[International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP), July 1992, Manchester]

# Rule-Based Translation as Constraint Resolution

Louisa Sadler
University of Essex,
Wivenhoe Park,
Colchester, C04 3SQ,
UK.
louisa@essex.ac.uk

July 6, 1992

**Abstract**

This paper is mainly concerned with the rule-based approach to Machine Translation. It begins with some general remarks on the relation between rule-based and empirical approaches to MT, stressing the importance of the diversity of goals for MT research and the contributions both types of approach can make. In the following section, an assessment is given of the major difficulties for rule-based MT. The major part of the paper discusses the notion of constraint-based translation and shows how some of these difficulties can be partly addressed with this approach.

**Keywords:** machine translation, transfer, constraint resolution

## 1 General Remarks

This conference provides a welcome forum for contact between the rule-based and empirical approaches to natural language processing, and I would like to start my contribution with some remarks which address the relationship between these rather different approaches to machine translation, and the various contributions that they can make to the practical and theoretical goals of MT.

In the main body of the paper I discuss some recent notions in rule-based approaches to machine translation which I think show substantial promise and indicate useful areas for further research. I begin that section with some remarks about the very real problems which rule-based approaches face or will face in any transition from research test-bed to industrial or semi-industrial prototype.

I should begin by stating what I understand by the terms 'rule-based' and 'empirical' in this context. Until very recently (perhaps the last five years) the vast majority of contemporary work in MT has been rule-based. A rule-based system is one which entirely relies on the formulation of explicit rules for relating (linguistic) objects. Traditional interlingual and transfer architecture MT systems are rule-based, they differ in the depth of linguistic analysis and whether there is any component of directly expressed bilingual information. The main tasks in this style of MT would seem to be (i) the specification and implementation of a translation engine, typically involving a number of discrete phases of processing and (ii) the encoding of linguistic information in appropriate formalisms. In extended rule-based architectures, some attempt may also be made to include some explicit expression of and computation over contextual and extra-linguistic information ([15]). It should be clear that a large number of quite widely divergent systems fall under this term.

Of course, rule-based approaches to MT can be further classified along a number of different dimensions. Until quite recently, the most fundamental distinction was perceived to separate transfer systems, which involve and explicit formulation of bilingual knowledge, from interlingual systems, which do not. In the light of the sort of approach I discuss below, this distinction becomes less important. Other dimensions of variation would include the approach to monolingual processing, the existence or otherwise of pre- and post- transfer or interlingual "re-alignment" or "patch-up" phases, adherence to (versions) of the compositionality principle, the reversibility of the approach, the role of interaction with the user, and so on.

I take *empirical approaches to MT* in a strict sense to refer to systems which do not rely on explicit linguistic knowledge at all, but make direct use of information automatically extracted from large corpus resources. A pure example of this approach to MT is the work of Brown et al, reported in [5] on statistical MT, in which there is no recourse whatsoever to explicit linguistic information.[1] Another example is phrase book translation, in which translation essentially consists of using keyword recognition to locate phrases in the target phrasal database corresponding to those in the source language input, with very little other linguistic processing. There are to my knowledge relatively few pure empirical approaches to MT, the majority of systems describing themselves as empirical actually being hybrid in the sense of mixing explicit linguistic processing with the direct use of knowledge extracted from corpus resources. The term *hybrid approaches to MT* refers to those systems which involve a traditional rule-based core and add-on modules based on empirical techniques. Examples of such extensions would be statistically based preference mechanisms (as opposed to preference mechanisms based on linguistic principles, on relations between rules such as *subsumption* [9], [25] or on linguist defined control grammars [2]) and large scale lexical and corpus

[1] Brown and his colleagues have weakened this position in more recent work [6].

resources deployed essentially as a bilingual lexical disambiguation components [22], and certain analogical approaches to MT which use a database of bilingual pairs in transfer [23].

In discussing the way forward in research in MT and the relationship between rule-based and empirical approaches to MT, it is important to bear in mind the diversity of goals underpinning this research. Broadly, one can distinguish between the practical or engineering goal, which I take to be to provide a better basis for future MT products, and a number of distinct scientific goals. The engineering goal in research is quite different from the practical development task of producing a system within certain set parameters and a short timescale, when ad hoc, generally inadequate solutions to problems may be adopted even though one knows they are incompatible with the goal of providing a better basis for future products.

A scientific goal for some research in MT is to articulate a formal theory of translation, allowing one to define the notions "possible translation" "best translation" etc, identifying the grammatical, stylistic, contextual and extralinguistic constraints that determine the limits of translation. This is an important goal to which research in MT can make a substantial contribution. Much work in especially transfer-based MT explicitly claims this translation theory goal.

Another goal (only derivatively concerned with translation as such) is the goal of devising a level of linguistic representation sufficiently abstract to encompass the expressive possibilities of all human languages, the universal semantics goal. Optimism about the ultimate success of this goal must surely underpin much work in interlingual rule-based systems.

These are both primarily linguistic goals, contributing in complementary ways to our understanding of how the form-function relation may *be* differently realised in languages. It would seem fairly uncontentious that nonlinguistic systems can have little to contribute to these goals, although statistical approaches may conceivably yield generalisations not yet thought of.

There is a further translation-oriented goal, which is to model how human translators translate themselves. Some researchers into translation by analogy suggest that this is a plausible model of the way competent translators work. The idea is that a translator finds the best translation by feel or successive approximation from reasonably close example cases gleaned from her previous experience. Systems which have some explicit metric for calculating "distance" between lexemes and structures would be quite straightforward models of this hypothesis about how human translators translate. One might also also argue that translation by negotiation or constraint (see below), in which a target string is figured out from a set of constraints derived from the input and the target grammar, is a plausible model of human translation [26, 3, 11, 4].

As regards the practical goal of MT research, it seems to be that hybrid systems show the most promise, and therefore that research in both empirical and rule-based techniques is crucial. The pursuance of the other, scientific goals that I have mentioned seems to me of equal importance, contributing to our overall understanding of language. Addressing these questions presupposes explicit descriptions of the properties of and relations between linguistic expressions, which empirical techniques are designed to avoid.

In the rest of the paper, I will first review the key problems for rule-based MT research, and then go on to discuss some interesting directions.

# 2 Bottlenecks for Rule-Based Systems

In this section I review what I take to be the major challenges in achieving the practical goals of MT within the rule-based paradigm, at least in the relatively short-term. This does not mean to imply that all or even most of these problems are limited to the rule-based approach, for while empirical approaches to MT seem quite well suited to helping with rather local problems like lexical choice, it is difficult to see how they can provide any answer to problems like the resolution of gross structural divergences between languages.

## 2.1 The Acquisition of Linguistic Knowledge

By far the most serious bottleneck for MT is the problem of the acquisition of linguistic knowledge. The unprecedented growth in the availability of cheap computing power and storage has thrown this problem into sharp relief by effectively removing the computational barrier to high-speed MT, emphasizing the need for linguistic expertise and the sheer volume of coding required to move a system from research prototype to anything approaching use on real applications.

The problem shows itself in different ways in transfer and interlingual systems. Adding a new language to a system of interlingual design is relatively easy because there is no bilingual information. But linguistically, there are many open questions in defining an interlingua, even for closely related languages. Prom a purely practical point of view, setting up the framework of conceptual primitives is a huge undertaking for any non-trivial domain. If the interlingua is truly to be language independent, it must 'multiply out' the distinctions that are made in any of the languages concerned: e.g. since Spanish distinguishes two kinds of corner (concave and convex, roughly), so must the interlingua.

If it is logical there will typically be a large number of equivalent representations for any source sentence. Although just one formula will be found in analysis, the system must have the ability to draw inferences over formulae in order to ensure that a formula is found which can be used in synthesis [13].

The attraction of transfer systems over interlingual systems is their supposed feasibility, but for anything other than a very small domain, the rule writing effort is likely to be considerable. If the level of representation which transfer rules operate on is rather superficial, as in McCord [14], then the analysis and generation task is correspondingly more tractable. But in this case, the number and complexity of structural and lexical transfer rules can be expected to be very large for most language pairs. The less superficial it is, the more the difficulties associated with the interlingual method will occur, coupled with the effort of specifying bilingual mappings, however simple. Adding a new language of course requires transfer components into and out of each existing language as well as the additional monolingual components.

For transfer systems, the acquisition and maintenance of bilingual knowledge appropriate for MT and the considerable difficulty of capturing explicitly the precise conditions under which elements stand in a translation relation are key problems. This problem is compounded by the scarcity of adequate bilingual knowledge banks.

The knowledge acquisition task is made harder by the problem of redundancy: transfer rules and the rules of the target grammar are both involved in characterizing the target structures. Ideally, one would like to divide work between them on some principled way; it is a serious waste of effort to describe the same facts twice (the same issue arises with respect to transfer and the source grammar, and between the grammars of each language when it is considered as source and target). Similarly, there will typically be a considerable overlap between the transfer component from (say) English to French, and that from French to English.

## 2.2 Ambiguity

The problems of the acquisition and expression of linguistic knowledge are compounded by the ambiguity of the natural language. In transfer systems, complete disambiguation of elements in context is not attempted and the shallowness of the level of representation used for transfer can lead to a serious problem with multiple output.

In an interlingual system, the difficulty will show up in the monolingual components, which will be complex, with problems occurring especially in word sense disambiguation in analysis. In synthesis, since the form of the target structure will be radically underdetermined, one will typically have to choose between a large number of alternative realizations. All in all, building a robust monolingual component for a significant vocabulary is likely to be a daunting task.

## 2.3 Rule interaction

It is widely acknowledged that the process of ensuring that rules do indeed have the intended effect in extensive rule-based systems is difficult. Controlling the operation and interaction of rules is a significant problem in both interlingual and transfer-based MT. In transfer systems, for example, a reasonable strategy for cases of lexical mismatch, in which one language makes explicit in terms of different lexical items a distinction which is not lexicalised in another language (e.g, English *pregnant* translates into German *schwanger* and *trächtig* depending on whether the gestational process is human or not), is to give the two lexical pairings together with conditions on the application of the transfer rule which refer to the argument of the adjective. In systems which use some form of recursive rule application which traverses and perhaps decomposes the source tree or structure, care must be taken that if this condition applies in the source, that the relevant information is actually available when the transfer rule for *pregnant* applies. If the rules are to operate "as intended" the transfer writer may be led to exploit contingent facts such as the known order of applications of rules with a given processing strategy, which in turn will make the grammars difficult to maintain and understand, or to use a control grammar, with the same effect. Similar difficulties may occur in dealing with cases of structural mismatches such as head switching, either monolingually or in a transfer grammar [2].

### 2.4 An organising level for translationally relevant information

In simple interlingual or transfer architectures, if information is thought to be required for translation, it must be explicitly represented in the representation that is output by analysis. This poses a problem, because it is widely assumed that such information will not be conceptually homogenous, but will relate to different levels of linguistic organisation. For example, one may want to refer both to surface properties of the source structure (e.g. what the *subject,* or *tense* is), and semantic properties (what the *agent,* or *time reference* is). This may be for reasons of descriptive convenience or robustness (surface properties can be computed more reliably and easily, and so provide a more reliable basis for transfer than more abstract properties). Whitelock [26] and Hobbs and Kameyama [11] provide some discussion of the diverse sorts of linguistic information that might be relevant. Creating a coherent design for such a hybrid level of representation is very problematic.

## 3 Discussion

In the proceeding section, I have pointed out what seem to me to be the major problems facing rule-based approaches to MT. The major bottleneck is the sheer effort involved in expressing linguistic information. Compliance with a number of design choices and desiderata make a substantial contribution to the solution of this problem, however.

For example, the need for the representations to be suitably expressive, with a well-defined theoretical and conceptual basis is partly met if one stays close to established mainstream computational linguistic theories, and does not tailor one's descriptions too closely to the intended application (e.g. the grammars should not be tailored to any particular language pair, and ideally the monolingual parts of the system should not even be specifically designed for MT). This also reduces the size of the descriptive problem (i.e. the number and complexity of the rules to be written if a system is to operate over a significant domain), since one can exploit descriptions developed for such theories, and for other applications.

The problem is reduced if the principles of *modularity, declarativity* and *monotonicity* are adhered to as much as possible, meaning broadly that information of different sorts should be clearly separated, so that the description of different languages and different levels of linguistic information are kept apart, as well as separating algorithms from data; rules should be interpretable without reference to the procedures which apply them, and the behaviour of the system should not be dependent on a particular execution order.

Adopting requirements such as these may rule out some short-cuts (for example, exploiting the use of a particular processing strategy to rule out certain derivations or providing a "solution" to multiple outputs differing in word order in generation by using a more restricted grammar in generation than in analysis), but these are arguably inappropriate strategies in any case.

It seems right also to attempt the use of bidirectional grammars, wherever possible, and at least to ensure that the grammar is reversible (this ensures the same coverage in analysis and generation). Although one can tailor the generation grammar so that it produces, e.g. only one or two of the alternative placements of adverbs, thereby acting as thought it contained a preference filter, this would seem to be wrong, since it is more important to maintain a strict separation between tasks - offering up the set of possible translations is the jobs of the core engine, choosing between them is the job of another component.

Closely related to these requirements, and especially to modularity, is the reusability of linguistic knowledge. In a transfer system, the only application-specific portion should be the bilingual pairings, all monolingual components should be straightforwardly reusable in other applications. (Reversible components are of course a first step towards general reusability). Reusability places a number of quite strong constraints on grammar writing, often changing the view one has of the task in hand. Firstly, grammar design must stay very close indeed to current mainstream computational linguistic formalisms. This in turn casts light on the problem mentioned above of the design of specially tailored hybrid levels of representation - that is, they should not be used unless they can be straightforwardly compiled into and out of such well understood formalisms. The need for reusability thus increases the importance of pure and modular linguistic descriptions. Secondly, since little is currently understood of the process of reusing grammatical descriptions, the grammar writing effort should probably be accompanied by a serious effort at expressing the scope and function of each rule in some appropriately neutral metalanguage. The same is of course true of lexical descriptions, and here the issue of reusability of resources has already begun to receive some attention [10]. For grammatical resources, very little is currently known about the intertranslatability of formalisms, which has emerged as a key area for research.

The use of mainstream formalisms and the reuse of linguistic resources are important, but only provide a partial solution to the problems discussed in Section 2. above. One should not underestimate the major linguistic problems in designing broad coverage systems. A serious issue is actually determining what the language in the intended input texts is actually like. That such a problem exists is due to the long-term split between corpus-based linguistic work and the vast majority of work in theoretical and computational linguistics. The singleminded concentration on problem cases and cute examples has not been without benefit in illuminating the limitations and possible inadequacies of theoretical proposals or proposed architectures, but has occupied centre stage to the detrimental of solid text-based legwork. If the sort of detailed grammatical and lexical information which can be gleaned from texts is to be used in a rule-based system, then ultimately one needs a way of semi-automatically or automatically deriving such information from corpus resources so that it can be directly and straightforwardly expressed as a resource for the transfer rule writers.

A partial solution to the linguistic knowledge problem (especially for lexical selection) is the use (in a hybrid model) of a corpus as a bilingual knowledge bank to guide lexical choice, bypassing the need for explicit statements of complex conditions on bilingual pairings (for the discrimination of readings, for example). With a sufficiently large corpus, one can extract statistical information about likely translations in contexts of various sizes, and on this basis derive probable translations for the input. Alternatively, one could try to find material that shares certain similarities (e.g. being in the same thesaural area for certain items), and thus compute a best match for translation (translate 'by analogy'). Several refinements are possible- for example the items in the database of translations derived from the original corpus can be linguistically analyzed (parsed) in some way. Taking this a step further, adding some statistical or analogical techniques to a basically rule based approach would alleviate some of the difficulties of constructing rule based systems. However sophisticated the tools and environment for the acquisition of knowledge for rule-based systems become, it is more than likely that such empirical techniques will remain the only solution to many cases of word sense discrimination, etc., in the medium term. But these techniques do not amount to a complete solution, even in combination with more traditional techniques.

Underlying traditional interlingual and transfer systems there are two assumptions (i) that rule-based MT is about defining **one level of representation** at which translation "takes place" (ii) that the basic strategy is that of applying rules by some recursive procedure to all the "parts" of the source representation in some way. There are a number of reasons for doubting these views, because much of the difficulty comes precisely from the need to formulate those rules.

For example, Hobbs and Kameyama [11] make the point that languages differ in what aspects of "the message" they make explicit (in terms of linguistic encoding). Taking their example, politeness considerations may be explicit in a wide variety of diffuse ways in English texts, or even totally absent, but they must be realised grammatically in Japanese. Differences like these will effect not just the appropriacy but also the acceptability of the translation, and increase the difficulty of finding one homogenous and adequate level of representation. They argue that one should look on the linguistic expression as partial and multi-dimensional, with translation being seen as a process of abduction or reasoning to the best explanation.

A similar conclusion, i.e. that you need to think in terms of partial specifications of translational correspondences, follows if you think that complete syntactic and semantic analysis and synthesis is simply impossible [7]. For reasons of robustness, too, a model which exploits partial correspondences between structures is likely to be useful.

For political and social reasons there will continue to be a need for translation between closely related and even mutually intelligible languages. For reasons of reusability, one would like to ensure that the same system can deal with these pairs as with highly divergent languages, perhaps with some specialisation. But applying traditional rule-based techniques with a level of highly abstract representation would be overkill in these cases. The point is made by Dyvik [7, page 67]:

"In cases where source and target language are similar in structure we want to be able to use grammatical information from the analysis of the source string in the search for a target string instead of finding (virtually) the same information again."

An enormous effort is expended, where lexical pairings and some simple manipulation might have sufficed (this is especially true when the aim is simply to make translation more cost-effective, and postediting can be simply performed on the output text). For such languages, a far more surface syntactic system might be appropriate, but by our argument for reusability above, what we would seem to need is an approach with enough flexibility to operate at whatever levels of linguistic description are appropriate for different language pairs, while still remaining mainstream. That is, an approach which can adapt to the complexity of the task at hand, rather than retrieving a lot of unnecessary information.

In the following section, I discuss an approach to rule-based MT which have some interesting contributions to make to the problems and issues discussed in this and the previous section.

## 4 Constraint-Based Translation

The specific challenge for transfer-based MT is that of extending mainstream computational linguistic formalisms with transfer components which exploit appropriate techniques. Many transfer systems do in fact use some sort of unification based attribute value grammar both for the monolingual components and also in transfer. In addition to the formal properties mentioned above, such formalisms and the associated linguistic theories of course have much to offer in terms of the description of a wide range of phenomena. In particular, re-entrance or structure-sharing in feature structures provides the basis for elegant and intuitively satisfying accounts of phenomena such as control, raising, ellipsis, anaphora and long distance dependencies.

Extending the traditional tree transducing transfer method to feature structures is straightforward. Assume that the monolingual component is organised to produce in analysis some sort of semantic representation as input to transfer. This source FS can be mapped to a target FS by means of a grammar of transfer rules whose lefthand sides are matched against the source representation and whose righthand sides indicate the content of the corresponding target FS. The information combining process in the target FS is unification. In a number of such formalisms, the transfer rules are in effect bilingual lexical entries. For example, a rule like (2) would apply to a source structure such as (la), producing, in conjunction with other rules, the structure in (lb). (2) states a bidirectional correspondence between a FS containing the PRED *like* and a FS containing the PRED *aimer* and also states that further correspondences (given by other transfer rules) must hold between the French and English ARG1s and ARG2s:

(la)

| | | |
|---|---|---|
| **PRED** | like | |
| **ARG1** | [ **PRED** | sam ] |
| **ARG2** | [ **PRED** | kim ] |

| | | |
|---|---|---|
| **PRED** | aimer | |
| **ARG1** | [ **PRED** | sam ] |
| **ARG2** | [ **PRED** | kim ] |

(2)

```
EN:                 FR:
PRED = like         PRED = aimer
ARG1 = El           ARG1 = Fl
ARG2 = E2           ARG2 = F2

CORRES: El <-> Fl, E2 <-> F2
```

Rule application is compositional in the source feature structure - in a top-down approach, rules are recursively applied to successively smaller collections of source language attribute value pairs. Like classical tree-based transfer systems, this is 'structural' or 'derivational' in working off representations in the familiar way. This is essentially the approach employed in ELU [9], and MiMo2 [16], in which transfer rules or correspondences hold or are stated *between feature structures,* and rules are applied recursively to feature structures. In the rest of this paper we will try to show the alternative strategy of interpreting AV grammars as systems of constraints which structures may or may not satisfy, with no commitment to recursion through (one dimension of) the source object itself helps overcome some of the problems discussed in the previous sections and in particular has some descriptive advantages in the specification of transfer.

By constraint based approaches, then, I mean approaches which are not derivational or structural in the sense of involving a traversal of or recursion through the source representations. The essence of constraint-based translation is the very simple idea that specifying transfer should simply be the statement of local equalities which are interpreted as constraints over the target structure. It is not possible, of course, to do justice here to the intricacies of particular analyses of linguistic data made available, nor to the relationship between different approaches, or even to address in an adequate manner the many open issues raised by these approaches, but I hope to provide a sketch of why I think these approaches are important. I begin by sketching out two different approaches to constraint-based translation.

## 4.1 Generation from Lexical Signs

A number of contemporary unification-based formalisms are *sign-based,* in that the basic unit of description is a sign which simultaneously describes a linguistic object along a number of different dimensions (e.g. in terms of the string, the phrase structure or parse tree, the intrinsic syntactic and semantic properties, and some representation of the context). The best known of these formalisms are HPSG and unification-based versions of categorial grammar [17], [27]. Typically, these dimensions occur as values of distinct attributes within a single structure.

The first and simplest model of constraint based translation that I will discuss views translation as a problem of generation from a bag of lexical signs. In parsing with a sign-based formalism, analysis begins with a string of underspecified signs from the lexicon (retrieved after morphological analysis). Parsing a source string produces a successively larger structures, unifying information in the component signs in various ways, and simultaneously further specifying the content of the lexical signs. Now one way of avoiding recursion through a source structure, often referred to as the 'shake-and-bake' approach [26, 4] is to take just the string of lexical signs (which have become instantiated by the parsing process) as the input to transfer, and map them to their target equivalents, preserving certain 'transfer' properties (e.g. their semantics). Notice that the output of transfer is **not a structure** at all. Now, the idea is that there are very few ways that these signs can be combined by the target grammar to produce a single sign - and the way the target grammar operates will mean that this sign will necessarily have essentially the same transfer properties as the source sign. Thus, all one needs to do is to process the 'bag' of target lexical signs with the target grammar, and one will produce a target sign which is equivalent to the source sign. The obvious way to do this is just to parse the target signs, that is, synthesis by analysis.

For example, parsing the string of English lexical signs *Sam*, *sees, the,* and *secretary* will produce a sign for the whole sentence, and instantiate the semantics on each lexical sign. These signs can then be looked up in a bilingual lexicon, giving the bag of corresponding French signs *Sam*, *voir, le, secrétaire,* with semantics unified with that of the corresponding source items (since *secrétaire* can be masculine or feminine in French, one would also have to consider the bag containing *la,* but we will ignore this here). Normally in parsing, one has an ordered list of signs, whose syntactic and semantic relations are underdetermined. In this case, one has an unordered bag of signs, whose semantics are determined. However, normal parsing techniques are still applicable – crudely, one simply tries all possible orders (one 'shakes' the 'bag' of target lexical signs to obtain alternative orders, and tries to 'bake' to produce a single target sign). This process will produce *Sam voit le secrétaire.* The identity between the semantics of *see* and *voir* that was established in transfer means that *Le secrétaire voit Sam* is not produced (which would require a different semantics). **voit Sam le secrétaire* is not produced because the target grammar rules do not permit this, and agreement between subject and verb is achieved by the target grammar rules in the same way.

## 4.2 LFG Co-description

The second example of constraint-based transfer that I will discuss involves the multi-dimensional but not sign-based theory of LFG. In LFG, the two levels of syntactic description, c-structure and f-structure, are related by a mapping function or projection $\phi$ from c-structure nodes to feature structures. Any number of functions can in principle be defined to simultaneously classify a linguistic object along a number of dimensions. Since projections are functions, they may be composed, opening up wide descriptive possibilities for relating levels. The theory also allows for the use of inverse functions (e.g. $\phi^{-1}$ from f-structure to the associated c-structure nodes). Kaplan *et al* [12] show how the LFG constraint language can be used to state bilingual correspondences. They define two translation functions $\tau$ (between f-structures) and $\tau'$ (between semantic structures). Semantic (s-) structures themselves are projected from f-structures by means of the mapping function $\sigma$. By means of these functions, one can 'co-describe' elements of source and target f-structures and s-structures respectively. Achieving translation can be thought of in terms of specifying and resolving a set of constraints on target language descriptions, constraints which are expressed by means of the $\tau$ and $\tau'$ functions.

Of course the availability of function composition opens up some rich expressive possibilities for stating bilingual correspondences: $\tau$ and $\phi$ can be composed, as can $\tau'$ and $\sigma$. For ease of exposition, we will initially limit attention to $\tau$, the projection from f-structure to f-structure. The basic idea of this approach to translation is as follows. A bilingual constraint is exactly like a monolingual constraint, except that it makes reference to levels of description in both source and target languages. For example:

(3)
$(\tau\ (\uparrow \text{SUBJ})) = ((\tau\ \uparrow)\ \text{SUBJ})$

which composes $\tau$ and $\phi$, states a (target side) equality between the $\tau$ of the source SUBJ f-structure and the value of the SUBJ attribute of the $\tau$ of the source f-structure containing the subject. The equation simply states an equality between two paths, hence any satisfying structure must be the value of these two paths. Although it is slightly misleading to talk in terms of structures, because there is no recursive rule application to the source f-structure, we can view (3) as saying that the translation of the value of the SUBJ slot in a source f-structure fills the SUBJ slot in the f-structure which is the translation of the source f-structure immediately containing that SUBJ slot. Monolingual target language constraints can also be stated, for example, (4) says that the value of the PRED attribute in the target f-structure is *voir.*

(4)
$((\tau\ \uparrow)\ \text{PRED FN}) = \text{voir}$

Constraints such as these are added to the lexicon and c-structure rules alongside the monolingual constraints. In parsing the source language string one gathers a set of constraints describing the source language f-structure and another set of constraints describing the target language f-structure. The solution of this latter set is a (probably incomplete) target f-structure which must then be completed and validated by the target grammar.

For concreteness, we give the set of equations for (5):

(5)
Sam saw Kim

(6)
see, V
$(\uparrow \text{PRED}) = \text{see}$
$((\tau\ \uparrow)\ \text{PRED FN}) = \text{voir}$
$(\tau\ (\uparrow \text{SUBJ})) = ((\tau\ \uparrow)\ \text{SUBJ})$
$(\tau\ (\uparrow \text{OBJ})) = ((\tau\ \uparrow)\ \text{OBJ})$

kim, N
$(\uparrow \text{PRED}) = \text{kim}$
$((\tau\ \uparrow)\ \text{PRED FN}) = \text{kim}$

sam, N
(t PRED) = sam
$((\tau\ \uparrow)\ \text{PRED FN}) = \text{sam}$

These constraints co-describe the structures in (7) and (8): [2]

(7)

$$
e1 \left( \begin{array}{lll}
\textbf{SUBJ} & e2 & \left[ \begin{array}{ll} \textbf{PRED} & \text{sam} \\ \textbf{NUM} & \text{sing} \end{array} \right] \\
\textbf{OBJ} & e3 & \left[ \begin{array}{ll} \textbf{PRED} & \text{kim} \\ \textbf{NUM} & \text{sing} \end{array} \right] \\
\textbf{PRED} & \multicolumn{2}{l}{\text{see} < \text{SUBJ, OBJ} >} \\
\text{TENSE} & \multicolumn{2}{l}{\text{past}}
\end{array} \right)
$$

(8)

$$
f1 \left( \begin{array}{lll}
\textbf{SUBJ} & f2 & [\ \textbf{PRED}\ \ \text{sam}\ ] \\
\textbf{OBJ} & f3 & [\ \textbf{PRED}\ \ \text{kim}\ ] \\
\textbf{PRED} & \multicolumn{2}{l}{\text{voir} < \text{SUBJ, OBJ} >.}
\end{array} \right)
$$

To reiterate, the crucial point is that there is no separate recursive application of a set of transfer rules to a source f-structure - the constraints stated simply codescribe the source and target structures. Although one could first build the source feature structure and then the (partially described) target feature structure, there is also no necessary commitment to any one order of evaluation for the two sets of constraints.

## 4.3 Multilevel Transfer

To be adequate, translation formalisms must be able to deal with linguistic information expressed in a very wide range of different forms. In traditional rule-based systems, such information, whatever its linguistic source, must be channeled through one level of representation. For various reasons to do with modularity, linguistic purity and reusability, however, this particular organisation is problematic, as well as being difficult to design. At first glance, sign-based formalisms would seem to allow for the direct use in transfer of information expressed in various linguistic dimensions (and associated with these different features in the sign or linguistic structure) without falling into the problems associated with the hybrid but abstract level of representation of the standard transfer.

However, the only straightforward way in which the transfer algorithm can apply rules to *signs* is by recursing through the structure of a feature or attribute which encodes the derivational history of the sign itself (such as the HPSG dtrs attribute). Alternatives involving the explicit threading of translational relevant information "to the right place" complicate the monolingual grammars and with unfortunate results for reusability, modularity, task independence, etc.

[2] We assume here that English is the source. Notice that information irrelevant information or information which can be recovered from the monolingual grammar is omitted.

Of course, writing transfer rules of the basis of what is in effect the derivational history or parse tree of the final sign is counterintuitive at best. For example, this means that the set of transfer rules will be unnecessarily complicated (the whole point of normalisation and abstract in standard transfer systems is precisely to minimise the differences between languages and therefore the complexity of transfer).

In a constraint-based approach on the other hand, it should be clear that it is possible to mix information pertaining to different levels of linguistic description. In the Shake and Bake approach, for example, any properties of the source sign at all can be used in choosing the target lexical signs. As we have described it, the requirement that source and target items have the *same* semantics makes it appear 'interlingual', but this is not essential, so long as the semantics instantiated on the target items is sufficient to constrain the synthesis process (target language parsing) to producing expressions which are equivalent to the source expressions. Identifying source and target semantics is one approach, but it would also be possible to manage with parts of the semantics, and parts of the syntax, for example.

In the LFG co-descriptional approach, the descriptive apparatus of projections allows for multiple levels of structure to be related by separate correspondences. Kaplan *et al* define $\tau'$ between semantic structures, where the $\sigma$ correspondence maps from f-structures to semantic structures. For example,

(9)
$\tau' \ (\sigma \uparrow \text{ARG1}) = (\sigma \uparrow \text{ARG1})$

asserts an identity between the values of ARG1 in source and target semantics. This would be appropriate if the values were, for example, semantic indices. The equation:

(10)
$\tau' \ (\sigma \uparrow \text{ARG1}) = (\sigma \ (\tau \uparrow \ \text{TOPIC})$

states an identity between the translation of the ARG1 in the source semantics and the semantics associated with the TOPIC of the target f-structure. Constraints such as these can also be given as further annotations to the c-structure rules and within the source lexicon, making possible the statement of constraints over multiple levels of both source and target structure, whilst still maintaining the coherence of the levels of representation in question. There are various ways in which one might want to use this facility for "multi-level" transfer, depending in part on the linguistic content of the various levels of representation related by different bilingual mapping functions. For example, one can imagine discourse-oriented information being used for certain types of disambiguation. Or transfer could be attempted at some level of semantic structure, with f-structure correspondences being used as a fall-back position, for reasons of robustness. Phenomena for which there are (at least the beginnings of) an adequate interlingual treatment (time and aspect, for example) could be factored out and dealt with interlingually, with time and aspect information in the other levels for which bilingual correspondences are stated simply ignored (see the section on completeness, below). There is little danger of translationese, since there is no one source structure as organising level for translation. A multi-level architecture also allows one to vary the depth of transfer according to the language pair in question, using only translational correspondences between surface-oriented information for very closely related languages, but relying on information from deeper levels of analysis for less closely related language pairs. In particular, this is possible without any consequences at all for the monolingual grammars, so that the same source grammar can be used irrespective of the level at which transfer correspondences are stated for any given pair. In these ways, then, this constraint-based approach seems to provide an interesting response to the need for multilevel transfer.

## 4.4 Translating Re-entrancies

A key advantage in terms of linguistic adequacy of feature structures and attribute value grammar formalisms is that they permit re-entrance or structure sharing. In structure-based transfer it is fairly straightforward to deal with 'local' re-entrancies (such as that occurring in control constructions), which fall within the scope of one transfer rule. On one approach (taken in ELU [9]), re-entrancies which fall within the domain of a single transfer rule can be translated by binding the re-entrant paths within the (input) structure to the same variable and stating a correspondence between the relevant source side and the target side variables. In this way the re-entrancy is translated as one structure. Of course, source side re-entrancies can be ignored (or effectively unfolded) in transfer, and target side re-entrancies created where there is no re-entrant source, by means of the same strategy. An alternative (taken in MiMo2, [16]) is to translate the re-entrant paths separately, but mention the re-entrancy explicitly on source and/or target sides, requiring token-identity (i.e. re-entrance) between the results of the separate translations.

However neither approach can be simply generalized to 'long-distance' re-entrancies, which are typically used to encode long-distance dependencies, and for such phenomena, these approaches can provide no general treatment, because the re-entrant paths can be structurally remote from each other, and thus will often fall outside the scope of any transfer rule.

Of course, there are a number of ways in which one might try to remedy this inadequacy. For example, one could unfold the re-entrances as type identities (i.e. reinterpret the FS as a tree), or 'thread' shared values through the structure, in such a way that they become local. However, none are satisfactory. The former is no more than a temporary expedient, for it loses the descriptive advantages of using feature structures, looses information from the source structure and causes problems in generation, where some method must be found for ensuring that lexical content is not duplicated, and appears in the right place. Threading techniques are unattractive because of the (often extreme) complication they introduce in grammars and representations.

The fact that there is no recursive decomposition of a source feature structure makes it possible to deal with the translation of long-distance re-entrancies without difficulty using the LFG $\tau$ projection.[3] In the case of a wh-question or relative clause, the (source side) wh-element is introduced by a c-structure rule which also associates it with an attribute (either FOCUS or TOPIC) in the f-structure. The value of this attribute is re-entrant with the value of some other attribute (e.g. with the value of OBJ). A translation correspondence for this attribute will (typically) be given in an equation in the lexical entry for the PRED of that f-structure (exactly as above). By annotating the relevant c-structure rule with a bilingual correspondence, we specify a constraint over the translation of the source FOCUS or TOPIC (for example, by means of the annotation $(\tau\ (\downarrow\ \text{TOPIC})) = ((\tau\ \downarrow)\ \text{TOPIC})$ on the S' which introduces a relative clause). Notice that since $\tau$ is a function and the same source language structure is the argument of $\tau$ in these two equations, the target f-structure is required to contain a re-entrancy between the two attributes specified in that target f-structure. Thus long-distance re-entrancy can be simply treated.[4]

## 4.5 Completeness

Although there are various approaches to structure-based transfer, they have in common the need to visit every node or attribute of the source representation. In the systems we have mentioned, transfer rules must be exhaustively applied to every substructure of the source structure. This is necessary to ensure the completeness of translation, that is, to make sure that no part of the source representation (and therefore the source text) has been missed. Various checks are incorporated into the rule application algorithm to carry out completeness checks. The consequence of this is that if an element in one language simply fails to translate, but is nonetheless present at the representational level at which transfer is defined, then explicit rules translating that element as "nil" must be written. In other systems (METAL [24] is an example) transfer rules are essentially attached as procedures to nodes in the source trees and it is also necessary to ensure complete traversal of the tree.

Because constraint-based approaches are not concerned with structures *per* se, and in particular because there is no recursive application of transfer rules to source feature structures, there is no requirement that every part of the feature structure be either translated or explicitly left untranslated. In the Shake and Bake approach, within certain limits, the information given in bilingual lexical entries can be partial. In the LFG approach, stating transfer relations at the rather superficial level of f-structure does not entail

[3] Transfer between TAGS also avoids this problem [1]

[4] There are some complications where the languages differ in the re-entrancies they permit. In these cases, the possibility of underspecification of the target f-structure must be exploited. This is discussed in [20] and [3].

the writing of transfer equations for elements such as pleonastics.

Consider the translation from French to English of (11)[5]:

(11)
Il est probable que Kim viendra
Kim is likely to come

The lexical entry for *probable* will contain equations stating a correspondence for the sentential argument, but no correspondence for the pleonastic SUBJ. The lexical entry for *venir* will state a correspondence for its SUBJ. The French SUBJ *il* is thus left untranslated.

(12)
probable A
(↑ PRED) = 'probable< *COMP* > *SUBJ'*
(↑ SUBJ FORM) = il
(($\tau$ ↑) PRED FN) = likely
($\tau$ (↑ COMP)) = (($\tau$ ↑) XCOMP)

(13)
venir V
(↑ PRED) = venir < *SUBJ* >
(($\tau$ ↑) PRED FN) = come
($\tau$ (↑ SUBJ)) = (($\tau$ ↑) SUBJ)

These $\tau$ equations (and further equations from the lexical entry for *Kim*) only give a partial description of the f-structure we require (for example, there is no value assigned for the SUBJ of *likely*). Further equations, such as that stating a re-entrance between the SUBJ and the XCOMP SUBJ, come from the English lexicon and grammar in generation.

Both the Shake and Bake approach and LFG translation by co-description allow a nice division of labour between monolingual and bilingual components, providing some solution to the problem of redundancy. In Shake and Bake, transfer stipulates the semantics, and the lexical signs to be used, but all other decisions are left to the target grammar. In LFG translation the output of transfer can be heavily underspecified, with no redundant repetition of information contained within the monolingual dictionary or grammar.

# 5 Open Issues

While there would seem to be a number of advantages to these constraint-based approaches to MT, there axe many open issues.

In Shake and Bake, it seems to be a requirement that all source-derived

[5] For simplicity, we do not consider the translation of tense information here.

information necessary for generating the correct target expressions must be present in a target lexical sign. This poses a problem with properties that are not realized lexically, but are expressed by word order. Unless these properties can be 'transfer properties' of lexical signs, abstract lexical items must be introduced to carry the information. This sort of contamination of monolingual grammars is precisely what we wish to avoid.

Furthermore, although it is clear how pieces of structure can be left untranslated in the LFG approach, it would seem to be a requirement in Shake-and-Bake that all lexical signs in the source language occur in at least one bilingual lexical sign. This means that lexical items which have no translation must be "translated" as nil, which will result in arbitrary insertion into the target string when the direction of translation is reversed. At the very least, this will lead to an explosion in the amount of work for generation.

While the Shake and Bake approach is bidirectional, the LFG projections are directional, and it is not clear how the inverse of these projections can be used. Of course the same monolingual grammar is still used for analysis and generation.

There are a number of interesting questions about the expressivity of constraint-based approaches compared with structure-based approaches to transfer. For example, Sadler and Thompson [19] show that head-switching is not possible in the LFG approach as currently formulated, while it is possible in structure-based approaches [18]. There are a number of open questions concerning the integration in generation of information from different levels in more than toy applications [8, 20, 3].

## References


[1] Anne Abeillé, Yves Schabes & Aravind K. Joshi, (1990). "Using Lexicalized TAGS for Translation", **Proceedings of COLING 90,** Helsinki, Vol 3, pp 1-6.

[2] D.J. Arnold & L. Sadler, (1990). "The Theoretical Basis of MiMo". **Machine Translation,** Vol 5, pp 195-222.

[3] Doug Arnold & Louisa Sadler, (1992). "Rationalism and the Treatment of Referential Dependencies". **Proceedings of the Fourth International Conference on Theoretical and Methodological** Issues **in Machine Translation,** Montreal, Canada, pp 195-204.

[4] John Beaven, (1992). "Shake-and-Bake Machine Translation". To appear in **Proceedings of COLING 92,** Nantes.

[5] Peter Brown, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer & Paul S. Roossin, (1990). "A Statistical Approach to Machine Translation". **Computational Linguistics,** Vol 16/2, pp 79-85.

[6] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, John D. Lafferty & Robert L. Mercer, (1992). "Analysis, Statistical Transfer and Synthesis in Machine Translation". **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation,** Montreal, Canada, pp83-100.

[7] Helge Dyvik, (1989). "The PONS Project: Features of a Translation System", Department of Linguistics and Phonetics, University of Bergen.

[8] Kurt Eberle, Walter Kasper & Christian Rohrer, (1992). "Contextual Constraints for MT". **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation,** Montreal, Canada, pp 213-224.

[9] Dominique Estival, Afzal Ballim, Graham Russell & Susan Warwick, (1990). "A syntax and Semantics for Feature Structure Transfer". **Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation,** Austin, Texas.

[10] Ulrike Heid & Jock McNaught, (1991). "Feasibility and Project Definition Study on the Reusability of lexical and terminological resources in computerized application", CEC, Luxembourg.

[11] Jerry R. Hobbs & Megumi Kameyama, (1990). "Translation by Abduction". **Proceedings of COLING 90,** Helsinki, Vol 3, pp 155-61.

[12] R.M.Kaplan, K. Netter, J. Wedekind & A. Zaenen, (1989). "Translation by Structural Correspondence". **Proceedings of the Fourth European ACL,** Manchester, pp 272-81.

[13] S.P.Jan Landsbergen, (1987). "Montague grammar and machine translation". In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod L. Johnson & Paul Bennett (eds) **Linguistic Theory and Computer Applications,** Academic Press, London, pp 113-147.

[14] Michael C. McCord, (1989). "Design of LMT: A Prolog-Based Machine Translation System". **Computational Linguistics,** Vol 15/1, pp 33-52.

[15] Sergei Nirenburg, (1989). "Knowledge-Based Machine Translation". **Machine Translation,** Vol 4/1, pp 5-24.

[16] Gertjan van Noord, Joke Dorrepaal, Pirn van der Eijk, Maria Florenza & Louis des Tombe, (1990). "The MiMo2 Research System". **Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation,** Austin, Texas.

[17] Carl Pollard & Ivan Sag, (1992, forthcoming). **Head-Driven Phrase Structure Grammar,** CSLI & University of Chicago Press.

[18] Louisa Sadler (1991). "Structural Transfer and Unification Formalisms". **Applied Computer Translation,** Vol 1/4, pp 5-21.

[19] Louisa Sadler & Henry Thompson, (1991). "Structural Non-Correspondence in Translation". **Proceedings of the Fourth European ACL,** Berlin, pp 293-298.

[20] Louisa Sadler & Doug Arnold (1992) "A Constraint-Approach to Translating Anaphoric Dependencies". To appear in **Proceedings of COLING 92,** Nantes.

[21] Louisa Sadler & Doug Arnold (to appear) "Unification and Machine Translation". To appear in **META.**

[22] Victor Sadler (1989) **Working with Analogical Semantics: Disambiguation Techniques in DLT,** Foris, Dordrecht.

[23] Satoshi Sato & Makoto Nagao, (1990). "Toward Memory-based Translation", **Proceedings of COLING 90,** Helsinki, Vol 3, pp 247-52.

[24] J. Slocum, W.S. Bennett, J. BEar, M. Morgan & R. Root, (1984) "METAL: The LRC MT System", LRC 84-2, Linguistics Research Center, University of Texas.

[25] Jun-ichi Tsujii & Kimikazu Fujita (1991) "Lexical Transfer Based on Bilingual Signs: Towards Interaction During Transfer". **Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics,** Berlin, pp 275-280.

[26] Whitelock, Pete, (1991) "Shake and Bake Translation" MS., Sharp Labs of Europe, Oxford.

[27] Henk Zeevat, Ewan Klein & Jo Calder (1987) "An Introduction to Unification Categorial Grammar". In N.J. Haddock , E Klein & G. Merrill (eds) **Categorial Grammar, Unification Grammar and Parsing,** *Edinburgh Working Papers in Cognitive Science Vol 1,* pp 195-222.