Bi-Lexical Rules for Multi-Lexeme Translation in Lexicalist MT

Arturo Trujillo
SCMS, The Robert Gordon University
St Andrew Street, Aberdeen AB1 1HG
Scotland
iat@scms.rgu.ac.uk

Abstract

The paper presents a prototype lexicalist Machine Translation system (based on the so-called 'Shake-and-Bake' approach of Whitelock (1992)) consisting of an analysis component, a dynamic bilingual lexicon, and a generation component, and shows how it is applied to a range of MT problems. Multi-Lexeme translations are handled through bi-lexical rules which map bilingual lexical signs into new bilingual lexical signs. It is argued that much translation can be handled by equating translationally equivalent lists of lexical signs, either directly in the bilingual lexicon, or by deriving them through bi-lexical rules. Lexical semantic information organized as Qualia structures Pustejovsky (1991) is used as a mechanism for restricting the domain of the rules.

1 Introduction

Transfer based approaches to machine translation (MT) involve three main phases: analysis, transfer and generation. During analysis, the syntactic and semantic structure of a sentence is made explicit through a source language (SL) grammar and semantic processing modules. The result of analysis is one or more syntactic and semantic representations which are used to construct a syntactic and/or semantic representation in the target language (TL) through a series of transfer rules and a bilingual lexicon. From this representation a TL sentence is generated based on some form of mapping procedure, usually exploiting the TL grammar [1].

In this paper we describe a prototype implementation of a transfer MT system based on the lexicalist MT (LMT) approach of Whitelock (1992), also known as 'Shake-and-Bake' (SB). For our implementation we have extended the original SB formulation by postulating bilingual lexical rules (bi-lexical rules henceforth) which dynamically

[1] While this definition of transfer systems is current in most MT discussions, it has been challenged (Kay et al. 1994) on the basis that the interlingua-transfer distinction, that is, the distinction between systems which construct language independent representations and systems which do not, is artificial and that in fact the two paradigms simply represent different aspects of the same problem. While we agree with this observation, many systems at present start with an interlingua or a transfer architecture and then incorporate solutions from the alternative paradigm. We therefore maintain the distinction, at least for the purposes of this paper.

expand the bilingual lexicon in order to extend its functionality. This allows us to uniformly treat mono- and multi-lexeme translations in a variety of contexts.

We describe the main characteristics of the LMT approach. This is followed by a description of the problems posed by certain multi-lexeme translations, and of how bi-lexical rules, in conjunction with lexical semantic information provide a framework for overcoming these problems. We then point out some limitations in our approach and give some idea as to the status of our implementation.

2 Lexicalist Machine Translation

In its original formulation, LMT consists of three main phrases: analysis, lexical-semantic transfer and generation. The analysis phase involves parsing the input sentence to produce an output bag or multiset of SL lexical signs instantiated with sufficient information to permit appropriate translation. Transfer maps these signs into a TL bag through the bilingual lexicon in which sets of source and target lexical signs are placed in translation correspondence. Generation consists of finding an ordering of the TL bag which satisfies the constraints imposed by the TL grammar. Normally, generation involves a modified parser which ignores ordering information (Brew 1992; Popowich 1995) although other approaches are also possible (Poznański et al. 1995).

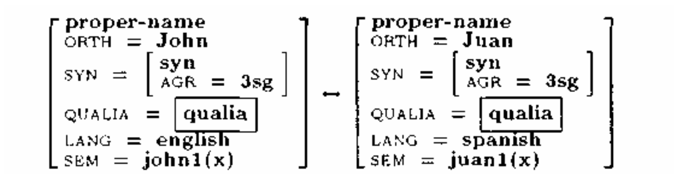
2.1 Notation

We introduce some notation through a simple example of our implementation. Since we will not be concerned with quantification nor scoping, we adopt a simplified transfer representation. If quantification and scope were to be included, however, a mechanism along the lines of Frank and Reyle (1995) and Copestake et al. (1995) may be followed in order to preserve the recursiveless nature of lexicalist transfer.

Our lexical signs broadly follow the signs of Pollard and Sag (1987) although our work seems adaptable to the signs of Pollard and Sag (1994), The implementation is based on the Typed Features Structures (TFSs) of the Acquilex LKB (Copestake et al. 1993) from where we borrow our notation. Consider the (simplified) lexical entry for ‘John’:

$$\begin{bmatrix} \textbf{proper-name} & \\ \text{ORTH} = \textbf{John} & \\ \text{SYN} = \begin{bmatrix} \textbf{syn} \\ \text{AGR} = \textbf{3sg} \end{bmatrix} & \\ \text{QUALIA} = \boxed{\textbf{qualia}} & \\ \text{SEM} = \textbf{john1(x)} & \end{bmatrix}$$

In this TFS, features are written in small capitals, while types are in bold face. To make TFSs easier to read, detail may be hidden by 'shrinking' a TFS; this is indicated with a box around the type of the TFS (e.g. **qualia** above). TFSs of type **qualia** encode lexical semantic information based on the Qualia structures of Pustejovsky (1991). For the semantic representation of proper names we assume a predicate treatment following the arguments of Devlin (1991:225). A bilexical entry for 'John – *Jua'* would be:



For reasons of space and convenience, we will abbreviate the above lexical sign and bilexical entry to

$\mathrm{john1}_x$
$\mathrm{johnl}_x \leftrightarrow \mathrm{juanl}_x$

respectively, where the subscripts correspond to the argument variable. It should be emphasised, however, that this abbreviated notation implicitly includes syntactic and semantic information which may be accessed during transfer or generation.

To exemplify LMT, consider the translation of 'John likes Mary'. Analysis results in a list[2] of lexical signs the semantics of which will contain shared variables:

$\mathrm{johnl}_x\ \mathrm{lovel}_{e,x,y}\ \mathrm{maryl}_x$

The (tenseless) FOL formula corresponding to this expression is $\exists exy.$ johnl(x) & lovel(e,x,y) & maryl(y), but since quantification and scope will be ignored they will be omitted from our examples; furthermore, coordination will be assumed between predicates unless otherwise stated.

Before transfer, a process similar to skolemization is applied to the transfer representation in order to replace variables by constants. The purpose of this operation is to prevent spurious bindings during lexicalist generation, as will become clearer later. The result of analysis is a list of lexical signs with translationally relevant relationships expressed by shared constants (indicated by integers in our notation):

$\mathrm{Johnl}_1\ \mathrm{lovel}_{2,1,3}\ \mathrm{maryl}_3$

The transfer step uses the source side of the bilexicon (possibly expanded by bilingual lexical rules as described below) to derive a total cover of the SL list (Garey and Johnson 1979:221) (a total cover is a

[2] We use lists of SL lexical items, instead of bags as is done in SB, to avoid certain inefficiencies caused by the nature of lexicalist transfer (Garey and Johnson 1979:221).

division of a set into a number of allowed subsets such that every element in the set is a member of exactly one subset; we extend the term here to apply it to lists). The bilexicon below enables construction of an appropriate TL bag:

johnl_x ↔ juanl_x
maryl_x ↔ maríal_x
love$l_{x,y,z}$ ↔ amar$l_{x,y,z}$ al_z

(Tense is omitted in this example; a simplistic model has been adopted in which an interlingua tense feature is passed from source to target verbs in the bilexicon.) Note that we include function words such as the Spanish case marker *a* in the bilingual lexicon (and therefore in the transfer representation). These words are treated as vacuous predicates (Calder et al. 1989) over the variable of the semantic head on which they depend. For the present example, transfer results in the following TL bag:

{juanl_1 , amar$l_{2,1,3}$, a_3 , maríal_3}

Lexicalist generation involves reordering the TL bag to construct a valid TL sentence. Since normally all permutations of the TL bag are attempted, the fact that variables are replaced by constants ensures that arguments not shared between predicates in the SL representation are not shared in the TL representation either. This prevents *María* from being the subject of the sentence. The result of generation, after morphological synthesis, is:

Juan ama a María

2.2 Other Properties of LMT

LMT encourages two useful properties: modularity and reversibility. From an engineering point of view, modularity is desirable because it can reduce development and maintenance costs. By using sets of lexical signs as their transfer representation, LMT systems can reduce the difficulties posed by structural mismatches between two languages, thus increasing the independence between source and target transfer representations. For example, transfer systems adopting a recursive representation for transfer (Kaplan et al. 1989), as opposed to a non-recursive one (Copestake et al. 1995), may need additional mechanisms for handling head switching (Kaplan and Wedekind 1993). By-contrast, under a lexicalist approach, head switching can be handled purely compositionally with minimal assumptions (Whitelock 1992). Reversibility is an important property in bi-directional systems as it reduces development costs. In LMT, grammars are fully reversible since they are used in similar ways for analysis and generation: the difference is that during lexicalist generation, ordering information is disregarded. However, the process is complete because the generator is guaranteed to generate all the strings accepted by the TL grammar which satisfy the constraints imposed by the TL bag. Lexicalist generation is also sound because only strings which satisfy the constraints of the TL grammar are constructed. In addition, termination is guaranteed if it is guaranteed for parsing since one can at worst construct a generation algorithm which simply attempts all permutations of the TL bag and then parses them in order to test whether they are appropriate TL sentences.

3 Multi-Lexical Translations

One of the reasons for transfer modules being expensive to construct is the presence of complex transfer relations (Arnold and Sadler 1992; Hutchins and Somers 1992). One type of phenomena that leads to complex transfer in a number of systems may be called multi-lexical translation. These are translations in which a phrase cannot easily be translated through the translation of its parts. The translation of idioms is an extreme case of this. For example, ‘kick the bucket’ translates as *estirar la pata* (Lit. ‘to stretch a leg’) in Spanish, even though there is no simple correspondence between the components of each phrase (all translations in this paper are between English and Spanish unless otherwise stated). For such constructions, structures corresponding to the source and target phrases need to be equated either in the transfer module (Schenk 1986) or in separate dictionaries (Sadler et al. 1990) in many systems. Other phenomena which may be loosely labelled multi-lexeme translations include: lexical gaps such as ‘piece of advice’ - *consejo* (Soler and Marti 1993); support verb and category differences such as ‘to be thirsty’ - *tener sed* (to have thirst) (Danlos and Samvelian 1992); lexicalization patterns like ‘swim across the river Dee’ - *cruzar el río Dee nadando* (Talmy 1985); conflational divergences as in ‘to stab someone’ - *darle puñaladas a alguien* (Dorr 1992).

Phenomena such as idioms, lexical gaps and conflational divergences can be tackled in LMT by equating sets of source and target lexical signs:

a) kickl $_{e,s,o}$, thel$_{o}$, bucket1$_{o}$ $\leftrightarrow$ estirarl$_{e,s,o}$, la1$_{o}$, pata1$_{o}$
b) piecel$_{x}$, ofl$_{x,y}$, advicel$_{y}$ $\leftrightarrow$ consejo1$_{x}$
c) stabl$_{e,s,o}$ $\leftrightarrow$ darl$_{e,s,p,o}$ lel$_{o}$ puñaladal$_{y}$ al$_{o}$

(We include lexical signs for determiners, clitics and accusative markers as predicates over the variable of their syntactic head; however, reasoning formalisms may dispense with them.) Note that we choose the variable of ‘piece’ on the English side as the argument variable on the Spanish side; if phrases such as ‘a piece of good advice’ are allowed, the Spanish side would be *consejo*1$_{x \cup y}$, whose semantic argument would be unifiable with both *x* and *y* to permit modifiers and heads to combine appropriately during generation.

To translate 'John kicked the bucket', the SL transfer representation:

Johnl$_1$ kick1$_{2,1,3}$ thel$_3$ bucketl$_3$

is covered by the bilexicon. The result is the union of the target side of all the bilexical entries used in this process:

$\{\text{juanl}_1\} \cup \{\text{estirarl}_{2,1,3}, \text{lal}_3, \text{pata1}_3\}$

(We ignore the literal translation of the idiom.) Generation then proceeds via the Spanish grammar and bag generator.

In the case of the other multi-lexeme translations mentioned the difficulties posed by varying lexical elements in part or all of the translation relation cannot be easily handled in the original SB formulation. Consider for example the case of 'John is thirsty'; its Spanish translation, *Juan tiene sed* (lit. 'John has thirst') differs from it in two main ways: the English adjective translates into a Spanish noun, while the verb is not intuitively felt to be the translation of *tener.* The problem for LMT based on one-to-one transfer is that a literal translation into Spanish is incorrect (* *Juan está sediento),* and that even if TL filtering (Alshawi et al. 1992) were used to eliminate such a sentence, the efficiency of the system would be compromised and translation of unseen sentences would be more error prone. Alternatively, an idiom-based translation in which the bilexicon relates 'be thirsty' and *tener sed* ignores important systematic differences between the two languages:

John **is thirsty**	Juan **tiene sed**
John **is hungry**	Juan **tiene hambre**
John **is lucky**	Juan **tiene suerte**
John **is angry**	Juan **tiene rabia**
John **is hot**	Juan **tiene calor**
John **is cold**	Juan **tiene frío**

We therefore argue that a one-to-one translation for such phrases is not adequate but instead consider the highlighted phrases above as the correct equivalences between the two languages. The task then, is to find a mechanism for efficiently capturing regularities of this sort in the present framework. There are a number of alternatives for achieving this. We will consider three.

3.1 Lexical Neutralization

The first possibility for handling multi-lexeme regularities in LMT is to eliminate support verbs from the SL transfer representation altogether, and to reintroduce them during generation. In this case, a semantic representation for the sentences must be proposed. For the sake of argument assume an adjective-like intersective semantics for both the Spanish nouns *Juan* and *sed* and the corresponding English noun and adjective:

SL: johnl$_1$ thirsty1$_1$
TL: juanl$_1$ sedl$_1$

Then, the bilexicon would include, among other things:

thirstyl$_x$ $\leftrightarrow$ sedl$_x$
hungryl$_x$ $\leftrightarrow$ hambrel$_x$
etc.

Lexicalist transfer would apply these equivalences to construct an appropriate TL bag. During Spanish bag generation, the appropriate support verb (i.e. *tener)* would be introduced by inspection of monolingual lexical information associated with *sed* (Danlos and Samvelian 1992), from which correct instantiation of the orthography of the TL sentence would ensue. A variation of this strategy would be to use a partially instantiated lexical sign corresponding to the English support verb:

{ johnl$_1$, support-verb$_{2,1,3}$, thirsty$_3$ }

During transfer, the support verb is translated as a partially instantiated support verb in Spanish. The generation algorithm would then be applied such that monolingual constraints in the Spanish grammar fully instantiated the semantics and orthography of this verb according to the support verb requirements of its complement noun.

3.2 Lexical Variables

The second mechanism for capturing multi-lexeme regularities assumes translation variables similar to those used in several transfer systems (Alshawi et al. 1992; Bech et al. 1991; Russell et al. 1991). If one represents transfer variables by **tr**(<*restrictions*>), then the necessary bilexical entry would be:

bel$_{x,y,z}$, **tr**(Adj$_z$) $\leftrightarrow$ tenerl$_{x,y,z}$, **tr**(Noun$_2$)

This entry states that ‘be’ translates as *tener* as long as its complement adjective translates as the complement noun of *tener.* The transfer algorithm is modified to accommodate the transfer variable by, for example, recursively calling itself on the value of tr(Adj$_z$). Generation, however, proceeds as before. A variation of this mechanism is to use contextual rather than transfer variables. In this case, a particular lexical context is specified which constraints translation equivalence in a manner analogous to the way left and right contexts are used in morphological rewriting rules (Kaplan and Kay 1994). Thus, the transfer relation

bel$_{x,yz}$, (Adj$_z$) $\leftrightarrow$ tenerl$_{x,y,z}$, (Noun$_z$)

would indicate that in the context of an adjective complement, 'be' may translate as *tener* or vice versa. The main difference between this and the transfer variable variant is that the contextual elements, Adj and Noun, can serve as context to multiple transfer relations within the same cover, whereas this would not be possible with transfer variables. We will appeal to contextual variables in Section 5.

The third mechanism uses bilingual lexical rules to map bilexical entries into new bilexical entries. We have adopted this mechanism for certain multi-lexeme translations because it allows the exploitation of monolingual lexical rules in a motivated manner which integrates naturally with the LMT architecture, and because it provides a framework in which to study differences between lexical processes in different languages.

4 Lexical and Bi-Lexical Rules

The lexicon has taken a prominent place in several linguistic theories (Pollard and Sag 1994; Oehrle et al. 1988), not least because, given appropriate tools, both general and idiosyncratic properties of language can be captured within a uniform framework. Among the tools normally employed one finds lexical rules (Dowty 1978; Flickinger 1987; Pollard and Sag 1994) and inheritance mechanisms (Briscoe et al. 1993; Flickinger and Nerbonne 1992). Lexical rules may be thought of as establishing a relationship between lexical items such that given the presence of one lexical item in the lexicon the existence of a further item may be inferred. The regularities captured by lexical rules might include changes in the subcategorization and control properties of a verb, the denotation of a noun or the interpretation of a preposition. With the advent of lexically oriented approaches to translation, it is worth considering whether and how the generalizations captured by lexical rules might be exploited in MT.

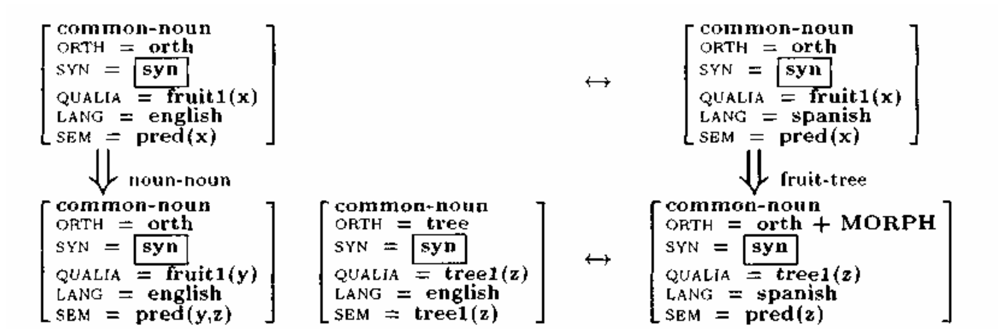
In order to investigate this issue we have adopted the notion of a bi-lexical rule. A bi-lexical rule (Trujillo 1992; Copestake et al. 1993) takes a bilexical entry as input, and outputs a new bilexical entry. These rules may be seen as expanding the bilexicon in order to increase its coverage; under this view, they are somewhat analogous to lexical rules in that they reduce the number of bilexical entries that need to be explicitly listed. Bi-lexical rules also serve to capture lexical, syntactic and semantic regularities in the translation between two languages by relating equivalent lexical processes cross-linguistically.

4.1 Simple Bi-lexical Rule

We give a simple example of a bi-lexical rule before addressing the multi-lexeme translations introduced earlier. Consider the relationship that exists in English-Spanish translations between the translation of fruits and the translation of their corresponding trees (Soler and Marti 1993):

Fruit		Tree	
English	*Spanish*	*English*	*Spanish*
almond	almendra	almond tree	almendro
apple	manzana	apple tree	manzano
cherry	cereza	cherry tree	cerezo
orange	naranja	orange tree	naranjo
plum	ciruela	plum tree	ciruelo
lemon	limón	lemon tree	limonero

The relevant relationship may be described by the following bi-lexical rule:



This bi-lexical rule says that if there is a bilexical entry translating English fruit nouns into Spanish fruit nouns, then there is a bilexical entry translating '*noun* tree' in English into a morphologically derived tree-denoting noun in Spanish.

We adopt Qualia structure (Pustejovsky 1991) as our lexical-semantic representation formalism. According to Pustejovsky Qualia structure is one of the four main types of information to be associated with a lexical entry (the others being Argument, Event and Inheritance structure). The information incorporated in a Qualia structure specifies the semantics of a lexical item by virtue of the relations and properties in which it participates. For this example we assume a simplified Qualia value (Pustejovsky 1991) indicating whether a noun denotes a tree or a fruit. Note that the morphology of the output Spanish lexical sign is left implicit since it depends on the actual noun used (see fruit-tree table above); in addition, the English rule mapping a noun into a noun modifier is a practical simplification of the complex issue of noun-noun modification which we do not address here (Pustejovsky and Boguraev 1993; Johnston et al. 1994). Another point to note is that we will be vague regarding the amount of information shared between the input and output lexical signs of lexical rules; a full treatment of this issue involves aspects of default unification which are beyond the scope of this paper (Meurers 1994; Lascarides et al. in press). Suffice it to say that in our implementation, an attempt has been made to share maximum information between input and output lexical signs, although values such as semantic variables are not shared between input and output lexical signs.

In the abbreviated notation introduced earlier, the above bi-lexical rule will be represented as:

$$\begin{array}{lll} \mathrm{Ne}_x & & \leftrightarrow \quad \mathrm{Ns}_x \\ \Downarrow \text{identity} & & \quad\;\; \Downarrow \text{fruit-tree} \\ \mathrm{Ne}_{y,z} & \mathrm{tree1}_z & \leftrightarrow \quad \mathrm{Ns}'_z \end{array}$$

Given the translation ‘apple – manzana’, for example, the rule would operate as indicated below:

$$\begin{array}{lll} \mathrm{apple1}_x & & \leftrightarrow \quad \mathrm{manzana1}_x \\ \Downarrow \text{identity} & & \quad\;\; \Downarrow \text{fruit-tree} \\ \mathrm{apple1}_{y,z} & \mathrm{tree1}_z & \leftrightarrow \quad \mathrm{manzano1}_z \end{array}$$

Its output is the additional translation relation 'apple tree - *manzano'.* Similar translations are achieved for other fruits.

Clearly this rule should only apply to fruits which grow on trees and not to fruits such as strawberries which are found in low growing plants. Such restrictions need to be incorporated in the monolingual lexical signs and rules.

Implementationally, bilexical rules may be applied off-line in order to expand the bilexicon before processing, or they may be applied during transfer to extend the bilexicon just sufficiently to enable transfer. We have opted for the latter approach.

4.2 Support Verbs

We now show how bi-lexical rules can be used in the translation of ‘thirsty’, basing our analysis on the classification of support verbs proposed by Danlos and Samvelian (1992) for English-French translation. Their proposal, implemented as part of a Eurotra project, involves transfer at the Interface Structure. The essence of their approach is similar to that for multi-lexeme translations given in Section 3.1: the support verb is deleted from the SL transfer structure, the adjective ‘thirsty’ is translated into the TL noun *(sed* in our case), and an appropriate TL support verb is incorporated into the TL sentence during generation. Information regarding which support verb a noun requires is encoded in its lexical entry.

Support verbs can be of five types: neutral (e.g. ‘is thirsty’), durative (e.g. ‘remain thirsty’), inchoative (e.g. ‘get thirsty’), terminative (e.g. ‘stop being thirsty’) and iterative (e.g. ‘be thirsty again’). We will consider neutral support verbs only although the other categories could also be handled through bi-lexical rules. One difference between the present approach and that of Danlos et al. in that we equate the noun 'thirst' with the noun *sed* in the bilexicon, rather than equating an adjective and a noun, thus factoring category and support verb differences:

$\text{thirstl}_x \leftrightarrow \text{sedl}x$

We believe this reflects more truly the translation relation that exists between the two lexical items. An English-Spanish bi-lexical rule is then introduced to derive the adjective on the English side and to include the neutral support verb ‘be’; on the Spanish side the support verb *tener,* for the noun *sed,* is introduced:

$$\begin{array}{lllll} & N_x & \leftrightarrow & & N[\text{ntrl=tener}]_x \\ & \Downarrow \text{adjective} & & & \Downarrow \text{identity} \\ \text{bel}_{e,s,y} & A[\text{ntrl=be}]_y & \leftrightarrow & \text{tenerl}_{e,s,y} & N[\text{ntrl=tener}]_y \end{array}$$

Note that we underspecify the support verb for the input English noun to allow ‘John has an unquenchable thirst’ and similar examples. The neutral (ntrl) control verb required by the English adjective is included in its lexical entry’s Qualia structure. Thus, a fuller TFS for ‘thirsty’ is:

$$\begin{bmatrix} \textbf{adjective} \\ \text{ORTH} = \textbf{thirsty} \\ \text{SYN} = \boxed{\textbf{syn}} \\ \text{QUALIA} = \begin{bmatrix} \textbf{qualia} \\ \text{SUPP-VERBS} = \begin{bmatrix} \textbf{supp-verbs} \\ \text{NTRL} = \textbf{be(e,s,y)} \\ \text{INCH} = \textbf{get(e,s,y)} \end{bmatrix} \end{bmatrix} \\ \text{LANG} = \textbf{english} \\ \text{SEM} = \textbf{thirsty1(y)} \end{bmatrix}$$

In designing an appropriate Qualia structure we have added to the roles proposed by Pustejovsky (1991) (Constitutive, Formal, Telic and Agentive) in order to incorporate information necessary for capturing particular phenomena (Johnston et al. 1994).

When translating ‘John is thirsty’, the analyser constructs the transfer representation:

$\text{Johnl}_1\ \text{bel}_{2,1,3}\ \text{thirstyl}_3$

We include the support verb ‘be’ in our representation, even though it has empty semantics, in order to encode scoping information - i.e. to prevent ‘John is a painter’ translating as ‘a painter is John’; this rather *ad hoc* solution could be replaced by a mechanism analogous to the labels used in Underspecified Discourse Representation Theory (Reyle 1995; Frank and Reyle 1995).

During transfer, the bi-lexical rule above is applied to the bi-lexical entry for ‘thirst’ to yield:

$\text{bel}_{e,s,x}, \text{thirstyl}x \leftrightarrow \text{tenerl}_{e,s,x}, \text{sedl}_x$

This multi-lexeme relation is used to translate 'is thirsty' into *tiene sed;* a separate entry translates 'John' into *Juan.* Bag generation then ensures that the TL bag yields a sentence which satisfies the constraints specified by the TL grammar.

The intuitive description of the above process is that we consider 'is thirsty' not to be translatable compositionally, but instead to require a multi-lexeme translation. The purpose of bi-lexical rules then is to minimize the repetition of information in the bi-lexicon while allowing the exploitation of monolingual lexical processes.

4.3 Lexicalization Patterns

There are other translation phenomena which can be described through the use of bi-lexical rules. Consider lexicalization patterns for example (Talmy 1985):

John **swims across** the river.
Juan **cruza** el río **nadando.**

In the English sentence, the main verb encodes manner (i.e. swimming) and motion, while in Spanish it encodes path (i.e. across) and motion; the remaining meaning component in each case is expressed through a modifier. Talmy attributes these distinctions to differences in lexicalization patterns between the two languages.

A previous approach to such translations has been to introduce the bilexical entries 'swim - *nadar* + *ando'* and 'across – *cruza'* (Beaven 1992). This approach, however, only implicitly acknowledges that theses two translations are only appropriate in conjunction, and that separately they are in fact unintuitive. This not only increases the non-determinism of transfer and generation, but can increase the likelihood of incorrect translations for unseen sentences. In the bi-lexical rule view, one relates verb translations to translations incorporating lexicalization patterns as follows:

$V_{e,s}$		$\leftrightarrow$		$V'_{e,s}$
$\Downarrow$ identity				$\Downarrow$ gerund
$V_{f,t}$	$\text{across1}_{f,x}$	$\leftrightarrow$	$\text{cruzar}_{f,t,x}$	$V[\text{vform= ing}]'_{f,t}$

This rule derives, for every (movement) verb translation, a multi-lexeme translation which includes 'across' as a modifier (we leave the restriction on verbs to movement events implicit; also, a simplified description of 'across' is assumed (Trujillo forthcoming)).

Application of this rule to 'swim - *nadar'* may be depicted as follows:

$\text{swim1}_{e,s}$		$\leftrightarrow$		$\text{nadar1}_{e,s}$
$\Downarrow$ identity				$\Downarrow$ gerund
$\text{swim1}_{f,t}$	$\text{across1}_{f,x}$	$\leftrightarrow$	$\text{cruzar}_{f,t,x}$	$\text{nadando1}_{f,t}$

Lexicalist translation of ‘John swims across the river’ can then proceed by translating ‘swims across’ with the output of this rule and the remaining elements of the input via other bilexical entries.

4.4 Head Switching

The phenomenon of head switching in translation can be exemplified by the following pair of sentences:

> John **just** arrived.
> Juan **acaba de** llegar.

The problem with such translations is that the syntactic head in the SL sentence is not the syntactic head in its translation. This is a major obstacle for syntactic and even some semantic based translation systems because of the recursive nature of their transfer representations.

Head switching has been given a number of solutions in a variety of systems (Kaplan et al. 1989; Sadler and Thompson 1991; Russell et al. 1991; Whitelock 1992; Kaplan and Wedekind 1993). In our framework, the solution is expressed by the following rule [3]:

$$\begin{array}{lllll} & V_{e,s} & \leftrightarrow & & V'_{e,s} \\ & \Downarrow \text{identity} & & & \Downarrow \text{infinitive} \\ \text{justl}_f & V_{f,t} & \leftrightarrow & \text{acabar\_del}_{f,t,f} & V'_{f,t} \end{array}$$

Application to the bilexical entry ‘arrive – *llegar’* results in:

> justl_f, $\text{arrivel}_{f,t} \leftrightarrow \text{acabar \_del}_{f,t,f}$, $\text{llegarl}_{f,t}$

Lexicalist translation progresses as before. To exemplify the use of bilexical rules in head switching, we consider translation in embedded contexts in more detail now. To translate between:

> Mary thinks John just arrived.
> María piensa que Juan acaba de llegar.

the parser constructs the following representation (again, ignoring issues of scope and quantification):

> maryl_1 , $\text{thinkl}_{2,1,4}$, johnl_3 , justl_4 , $\text{arrivel}_{4,3}$

Assuming appropriate transfer of ‘Mary’ and ‘John’, translation of the embedded clause obtains as follows. ‘Thinks’ is translated by the following entry:

> $\text{Thinkl}_{e,s,f} \leftrightarrow \text{pensar\_quel}_{e,s,f}$

In addition, the output of the previous bi-lexical rule serves for multi-lexeme transfer of ‘just arrive’ to give the incomplete bag:

> { $\text{pensar\_quel}_{2,1,4}$, $\text{acabar\_del}_{4,3,4}$, $\text{llegarl}_{4,3}$ }

The final result of transfer is the TL bag:

[3] We ignore the (complex) issue of tense for this type of example of head switching; we expect that it can be tackled independently of the present approach.

{ maríal$_{1}$, pensar_quel$_{2,1,4}$, juanl$_{3}$, acabar_de1$_{4,3,4}$, llegarl$_{4,3}$ }

During generation, *acabar de* is made the syntactic head of the sentence through grammatical constraints in the Spanish grammar. Illustrative rules might be:

$S_{e,s} \Rightarrow NP_S\ VP_{e,s}$
$VP_{e,s,c} \Rightarrow Vvp_{e,s,c}\ VP_C$
$VP_{e,s,c} \Rightarrow Vs_{e,s,c}\ S_c$

If *pensar_que* has category Vs, and *acabar_de* has category Vvp, there is only one ordering of the TL bag by which the constraints indicated by this small grammar can be satisfied, namely, the order given by its translation:

María piensa_que Juan acaba_de llegar.

It may be noticed that head selection by the TL grammar is possible because the event semantic constants in *acabar_de* and *llegar* are the same. The consequence of this is that modifiers which apply to 'just arrived' and 'arrived' separately will be indistinguishable during TL generation. Avoiding this problem entails transferring scoping domains for modifiers in order to constraint generation. However, we have no readily implementable mechanism for achieving this in LMT as yet.

This concludes our overview of the different translation mismatches that may be handled through bi-lexical rules. We now consider some unresolved issues arising from their use.

5 Bi-lexical Rule Interaction

One difficulty we have found with bilexical rules has been their composition. For example, consider the following translation:

l) John **marched** the soldiers **across** the valley.
1') Juan le **hizo cruzar** el valle a los soldados **marchando.**

In our framework, two bi-lexical rules should be applied in such cases: one to construct causative translations (Comrie 1985; Levin and Rappaport 1995):

march1	↔		marchar1
⇓ causative			⇓ infinitive
march1$_{causative}$	↔	hacer1	marchar1$_{infinitive}$

The other to deal with differences in lexicalization patterns such as ‘march across - *cruzar marchando’.* The problem is that in isolation neither of these rules could perform the above translation. Ideally one should be able to use the output of one as input to the other to derive ‘march across - *hacer cruzar marchando’,* but this is not possible because both bi-lexical rules expect a mono-lexeme bilexical entry.

One possible solution is to manually add further bi-lexical rules which incorporate the composition of other rules:

$$\begin{array}{llll} V & & \leftrightarrow & V' \\ \Downarrow_{\text{causative}} & & & \Downarrow_{\text{gerund}} \\ V & \text{across1} & \leftrightarrow \text{hacer1 cruzar1} & V' \end{array}$$

However, this solution leads to a combinatorial explosion in the number of bi-lexical rules.

The line of work we are investigating combines bi-lexical rules with the context variables given in Section 3.2. There remain problems in our implementation, however, which will be evident from the following description. In our proposed approach either the causative or the lexicalization pattern bi-lexical rule, or both, incorporate a context variable in their output bilexical entry. For example, assume that the variable is included in the causative rule:

$$\begin{array}{llll} V & \leftrightarrow & & V' \\ \Downarrow_{\text{causative}} & & & \Downarrow_{\text{infinitive}} \\ (V) & \leftrightarrow & \text{hacer1} & (V') \end{array}$$

This rule says that whenever there is a verb bilexical entry, there is also an entry which in the context of a causative verb introduces *hacer* in the TL bag. Applying the rule to 'march – *marchar'* gives:

$$\begin{array}{lllll} & \text{march1} & \leftrightarrow & & \text{marchar1} \\ & \Downarrow_{\text{causative}} & & & \Downarrow_{\text{infinitive}} \\ 2) & (\text{march1}_{causative}) & \leftrightarrow & \text{hacer1} & (\text{marchar1}_{infinitive}) \end{array}$$

Lexicalist transfer of '$\text{march}_{\text{causative}}$ ... across ...' via the output of this rule and that for lexicalization patterns proceeds as follow: the causative reading of 'march' unifies with the context lexical sign in 2) but is not translated by it. The TL side therefore only contributes *hacer* to the final TL bag. Via the bi-lexical rule given in Section 4.3, 'march across' is transferred such that *cruzar* and *marchando* form part of the final TL bag. The result is therefore *hacer cruzar marchando,* which, in combination with the translation of the rest of the sentence can form the basis for bag generation.

Our main problem is that of resolving conflicts between the syntactic constraints imposed by each bi-lexical rule. The causative rule requires the Spanish side to include an infinitive verb, while the lexicalization pattern rule requires a gerundive verb. Clearly both constraints cannot be satisfied for the same lexical sign *marchar1.* The problem reflects itself in our proposal in that the rule which includes the contextual pattern must be chosen carefully. If the lexicalization pattern rule rather than the causative rule had included the contextual verb lexical sign, the gerundive *marchando* could not have been generated. Instead, a sentence analogous to 'John made the soldiers march crossing the valley' would result, which is perhaps not desirable. In other words, the conflict between gerundive and infinitive morphology for ‘march’ is decided manually in advance. The interaction of such decisions with other bi-lexical rules therefore might be unpredictable, and hence is left for further investigation.

6 Implementation

The implemented prototype system contains approximately 250 bilexical entries; this figure includes 20 proper names, 20 multi-lexeme translations and 6 contextual rules. The following translations were done on a SUN Spare workstation using Allegro Common Lisp. The time taken to find all possible TL sentences is given in seconds; total times are for CPU + typical garbage collection times.

Translation	Total (CPU)
John thinks Mary just arrived Juan piensa_que María acaba_de llegar	50 (28)
John swam across the river Juan cruzó el río nadando	19 (16)
John marched the soldiers Juan hizo marchar a los soldados	19 (17)

These timings are only intended to give some idea of the type and stage of our implementation, rather than reflect the performance of an optimized system.

7 Conclusion

We have introduced the mechanism of bi-lexical rules for incorporating lexical rules in MT. These rules establish correspondences between bilexical entries such that given the presence of one entry, the existence of another bilexical entry can be inferred. We presented various phenomena that can be described using such rules: noun sense extensions, support verbs, lexicalization patterns and head switching. The rules provide a useful and motivated extension to the LMT paradigm by providing it with a uniform approach to the description of a number of translation phenomena.

The problems arising from conflicting constraints imposed by different translation relations are described, and a partial solution to these was offered involving the combined use bi-lexical rules and contextual variables.

Future work could consider implementing Mel'cuk's lexical functions (Heylen et al. 1994) in a manner similar to the way bi-lexical rules were used in the translation of support verbs.

Acknowledgements

Thanks to two anonymous reviewers for their valuable comments. The LKB was implemented by Ann Copestake as part of the ESPRIT ACQUILEX project. Remaining errors are mine.

References

Alshawi, H., ed. (1992). *The Core Language Engine.* MIT Press, Cambridge, MA.

Alshawi, H., Carter, D., Gamback, B., and Rayner, M. (1992). Swedish-English QLF translation. In Alshawi (1992), ch. 14, pp. 277-309.

Arnold, D. and Sadler, L. (1992). Rationalism and the treatment of referential dependencies. In *Proceedings of the Fourth TMI,* pp. 195-204, Montreal, Canada.

Beaven, J. L. (1992). Shake-and-Bake machine translation. In *Proceedings of COLING '92,* pp. 602-09, Nantes, France.

Bech, A., Maegaard, B., and Nygaard, A. (1991). The Eurotra MT formalism. *Machine Translation,* 6(2) pp. 8.3-101.

Brew, C. (1992). Letting the cat out of the bag: Generation for Shake-and-Bake MT. In *Proceedings of COLING '92,* pp. 610-16, Nantes, France.

Briscoe, E., Copestake, A., and de Paiva, V., eds. (1993). *Inheritance, Defaults and the Lexicon.* Cambridge University Press, Cambridge, UK.

Calder, J., Reape, M., and Zeevat, H. (1989). An algorithm for generation in unification categorial grammar. In *Proceedings of the Fourth European Conference of the ACL,* pp. 233-40, Manchester, England.

Comrie, B. (1985). Causative verb formation and other verb-deriving morphology. In Shopen (1985), ch. 6, pp. 309-48.

Copestake, A., Flickinger, D., Malouf, R., Riehemann, S., and Sag, I. (1995). Translation using minimal recursion semantics. In *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation,* Leuven, Belgium.

Copestake, A., Sanfilippo, A., Briscoe, E., and de Paiva, V. (1993). The ACQUILEX LKB: An introduction. In Briscoe et al. (1993), ch. 9, pp. 148-63.

Danlos, L. and Samvelian, P. (1992). Translation of the predicative element of a sentence: category switching, aspect and diathesis. In *Proceedings of the Fourth TMI,* pp. 21-34, Montreal, Canada.

Devlin, K. (1991). *Logic and Information.* Cambridge University Press, Cambridge, UK.

Dorr, B. J. (1992). The use of lexical semantics in interlingual machine translation. *Machine Translation,* 7(3) pp. 135-93.

Dowty, D. R. (1978). Applying Montague's views on linguistic metatheory to the structure of the lexicon. In Farkas et al. (1978), pp. 97-137.

Farkas, D., Jacobsen, W. M., and Todrys, K. W., eds. (1978). *Papers from the Parasession on the Lexicon.* Chicago Linguistic Society, University of Chicago, IL.

Flickinger, D. (1987). *Lexical Rules in the Hierarchical Lexicon.* PhD thesis, Stanford University, Stanford, CA.

Flickinger, D. and Nerbonne, J. (1992). Inheritance and complementation: A case study of *easy* adjectives and related nouns. *Computational Linguistics,* 18(3) pp. 269-310.

Frank, A. and Reyle, U. (1995). Principle based semantics for HPSG. In *Proceedings of Seventh Conference of the European Chapter of the ACL,* pp. 9-16, Dublin, Ireland.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, New York.

Heylen, D., Maxwell, K. G., and Verhagen, M. (1994). Lexical functions and machine translation. In *Proceedings of the 15th COLING,* pp. 1240-44, Kyoto, Japan.

Hinrichs, E. W.. Meurers, W. D., and Nakazawa, T. (1994). Partial-VP and Split-NP Topicalization in German - An HPSG Analysis and its Implementation. Technical Report Arbeitspapiere des SFB 340 Nr. 58, Universität Tübingen, Germany.

Hutchins, W. J. and Somers, H. L. (1992). *An introduction to Machine Translation.* Academic Press, London.

Johnston, M., Boguraev, B., and Pustejovsky, J. (1994). The acquisition and interpretation of complex nomináis. Paper presented at the *Workshop on the Future of the Dictionary,* Uriage-les-bains, France.

Kaplan, R. M. and Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics,* 20(3) pp. 331-78. Special Issue on Computational Phonology.

Kaplan, R. M., Netter, K., Wedekind, J., and Zaenen, A. (1989). Translation by structural correspondences. In *Proceedings of the Fourth Conference of the European Chapter of the ACL,* pp. 272-81, Manchester, UK.

Kaplan, R. M. and Wedekind, J. (1993). Restriction and correspondence-based translation. In *Proceedings of the Sixth Conference of the European Chapter of the ACL,* pp. 193-202, The Netherlands. OTS, Utrecht University.

Kay, M., Gawron, J. M., and Norvig, P. (1994). *Verbmobil: A Translation System for Face-to-Face Dialog.* Number 33 in Lecture Notes. Centre for the Study of Language and Information, Stanford, CA.

Lascarides, A., Briscoe, T., Asher, N., and Copestake, A. (in press). Order independent and persistent typed default unification. *Linguistics and Philosophy.*

Levin, B. and Rappaport, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface.* MIT Press, Cambridge, MA.

Meurers, W. D. (1994). On implementing an HPSG theory - aspects of the logical architecture, the formalization, and the implementation of head-driven phrase structure grammars. In Hinrichs et al. (1994).

Oehrle, R., Bach, E., and Wheeler, D., eds. (1988). *Categorial Structures and Natural Language Structures.* D. Reidel, Dordrecht, The Netherlands.

Pollard, C. and Sag, I. (1987). *Information Based Syntax and Semantics: Vol. 1.* Lecture Notes. CSLI, Stanford, CA.

Pollard, C. and Sag, I. (1994). *Head Driven Phrase Structure Grammar.* Chicago University Press, IL.

Popowich, F. (1995). Improving the efficiency of a generation algorithm for Shake and Bake machine translation using Head-Driven Phrase Structure Grammar. In *Proceedings of Natural Language Understanding and Logic Programming V,* Lisbon, Portugal.

Poznański, V., Beaven, J. L., and Whitelock, P. (1995). An efficient generation algorithm for lexicalist MT. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics,* Boston, MA.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics,* 17(4) pp. 409-441.

Pustejovsky, J. and Boguraev, B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence,* 63(1-2) pp. 193-224. Special Volume: Natural Language Processing.

Reyle, U. (1995). On reasoning with ambiguities. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics,* pp. 1-15, Dublin, Ireland.

Russell, G., Ballim, A., Estival, D., and Warwick, S. (1991). A language for the statement of binary relations over feature structures. In *Proceedings of the Fifth Conference of the European Chapter of the ACL,* Bonn, Germany.

Sadler, L., Crookston, I., and Way, A. (1990). LFG and translation. In *Proceedings of the Third TMI,* University of Texas at Austin. LRC.

Sadler, L. and Thompson, H. (1991). Structural non-correspondence in translation. In *Proceedings of the Fifth Conference of the. European Chapter of the ACL,* pp. 293-98, Bonn, Germany.

Schenk, A. (1986). Idioms in the Rosetta machine translation system. In *Proceedings of COLING '86,* Bonn, Germany.

Shopen, T., ed. (1985). *Language Typology and Syntactic Description Vol. III: Grammatical Categories and the Lexicon.* Cambridge University Press, Cambridge, UK.

Soler, C. and Marti, M. A. (1993). Dealing with lexical mismatches. Technical Report 4, ESPRIT BRA-7315 ACQUILEX II Working Paper, Publishing Division, Cambridge University Press, UK.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In Shopen (1985), ch. 2, pp. 57-149.

Trujillo, A. (1992). Spatial lexicalization in the translation of prepositional phrases. In *Proceedings of the 30th Annual Conference of the ACL, Student Session,* pp. 306-08, Newark, Delaware.

Trujillo. A. (forthcoming). Towards a cross-linguistically valid classification of spatial prepositions. *Machine Translation.*

Whitelock, P. (1992). Shake-and-Bake translation. In *Proceedings of COL1NG '92,* pp. 784-91, Nantes, France.