two languages with regard to such alternation may be attributed to the lack of an 'exact' counterpart of 'detach' (or 'separate') in Korean. In other words, there is no guarantee that the verbs used in (76) and (77) correspond exactly to the English verbs 'separate' and 'detach', respectively. The same question arises when we apply other alternation tests from Levin (1993) to Korean. In sum, it might be difficult to find out alternation patterns observed in English in another language since there may not be exact counterpart of (certain) English verbs in that language. (If identical alternation patterns are observed across different languages, then we can say that the validity of verb classification based on such alternation patterns is confirmed.)

A similar question arises with regard to the translation of prepositions. For example, the word 'from' (and also other English prepositions) could be translated in more than one way in Korean. Similarly, certain Korean postpositions are translated in more than one way in English. This issue becomes crucial when we consider alternation tests such as the above. That is, an exact equivalence between prepositions (or postpositions) in the two languages is an important prerequisite for the application of the alternation tests. It could be the case that many alternation tests are not applicable to Korean simply because a certain preposition (or an expression involving a certain preposition) in English does not have a Korean counterpart (at least for a certain usage of that preposition). The lack of a Korean counterpart to the following English sentence is relevant to this point:

(78) `John cut at the bread.`

10.3 Toward a Universal Semantic Classification

The previous section appears to indicate that the distinctions between meanings of verbs and prepositions across languages poses a great obstacle to the development of a universal semantic classification. In fact, there has been work by Mitamura (1990) indicating that in a language like Japanese, which is similar in structure to Korean, syntactic patterns are indeed associated with semantic classes, but these semantic classes are not the same as those of English.

An example of a syntactic alternation in Japanese that delineates a semantic class is the 'ga/kara' alternation, which appears with 'giving' verbs such as 'tutaeru' (report to) and 'okuru' (send):

(79) (i) `Watasi ga/kara Kimura-san ni tutaeta`
`'I reported to Mr. Kimura'`

(ii) `Kodomotati ga/kara sensei ni kaado o okutta`
`'The children sent a card to the teacher'`

The subject is normally marked by 'ga', but when the subject is a giver (source), it can be marked by 'kara'. When 'kara' is used, the meaning of giver or source in terms of original place is emphasized.

In the English-based framework of Levin (1993) there is no semantic class corresponding to verbs of 'giving'. However, we do not see this to be an absolute obstacle to using syntactic alternations as the basis for the development of a universal semantic classification system. There are two hypotheses that we are currently testing along these lines; either: (1) there is a single semantic classification that provides a union of the underlying features of Levin's and Mitamura's semantic frameworks (as well as others proposed for other languages) or (2) languages carve up the world into different semantic features and, while syntactic alternations might be used to divide language into semantic classes, there is no single division that applies across all languages.

If we are to test the first hypothesis, then we need to isolate the components of meaning that underlie Levin's semantic classes. Although Levin's semantic classification provides a precise characterization of an extensive number of verbs, no attempt is made to explain the semantic features inherent to each class. For example, Levin considers the verbs 'separate' and 'detach' to be in distinct semantic categories, the former falling into class 23.1 (the 'separation' class) and the latter falling into class 23.3 (the 'disassemble' class). The reason for this distinction is that the former class includes verbs that participate in the reciprocal alternation (as illustrated in (74) above) whereas the latter includes verbs that do not participate in this alternation (as illustrated in (75) above). In order to test our hypothesis, we need to identify the precise reason behind the existence of this distinction. In the case of 'separate' vs. 'disassemble', it could be argued that the verbs in the 'separate' class are ones concerned with the separation of two objects that, in some sense, have an equal status (along some dimension such as size), whereas the verbs in the 'disassemble' class are concerned with objects that are of unequal status (i.e., one is generally considered to be subordinate to the other). Identifying this type of semantic information on a per-class basis would allow us to work toward the identification of a universal semantic classification system. That is, once we identify the semantic features that delineate the semantic classes in Levin's framework, we can determine how this feature specification applies to other languages such as Japanese and Korean. These features can then be merged with those that characterize verbs in other languages in order to arrive a 'universal' semantic classification.

Regarding the second hypothesis, it might be the case that the semantic features across languages are different enough that no 'universal' semantic classification can be found. If this is the case, then we need to devise a set of systematic mapping rules between the semantic classes of each language and the set of LCS primitives that exist in our interlingual representation. For example, the class of verbs of 'giving' described in Mitamura's framework are linked systematically to the GO and TO primitives in the Possessional field, whereas the class of verbs of motion with location are linked to the GO and TO primitives in the Locational field. Devising a set of systematic mapping rules allows us to link semantic classes together cross-linguistically, by using the primitive units of meaning underlying the LCS. The LCS thus becomes our basis for defining a 'universal' semantic classification if our second hypothesis is correct.

Once we have determined the nature of our semantic classification system, we will have established a link between our interlingual representation (i.e., the LCS that is retrieved from the dictionary and composed during on-line processing) and the potential syntactic realizations of the underlying conceptual information. This link allows us to build a framework for automatic acquisition of LCS-based lexical entries. The basic idea would be to search through corpora for occurrences of verbs in particular syntactic

Syntactic	Semantic Classes				
Alternations	9.1	9.2	9.3	9.4	...
LOCATIVE	N	N	N	N	...
MIDDLE	N	N	N	N	...
CAUSATIVE	N	Y	N	N	...
DATIVE				N	...
ZERO-REL NOMINAL	Y	Y	Y	Y	...
into			N		...
onto				N	...
NP	N	N	N	Y	...
PP	N	N			...
NP from	N		N		...
NP from-to	N		N	Y	...
NP in			Y		...
NP into			Y		...
NP near	Y				...
NP next to		Y			...
NP on	Y	Y		N	...
NP onto				Y	...
NP out of				Y	...
NP to	N		N		...
NP under	Y				...
⋮	⋮	⋮	⋮	⋮	

Figure 10: Matrix of Syntactic Frames Crossed with Semantic Frames

alternations and to then to build LCS representations for these verbs according by using our pre-established link between syntactic realizations and underlying conceptual information.

As an initial step toward building this framework for automatic acquisition, we have been working on the construction of a matrix of syntactic frames crossed with semantic classes. A preliminary version of this table is given in figure 10.[19]

Once we have built this table, we intend to use the result on a Korean text corpus to build LCS representations automatically. As our initial testbed, we intend to use a military message corpus which we have translated into Korean. The translation is shown in Appendix C in the Yale romanized

[19]The status of the blanks still needs to be established. Many of these will be filled in by hand after a thorough analysis of Levin's book.

format.

We are currently in the initial stages of negotiation with the Linguistic Data Consortium (LDC) for a much larger online corpus of Korean text called the YONHAP corpus. (This is a bilingual Korean-English corpus in their category of 'Newswire Negotiations'.) In addition, the LDC is currently negotiating with publishers to provide us a machine-readable Korean dictionary.

11 References

Antworth, Evan L. (1990) *PC-KIMMO: A Two-Level Processor for Morphological Analysis*, Occasional Publications in Academic Computing No. 16, Summer Institute of Linguistics, Dallas, TX.

Chomsky, Noam A. (1981) *Lectures on Government and Binding*, Foris Publications, Dordrecht.

Chomsky, Noam A. (1986a) *Knowledge of Language: Its Nature, Origin and Use*, MIT Press, Cambridge, MA.

Chomsky, Noam A. (1986b) *Barriers*, MIT Press, Cambridge, MA.

Dorr, Bonnie J. (1990) "Solving Thematic Divergences in Machine Translation," *Proceedings of the 28th Annual Conference of the Association for Computational Linguistics*, University of Pittsburgh, Pittsburgh, PA, 127–134.

Dorr, Bonnie J. (1992a) "A Parameterized Approach to Integrating Aspect with Lexical-Semantics for Machine Translation," *Proceedings of 30th Annual Conference of the Association of Computational Linguistics*, University of Delaware, Newark DE, 257–264.

Dorr, Bonnie J. (1992b) "Lexical Semantics for Interlingual Machine Translation," *Machine Translation* 7:3, 135–193.

Dorr, Bonnie J. (1993a) *Machine Translation: A View from the Lexicon*, MIT Press, Cambridge, MA.

Dorr, Bonnie J. (1993b) "Interlingual Machine Translation: A Parameterized Approach," *Artificial Intelligence* 63:1&2, 429–492.

Dorr, Bonnie J., and Clare Voss (1993) "Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System," *Proceedings of Twelfth Conference of the American Association for Artificial Intelligence*, Washington, DC.

Dorr, Bonnie J., Jye-hoon Lee, Dekang Lin, and Sungki Suh (1994) "Parameterization of a Message Passing Parser for MT of Korean and English," *submitted to 32nd Annual Conference of the Association for Computational Linguistics.*

Dowty, David (1979) *Word Meaning and Montague Grammar*, Reidel, Dordrecht, Netherlands.

Fillmore, Charles J. (1968) "The Case for Case," in *Universals in Linguistic Theory*, Bach, E., and R. T. Harms (eds.), Holt, Rinehart, and Winston, 1–88.

Fukui, N (1986) "A Theory of Category Projection and its Applications," Ph.D. thesis, Massachusetts Institute of Technology.

Grimshaw, Jane (1990) *Argument Structure*, MIT Press, Cambridge, MA.

Grosz, B., K. Sparck-Jones, and B. Webber (1987) *Readings in Natural Language Processing*, Morgan Kaufmann, Los Altos, CA.

Gruber, J. S. (1965) "Studies in Lexical Relations," Ph.D. thesis, Department of Information Science, Massachusetts Institute of Technology.

Hale, Kenneth and S. Jay Keyser (1986a) "Some Transitivity Alternations in English," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, Lexicon Project Working Paper 7.

Hale, Kenneth and S. Jay Keyser (1986b) "A View from the Middle," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, Lexicon Project Working Paper 10.

Hale, Kenneth and S. Jay Keyser (1989) "On Some Syntactic Rules in the Lexicon," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, manuscript.

Hale, Kenneth and Mary Laughren (1983) "Warlpiri Lexicon Project: Warlpiri Dictionary Entries," Massachusetts Institute of Technology, Cambridge, MA, Warlpiri Lexicon Project.

Hutchins, J. (1993) "After the Second Generation ...," *Machine Translation* 7:4, 259-264.

Ihm, Hong, and Chang (1988) *??*.

Jackendoff, Ray S. (1983) *Semantics and Cognition*, MIT Press, Cambridge, MA.

Jackendoff, Ray S. (1990) *Semantic Structures*, MIT Press, Cambridge, MA.

Karttunen, L., and K. Wittenburg (1983) "A Two-Level Morphological Analysis of English," *Texas Linguistic Form* 22, 163–278.

Kim, Soo-won (1989) "Wh-phrases in Korean and Japanese are QPs," *Proceedings of WCCFL 8*.

Levin, Beth (1985) "Lexical Semantics in Review," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, Lexicon Project Working Paper 1.

Levin, Beth (1993) *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, Chicago, IL.

Levin, Beth and Malka Rappaport (1986) "The Formation of Adjectival Passives," *Linguistic Inquiry* 17, 623–662.

Lin, Dekang (1993) "Context-Free Grammar Parsing by Message Passing," *Proceedings of the 31st Annual Conference of the Association of Computational Linguistics*, Ohio State University, Columbus, OH, 112–120.

Lin, Dekang and Randy Goebel (1993) "Context-Free Grammar Parsing by Message Passing," *Proceedings of PACLING-93*, Vancouver, BC.

Mitamura, Teruko (1990) "The Hierarchical Organization of Predicate Frames for Interpretive Mapping in Natural Language Processing," Center for Machine Translation, Carnegie Mellon University, Ph.D. dissertation, CMU-CMT-90-117.

Moon, G.-S (1989) "The Syntax of Null Arguments with special Reference to Korean," Ph.D. thesis, University of Texas.

Nirenburg, Sergei, and Lori Levin (1989) "Knowledge Representation Support," *Machine Translation* 4:1, 25–52.

Nirenburg, Sergei, Jaime Carbonell, Masaru Tomita, and Kenneth Goodman (1992) *Machine Translation: A Knowledge-Based Approach*, Morgan Kaufmann, San Mateo, CA.

Olsen, Mari Broman (1991) "Lexical Semantics, Machine Translation, and Talmy's Model of Motion Verbs," Northwestern University, Evanston, IL, Linguistics Working Paper, Volume 3.

Park, Chang Ho (1992) "Korean Generation in MATES/EK," *MT Conference*, Malaysia.

Pustejovsky, James (1988) "The Geometry of Events," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, Lexicon Project Working Paper 24.

Pustejovsky, James (1989) "The Semantic Representation of Lexical Knowledge," *Proceedings of the First International Lexical Acquisition Workshop*, IJCAI-89, Detroit, MI.

Pustejovsky, J. (1991) "The Generative Lexicon," *Computational Linguistics* 17:4, 409–441.

Pustejovsky, J., S. Bergler, and P. Anick (1993) "Lexical Semantic Techniques for Corpus Analysis," *Computational Linguistics* 19:2, 331–358.

Pustejovsky, J. (1994) "Linguistic Constraints on Type Coercion," in *Computational Lexical Semantics*, Patrick Saint-Dizier and Evelyn Viegas (eds.), Cambridge University Press, Cambridge, MA.

Rappaport, Malka and Beth Levin (1988) "What to Do with Theta-Roles," in *Thematic Relations*, Wendy Wilkins (ed.), Academic Press.

Rappaport, Malka, Mary Laughren, and Beth Levin (1987) "Levels of Lexical Representation," Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, Lexicon Project Working Paper 20.

Schank, Roger C. (1972) "Conceptual Dependency: A Theory of Natural Language Understanding," *Cognitive Psychology* 3, 552–631.

Suh, Sungki (forthcoming) "Constituent Structure Processing and Garden Path Phenomena in Korean," Ph.D. thesis, Department of Linguistics, University of Maryland.

Vauquois Bernard, and Christian Boitet (1985) "Automated Translation at Grenoble University," *Computational Linguistics* 11:1, 28–36.

Voss, Clare (forthcoming) "Encoding and Acquiring Lexical Representations for Machine Translation of Locative Expressions," Ph.D. thesis, Department of Computer Science, University of Maryland.

Wall, R. (1972) *Introduction to Mathematical linguistics*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Wexler, Kenneth, and M. Rita Manzini (1987) "Parameters and Learnability in Binding Theory," in *Parameter Setting*, Thomas Roeper and Edwin Williams (eds.), D. Reidel Publishing Company, Dordrecht, Holland.

Wilks, Yorick (1973) "An Artificial Intelligence Approach to Machine Translation," in *Computer Models of Thought and Language*, Roger C. Schank and K. M. Colby (eds.), Freeman, San Francisco, CA, 114–151.

Wilks, Y., D. Fass, C. Guo, J. McDonald, T. Plate, and B. Slator (1990) "Providing Machine Tractable Dictionary Tools," Machine Translation," *Machine Translation* 5, 99-154.

Winograd, Terry (1973) "A Procedural Model of Language Understanding," in *Computer Models of Thought and Lnaguage*, R. Schank and K. Colby (eds.), W. H. Freeman, San Francisco, CA.

Yoon, J.-Y. (1990) "Korean Syntax and Generalized X-bar Theory," Ph.D. thesis, University of Texas.

Zubizarreta, Maria Luisa (1982) "On the Relationship of the Lexicon to Syntax," Ph.D. thesis, Department of Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Zubizarreta, Maria Luisa (1987) *Levels of Representation in the Lexicon and in the Syntax*, Foris Publications, Dordrecht, Holland/Cinnaminson, USA.

A Summary of Locative Markers in Korean

- Argument/ Adjunct distinction ('ey' vs. 'eyse'):

(a)
```
John-i    cip-ey/*-eyse     kassta (went)
    -Nom    home-Loc
                            tochakhayssta (arrived)
                            tulekassta (entered)
                            namassta (remained)
```
(b)
```
John-i    cip-*ey/-eyse     nolayhayssta (sang)
    -Nom    home-Loc
                            cassta (slept)
                            wulessta (cried)
                            wuntonghayssta (exercised)
```

Verbs in 1(a) take locative PP as their argument, whereas verbs in 1(b) do not.

- Deriving [+animate]/[+source] Locatives:

(a) 'ey': argument locative marker

```
John-i namwu-ey      tol-ul   tenciessta
    -Nom  tree-Loc   stone-Acc     threw
'John threw a stone to a tree'
```

(b) 'ey+key': [+animate] argument locative marker

```
John-i Mary-eykey     tol-ul  tenciessta
    -Nom     -Loc     stone-Acc     threw
'John threw a stone to Mary'
```

(c) 'ey+se': [+source] argument locative marker

```
John-i  namwu-eyse tteleciessta
    -Nom  tree-Loc     fell
'John fell from the tree'
```

(d) 'ey+key+se': [+animate]&[+source] argument locative marker

```
John-i Bill-eykeyse  ton-ul   kkwuessta
    -Nom     -Loc    money-Acc   borrowed
'John borrowed some money from Bill'
```

Note [+animate] and [+source] features are irrelevant for adjunct locatives.

- Interpretation of argument locatives:

(a) PATH:

```
John-i  namwu-ey     tol-ul  tenciessta
    -Nom tree-Loc  stone-Acc    threw
'John threw a stone to a tree'
```

(b) NO PATH:

```
John-i  namwu-ey      ilum-ul  saykyessta
    -Nom tree-Loc     name-Acc   engraved
'John engraved his name on a tree'
```

Note: '-lo' always implies that a path is involved.

- Other Properties:

(a) (i)
```
John-i hankwuk-ey memwulko issta
    -Nom  Korea-Loc   stay    Aux
'John is staying IN Korea'
```

(ii)
```
John-i cikum mwunkan-ey   se    issta
    -Nom  now    door-Loc stand   Aux
'John is standing AT the door now'
```

(a) (i)
```
*?John-i   pyek-ey     kulim-ul  poassta
      -Nom  wall-Loc  picture-Acc   saw
```

(ii)
```
John-i   pyek-ey  iss-nun      kulim-ul  poassta
    -Nom  wall-Loc exist-Rel  picture-Acc   saw
'John saw the picture on the wall'
```

(a) (i)
```
*?John-eykeyse   phyenci
      -Loc(from)  letter
```

(ii)
```
John-eykeyse-uy     phyenci
    -Loc(from)-Gen  letter
'the letter from John'
```

B LCS Editing Capability

Manual construction of dictionaries for each of the languages handled by the MT system is a very tedious task; thus, we have designed and implemented a LCS Editor that makes the construction process much faster, easier, and less error-prone.

The heart of the LCS Editor is the LCS Entry Window shown in figure 11. This window is where LCS's are built and modified. During editing, the current state of the LCS is continuously updated and displayed in two windows, the 'Tree Representation' window and the 'Text Representation' window.

At the top of the Editor window is a browser which allows the author to set the Type, Primitive, and Field of the selected (highlighted) node in the current LCS. The selected Type determines which Primitives are available, and the selected Primitive determines the Fields that are allowed. Some Primitives have no allowed Fields. The Editor only allows valid choices to be made.

The Editor also displays sample sentences for various combinations of Primitive and Field, to guide the author with respect to the meanings of the various terms. In figure 11, the words 'We moved the statue from the park to the zoo' represent an example of a sentence using a word (i.e., 'move') that has the same Type, Primitive, and Field as the currently-selected node.

The LCS of this example contains nodes labeled X, Y, and Z. There are no Primitives by these names; rather, these nodes represent variables. Any nodes with the same variable name will eventually refer to the same item in the natural language sentence that this LCS is used to represent. For each Type, there is a Primitive called 'var_Type'. For example, under the Event type there is a primitive called 'var_Event'. If this were chosen as a Primitive, a window would appear asking for a variable name.

Under the browser is the Add Primitive button. Clicking this button allows the author to add a new primitive to the set that is currently defined, as well as specify which Type the primitive belongs to and which Fields are applicable.

The next item down is a group of controls labeled 'LCS Characteristics'. These are items that apply to the LCS entry as a whole. The author enters the word that the entry represents in the Text Field labeled 'The Word'. The Usage Comment button allows the author to enter a comment about the usage of this particular entry for documentation purposes. (It has no effect on the correctness of the entry.) Below this button is a popup menu labeled 'English', reflecting the fact that this is the LCS entry for the English verb, 'go'. Next to this is another popup menu labeled 'Root', indicating that this LCS is for a word with root causality. If, for example, the LCS

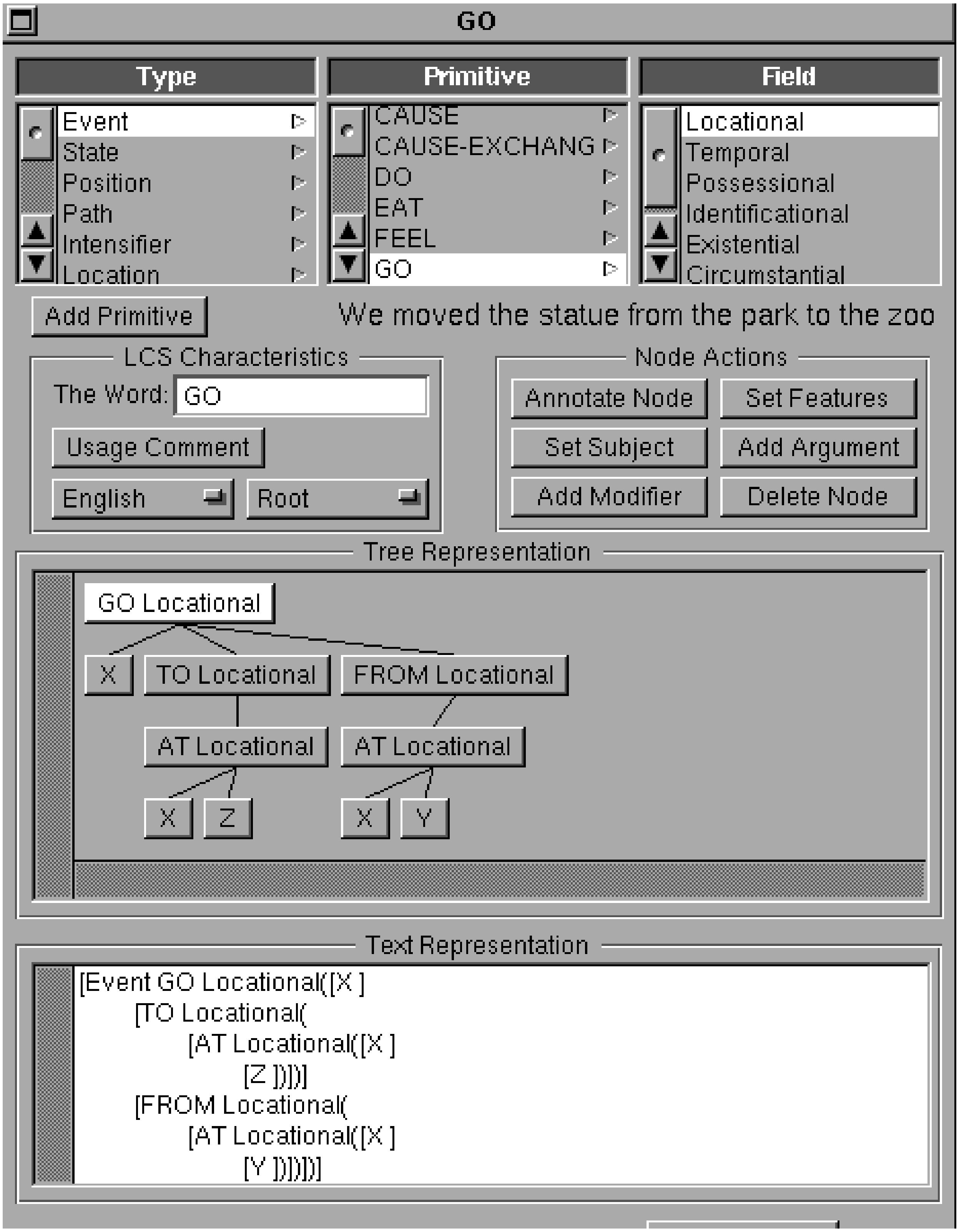

Figure 11: LCS Entry Window

were representing the verb, 'shove', then this popup would have been set to 'Causative'.

Next to the group labeled 'LCS Characteristics' is another group of controls labeled 'Node Actions'. These items apply solely to the node that is currently selected. In the current example, the root node of the tree is selected, so any of the 'Node Actions' items that are chosen would apply to the root node. The selection of a node action would open up a secondary window (palette) that supports the annotation of LCS items with classifying markers, semantic and syntactic feature information, and usage comments. If the tree grew large enough, vertical and/or horizontal scrollers would appear to allow the author to see different regions of the tree.

In addition to entering words for English and Korean, the LCS Editor is currently being used to enter vocabulary into the dictionaries for Spanish and Arabic.

C Translation of Military Message Corpus into Korean

REQUEST STATUS ON CLASS 1 ITEMS 103 FSB.

1kup co.hang 103 cen.wi.ci.wen.tay.tay.ey kwan.han sang.hwang yo.cheng.

WE CURRENTLY SHOW ZERO STATUS ON COMMANDERS REPORT.

hyen.cay wu.li.nun ci.hwi.kwan po.ko hyen.hwang.ey kwan.han han yeng.wi.ey.iss.ta.

LTC LEEH, XO DISCOM.

LEEH yuk.kwun.cwung.lyeng, sa.tan.ci.wen.sa.lyeng.pwu hayng.ceng.kwan.

PLEASE UPDATE 103 FSB, AND SEND IT TO THIS LOCATION.

103 cen.wi.ci.wen.tay.tay.ey kwan.han choy.kun ceng.po.lul swu.cip.ha.ye, ku.kes.ul i.kos.u.lo po.nay tal.la.

NEED AN UPDATED COMMANDERS REPORT.

choy.kun.uy ci.hwi.kwan po.ko.ka phil.yo.ha.ta.

SPEEDY, CAN WE GET A COMM REP ON ALL BDE?

Speedy, mo.tun ye.tan.uy (ci.hwi.kwan po.ko - thong.sin ung.tap) et.ko.siph.ta.

WAS THIS PROBLEM SELF-INFLICTED?

i mwun.cey.nun ca.sayng.cek.in kes.in.ka?

SEND REPORT AGAIN, HAVING TROUBLE ENTERING THE DATA BASE.

po.ko.lul.ta.si hay.tal.la tey.i.tha.pey.i.su.ey mwun.cey kyess.ta.

COMMO IS UP. PLEASE SEND A CURRENT CMDRS REPORT ON ALL UNITS.

thong.sin.i i.cey ka.nung.ha.ta. mo.tun pwu.tay.ey hyen.cay ci.hwi.kwan po.ko.lul.po.nay.la.

LOST YOUR REPORT. PLEASE SEND AGAIN.

po.ko.lul pwun.sil.hayss.ta. ta.si po.nay.la.

DID YOU RECEIVE OUR LAST REQUEST? WE NEED AN UPDATED OR CURRENT CMDRS REPORT ASAP.

wu.li.uy choy.kun yo.cheng.ul swu.sin.hayss.nun.ka? wu.li.nun choy.kun hok.un hyen.cay.uy ci.hwi.kwan po.ko nay.yong.ka ka.nung.han.han ppal.li phil.yo.ha.ta.

THE ONLY UNIT WE HAVE A CMDR REPORT ON IS THE 149, WE ARE NOT IN CONTACT VIA MCS WITH THE 67, 69, OR 1-167.

ci.hwi.kwan po.ko.ka ip.swu.toyn yu.il.han pwu.tay.nun 149i.ta. 67, 69, ku.li.ko 1-167kwa.nun ki.tong.cey.e.si.su.theym.ul thong.han cep.chok.i toy.ko iss.ci anh.ta.

WHO IS IN CONTACT WITH 67, 69, AND THE 1-167 IF YOU AREN'T?

67, 69, ku.li 1-167kwa cep.chok.ul kwi.ha.ka a.ni.la.myen nwu.ka ha.ko iss.nun kes.in.ka?

WE'VE TALKED TO THEM ON THE PHONE TO GET UNIT LOCATIONS, BUT BEYOND THAT WE HAVE NOTHING ON THEM.

wu.li.nun pwu.tay.wi.chi.lul al.ki.wi.hay ku.tul.kwa cen.hwa.lo kyo.sin.hayss.ta. ha.ci.man ku i.sang.un han il.i eps.ta.

REFERENCE TO YOUR REQUEST FOR CURRENT CMDS REPORT. WILL SEND BY 0600.

kwi.ha.uy hyen.cay ci.hwi.kwan.po.ko yo.cheng.ey kwan.han cham.ko ca.lyo.i.ta. 0600Kka.ci po.nay.cwu.keyss.ta.

WE ARE IN THE PROCESS OF FULFILLING YOUR REQUEST. WILL SEND WHEN COMPLETED.

wu.li.nun hyen.cay kwi.ha.uy yo.cheng.ul swu.hayng.ha.ko iss.nun cwung.i.ta. wan.lyo.toy.nun tay.lo po.nay cwu.keyss.ta.

CONTACT BY LAN WITH DIVARTY ACCOMPLISHED AT 0706.

ci.yek.thong.sin.mang.u.lo 0706ey wan.swu.han pho.pyeng.tay.wa cep.chok.ha.la.

PLEASE UPDATE CMDR'S REPORT FOR 230600FE91.

230600FE91ey kwan.han choy.kun ci.hwi.kwan po.ko.lul al.lye.tal.la.

DO YOU HAVE ANY INFO FOR US YET?

wu.li.ey.key al.lye.cwul e.tten ceng.po.la.to kat.ko iss.nun.ka?

IF YOU HAVE TROUBLE RETRIEVING THE IBP FILE WE JUST SENT, PLEASE CONTACT YOUR MC REP. FOR ASSISTANCE.

wu.li.ka pang.kum.po.nayn IBP hwa.il.ul kem.sayk.ha.nun.tey e.lye.wum.i sayng.ki.myen MC REPwa cep.chok.ha.ye to.wum.ul et.u.la.

REQUEST FOR CLARIFICATION FORM G5 X CORP THE LEGAL JURISDICTION ON A SOLDIER BEING HELD BY LOCAL GERMAN POLICE FOR ATTEMPTED ROBBERY AND MURDER.

G5 CORPlo.pwu.the.uy sang.hwang.ul myeng.hwak.hi ha.ki.wi.hay kang.to.sal.in mi.swu.lo tok.il kyeng.chal.ey cap.hye.iss.nun han pyeng.sa.ey kwan.han sa.pep.kwen.ul yo.cheng.ha.la.

WHAT ACTIONS, IF ANY, ARE REQUIRED?

phil.yo.ha.ta.myen, e.tten hayng.tong.ul chwi.hay.ya ha.nun.ka?

PLEASE RESPOND TO 35 DISCOM IF YOU HAVE RECEIVED THIS MESSAGE AT NODE CD.

kwi.ha.ka i so.sik.ul NODE CDey.sey swu.sin.hayss.ta.myen, 35 sa.tan.ci.wen.sa.lyeng.pwu.ey ung.tap.ha.la.

PLEASE SEND PREVIOUS QUERY TO 35 DIVARTY CAN NOT REACH ON LAN.

i.cen.uy uy.mwun.sa.hang.ul ci.yek.thong.sin.mang.u.lo.nun kyo.sin pwul.ka.nung.han 35 sa.tan.pho.pyeng.tay.ey po.nay.la.

HAVE RECEIVED ACK FROM YOUR NODE. HAVE NOT RECEIVED ACK OR MSG ABOUT FROM 35 G-3 OPS. IS OS OPERATIONAL AT G-3?

kwi.ha.uy NODElo.pwu.the.uy cep.swu.lul thong.po.pat.ass.ta.35 G-3 cak.cen.ey kwan.hay.sen cep.swu thong.po.na ki.tha.yen.lak.ul pat.ci mos.hayss.um.

CPT JACKSON

tay.wi JACKSON

MCS TEAM LEAD

ki.tong.cey.e.si.su.theym.cang

CMDR REPORT FOR THE 6 BDE.

6ye.tan ci.hwi.kwan.po.ko

69TH, WE GOT YOUR MESSAGE CMDR REPORT FOR THE 69 BDE, BUT WE DIDN'T GET THE ACRUAL REPORT, EITHER TRY AGAIN OR TURN ON YOUR AUTO LIGHT.

kwi.ha.lo.po.the.uy thong.po 69ye.tan ci.hwi.kwan po.ko.lul pat.ass.um. ku.le.na sil.cil.cek.in po.ko.lul pat.ci mos.hayss.um, ta.si po.nay.cwu.ke.na a.ni.myen kwi.ha.uy AUTO LIGHTlul khyel.kes.

SEND US A CMDR'S REPORT.

ci.hwi.kwan po.ko.lul po.nay.tal.la.

REQUEST COMMANDER REPORT CURRENT FROM DIVARTY UNITS WITHIN YOUR AREA.

kwi.ha ci.yek.uy sa.tan.pho.pyeng pwu.tay.lo.pwu.the.uy hyen.cay ci.hwi.kwan po.ko.lul yo.cheng.

THE ONLY INFORMATION IN THIS TERMINAL IS THE 70 FA BD. ALL OTHER DIVARTY UNITS HAVE NOT BEEN INPUT INTO THIS MACHINE AND WITH THE LIMIT OF MY KNOWLEDGE AT THIS TIME, I DON'T KNOW WHERE TO START, BUT WILL TRY TO DO THE VERY POSSIBLE BEST TO COMPLY WITH YOUR WISHES, SWEET PEA.

i the.mi.nal.ey.se.uy yu.il.han ceng.po.nun 70 ya.cen.pho.pyeng.ye.tan.ey kwan.han.kes.i.ta. ta.lun mo.tun sa.tan.pho.pyeng.tay.tul.ey kwan.han ceng.po.nun i ki.kyey.ey ip.lyek.toy.e.iss.ci anh.ta. hyen.cay.lo.se.nun a.nun.pa.ka eps.e.se e.ti.se si.cak.hay.ya hal.ci mo.lu.keyss.ta. ha.ci.man kwi.ha.uy ki.tay.ey pwu.ung.ha.ki.wi.hay choy.sen.ul ta.ha.keyss.ta.

REQUEST ALL NBC REPORTS.

mo.tun hwa.sayng.pang po.ko.lul yo.cheng.han.ta.

LAST MSG?

choy.kun thong.po.nun?

FSE NEED A GUMBALL REPORT ON ALL DIVARTY UNITS.

hwa.cay.ci.wen.phyen.tay.nun mo.tun sa.tan.pho.pyeng.tay.ey kwan.han Gullball po.ko.lul phil.yo.lo han.ta.

DTAC HAS JUST RECEIVED 3 MESSAGES ABOUT AIR MISSION DET REPORTS AND 2 MESSAGES ABOUT UNIT LOCATIONS FROM THE 135 MI BN THAT ARE 2 TO 3 HOURS OLD. BE CAREFUL WHEN FORWARDING THESE MESSAGES.

sa.tan.cen.swul.pwu.nun 135 ceng.po.tay.tay.lo.pwu.the twu.sey.si.kan ceng.to ci.nan myech.ka.ci thong.sin.ul pat.ass.nun.tey, kong.cwung.cen im.mwu po.ko.ey kwan.han say.ka.ci thong.sin.kwa pwu.tay wi.chi.ey kwan.han twu.ka.ci thong.sin.i.ta. i.tul thong.sin.nay.yong.ul cen.song.hal.ttay cwu.uy.lul yo.han.ta.

ON THE RADIO I HEARD SOMEONE AT NB 440191. WHO IS THAT?

la.ti.o.sang.ey.se NB 440191uy nwu.kwun.ka.uy mok.so.li.lul tul.ess.ta. nwu.kwu.in.ka?

DID THE BDE CDR RETIRE THE FLAG FOR 2/635 AR? REQUEST FROM PREVIOUS BN CDR FOR THE FLAG HAS BEEN SUBMITTED.

ye.tan.ci.hwi.kwan.i 2/635 ki.kap.pwu.tay.uy ki.lul chel.hoy.say.khyess.nun.ka? i.cen tay.tay ci.hwi.kwan.u.lo.pwu.the ki.lul cey.chwul.hay.tal.la.nun yo.cheng.i iss.um.

COULD YOU CHECK?

hwak.in.hay cwul swu iss.nun.ka?

NEED VERIFICATION ON MESSAGE FROM X CORP ABOUT INFO FROM CEWI BEING BOGUS ABOUT 79 TD WILL NOT HIT IGB UTIL MID-NIGHT.

ca.ceng.Kka.ci.nun 79 thayng.khu.sa.tan.i tok.il.kwuk.kyeng.ul kong.kyek.ha.ci alh.ul kes.i.la.nun thong.sin.cen ceng.po.pwu.lo.pwu.the.uy ceng.po.ka he.wi.la.nun X CORPlo.pwu.the.uy thong.sin.ul hwak.in.ha.ko siph.ta.

PLEASE SEND YOUR LATEST COMMANDER'S REPORT.

choy.kun.uy ci.hwi.kwan.po.ko.lul po.nay.tal.la.

YOUR CMDRS REPORTS ARE THE BEST I HAVE SEEN ALL WEEKEND. ONE SMALL DETAIL IS ON THE SECOND PAGE, THE 1ST LINE OF THE PERSONNAL REPORTS. THERE ARE NUMBERS IN THE LAST TWO COLUMNS BUT THE PERSONELL TYPE ARE NOT SHOWN. ALSO WHEN YOU TRANSMIT THE REPORT YOU CAN PUT AD IN THE DB STATUS AND YOUR INFORMATION WILL GO INTO MY DATA BASE AUTOMATICALLY. THANKS AGAIN FOR THE GOOD WORK KEEP IT UP.

kwi.han.uy ci.hwi.kwan.po.ko.ka i.pen cwu.mal.ey cep.swu.toyn kes.cwung ka.cang hwul.lyung.ha.ta. han.ka.ci cwu.mok.hal sey.pwu.sa.hang.i 2 phey.i.ci.uy kay.in sin.sang.po.ko.lan ches.ccay cwul.ey iss.ta. ma.ci.mak twu.nan.ey.nun swus.ca.nun na.wa.iss.u.na kay.in.yu.hyeng.i na.wa.ayss.ci.anh.ta. ku.li.ko kwi.ha.ka po.ko nay.yong.ul cen.song.hal.ttay pang.kong.ey kwan.han kes.un DB Statusey neh.u.myen na.uy tey.i.tha.pey.i.su.ey ca.tong.cek.u.lo ku ceng.po.ka tul.e.ol kes.i.ta. aph.u.lo.to hwul.lyung.hi im.mwu.swu.hayng hay.cwu.ki pa.lan.ta.

WE AT THE A2C2 ARE RECEIVING NO COMBAT STATUS REPORTS FROM THE AV. BDE. PLEASE INCLUDE THIS STATION FOR THOSE REPORTS. LAST STATUS UPDATE WAS 1200 LOCAL. WE NEED THOSE COMBAT STATUS REPORTS BY UNIT BREAKDOWN.

wu.li.nun A2C2 kong.kwun ye.tan.u.lo.pwu.the a.mwu.len cen.thwu.sang.hwang.po.ko.lul pat.ci mos.hayss.ta. wu.li stationu.lo ku po.ko.ka cep.swu ka.nun.hay.ci.to.lok co.chi.hay.tal.la. ka.cang choy.kun po.ko.nun 1200 Locali.ess.ta. wu.li.nun ku cen.thwu sa.hwal.po.ko.lul Unit BreakdownKka.ti.nun ip.swu.hay.ya han.ta.

REMEMBER THAT LOG SPOT ARE DUE BY, 1-161, 1-168, 2-138, 1-127TH TO THEIR FSB BY 1930 AND INFO COPY TO S4 35T DIVARTY.

ta.si al.li.nun pa, wi.chi.po.ko.lul so.hi.hay.tal.la. 1-161, 1-168, 2-138, 1-127 kak.ca.uy cen.wi.ca.wen.tay.tay.lo 19si 30pwun.Kka.ci, ku.li.ko ceng.po pok.sa.to S4 35T sa.tan.pho.pyeng.tay.tay.lo sok.hi.hay.tal.la.

RECEIVE BY SPC HOKE, 1914 23FEB91.

SPC Hokeey uy.hay 1991nyen 2wel 23il 19si 14pwun swu.sin.

PLEASE SEND CURRENT LOCATIONS OF BNS. THANKS FOR YOU SUPPORT.

tay.tay.uy hyen.cay.wi.chi.lul al.lye.tal.la. ci.wen.hay cwu.e.se ko.map.ta.

DO YOU HAVE THE CURRENT FRONT LINE TRACE? IF SO PLEASE SEND.

cen.wi pwu.tay.uy hun.cek.ul pha.ak.ha.ko iss.nun.ka? al.ko iss.u.myen al.lye.tal.la.

DID YOU GET LAST MESSAGE? IF NOT HERE IS A REPORT. WE WANT TO KNOW IF YOU HAVE THE CURRENT FRONT LINE TRACE IF SO PLEASE SEND.

pang.kum thong.sin.toyn kes.ul swu.sin.hayss.nun.ka? swu.sin mos hayss.ta.myen ci.kum ta.si po.nay cwu.keyss.ta. kwi.ha.ka cen.wi.pwu.tay.uy hun.cek.ul pha.ak.ha.ko iss.nun.ci al.ko siph.ta. al.ko iss.u.myen al.lye.tal.la.

REQUEST GRID OF ALTERNATE POSITION FOR CHEMICAL MUNITIONS ASAP. WE DO NOT HAVE EARLIER MESSAGE LISTRING GRID.

ka.nung.han han ppal.li hwa.hak kwun.swu.phwum kyo.lyu gridlul yo.cheng.han.ta. ci.nan.pen mey.si.ci.lul yel.ke.ha.ko.iss.nun gridlul wu.li.nun ka.ci.ko iss.ci.anh.ta.

WHAT ARE YOUR CURRENT LOCATIONS FOR 149 BDE? ARE THEY MOVING?

149ye.tan.uy hyen.cay.wi.chi.ka e.ti.la.ko sayng.kak.ha.nun.ka?
ku.tul.i i.tong.cwung.in.ka?

PLEASE SEND NEW LOCATIONS OF MLRS PLATOONS OF 2-675.

2-675 ta.pal lo.kheys si.su.theym so.tay.uy wi.chi.ka e.ti.in.ci
al.lye.tal.la.

NEED TIME AND LOCATION FOR CLASS IV CCL I ASAP.

ka.nung.han han ppal.li class IVuy CCL Iuy si.kan.kwa wi.chi.lul al.ko
siph.ta.

SEND PRESENT LOCATION AND DIRECTION AND SPEED OF MOVEMENT, IF ANY.

ka.nung.ha.ta.myen hyen.cay.uy wi.chi.wa i.tong pang.hyang ku.li.ko
sok.to.lul al.lye tal.la.

PLEASE SUBMIT CENTER OF MASS FOR HQ 69 BDE, HQ 67 BDE, HQ 149 BDE.

69 ye.tan pon.pwu, 67 ye.tan pon.pwu, ku.li.ko 149 ye.tan pon.pwu.uy
center of masslul cey.chwul.hay tal.la.

PLEASE SEND GRID CORD. OF CENTER MASS OF DE. AND GIVE DIRECTION AND
SPEED OF MOVEMENT. ACKNOWLEDGE WHEN RECEIVED.

DE.uy center massuy Grid CORD.lul po.nay.tal.la. ku.li.ko i.tong
pang.hyang.kwa sok.to.to al.lye tal.la. swu.sin.si hwak.in ung.tap.hal
kes.

REQUEST INFO AND LOCATION OF ALL 9TH TD ASSETS AS WELL AS ASSETS OF
THE 79TH TD. THESE UNITS APPEAR TO HAVE FLOWN FORWARD OF LAST
POSITION.

9 thayng.khu sa.tan.kwa 79 thayng.khu sa.tan.ey tay.han ceng.po.wa ku wi.chi.lul al.ko siph.ta. i.tul pwu.tay.tul.un ci.nan.pen.po.ta te cen.pang.u.lo i.tong.han.kes kath.ta.

STILL WAITING ON LOCATIONS OF FRIENDLY UNITS.

kyey.sok a.kwun.uy wi.chi.lul ki.ta.li.ko iss.nun.cwung.i.ta.

NEED EXACT LOCATION AND ATTITUDES OF TAB E/161 RADARS.

phyo.cek chwi.tuk pho.tay E/161 ley.i.ta.uy ceng.hwak.han wi.chi.wa ko.to.lul al.ko siph.ta.

NEED CURRENT LOCATION OF TP 25.

TP 25uy hyen.cay.wi.chi.lul al.ko siph.ta.

WHAT IS STATUS OF PREVIOUS REQUEST FOR CENTER MASS OF FA BNS AND HQ CP OF MANEUVER BDE?

cen.swul.ye.tan.uy pon.pwu.ci.hwi.cho.so.wa ya.cen pho.pyeng.tay.tay.uy Center Massey tay.hay ci.nan.pen.ey yo.cheng.han kes.un e.tteh.key toy.ness.nun.ka?

PLEASE PASS ABOVE MESSAGE ONTO DIVARTY S3. I CAN NOT REACH THEM ON THE LAN.

wi.uy Messagelul sa.tan.pho.pyeng.tay S3ey cen.hay.tal.la. wu.li.nun ci.yek.thong.sin.mang.u.lo.nun ku.tul.kwa thong.sin.i pwul.ka.nung.ha.ta.

MESSAGE RECEIVED ON FSB UNIT LOCATIONS.

cen.wi.ci.wen.tay.tay.uy wi.chi.ey kwan.han Message swu.sin.

CURRENT LOCATION AND STATUS OF 175 TK REG.

175TK yen.tay.uy hyen.cay wi.chi.wa sang.hwang.

CURRENT STATUS AND LOCATION OF 180 MRR. ANY INFO ON ENEMY ACTIVITY IN THE VICINITY OF NB4524.

180ca.tong.hwa.ki yen.tay.uy hyen.cay.wi.chi.wa sang.hwang NB 4524 kun.pang.uy cek.kwun hayng.tong.ey kwan.han ceng.po yo.mang.

REQUEST CURRENT STATUS AND LOCATION FOR 12 MRR, 120 GMRD.

12 ca.tong.hwa.ki yen.tay, 120 ca.tong.hwa.ki pang.wi sa.tan.uy hyen.cay.wi.chi.wa sang.thay.lul al.ko siph.ta.

ACKNOWLEDGEMENT NEEDED.

swu.sin hwak.in yo.mang.

REQUEST ASPS TO ADD NAI'S FOR THE 149BDE TO GIVE THEM MORE INFORMATION.

te manh.un ceng.po.lul cey.kong.ha.ki.wi.hay 149ye.tan.ey NAI'slul po.kang.hal kes.ul kwun.swu.phwum ci.wen.tay.ey yo.cheng.han.ta.

REQUEST LOCATIONS FOR THE FOLLOWING UNITS:

ta.um pwu.tay.uy hyen wi.chi.lul al.ko siph.ta.

FROM: MAJ SHAIN ANTICIPATED MOVEMENT OF THE 79TH TANK DIV INTO EA SATURN MAKES SITUATION IMPERATIVE THAT INFORMATION ON MOVEMENT OF ALL UNITS OF 79TH TANK DIV BE FIRST PRIORITY. REQUEST ADDITIONAL EMPHASIS ON HIGHWAY 62 BETWEEN IMMELBORN (NB 9028) TO VACHA (NB 7231). ALSO ALONG THE ROADWAY BETWEEN SCHMALKALDEN (PB 022) THROUGH DERMBASH (NB 7414) THROUGH GEISA (NB6814). REQUEST PRIORITY ON MTI'S. ALSO, REQUEST

INDICATORS OF ENEMY MECH FORCES MOVING BETWEEN EISENACH (NB 9248) TO BAD HERSFELD (NB 5036).

Shainso.lyeng.u.lo pwu.the: yey.sang.toy.ess.ten EA Saturnu.lo.uy 79thayng.khu.sa.tan.uy i.tong.un sang.hwang.ul kup.pak.ha.key man.tul.ko iss.nun.pa, mo.tun 79thayng.khu.sa.tan pwu.tay.uy i.tong.ey kwan.han ceng.po.lul choy.wu.sen.u.lo ta.lwu.e.ya.han.ta. Immelborney.se Vachasa.i.uy 62pen ko.sok.to.lo.ey te.wuk cwung.cem.ul twul kes.ul yo.cheng.han.ta. Schmalkaldeney.se Dermbach, GeisaKka.ci.uy to.lo.ey.se.to yek.si cwung.cem.ul twul.kes. MTI'slul choy.wu.sen.u.lo ta.lwul.kes.ul yo.cheng.han.ta. tto.han Eisenachey.se Bad Hersfeldey.se i.tong.cwung.in cek.kwun ki.kyey.hwa pwu.tay.uy Indicatorlul yo.cheng.han.ta.

I HAVE NOT RECEIVED ENEMY POSITIONS OR ELEMENTS. CAN YOU SUPPLY THAT INFO SO I CAN PLOT ON GRAPHIC?

cek.kwun.uy wi.chi.na yo.so.ey kwan.han yen.lak.ul a.cik pat.ci mos.hayss.um. nay.ka Graphicul hal swu iss.to.lok hay.tang ceng.po.lul po.nay.cwul swu iss.nun.ka?

PLEASE FORWARD THE MESSAGE ABOUT THE MOVEMENT OF THE 79TH TANK DIV.

79thayng.khu sa.tan.uy i.tong sang.hwang.ey tay.han mey.si.ci.lul cen.song.ha.la.

NEED LOCATION OF EPW COLLECTION POINT FOR BRIGADE.

ye.tan.uy cek.kwun pho.lo swu.yong ci.cem.uy wi.chi.lul al.ko siph.ta.

REQUEST LOC OF 79 HQ, ALL REGT. HQS, AND ALL RECON UNITS, AND UNIT STRENGTHS.

79thayng.khu sa.tan pon.pwu, mo.tun yen.tay pon.pwu, ku.li.ko mo.tun swu.sayk.pwu.tay.uy wi.chi.wa pyeng.lyek.ul al.ko siph.ta.