

Reuse of linguistic resources in MT

Daive Turcato Fred Popowich Paul McFetridge Olivier Laurens

Natural Language Lab
School of Computing Science
Simon Fraser Univ.
Burnaby, BC
Canada V5A 1S6

{turk,popowich,mcfet,laurens}@cs.sfu.ca

Abstract

Machine translation (MT) draws more heavily on lexical resources than most other NLP applications. First, grammars of both source and target languages require lexicons, Second, some sort of mapping between lexicons is required in order to transfer information from a source to a target language. The MT system described here is based on Shake-and-Bake technology and uses lexical transfer as the interface between language pairs, necessitating a third lexicon for each language pair. Consequently, efficient reuse of existing lexical resources can reduce development cost and time. We describe methods of reuse from three perspectives: adapting NLP resources to a new application, adapting electronic reference resources to an NLP application and designing NLP resources for reuse in related but novel domains. The effort required to adapt a lexicon to a new application is found to be a function of both the amount of information available from the lexicon and its role in the new application. When system design isolates a module from its function, integration of a new lexicon can be virtually automatic. When the module is application dependent - as is the case with bilingual lexicons - integration is semi-automatic with an obligatory post-editing phase.

1. Introduction

Reusability is widely recognized as a desirable property for Natural Language Processing (NLP) resources. The issue is particularly relevant in commercial applications, where time and money are critical issues. Such relevance is further increased when Machine Translation (MT) is the domain involved. MT is a complex task which involves virtually every aspect of NLP. MT applications need multiple components and large scale linguistic resources.

Despite the importance attributed to reusability in the literature and the description of many resources which are claimed to be reusable to some extent, there are very few descriptions of applications actually reusing pre-existing linguistic resources (among them (Rayner et al., 1996) and, at an experimental level, (Arnold et al., 1993)).

Reusability of existing resources is limited by two problems:

1. Resources can be application domain dependent. The specific purpose for which a resource is developed is often reflected in its structure and content, thus making its reuse difficult in a different application domain.
2. Linguistic information is always expressed in some specific formalism reflecting some underlying theoretical approach. The re-statement of the same information under a different formalism and theoretical approach is far from trivial.

The solutions to the problem vary considerably, ranging from the specification of standards for resource development (Calzolari & Zampolli, 1994) to the implementation

of migration procedures (Arnold et al., 1993). In the former case, the gap between different application domains and theoretical approaches is avoided at the source, by recommending a *multifunctional* approach to resource development (Calzolari & Zampolli, 1994, p. 4) and the adoption of some abstract, theory neutral formalism. In the latter case, the gap is overcome by defining sophisticated mappings between different formalisms and theoretical backgrounds.

In this paper a different approach to the issue of reusability is proposed. The focus is shifted from the reusable resources to the applications in which the resources are to be incorporated. We emphasize how the adoption of appropriate knowledge engineering principles in developing applications is relevant to an effective reuse of external resources. Instead of, or in addition to, placing restrictions on formalisms or defining complex migration procedures, in order to adjust resources to a range of specific applications, we advocate the design of NLP applications whose architecture is flexible enough to accommodate resources expressed in a wide range of unrelated formalisms, while keeping the migration procedures reasonably simple. We discuss how the ideas presented have been concretely applied to a commercial MT application in which pre-existing linguistic resources have been massively and successfully reused.

Here is an outline of the paper. In section 2 we provide some background about our application domain and MT system. In section 3 we discuss the architecture of the system and its relevance to resource reusability. In sections 4 through 6 we discuss the actual resources reused in our system, the related porting procedures, the specific problems posed by each of them and the way our knowledge engineering decisions affected the porting procedure. In section 7, conclusions are drawn.

```

morning.
morning, mj .
hi.
wow.
uh-huh.
pretty wild.
how you doing?
got an estimate for me?
about eight hours.
mm-hmm.
i am seeing ligature marks.
petechia.
strangled, huh?
same as the others.
yeah.
bag all the stuff on the bedside table?
thanks, bill.
that's the only one i need.
mike.
yes, madam.
were you the first one here?
yes.

```

Table 1: Script fragment

2. System overview

Large scale MT systems are typically designed for the translation of documents, or for some very narrow semantic domain. Our goal was to design an MT system with a very wide semantic domain, but with a narrow syntactic domain. Specifically, the goal was to translate the English found in North American closed captions into understandable Spanish.

The sentences and phrases used in closed captions tend to be relatively simple, with the average sentence length being approximately six words. The text tends to contain many idioms, and frequently contains expressions like those encountered in speech rather than in documents, as illustrated in the closed captioned text in Table 1 from (Popowich et al., 1997). Although the amount of syntactic information is relatively constrained and different from that used in traditional MT systems, a large amount of lexical information is needed. Thus there was an incentive to reuse existing lexical resources as much as possible.

We developed our MT system according to the lexicalist approach known as Shake-and-Bake (henceforth S&B). Like traditional transfer systems, an S&B system can be described in terms of analysis, transfer and generation. The goal of the analysis phase in the S&B approach is to produce a set of enriched lexical entries associated with the different words in the source language sentence. They are enriched in the sense that they contain additional syntactic and semantic information above what was present in the original lexical entries retrieved from the lexicon. During the analysis phase, lexical entries get combined according to the grammar rules and, through unification, values that were originally unspecified in the feature structures from the lexicon, become instantiated in the feature structures in the parse tree. After analysis, these enriched feature struc-

tures for the lexical entries are used as input to the transfer module, which uses lexical transfer rules, contained in a bilingual (or multilingual) lexicon, to create a corresponding set of feature structures for the target language. The generator then uses this set of feature structures as input, together with the target language lexicon and grammar, to produce an output sentence. Generation is performed by ordering the input bag of lexical items in some way consistent with the constraints expressed by the target grammar. In (Whitelock, 1994), for instance, this is done by parsing every possible permutation of the target bag until a successful parse is found.

Typically, a transfer based system requires structural transfer rules as well as lexical transfer rules. With the lexicalist approach, only lexical transfer rules are needed. There is sometimes redundancy between lexical transfer rules and structural transfer rules, so by placing all the information in the lexical transfer rules, this redundancy can be eliminated. The lexicalist approach is also attractive for translating idiomatic expressions, where the translation of complex phrasal structures can be simply specified in the transfer lexicon.

3. Knowledge engineering and reuse of resources

The S&B approach to MT was explicitly designed with awareness that portability is one of the most important requirements that any MT system, or, more generally, any NLP system, should fulfill (Whitelock, 1994, pp. 340-341). Whitelock (1994) lists three knowledge engineering principles concerning portability:

1. Linguistic knowledge should be independent of the programs that use it. Therefore, linguistic knowledge and underlying algorithms should be clearly kept separate.
2. Linguistic knowledge should be independent of translation direction and should be usable in both directions. Therefore should be expressed in a declarative formalism.
3. Linguistic knowledge should be independent of specific language pairs. Therefore, its development should be guided as much as possible by purely monolingual considerations, with no regard to the other language in the system.

While the first requirement can be directly extended to any other NLP application domain, the other two are more MT specific. However, their underlying motivations should be intuitively clear and it should be possible to find equivalent requirements for different application domains.

What features of the S&B architecture are relevant to the fulfillment of the above requirements? The key feature, in terms of bidirectionality of grammars, is that grammars and lexicons are expressed in a unification-based formalism that uses feature structures to represent linguistic objects (lexical entries and grammar rules). In accordance

with mainstream linguistic approaches (like HPSG), lexical entries are multidimensional signs, encoding different levels of representation at the same time (Pollard & Sag, 1994). In this way, sequential mappings between different levels of representation, typical of procedural models of MT, can be replaced by monostratal grammars, where the relation between different levels of representation is declaratively expressed in the form of constraints. Both parsing and generation aim at finding a relation between a string and a representation of some sort, be it a bag of instantiated lexical items, as in S&B, or a tree-like structures, as in other bidirectional approaches (Van Noord et al., 1991). The only difference between parsing and generation is which term of the relation is given first and which one needs to be computed. However, both processes use the same sort of linguistic resources, while differing, of course, in the algorithm.

The independence of linguistic resources from specific language pairs rests on the peculiar architecture of the transfer component. Transfer is a mapping between bags of lexical items. No structural transfer is performed. Transfer is minimally performed by equating variables (indices) on source and target lexical items. Such indices express semantic dependencies among lexical items. The linguistic knowledge used by transfer can thus take the form of a bilingual lexicon. Again, linguistic equivalences are declaratively expressed, with no regard to a specific direction. Moreover, the crucial aspect, in terms of independence from language pairs, is that the absence of any mapping between structures (i.e. trees associated to by grammatical rules), removes the necessity of tuning the grammars to each other. Grammars can be developed independently from each other and according to different theoretical approaches, because no assumption about what the other grammar should look like is necessary.

In addition to the foregoing, we adopted the further methodological principle of systematically using the macros typically found in grammar development systems as a means to interface different components (monolingual lexicons and grammars, grammars and bilingual lexicon, source and target sides of a bilingual lexicon). This approach was very effective in favouring the porting of external resources, as will be discussed later. A detailed description of the use of transfer macros, with particular reference to inflectional information, can be found in (Turcato et al., 1997). Some further remarks related to knowledge engineering aspects are added here.

A bilingual lexical entry contains three sorts of information: words, syntactic descriptions and bilingual constraints. Both syntactic descriptions and bilingual constraints are clearly grammar dependent, in the obvious sense that they use the same feature structures as used by either grammar. However, syntactic descriptions have the only purpose of matching monolingual lexical entries, hence they are not language pair dependent, whereas bilingual constraints clearly are. Bilingual constraints, aside from indexing, typically take the form

$$(F_{1,1}:V_{1,1} \dots F_{1,n}:V_{1,n}) \leftrightarrow (F_{2,1}:V_{2,1}, \dots, F_{2,m}:V_{2,m}).$$

where F and V stand, respectively, for *Feature* and *Value*

and the first of each pair of subscripts refers to one of the two languages at hand (*Language1* or *Language2*). If a lexical item on one side of a bilingual entry satisfies the respective description in a constraint, the lexical item on the other side must also satisfy the respective description in the constraint. Aside from idiosyncratic, word specific constraints, the set of all the constraints in a bilingual lexicon can be regarded as a language pair dependent and linguistic framework dependent mapping between linguistic phenomena in the two languages (such as tense, aspect, gender and number). In our bilingual lexicon, we adopted the decision of expressing all such bilingual constraints by means of transfer macros, created on a linguistically principled basis. The purpose of a transfer macro is to express linguistic generalizations over pairs of languages in a compact form. In this sense, its purpose is similar to that of transfer rules in traditional transfer approaches, with the crucial advantage that, in the case of transfer macros, no structural mapping is performed. The methodological decision of systematically using transfer macros introduced two clearcut separations:

- between language pair dependent information (bilingual constraints) and language pair independent information (syntactic descriptions);
- between linguistic framework dependent implementations (the content of transfer macros) and the linguistic phenomena they accounted for (represented by the transfer macro labels associated to bilingual entries).

In the following, we discuss the implications of the described methodological principles in the effective reuse of external resources. In doing so, we focus on lexicons (both monolingual and bilingual), for this is the kind of resources that we most widely reused in our MT system. The foregoing methodological principles are relevant to the reuse of lexicons in several ways:

1. It has been shown that the described architecture guarantees a wider capability of reusing existing grammars and a greater liberty in developing new grammars. As grammars and monolingual lexicons are clearly interdependent, because they use the same linguistic representations, the foregoing considerations about the portability of grammars obviously extend to lexicons. Moreover, one can more easily take advantage of the situation in which a reusable lexicon is available and a related grammar has to be developed, by adapting the grammar design to the characteristics of the available lexicon.
2. Although interdependencies among resources obviously still exist, for each resource needs to interface with some other resource (e.g. a lexicon with a grammar), such interdependences are limited to binary ones, i.e. no interdependency involves more than two resources at one time. Therefore, the porting of an external resource has to take into account only one module with which to interface, at most. The independence of modules described above carries over to the independence of porting procedures.

3. A constraint-based approach makes easier to effectively reuse partial or incomplete resources. A constraint-based approach is inherently incremental, hence it is conceptually straightforward to integrate an incomplete resource with further constraints.
4. Where language pair and linguistic framework dependent information is present, its clearcut and principled separation makes its replacement easier when porting a resource to a different language pair or linguistic framework.
5. The systematic use of macros allows one to express generalizations and to distinguish between abstract linguistic phenomena and their specific, language dependent implementations. Abstract phenomena are expected to be cross-linguistically more invariant than implementations.

We were able to reuse pre-existing resources for all the lexical components of our system (monolingual lexicons and bilingual lexicon). The migration procedures and their degree of success will be discussed in some detail. Our discussion of reusability will consider three different classes of reuse:

1. the reuse of existing resources that were originally developed for NLP;
2. the reuse of existing linguistic or lexicographic resources that were not originally developed for NLP;
3. the development of resources capable of being reused in future applications.

4. Reuse of NLP resources

4.1. An existing monolingual lexicon

A monolingual Spanish lexicon was created from an existing commercial Spanish lexicon for our system. A conversion procedure was developed which mapped source lexical items onto lexical items expressed in our formalism. The migration didn't pose any particular problems. Information in the source lexicon was stated in a feature structure formalism similar to that of our system. As well, the kind of information contained in the source lexicon could be easily reformulated in terms of the categories used by our system. The migration process was carried out according to two main guidelines:

- The conversion was performed *incrementally*. The rationale behind this decision was that the conversion was done at the same time as the Spanish grammar was developed. Therefore the migration was done in several steps. At each step more information was converted, depending on the capability of the grammar to make use of such additional information. For instance, the source lexicon contained semantic selectional restrictions. Since no selectional restrictions were initially handled by the grammar, this kind

of information wasn't converted in the initial stage. However, the information was successively utilized by the migration procedure, after mechanisms for verb argument selection had been implemented in the grammar.

- At each stage, the whole conversion had to be *reproducible*. When adjustments to the output Spanish lexicon were required, they were never made by manual post-editing. Instead, the adjustments were incorporated as rules in the conversion procedure, **SB** as to be automatically redone every time the conversion procedure was rerun. Such adjustments, which were very limited in number, typically involved single lexical items but in some cases involved classes of lexical items. For instance, the source lexicon contained a class of proper names like *Navidad* ('Christmas'), semantically marked as temporal expressions. Because items in this class can take quantifiers (e.g. *cada Navidad*, 'every Christmas') and quantifiers are not allowed on proper names in our system, such items were turned into common nouns.

4.2. Bilingual lexicons

The bilingual lexicon for the English-Spanish language pair has been partially implemented by converting a pre-existing bilingual lexicon developed for a commercial MT application. Such an application was different from and unrelated to the one from which the Spanish monolingual lexicon was taken. The original bilingual lexicon was basically a bilingual list of words or phrases, with the English side tagged for part of speech. A full conversion was limited by several problems:

1. Although drawn from an NLP application, the source lexicon suffered from several inconsistencies in the way information was coded. Specifically, differed formats were used for verbal entries (infinitive form vs. inflected form, infinitive with or without to, phrasal verbs with or without their argument expressed, e.g. *take advantage of* vs. *take advantage of someone*).
2. There was insufficient linguistic information added to either side of a bilingual entry (part of speech for English, none for Spanish). Consequently, when homographs were present, it was not possible to get an exact match for a corresponding monolingual lexical entry.
3. The automatic association of indices to bilingual entry items was problematic, particularly when verbs or multi-lexeme entries were involved.

In light of the foregoing difficulties, the conversion could only be semi-automatic. A bilingual lexicon was automatically created and then post-editing was performed to augment entries with the necessary additional information. The constraint-based approach to transfer made significantly easier and more effective this two-stage process, where the

partial information coming from the external resource was augmented with further constraints.

The success of the conversion can be related to two parameters: the part of speech involved and whether the entries were mono-lexeme or multi-lexeme. No or minimal post-editing was required on mono-lexeme entries for nouns, adjectives and adverbs. Heavier intervention was required on verbs and multi-lexeme entries, for the reasons outlined above. However, the success of the conversion can be appreciated by taking into account the relative weight that the two groups of entries (nouns/adjectives/adverbs vs. verbs and multi-lexeme entries) have in lexicons. For instance, in our source lexicon, simple noun, adjective and adverb entries covered 69% of the total number of entries (11711 entries out of a total of 16967).

5. Reuse of non-NLP resources

There are many on-line dictionaries of English that were not designed to be used in computational applications, but which nevertheless contain information that would be useful in developing linguistic modules for NLP applications. We have developed routines for the creation of feature structures from the lexical and syntactic information (particularly information relating to syntactic category and verb subcategorization) from an electronic version of one such dictionary. The entries from the electronic dictionary are also automatically enriched to contain the base form of the dictionary entry, thus eliminating the need for on-line morphological analysis. The resulting feature structures have been incorporated into an HPSG-based English grammar, written in the style of ALE, or Attribute Logic Engine (Carpenter and Penn, 1992).

The mapping from an electronic dictionary entry to a feature structure is done in two parts:

1. a mapping from electronic dictionary information to macros using a conversion table;
2. the definition of the macros.

Each subcategorization class can be captured by a macro like `@verb1` or `@verb2` above when the subcategorization information is explicit in the lexicon. In the case where there is not a one-to-one correspondence between subcategorization frames or syntactic categories and the feature structure we want to obtain, it is necessary to take into account the co-occurrence of several lexical or syntactic features from the electronic dictionary to generate one feature structure. Again, this is done with a conversion table. Such an approach is successful as long as the electronic dictionary is consistent in the format assigned to lexical and subcategorization information. Incorporation of information from other dictionaries or movement to a different dictionary would of course involve the reformulation of the conversion table.

For example, for an electronic dictionary entry for the word *ate*, we could obtain several ALE style lexical entries which make full use of macros that can be expanded into full feature structures.

`ate` → `@verb1`, `@preterit`

`ate` → `@verb2` , ...

:

The conversion exemplifies two aspects of the relevance of the knowledge engineering issues previously discussed:

- The dynamic relationship between lexicon and grammar development. In this case, the availability of an external lexicon with rich subcategorization information was an additional reason (although not the only one) for adopting an HPSG-like approach to grammar development.
- The use of macros as a means to interface linguistic resources. In the conversion procedure described above, macros are a convenient way to bridge the gap between the different categorizations used in the electronic dictionary and in the computational lexicon, due to the different nature of the two resources (human-oriented vs. computer-oriented).

6. Reuse in future applications

In this section we discuss the portability of a bilingual lexicon, like the one we developed in our system, to a new language pair. Since a bilingual lexicon has no inherent directionality, each of the two languages can be replaced to obtain a new bidirectional lexicon. However, for convenience, we conventionally refer to the unchanged language as the 'source' language and the changed language as the 'target' language. Our discussion is supported by some experiments we did in porting our English-Spanish lexicon to English-Brazilian Portuguese. We first discuss the porting procedure at the level of a single entry and then with respect to a whole lexicon.

6.1. Bilingual entry reuse

We assume, as earlier described, that a bilingual entry contains lexemes, syntactic descriptions and bilingual constraints (either lexeme-specific or general).

The first consideration is that the source side of a bilingual entry can be left unchanged, with the possible exception of lexeme-specific constraints, when their specificity refers to the target word (incidentally, experience shows this case to be very infrequent). Changes are thus limited to the target side and, possibly, to the bilingual constraints (here expressed in the form of transfer macros). The changes could range from the simple replacement of target words to the complete replacement of all the information. This depends on the similarity between the linguistic frameworks used for the old and new target modules, which in turn is closely related to the similarity between the two languages at hand. In any case, as pointed out earlier, when such changes concern parochial variations of cross-linguistic phenomena, their implementation is limited to changing the content of macros. In this way, the changes only need to be done in one place and the bilingual entries

can be left unchanged. In our experiments, the use of transfer macros in bilingual entries has proven to pay off with respect to later porting the system to new language pairs.

To give an example, a generalization for the English-Spanish pair is that a nominal indirect object (e.g. '*I tell the man that...*'), is mapped to a prepositional phrase plus a redundant personal pronoun ('*Yo le digo al hombre que ...*'). This type of transfer is performed in the English-Spanish bilingual lexicon by means of a specific transfer macro associated with verbs taking a dative complement. In porting the bilingual lexicon to English-Brazilian Portuguese ('*Eu digo ao homem que ...*') it was sufficient to change the specific transfer macro, removing the redundant pronoun requirement, without modifying anything from the actual transfer rules apart from the right hand side words.

6.2. Bilingual lexicon reuse

With respect to a whole lexicon, the simplest procedure would be a one-to-one mapping. For each entry in the original bilingual entry, a corresponding entry is created in the new one. Unfortunately, such a procedure would be very likely to give rise to redundancies or gaps.

In the case of mono-lexeme entries, translation mismatches between, say, English and Spanish may result in two possible Spanish translations for a particular English lexeme (thus, two bilingual entries), whereas the same lexeme might have only one possible translation in Portuguese. In this case the one-to-one procedure would give rise to redundancy. In the reverse case (one Spanish equivalence but two Portuguese equivalences) a gap would arise. Therefore, as for mono-lexeme entries, the mapping must be performed at the level of sets of entries with the same base-form, rather than at the level of single entries. Entries could be removed or added.

Things are more complex with respect to multi-lexeme entries. In addition to the remarks above, let's consider the case of an English idiomatic expression translating to a single Spanish word (thus requiring a specific multi-lexeme entry). It might well be the case that the same English expression corresponded to an isomorphic Portuguese expression. In this case it could be compositionally translated to Portuguese, thus requiring no specific multi-lexeme entry. Conversely, a specific multi-lexeme entry might be required for Brazilian Portuguese, whereas the same English expression might translate compositionally to Spanish. Therefore, multi-lexeme entries should require some specific consideration before porting.

However, the extent to which the foregoing theoretical considerations apply in practice largely depends on the pair of target languages under consideration. Our specific experience with Spanish and Brazilian Portuguese, showed a high rate of success for the automatic porting, given the greater similarity between the two languages than between each of them and English.

Consider, for instance, the following English-Spanish entries:

English-Spanish:
play ↔ jugar
play ↔ tocar

touch ↔ tocar

If we ported them to English-Portuguese on a one-to-one basis, we would correctly end up with the following entries:

English-Portuguese:
play ↔ jogar
play ↔ tocar
touch ↔ tocar

However, if we were instead to replace English with Portuguese, we would end up with the following, redundant entries:

Portuguese-Spanish:
jogar ↔ jugar
tocar ↔ tocar
tocar ↔ tocar

To sum up, although automatic porting of a bilingual lexicon to a new language pair seems hardly feasible, an S&B style bilingual lexicon, together with a systematic use of macros, allows one to minimize the amount of information which needs to be changed. Moreover, a suitable choice of target languages is another factor which improves the chances of a successful porting.

7. Conclusions

There is a sameness to NLP applications that creates a functional overlap of grammatical resources among those applications. Whatever the application, an NLP system requires grammars and lexicons. Furthermore, while grammars may vary with the linguistic theory they implement, they will always require information about such grammatical notions as syntactic category, grammatical number and gender and case. The repository of this information is the lexicon and, consequently, the lexicon is a likely source of information that can be transferred across applications.

A significant feature of a system that relies on existing resources is a modular design that permits, as much as possible, integration of a resource into a single module without adapting it to the requirements of the overall application. Declarative representations facilitate the task of decoupling a resource from its original function and extracting requisite information from it for the new application. Since both these properties are features of contemporary NLP design, the flow of resources among applications should increase with experience.

Even when a lexicon is tightly coupled with the application and the transferred lexicon contains little information, there are still considerable savings to be gained from reusing an existing lexicon. Our bilingual lexicon was adapted from what was not much more than a word-pair list with minimal syntactic information about the English side. From this list, more than two-thirds of the bilingual lexicon

was constructed automatically. Manual post-editing was required only for the remaining third, consisting of classes that would normally be classed as semantically complex.

Acknowledgements

This research was supported by TCC Communications, by a Collaborate Research and Development Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the Institute for Robotics and Intelligent Systems. The authors would like to thank Dan Fass and John Grayson for their comments on earlier versions of this paper.

References

- Doug Arnold, Toni Badia, Josef van Genabith, Stella Markantonatou, Stefan Momma, Louisa Sadler and Paul Schmidt. 1993. Experiments in Reusability of Grammatical Resources. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, pages 12-20, Utrecht, The Netherlands.
- John L. Beaven. 1992. Shake and Bake Machine Translation. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 603-609, Nantes, France.
- Nicoletta Calzolari and Antonio Zampolli. 1994. Standards to make natural language resources shareable resources. In *Proceedings of the International Workshop on Sharable Natural Language Resources*, pages 15-21, Nara, Japan.
- Bob Carpenter and Gerald Penn. 1994. *ALE 2.0 User's Guide*, Laboratory for Computational Linguistics Technical Report, Carnegie Mellon University, Pittsburgh, PA.
- Carl Pollard and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Centre for the Study of Language and Information, Stanford University, CA.
- Fred Popowich, Davide Turcato, Olivier Laurens, Paul McFetridge, J. Devlan Nicholson, Patrick McGivern, Maricela Corzo-Pena, Lisa Pidruchney and Scott MacDonald. 1997. A Lexicalist Approach to the Translation of Colloquial Text. In *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 76-86, Santa Fe, NM, USA.
- Manny Rayner, Ivan Bretan, Mats Wirén, Sara Rydin, and Eleonora Beshai. 1996. Composition of Transfer Rules in a Multi-Lingual MT System. In *Proceedings of Workshop on "Future Issues for Multilingual Text Processing"*, Cairns, Australia.
- Davide Turcato, Olivier Laurens, Paul McFetridge and Fred Popowich. 1997. Inflectional Information in Transfer for Lexicalist MT. In *Proceedings of the International Conference 'Recent Advances in Natural Language Processing'*, pages 98-103, Tzigov Chark, Bulgaria.

Gertjan Van Noord, Joke Dorrepaal, Pim Van Der Eijk, Maria Florenza, Herbert Ruessink and Louis Des Tombe. 1991. An Overview of MiMo2. *Machine Translation*, 6, pages 201-214.

Pete Whitelock. 1994. Shake and Bake Translation. In C.J. Rupp, M.A. Rosner, and R.L. Johnson (eds.) *Constraints, Language and Computation*, pages 339-359, London, Academic Press.