An Approach to Paraphrase in MT Systems Based on a Study of Chinese Verbs

Bonnie Chiu
IBM Los Angeles Scientific Center,
2525 Colorado Ave, Los Angeles, CA 90404, and
UCLA Department of Linguistics
Paula Newman
IBM Los Angeles Scientific Center
Shelley Smith
IBM Los Angeles Scientific Center and
USC Department of Linguistics

Abstract

In contrast to translation between related languages, translation among different language families often involves radical transformation of lexical items and their encompassing syntactic structure. As a result, the types of mechanisms involved in this transformation become a central concern. In this paper, we propose a generalized approach to lexically triggered structural transformations based on semantic features of lexical items, and relationships among lexical items, both encoded in the lexicon. The approach is motivated by a study of the requirements for translation from English to Chinese motion and resultative verbs.

1. Introduction

In order to improve translation quality and cope with the large volume of lexical and syntactic information required, many contemporary research efforts in translation divide the translation process so as to restrict the amount of bilingually dependent work required. These systems analyze inputs into relatively neutral forms, after which they may perform some explicit transfer, and from the result, generate target structures. The analysis stage of this process is the source of much debate in the field, with one of the main issues being the optimal depth of analysis. However, whatever position is taken on this issue, and whether or not an 'interlingua' is used, any translation system must deal with the problems posed by the following facts:

- In any language there are often many ways of saying the same thing, in terms of both lexical selection and structure.
- Often some or all of those ways will not closely correspond to the choices available in some other language.

These facts imply the need for paraphrase - transformations involving both word-senses and the structures (at any representational depth) in which they are embedded. In translation systems developed to deal only with closely related languages, the translation

process is assumed to be essentially compositional, with the division of lexical material among constituents roughly the same in both source and target, and with paraphrase requirements either ignored or treated as special cases. In translating among less closely related languages, however, paraphrase facilities must be more central, and a more systematic approach is required.

In this paper we suggest a direction for handling significant classes of paraphrase requirements in a more principled way, based on a study of translation to Chinese resultative and directional verbs. This work grew from efforts to extend the design of the LASC/MT machine translation system, from one for translation between and among English and Romance languages, to one with more general capabilities.

We begin by developing abstract rules for translation to Chinese resultative and directional verbs based on an analysis of general semantic features associated with word-senses. We then suggest how those rules might be incorporated in a machine translation system, taking as a concrete setting the Machine Translation system currently being developed at the IBM Los Angeles Scientific Center (LASC/MT).

2. Abstract Rules for Translation to Chinese Verbs

We begin the analysis by considering the following sentences:

```
1. He went.     ta qu  le.
                he go (past)

2. He went up.  ta shang  lai/qu  le.
                he top   come/go (past)
```

In (1), the translation appears quite straightforward, with 'went' translated as 'go+past'. When we add 'up', however, the sentence is quite different in Chinese. In English, the meaning expressed by 'up' in (2) appears as a preposition; in Chinese, not only is this expressed by a verb, 'shang', rather than a preposition, but it is also the main verb of the sentence. ('Lai' or 'qu' is added depending on whether the action is toward or away from the speaker.) At first glance, we might propose a surface treatment for the translation in (2), an 'Up-rule', which converts the preposition 'up' into a main verb 'shang'. However, it becomes apparent that this is not an adequate solution when we consider further examples such as the following:

```
3. He ran up.            ta pao shang lai/qu.
                         he run top  come/go

4. He ran up the hill.   (a)  ta pao shang shan   lai/qu.
                              he run top   hill   come/go

                         (b)  ta pao dao    shan-shang  lai/qu.
                              he run arrive hill-top    come/go
```

(3) is very similar to (2) in that we have the verb 'shang' appearing again, but in this case it is not the main verb (see note 1). To account for these facts, we could allow the 'Up-Rule' to apply, and add another rule to prevent 'shang' from being the main verb in this particular case. Then we come to (4). Even though the ambiguity is very subtle, this sentence is ambiguous between: (a) a 'path' reading, in which someone was simply taking an 'up-the-hill' course, without necessarily reaching the top, and (b) a 'goal' reading, in which this person actually reached the top of the hill. These two meanings are encoded differently in Chinese as (4a) and (4b) respectively. To account for both interpretations, further revision to the 'Up-Rule' is needed. Sentence (4a) can be obtained using much the same resources as those used for (3). For (4b), one might propose an alternative process which converts the 'up' of the PP 'up the hill' into 'shang' of the NP 'shan-shang', and adds 'dao' to 'shan-shang'.

Although it might be possible to continue to develop rules for each word/context combination, one should be skeptical about any approach in which the rules have to be modified or supplemented each time a new sentence is encountered. Moreover, with such word-by-word rules, it would be very difficult to accommodate cases where there is no one-to-one correspondence of lexical items between source and target language. Consider:

5. He ran across the street.	ta pao dao lu duimien	
	he run arrive street opposite-side	

'Across' encodes not only the path of the motion, but also the final goal of the motion. The Chinese counterpart, i.e. the 'dao-phrase', encodes only the goal aspect, and the path is not realized at all. Moreover, while the end point of the goal need not be explicitly stated in English, it must be in Chinese. In the approach outlined above, there is no principled way to account for these facts.

Semantic Features

It is apparent from the problems mentioned above that a more detailed analysis of the differences between the source and target languages is needed to avoid an unmanageable proliferation of rules and of reconciliation among rules. Moreover, this analysis must be based on a consideration of word-senses, not words, since a word can have various senses which impose different restrictions on their realizations in Chinese. Working within these premises, we will outline our analysis of motion verbs, inspired by the work of Talmy (1985), and then our analysis of causative/resultative verbs.

Based on the classification given in Talmy (1985), we define a motion verb as one expressing a situation involving movement of an object (see note 2).

These verbs are marked for the [Motion] feature. In addition, we also postulate certain semantic features which are either incorporated into the meaning of the motion verb, or

expressed within the same sentence, but separate from the motion verb itself. They are: [Manner], [Direction], [Goal], [Path] and [Location] (see note 3).

Assuming that these semantic features are shared by the languages under consideration, the task now is to identify generalizations about the way combinations of these features are realized in various languages. We can begin by looking at how semantic features are lexically encoded in English and how they are realized in Chinese.

Motion verbs express the motion meaning of the verb and can also express either the manner or the cause of the motion in both English and Chinese, as observed in Talmy (1985). Some examples are given below.

```
V[Motion, Manner]  ==> roll, swim,    walk, run...
                       guen, youyong, zou,  pao

V[Motion, Cause]   ==> pull, push, throw, kick...
                       la,   tui,  diou,  ti...
```

[Path], [Location] and [Direction] can also be encoded into the meaning of motion verbs, as shown by the following examples:

```
V[Motion, Direction]      ==> ascend, descend...
                              shang,  xia...

V[Motion, Path]           ==> traverse...
                              heng-guo...

V[Motion, Location]       ==> arrive, leave...
                              dao,    likai...
```

The conflation of a given feature with a motion verb is language-dependent. In English, the [Goal] feature is rarely conflated with the motion verb, but is instead usually encoded as a preposition. This strategy is widespread in English, and most of the semantic features proposed so far can be encoded as prepositions when they are used with a motion verb (see note 4).

```
P[Direction]  ==> up, down...

P[Path]       ==> across, through, under...

P[Location]   ==> at, in...

P[Goal]       ==> to...
```

In Chinese, these semantic features can be associated with the main verb of the sentence (as is the case in (1) and (2)), or as a formative compounded to the main verb (as in (3) and (4b)) (see note 5).

Rules for [Direction], [Path], & [Goal] Features

Using these features to analyze our earlier examples, the first generalization we can make is that [direction] can be realized as a main verb in Chinese. In example (2), 'up' expresses the [direction] feature in the English sentence. Its Chinese counterpart has conflated the [motion] and [direction] features into a single lexical item, namely 'shang'. 'Lai'/'qu' (come/go) explicitly encodes whether the motion is away from or towards the speaker, information that is usually missing in English (see note 6). We can sketch the rule converting English to Chinese as follows:

```
(A)   V[motion]   P[direction]  ===> V[motion, direction].
```

Roughly, rule (A) states that if there is a motion verb and a preposition expressing direction present in an English sentence, then in Chinese we must find a verbal lexical item in which the value of the [motion] and [direction] features are conflated. The motion verbs that undergo this rule are intransitive verbs. The argument of the preposition expressed in English will become the direct object of the main verb in Chinese (cf. example 2).

The structural description in Rule (A) does not apply to motion verbs which also incorporate the [manner] feature. In English, verbs of this type can combine with a prepositional phrase (PP) to express additional semantic features in a sentence, as illustrated in (4) and (5). In (4), 'up the hill' expresses the direction of the running, and in (5), 'across the street' encodes the path of the running. In the Chinese sentence (4a), we can analyze 'shang' (top) as a verbal component to the main verb 'pao'. 'Shan' (hill/mountain) can be regarded as the end point, or final destination of 'pao-shang' (run-up) (see note 7).

The second rule we propose deals specifically with this construction. Rule (B) states that the realization of the [direction] feature will be compounded with the realization of a main verb with [motion, manner] features. It accounts for the pattern in (3) and (4a).

```
(B)  V[motion, manner] P[direction]  ==> V[motion,manner]+V[direction]
```

Note that, both in Chinese and in English, it is also possible to express the [manner] feature as a constituent in a sentence, instead of conflating it with the main verb.

```
6.  He came running.   ta yong pao de lai.
                       he use  run DE come
```

In this example, the [motion] feature is expressed by the main verb in both languages, 'running' expresses the [manner] of the motion in English, and 'yong pao de'(see note 8) encodes the same [manner] feature in Chinese. This case is not treated further here.

A rule similar to rule (B) can be used in the treatment of examples such as the following:

```
 7. He ran to the school.  ta pao dao    xuexiao  lai/qu
                           he run arrive school   come/go
```

Here the English preposition encodes not [direction] but [goal], and the translation compounds a goal formative 'dao' with the main verb. This rule can be expressed as:

```
 (C) V[motion, manner]  P[goal]  ==> V[motion, manner]+V[goal]
```

Now let's consider the interpretation of example (4), repeated in (8) below, which involves the [goal] feature, together with example (9).

```
 8. He ran up the hill.   (a)ta pao-shang shan  lai/qu.
                             he run up    hill  come/go

                          (b)ta pao dao    shan-shang lai/qu.
                             he run arrive hill-top   come/go

 9. He ran across the street.     ta  pao dao    malu    duimien
                                  he  run arrive street  opposite-side
```

As mentioned earlier, there are two interpretations in (8): one interpretation of someone taking an 'up-the-hill' course without necessarily reaching the top, and another indicating that the hill-top has been reached. These two readings are realized differently in Chinese, as shown by the Chinese counterparts of (8), with (8a) having the 'up-the-hill course' reading, and (8b) the 'hill-top' reading (see note 9).

In our discussion above, we stated that 'up' encodes [direction] and showed that rule (B) can account for the pattern observed in (3) and (8a) (=4a). In the (8b) interpretation of the English sentence, 'up' has a different meaning than it does in (8a) as it includes the [goal] feature. The Chinese sentence (8b) encodes this feature as follows: The goal is expressed by the entire phrase 'dao shan-shang' (arrive hill-top); within this goal, 'shang' (hill) is the goal-object, 'shang' (top) is the point of the goal-object reached, and 'dao' (arrive) signals that this sentence has a goal reading. A similar phenomenon is seen in (9) where the 'dao-phrase' encodes the [goal] feature, with the goal-object being 'malu' (street). The position reached is 'malu duimen' (street other-side). Notice that there is no verbal component for 'across' in Chinese, as there is for 'up' which can be translated as the verbal component 'shang' (top). As a result, rule (B) will not be able to apply in (9). The only possible translation must include the 'dao-phrase'.

It is useful to divide the provisions needed to handle cases such as (8b) and (9) into two parts. First, the meaning of the preposition is paraphrased. Such paraphrasing is word-specific: thus the rule for the [goal] sense of 'up' would substitute 'to the top of' for 'up', and the rule for 'across' would substitute 'to the other side of'. It should be noted that the [goal] feature is retained in the substituted phrases. The presence of the [goal] feature enables rule (C) to apply and the necessary abstracted forms for Chinese are obtained. Finally, we derive the correct Chinese counterparts of the sentences by rules which spell out 'shang' (top) and 'duimien' (opposite-side) compounded to the nouns 'shan' (hill) and 'malu' (street) respectively, and compound the verbal component 'dao' (arrive) to the main verb.

Consider now the following sentences, to be contrasted with the above examples, which will need to be accounted for in a complete machine translation system.

```
 10. He crossed the street.      ta guo             malu
                                 he passed/crossed  street

 11. He ran across the street.   ta yong pao de guo  malu
                                 he use  run DE pass street

 12. He ran along the river.     ta ianzhe   he    pao
                                 he follow   river run
```

In (10), the [goal] feature is realized directly on the verb 'guo', so that the 'dao-phrase' is unnecessary. The goal-object 'malu' (street) is realized as the direct complement of 'guo'.

(11) and (9) are two different ways of expressing the same English sentence in Chinese. In (9), the main verb 'pao' (run) has the [manner] feature encoded in its verbal meaning, whereas in (11) the [manner] feature is encoded on the phrase 'yong pao de' (use run DE, literally 'using running') and instead of 'pau', 'guo', which does not carry the [manner] feature, appears as the main verb. Therefore, a more transparent English counterpart to (11) would be 'He crossed the street running' (see also (6) for a similar case).

In (12), 'along' encodes the [path] feature, but not the [goal] feature. As a result, there is again no 'dao-phrase' in Chinese, and the [path] feature is encoded on 'ianzhe he' (follow river).

For the purposes of this paper, we do not discuss rules relevant to these additional cases.

Rules for the [Cause] Feature

Moving now to another kind of construction, some motion verbs include the [cause] feature.

```
 13. I kicked John down the stairs.  wo ba John ti-xia    lou   qu.
                                     I  BA John kick-down floor go

                                     wo ba John ti-dao      lou-xia
qu
                                     I  BA John kick-arrive stair-down
go

 14. I threw the book on the table. wo ba shu  dio-dao      zuozi-shang
                                    I  BA book throw-arrive table-
on/top
```

In (13) and (14), the actions 'kick' and 'throw' have caused 'John' and 'the book' to undergo a certain motion with the result that John went down the stairs and the book landed on top of the table. Based on these facts, we will say that 'kick' and 'throw' can therefore include the [cause] feature. The results of the actions caused are expressed as constituents in the sentences. We will name verbs that have the [cause] feature encoded

in them 'resultative verbs', with the intended implication that these verbs are able to cause certain results to occur.

The Chinese counterparts of 'kick' (ti) and 'throw' (dio) can also include the [cause] feature. However, there are cases where the Chinese verb closest to the English source cannot have the [cause] feature. In these cases, a lexical item carrying the [cause] feature is added in the sentence, as illustrated by the following examples (see note 10).

```
15. The vase broke.       huaping puo   le
                          vase    break LE

16. I broke the vase.     wo ba huaping da-puo    le
                          I  BA vase    hit-break LE
```

In (15), the vase is broken, but the cause and causer are not specified, so the verb 'puo', which cannot have the [cause] feature, is used to encode the resulting state of the vase (that it is broken). In (16), where the causer is specified, 'da' is compounded to the verb for the purpose of encoding the [cause] feature. The rule accounting for this alternation can be written as follows:

```
(D). Vi[cause]    ==> V[cause]+Vi[-causative]
```

Rule (D) says that if a verb in English has the [cause] feature encoded and its Chinese counterpart doesn't, then a verb with the [causative] feature will have to be added to it. The verb that is added varies depending on the type of cause involved. For example:

```
17.  He kicked the vase and broke it. ta ti-puo     le neige  huaping
                                      he kick-break LE that   vase
```

When there is no other specific action encoded, 'da' (hit) appears as the default verb.

We have proposed semantic features which underlie sentences of motion verbs and resultative verbs, and have proposed four general rules which convert from English to Chinese patterns of encoding these features. By doing so, we have laid the foundation for achieving a tractable set of machine translation rules for these cases, and have identified a general method of obtaining other such rules. In the next section, we will discuss how the analysis proposed can be integrated into an experimental machine translation system.

3. Incorporating the Analysis into an MT System Design

The MT System

We first describe the computational context into which the above rules are to be incorporated, and then how the incorporation might be accomplished. The basic design of the LASC/MT system is semi-interlingual, and was arrived at mainly via a consideration of English and closely-related Romance languages. In this system, every language under consideration has a unilingual lexicon containing entries for stems, their morphological properties, and alternative syntactic context information (i.e., expected dependents of the

stem). Two-way links connect the unilingual lexicons to a common lexicon, which contains entries for the word senses/concepts (some shared and some unique) of all the languages involved.

A common-lexicon entry for a word-sense describes its expected dependents in abstract terms, its membership in a semantic domain lattice, connections to more specialized and general concepts, and paraphrase specifications. Specialization connections can be used when a word sense has no equivalent in a particular target language, but a more limited sense does have an equivalent, and its specified dependent types are consistent with those of the utterance. If such connections cannot be used, paraphrase specifications are employed.

The syntactic context information in the unilingual lexicons is expressed in terms of the syntactic forms of the expected dependents, and two-way mappings between those forms and more neutral forms consistent with those specified in the common lexicon for the for the associated concepts. For example, a mapping would be specified to delete/add the particle up from the English-specific form bring up to obtain the set of dependents expected by the common lexicon entry for bring-up.

The translation process uses the lexicons in the following way: First, morphological processing of the source is done to identify the stem of a word and associated affixes. The unilingual entries for that stem are modified by lexical transformation (e.g., for passive) before parsing takes place. The parser design is a lexicalist one inspired by HPSG (see Pollard and Sag 1987), Slot Grammar (see McCord 1989) and other related grammars. After parsing, the head/dependent structures are mapped to forms matching the common lexicon entries.

The primary function of the transfer phase is to perform specialization & generalization replacements, and paraphrase paraphrase transformations. Because of the close relationship between English and Romance (and other Indo-European) languages, many morphological and syntactic differences can be abstracted away during analysis, and the major function of the paraphrase transformations in translating among these languages is to account for lexical gaps. However, in translating among more widely separated language families, paraphrases must account for more specifically structural modifications, for example, to create or eliminate complex verb constructions in translating to and from Chinese.

Generation performs the head/dependent transformations specified in the target language lexicon, and then uses rules specific to the target language (independent of the source language) to further order, inflect, and otherwise modify constituents. For a more detailed description of the system, see Newman (forthcoming).

Integration of Motion & Resultative Verb Analysis

The analysis proposed above for motion and resultative verbs has specific consequences with respect to the formulation of paraphrasing rules in the above design and in other

designs for machine translation systems. In the initial design of the LASC/MT system, paraphrase rules accounting for lexical gaps were regarded as word-sense-specific, although some methods of generalization were explored. In expanding the design of the system to incorporate Chinese, it became obvious that generalizing lexically-triggered paraphrase rules was not an option, but a requirement (as it would be for any translation system involving more than one language family).

The analysis in the first section provides the basis for accomplishing this generalization. It suggests that many paraphrasing needs can be satisfied by associating a set of semantic features with common lexicon word-sense entries, and using them as triggers for paraphrase transformations. However, we need additional mechanisms to specify these transformations.

A major feature of the transformations in our analysis has been the apparent change in part-of-speech when going from one language to another. For example, in sentence (3), the preposition up becomes a verb-component. Such changes are in fact relevant for transformation to Romance languages as well as to Chinese. For example, the use of path prepositions, such as across, and into, is acceptable in languages such as English and some Latin-American Spanish variants, but not in others, such as French and European Spanish. Consider:

```
17. He ran across the street.  Il a   travers la  rue    en  courant.
                               He AUX crossed  the street via running
```

One way of handling this type of example is suggested by aspects of Mel'cuk's Meaning Text Theory (see Mel'cuk and Polgu, 1987, and Mel'cuk 1988) which includes provisions for lexical functions relating lexicon entries. Mel'cuk uses these functions for standard types of relationships, such as synonyms and antonyms, as well as for other relationships among lexical entries. These functions can be represented as attribute/value combinations within common lexicon entries, and can be referenced in generalized paraphrase rules. This would allow, for example, paraphrase rules for path-prepositions, such as are needed in (17), to be generalized to a single rule referencing a lexical function, e.g., v_path. The common-lexicon entry for each path preposition would contain an attribute/value pair v_path= X, with X being the name of the concept (labelling another common lexicon entry) representing the verbal equivalent. The single rule performing the required transformations would have the abstracted form:

```
    motion-verb + [path-prep + obj]  -->
        v_path(path_prep) + obj + manner-clause-headed-by-verb
```

where v_path(path-prep) indicates the value of the vpath attribute for path-prep.

It might be noted that another way to handle the across type of example is to use just one common lexicon sense for words which can be understood to have the same fundamental meaning but which occur in different parts of speech in different languages. The common lexicon entry for this sense would have sub-entries for alternative contextual expectations for the occurrence of the sense in different parts of speech. In this formulation, the sense of across and cross would be conflated in the common lexicon, and the rule would be

similar to that above except that the lexical function reference v_path(path_prep) would be replaced by the use of a more general transformation. This transformation would change the representation to conform to the needs of the new head category type.

To integrate generalized transformation rules into the lexicon, the lattice structure of the lexicon is utilized. Common lexicon entries are arranged in a domain lattice in which each word-sense may be associated with more than one superordinate domain. These domains allow lexical material to be inherited and they can represent semantic classes used in the contextual constraints. Much of the lexical feature information discussed in the earlier sections of this paper would be expressed as domain membership information in common lexicon entries. Transformation rules would be incorporated into this structure by associating them with the most general domain to which they might be applicable. If there are rules which must be tested for any occurrence of a particular constituent type (e.g., clause), they would be associated with the domain representing the syntactic category of the head of that constituent type (e.g., verb). Rules triggered by more specific word-classes would be associated with the domain representing those classes, e.g., motion verb, itself belonging to the domain verb.

The control mechanism for rule application is outside the scope of this paper but, in general, all rules are applied, with iteration after applying a rule, producing more than one transfer output. Outputs resulting from more specific rules are preferred to those resulting from more general ones.

We now consider how the specific rules developed for motion and resultative verbs can be integrated into this structure. The four rules proposed in the previous sections are the following:

```
(A)  V[motion]  P[direction ]  ==> V[motion, direction]
(B)  V[motion, manner]  P[direction]  ==>V[motion,manner]+V[direction]
(C)  V[motion, manner]  P[goal]  ==> V[motion, manner)+V[goal]
(D)  Vi[cause]  ==> V[cause]+Vi[-cause]
```

Before stating the precise equivalents of these rules in a form suitable to transfer processing, some additional explanation of the computational context is in order. The structures used as input to and output from transfer consist of nested frames, where the topmost frame represents the main clause, and the frame itself is a set of attribute-value pairs. One attribute of a frame is the sense of its head (sense =...). Frames also have derived attributes taken from the lexicon entries for the head sense. One is $domain, whose value is the list of domains of which the head sense is a member. Others are attributes representing various lexical functions (see below). For presentation purposes we will simplify the representation, and associated transformation rules by assuming that all dependents are core dependents, and are represented by predefined attributes, e.g., a1 = nested frame for canonical subject. In the actual representation, dependents are given within a list-valued attribute deps, and the rules are somewhat more complex.

The abstract rules stated above can be re-stated as conditional conversions of the transfer structures. Rule (A), for example, can be expressed as a rule associated with the domain of motion verbs:

```
Ch: cond
        (#in: ($domain ^a> 'manner',
               a2: (type = prep,
                    $domain > 'direction',
                    $domain ^a> 'goal',
                    $verbeq = #veq,
                    a1 = #dirobj));
    convert
        (#out <= #in).
        (#out: (sense <- #veq,
                a2 <- #dirobj);
```

This rule consists of two parts: conditions (cond) and transformations (convert). Cond specifies a list of conditions (here with a single element) which must be met by the input for the rule to apply. A condition specifies a set of unification and test operations to apply to a variable or components thereof. #in always represents the input frame, but other arbitrarily named variables may be set and/or tested as necessary. Within conditions, : indicates that the following operations are to be applied to the named variable or component. The operations =, ^a=, >, ^a>, mean respectively unify-with, doesn't-unify-with, contains, doesn't-contain. So the above condition states that the rule applies if the domain list for the verb does not contain 'manner', and that its second complement is a prepositional phrase where the domain list of the preposition contains 'direction' but not 'goal'. By unification also sets the variables '#veq' to the concept giving the verbal equivalent of the preposition, and '#dirobj' to the object of the preposition if any.

If a condition section succeeds, then the convert part of the rule is used to perform the required transformation. When all convert actions are executed, the value of the variable #out is taken as the result. In the example above the convert section consists of two actions. The first uses the assign operation <- to set the value of #out to that of #in. The second action modifies #out so that the head sense takes on the value of #veq, and the second complement is modified to take on the value of #dirobj, which may be null.

To discuss the expression of rule (B) we first note that at the transfer interface, Chinese compound verbs are represented as conjunctions of clauses with some common arguments, using some artificial conjunctions established for that purpose. For example, given a structure satisfying the input conditions for rule (B), i.e., a motion-manner verb with a specified direction, one would like to obtain a conjunction of clauses such as John run @ John go-up hill, where @ used to represent a common-subject conjunction. Chinese generation rules would then perform a further restructuring. Given these assumptions, rule (B) can be expressed by the following paraphrase rule, associated with motion-manner verbs.

```
Ch: cond
        (#in: (a2: (type = 'prep'
```

```
                    $domain ª> 'goal',
                    $domain > 'direction'
                    $verbeq = #veq,
                    a1 = #dirobj));
    convert
        (#newa1 <- #in).
        (#newa1: (a2 <- null)
        (#newa2 <- #in).
        (#newa2: (sense <- #veq,
                  a2 <- #dirobj)).
        (#out: (type <- 'compound'
                sense <- 'common-subject'
                a1 <- #newa1,
                a2 <- #newa2).
```

In other words, the first part of the resultant compound has the same subject and verb as the original, but no second argument. The second part has the same subject as the original, the verbal equivalent of the original preposition, and the object of the original preposition as its object.

Rule (C) when restricted to prepositions expressing pure goal ('to') can be expressed similarly. The cases of prepositions expressing [goal] plus [direction], such as across and up in its second meaning (as in example 4b) are handled by a combination of two kinds of rules. The first kind would be general, i.e., would be applied independent of target language - and might specify alternative transformations. For the sentence He ran across the street, these rules would generate such paraphrases as He crossed the street running and He ran to the other side of the street respectively. Following the application of these general paraphrase rules, adjustments needed for Chinese would apply - for He ran to the other side of the street, the formalization of rule (C).

We now turn to Rule (D) which adds an explicit cause to a clause if (a) the basic word-sense of the verb has the [cause] feature, (b) the interpretation is agentive rather than middle (he broke the vase rather than the vase broke), and (c) the only available Chinese realization of the main verb has a middle reading.

Here we also assume conjoined clauses at the interface between transfer and the Chinese generator, here using a cause-effect conjunction which shares the object of the first verb with the subject of the second, e.g., He hit the vase @ the vase broke. We also assume that the Chinese lexicon specifies only a single argument for break, and information on the link between the common lexicon (specifying a maximal number of dependents) and the Chinese lexicon would make a correspondence between common lexicon argument 2 for break, and Chinese lexicon argument 1, and no correspondence for common lexicon argument 1.

Given this formulation, an attempt to find a Chinese equivalent of a causative break would fail, because of the presence of a first argument, and the following paraphrase rule would then apply:

```
Ch: cond
```

```
   (#in: (a1 ª= null,
          $default-cause = #def).
convert
    (#newa1 <- #in).
    (#newa1: (sense <- #def).
    (#newa2 <- #in).
    (#newa2: (a1 <- null).
    (#out, (type <- 'compound',
            sense <- 'cause-effect',
            a1 <- #newa1,
            a2 <- #newa2).
```

The $default-cause attribute for break would be hit. It should be emphasized that the design for the addition of provisions for Chinese to the overall MT design is incomplete. While we have some confidence that the above rules are functionally correct, their precise formulation, as well as the precise partitioning of functions between transfer and generation for Chinese, are subject to change.

4. Summary

We have proposed a feature analysis for Chinese motion and resultative verbs, and abstract rules to account for differences in the realization of those features in English and Chinese. We then suggested a means for incorporating this analysis into a computational context originally designed for translation between and among English and Romance languages. Although the comparative study involved a small amount of data, and the incorporation of the resulting analysis is still at a preliminary stage, it has provided a new perspective on the overall design. Paraphrase rules were treated as word-specific in the original formulation. While this might be adequate for translation among related languages, it is less than optimal, and in other cases will result in considerable rule proliferation. The use of semantic features, as shown in this analysis of motion verbs and resultative verbs in Chinese, allows paraphrasing rules to be generalized. This enables us to integrate provisions for Chinese in a way consistent with the basic design, and also improves the handling of related phenomena in a more general translation context.

5. References

1. Chao, Yuen Ren (1968), *A Grammar of Spoken Chinese*, University of California Press, Berkeley and Los Angeles.
2. Li, C. and S. Thompson (1981), *Mandarin Chinese: A Functional Reference Grammar*, University of California, Berkely and Los Angeles.
3. McCord, Michael (1989), A New Version of Slot Grammar, IBM Research Report RC 1450.
4. Mel'cuk, Igor (1988), *Dependency Syntax: Theory and Practice*, State University of New York Press.
5. Mel'cuk, Igor and Alain Polguere (1987), A Formal Lexicon in Meaning-Text Theory (or How to Do Lexica with Words), *Computational Linguistics*, Vol 13, Nos 3-4.

6. Newman, Paula, (forthcoming), Symmetric Slot Grammar: A Bidirectional Design for MT.
7. Pollard, Carl and Ivan A. Sag (1987), *Information Based Syntax and Semantics,* Center for the Study of Language and Information.
8. Talmy, Leonard (1975), 'Semantics and Syntax of Motion' in J. Kimball, ed., *Syntax and Semantics,* Academic Press, New York.
9. Talmy, Leonard (1983),'How Language Structures Space' in H. Pick and L. Acredolo, eds., *Spatial Orientation: Theory, Research, and Application* Plenum Press, New York.
10. Talmy, Leonard (1985), 'Lexicalization Pattern: Semantic Structure in Lexical Forms' in T. Shopen, ed., *Language Typology and Syntactic Description*, Vol 3., Cambridge Uniiversity Press.

6. Notes

1. Many verbs in Chinese are of the form V1-V2, which we will call compound verbs following Li and Thompson (1981). Two of the basic semantic relations that can hold between V1 and V2 are resultative, with V2 being the resultative state of V1, and parallel, with both V1 and V2 being synonymous or signaling the same type of predicative notions. The following examples illustrate resultative verb compounds and parallel verb compounds respectively.

```
2.
3.      pull-open
4.
5.      zhi-liao    'to cure'
6.      cure-cure
```

   For further details, see Li and Thompson (1981).

7. Talmy defines a 'motion event' as 'a situation containing movement or the maintenance of a stationary location'. In this paper, we discuss verbs involving movement only. For further discussion, see Talmy 1975, 1983 and 1985.
8. The features employed here are not identical to those in Talmy (1985). However, the basic concepts are similar. It should also be noted that the features proposed here are approximations rather than precise definitions of the semantic concepts involved in the class of motion verbs under consideration. Precise definitions may be developed later, but they would have to be partly based on prescriptive criteria in order for them to apply consistently through various contexts.
9. Prepositions encoding the [manner] feature are rare in English. One possible candidate may be 'by' as in He explained it by drawing a map. However, the [manner] feature can be seen as being encoded on 'drawing a map', rather than on 'by', so 'by' in itself does not appear to encode [manner]. Furthermore, unlike other prepositions discussed in the text, 'by' here is not followed by a noun, but by a gerundive verb which can be regarded as directly contributing to the encoding of the [manner] feature. We will leave open the question of whether English has prepositions with the [manner] feature encoded in them. The issue of deciding which semantic features are present in certain contexts remains problematic.

10. The exact syntactic status of 'shang', 'xia' and similar words will not be discussed in this paper. We will simply refer to them either as the 'main verb' of a sentence, or when they occur attached to the main verb of a sentence, as a verb-compound/verb-component.
11. This information is almost always encoded in Chinese motion-event sentences. Although English doesn't usually encode this meaning, it can appear, conflated with the verb (e.g. 'come', 'go'...).
12. In Li and Thompson (1981), this construction is called a 'Directional Resultative Compound'. Chao (1968) calls 'shang' and other formatives occurring in this structure 'Directional Complements'.
13. Mandarin spoken in Taiwan uses yong... to express the [manner] feature of the motion. It is also used to express instrumental meaning, as shown by the following example.
14.
15. `He killed Z with a knife.  ta yong dao  sha-le  Zhangshan`
16. `he use  knife kill-LE  Zhangshan`
17. We should note that for native speakers of Chinese, (8a) and (8b) are often not considered semantically distinct from one another, which parallels the fact that both readings are associated with the English counterpart. As a result, we might want to claim that 'shang shan' in (8a) can be interpreted as having either the [direction] or the [goal] feature. Nevertheless, if we want to make the goal reading of sentence (8) explicit in Chinese, then the [goal] feature must be realized by the 'dao-phrase'.
18. 'Break' in English is ambiguous between a transitive and a middle reading:
19.
20. `John broke the glass.  (transitive reading)`
21. `The glass broke.  (middle reading)`

    Our [cause] feature distinguishes the two readings of 'break' (and verbs of the similar types: open, close...etc.). The [cause] feature is present in the transitive 'break', whereas it is not in the middle reading. It should also be noted that there are various definitions of the term 'middle' depending on different syntactic theories. In this paper, we use the term to refer to a resultative state in which the cause and causer are not explicitly or implicitly encoded.