MT Summit VIII

Santiago de Compostela (Spain), 18 - 22 September 2001

Resource Alignment and Implicit Transfer

Jean Senellart, Mirko Plitt, Christophe Bailly, Francoise Cardoso

Systran, Autodesk

1, rue du Cimetière - BP.7
95237, Soisy-sous-Montmorency Cedex
France
senellart@systran.fr, mirko.plitt@eur.autodesk.com, bailly@systran.fr, cardoso@systran.fr

Abstract

In this article we present the concept of “implicit transfer” rules. We will show that they represent a valid compromise between huge direct transfer terminology lists and large sets of transfer rules, which are very complex to maintain. We present a concrete, real-life application of this concept in a customization project (TOLEDO project) concerning the automatic translation of Autodesk (ADSK) support pages. In this application, the alignment is moreover combined with a graph representation substituting linear dictionaries. We show how the concept could be extended to increase coverage of traditional translation dictionaries as well as to extract terminology from large existing multilingual corpora. We also introduce the concept of "alignment dictionary" which seems promising in its ability to extend the pragmatic limits of multilingual dictionary management.

Keywords

syntactic alignment, customization, dictionary management, automatic translation, implicit transfer

Introduction

Everyday, feedback on Systran’s free translation services on the Web shows that despite the fact that Systran dictionaries contain an impressive number of entries, the size of the linguistic resources is still a bottleneck before high quality of general translation can be achieved. Continuous work on enrichment is thus a requirement in order to satisfy the more and more demanding translation user and to increase the translation quality.

In a “classic” incremental model this task is complex; a great number of rules is accumulated and must be maintained and coordinated. It is not easy to always understand the final behaviour of the whole set when new rules are added. If we consider, in addition, parallel maintenance of several target languages, maintenance and quality control becomes even more complex.

At the same time, experience with customization service shows that a single customization in a very narrow domain may be asking for very large glossaries, leading quickly to the same type of complexity in the handling of the resources mentioned above.

If we analyse the distribution of entry complexity in an average Systran main dictionary of 200,000 entries, we see that about 80% of the entries are simple lexicalized entries. For part of these entries, we could have considered more complex coding, extending the capacity of variation of the entry, or restricting its application to a given context. Practically, on a large scale, this fine-tuning is not feasible considering human limits of organization, let alone any cost considerations.

For such dictionary size, high-level organization of the information is thus a requirement. In this article we present the basic structuring work in progress on our multilingual resources. We will show, with examples, the following sources of complexity in dictionary management and our solutions:

- Great redundancy in the linguistic information. This can be reduced by giving the system a real capacity to handle implicit information. via our alignment module.
- Lack of factorisation of entries. This can easily be corrected using finite state representation corresponding to a local grammar of lexical units.

We present as an application, but also as a basis for explanation, a real-life customization project for machine translation. This presentation is followed by a generalization of the ideas applied in this project.

In particular, we show that the concept of alignment dictionary, which is not suitable for translation, contains an overview of the information spread all over the resources. This meta-information allows an implicit transfer capacity. This alignment has some common points with bilingual corpus alignment algorithms, although it does not serve the same purpose.

In conclusion, we will also show that the re-organization of resources combined with the alignment module allows us to reconsider enrichment based on newly derived tools: systematic description of lexical units based on corpus, extraction of potential equivalencies for source entries, and supervised validation using a target corpus.

Figure 1 gives an overview of the reorganization.

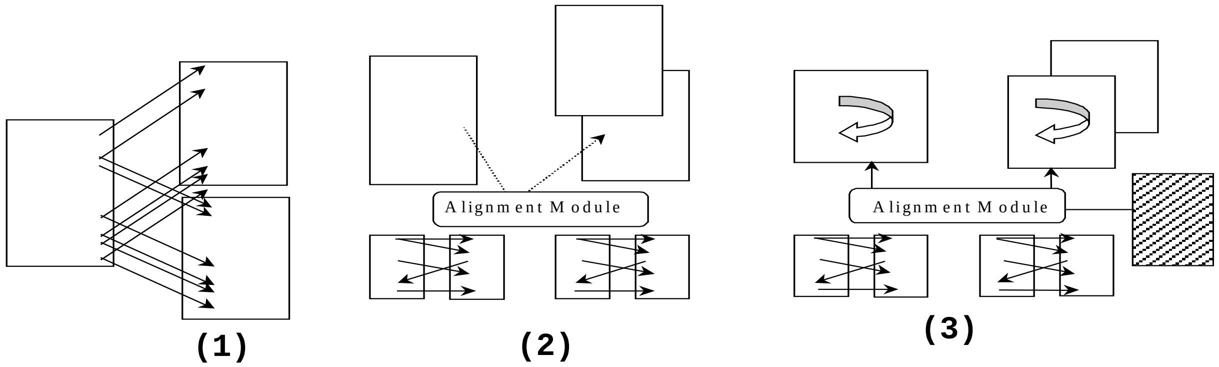

Figure 0 - Evolution of the information structure. (1) represents a multi-target standard dictionary. In (2), the content is the same, but the explicit equivalencies have been replaced by alignment dictionaries (and are dynamically retrieved by the alignment module). In (3), the monolingual resources are factorized offering more possibilities of being enriched, and using alignment resources : extraction and corpus alignment strategies are applied.

Toledo Project Presentation

The Toledo project is a project of dynamic automatic translation of support pages for the AutoCAD software family. A demonstration of the service is available at: `http://www.systranlinks.com/systran/cgi?partner=systran-ADSK-en`

An initial corpus was provided (about 4000 pages or 25 Mb of non-redundant clean English text) corresponding to an existing support database given at the beginning of the project. The goal of this project was to provide Autodesk with a customized translation service in order to translate dynamically any support page (present or not in the initial corpus) from English into French, German, Italian, and Spanish. The structure of this corpus is essentially step-by-step solution-oriented as we can see in the following sample:

We set a qualitative goal of "understandability" and a quantitative one of 70% coverage of the source text by a specialized grammar. These criteria are a good indicator of translation output quality, as we will show below.

We present here the technical solution, and we show that organization of linguistic information is at the heart of the customization.

```
There are two methods for preventing the
viewport border from plotting.
```

Method #1

```
1.Switch to the Layout that contains the
viewport border you do not want plotted.
2.Choose the Layers toolbar button from the
Object Properties toolbar.
3.Choose the New button in the Layers
dialog.
4.Name the new layer.
5.Select the Freeze Layer icon for the new
layer. (The icon toggles between a sun and a
snowflake).
```

Text 2 - Sample of support pages text.

ADSK Multilingual Resources

The Autodesk localization team has provided us with the following multilingual resources:

- Translation memory from localization of software documentation
- Software localization glossaries: software references, such as button name, menu name, alert message, dialog-box content, etc. (For simplicity's sake, we will call these references to software **token names**, and the object to which the **token** refers **token identifier**)
- General Autodesk terminology (drawing, layer, etc.)

The software localization glossary is an unstructured list of **token names**: some information about the origin of the product is given, but very little information about the **token identifiers**.

Standard Terminology/Transfer Approach

The traditional translation customization approach would have been to construct multilingual "user" dictionaries that are applied (with priority rules) in combination with main dictionaries. In our situation, this direct approach was not possible because of the high ambiguity of the **token** glossary. Indeed, such common words as "on", "new", "in", "add", "and" were in these glossaries. Moreover, the absence of any information concerning the **token** identifiers make any attempt to write contextual transfer entries unproductive (taking into account the fact that the associated **token** identifier could be absent as in: `Click on File>>New` (where menu is implicit: the complete form is `Click on File menu>>New`). Finally, the structure of the source text does not have any strict typesetting rules concerning use of **tokens** and formatting of sentences.

These reasons led us to extend the scope of the description to whole nominal phrases and up to whole sub-clauses concerning these **tokens**. This approach has been combined with the addition of traditional dictionaries for specialized terminology (see Figure 2 for the organization of external glossaries).

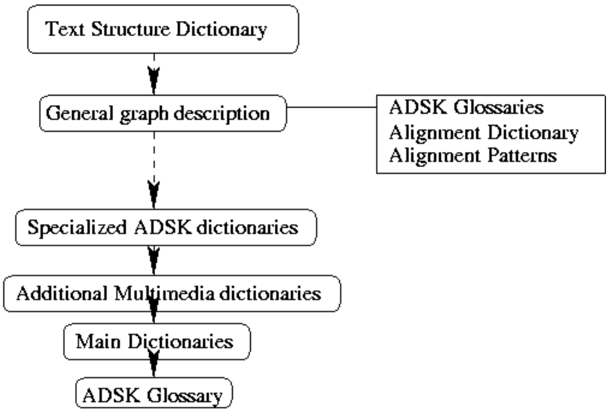
Text Structure Dictionary
General graph description
ADSK Glossaries
Alignment Dictionary
Alignment Patterns
Specialized ADSK dictionaries
Additional Multimedia dictionaries
Main Dictionaries
ADSK Glossary

Figure 2. Resource sequence. The Text Structure Dictionary contains all meta-format information on documents. The ADSK Glossaries are applied twice; the second time is a safety lookup. All tokens, whose context is not described, (e.g., structure errors) are translated at this level.

Principle of Customization

The principle of this customization project relies on a library of lexical/contextual/syntactic graphs. The first part of this work was to build *an accurate source description* of the grammar of the text based on **token** identifier contexts.

Build contextual graphs on text

The source description was performed using a graphical representation (Gross, 1997) formally equivalent to a finite state automaton (Roche, 1997). The choice of this representation was based on:

- The combinatory of syntactic structures described (Figure 3)
- Human linguistic intuition in the use of such representation
- The capacity of organizing information with sub-graphs and multiple boxes
- The capacity of building such graphs based on dynamic concordances on a corpus
- Finally, building of such a description based on a corpus is a direct method with a sample of the text, using a bootstrap methodology and intuitive completion (Gross, 2000)

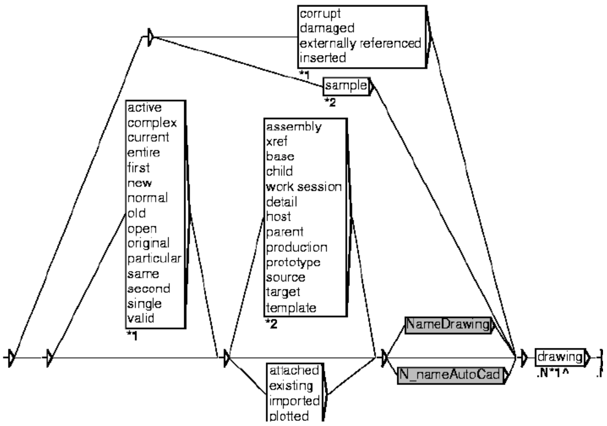

Figure 3. grammar of "drawing" in ADSK corpus. Gray boxes refer to sub graphs.

Figures 3 and 4 give examples of such graphs for nominal and verbal constructions. For example, in the graph in Figure 3, we have the description of all nominal constructions concerning "`drawing`" in the ADSK corpus.
In fact, this graph represents a shallow semantic structure linked with the associated syntactic construction. The *semantic* is in the gathering in boxes of the modifiers. For example, in the same boxes we have "historic" modifiers such as "`attached`", "`existing`", etc. In another box, we have the "type" modifiers:

```
source/target/host/...
```

Moreover, the degree of specialization of the modifiers is variable: this graph recognizes "new `drawing`", at the same level as "`xref drawing`". The intuition in constructing those graphs was to represent any term whose syntactic construction is not totally free. The modifier "`new`" in an editing software context has a special meaning related to `blank`, `empty` but also to `unnamed`...

This graph describes for instance the phrases: "`entire source Actrix drawing`", and "`damaged AutoCad2000i drawing`". With a deeper knowledge of the "`drawing`" concept in AutoCAD products, we could probably organize this particular graph better, but this will have a very slight effect on translation output. In fact, without any additional semantic input, one strategy could be to use the benefit of the factorization allowed by the graphs to over-describe the source text. Nominal expressions such as "`source target template host drawing`" would thus be accepted; but this will not have any real impact on the translation of real sentences. Our strategy was to intuitively organize the database, and submit the description to a technical expert for fine-tuning. In fact, most of the organization of the database is in the choice of the graph hierarchy.

Different kinds of graphs

We have named these kinds of graphs *lexical/contextual/syntactic* graphs. In fact, most of them will be lexical because they are based on one particular lexical item, and contextual because they describe this lexical item in its immediate syntactic context. Verbal description is probably more syntactic since these graphs are not applied directly, but can be "transformed" with generic transformation patterns (Senellart, 1999) to allow recognition of nominalized expressions, negative sentences, passive sentences, etc. In the framework of this project and because of the very simple structure of sentences, very few transformations were applied (apart from those regarding inflection patterns).

Using the Description to auto-organize the Token Glossary

The graph of Figure 3 relies on the `NameDrawing` sub-graph, which is a list of the token references whose identifier is the word "`drawing`". To avoid potential ambiguity, we have extracted the corresponding sub-glossary from the main glossary. This operation was simply performed by replacing the `NameDrawing` box by a joker, by applying this graph to the whole corpus, and by manually supervising the extraction of the joker part.

This monolingual source description is thus structured in different layers of graphs corresponding more or less to different concepts, and has a structuring effect on external glossaries. The description needs and provides a very precise structure of the grammar of the text: used as a tool during writing of new support pages, this description is not far from behaving like an authoring tool since it can highlight recognized patterns, and even propose new glossary candidates using the "joker" capacity.

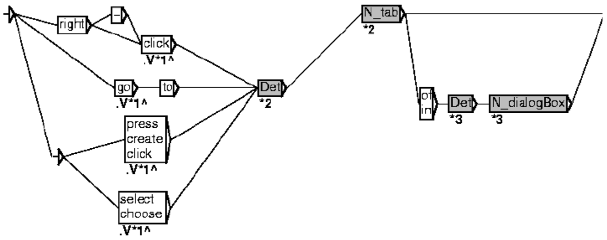

Figure 4. Verbal expression based on free determiner description (here Det), and nominal description of dialog box (here N_dialogBox).

Describing Target Grammar

The second part of the customization work was to associate each of the source descriptions with an equivalent target description. In this project, the "translation" stage was manually performed using essentially translation memory as a reference. See for example the graph in Figure 5 representing the French translation of the graph in Figure 3. The runtime alignment process relies on the fact that for each source path we can find at least one equivalent target path, as we explain below. The only constraints on the structure of the target description are thus constraints of syntactic correctness (contrary to source description, we have to be more precise in the order of modifiers for example, and in the agreement between modifiers and head noun), and of graph-to-graph equivalency (which is a constraint applied for this project, but that could be relaxed in another context).

Alignment.

The translation process (for details on technical issues see the following section) performs a parallel lookup of -general dictionaries, -additional ADSK dictionaries, and -expressions in the contextual graph library. After resolution of various potential "ambiguities" and after applying a standard heuristic of longest match, we obtain a list of expressions that we dynamically align with the corresponding target descriptions. The result of alignment is considered like any transfer result and is afterwards reprocessed according to synthesis and rearrangement rules.

Results

Il y a deux méthodes pour empêcher la bordure de fenêtre de tracer.

Méthode #1

1.Basculez sur la présentation qui contient la bordure de fenêtre que vous ne voulez pas tracé.
2.*Sélectionnez le bouton Calques de la barre d'outils dans la barre d'outils Propriétés de l'objet.*
3.*Sélectionnez le bouton Nouveau dans la boîte de dialogue Calques.*
4.Nommez le nouveau calque.
5.*Sélectionnez l'icône "@Freeze Layer"* pour le nouveau calque. (l'icône bascule entre un soleil et un flocon de neige).

Text 3. French translation of source sample using alignment of graph description.

Text 2 shows the translation of the reference sample (Text 1). In this translation output, underlined words are translated using additional Autodesk terminology, and italic words translated using the alignment process (in that case full sentences).

Some significant facts:
Coverage of 65% of the source text has been achieved.
Approximately 100 graphs have been built, corresponding to 5 levels of graph organization.

Because of factorization, these graphs represent about 32,000 different paths (without any sub-graph expansions and not counting all cycles).
This number is roughly the number of direct transfer entries that would have been needed for the same description (for which we would have needed to give explicit transfer equivalents). To complete the comparison, the number of conditional transfer rules needed to obtain the same description is approximately for each graph the number of lexical references in the graph: in that case about 24 rules, i.e. 2400 for the whole description.

The cost for introducing a new target language is minimal, as we only need to translate these 100 graphs, and check that the alignment is complete (see Generalization, below).

Nevertheless, the description is not really reversible, as we have over-description in the source language. In translating to English from French, we would need to add more constraints on the source description.

Generalization

In this section we generalize and formalize the implicit transfer methodology. We mainly focus on the "alignment" concept that can be applied to graph libraries as well as to raw "classical" dictionaries.
In fact, this methodology is an answer for translating expressions described with graphs. This problem was studied in Senellart (1998 and 1999). The solutions with which we experimented were direct translation of dates, and translation using a kind of "interlingua" description

for nominal phrases describing occupations. These experiments have proven their limits for general applications:

- The direct translation (using extended transducer) is very complex to write because general target properties (such as position of adjective, or agreement) have to be re-described for any path in the automata. Moreover, conjunction is almost impossible to describe.
- The Interlingua approach was based on the same principle of alignment. In that case, it was the interlingua (output of transducer) itself that was aligned between source and target description. The complexity was in the capacity of defining this Interlingua (in that case, the choice was a set of properties "prime_minister", "conservative", "French”: but this was not sufficient to describe complex combination of modifiers. For instance in phrases like: "the French deputy mayor colleague")

In the following, we leave the "graph" issue aside, because the alignment process is the same. Since we use a finite automaton to store dictionaries, there are no technical differences of implementation between a graph and a raw dictionary (apart from the possibility of having a cycle in a finite state automaton dictionary representing a graph).

The basic principle of implicit transfer is to suppress an important source of redundancy present in all multilingual linguistic resources and to calculate the transfer information dynamically. Note that this dynamic approach implies some maintenance work on the resource to keep coherency of the database.

The alignment process essentially relies on three resources: source and target description, and alignment dictionaries. The alignment process is the process that allows implicit transfer between source and target description. In order to simplify the description, we will consider here only the alignment of nominal phrases. This alignment can be compared with a lexicon-based alignment algorithm (Catizone et al., 1989) in bilingual corpus. Our approach is nevertheless different in the aim (we want to align resources) and more syntactic since the compounds we align are morphologically totally described. Note that this similarity explains the capacity we describe below to use corpus for validating translation.

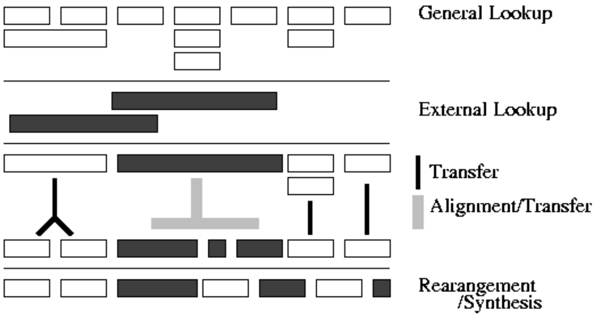

Figure 1. Integration of alignment in the main translation process.

In the following section we will deal with these points:

- Integration of alignment in main Translation Process
- Alignment vs. Redundancy
- Building of the alignment dictionary
- Alignment algorithm
- Enriching resources for alignment, and extending them.

Integration in the main Translation Process

Technically, this procedure is integrated in the whole translation process as we can see in Figure 6.

In this figure, the gray area represents the matching entry treated by the alignment procedure. We see that these entries follow the natural rearrangement and synthesis cycle.

Redundancy versus Alignment

In the EEC Eurodicautom dictionary we have 790 different entries based on the head noun "voltage". Here are the first entries of this list with their French equivalents:

absolute voltage level	niveau absolu de tension
absorbed voltage	tension absorbée
AC voltage	tension alternative
accelerating voltage	tension d'accélération
acceptance voltage	tension admissible
accidental voltage transfer	transfer accidentel de tension

In most of these entries, we have an intuition of sub-alignment between source and target. Indeed, this intuition is confirmed by the syntactic variants that most of these "frozen" expressions can have: AC accelerating voltage, or "the voltage that was absorbed". This glossary, which is typically the kind of specialized terminology that we want to introduce in customized dictionaries (here in the electric domain), we see that there is a lot of information redundancy in the equivalencies.

In order to evaluate the amount of information in each entry, we can evaluate the surprise of obtaining each translation: *what is the surprise of translating "absolute voltage level" by "niveau absolu de tension”?* This surprise is very limited: "level" is almost always translated by "niveau", "voltage" by "tension", and "absolute" by "absolu". Moreover, the patterns *AN→NA* and *NN→N de N* are equally very frequent. All this makes that the probability of getting the given translation was very high.

Finally, knowing these entries does not help to translate AC accelerating voltage, or "the voltage that was absorbed"... In that case, the only information is in the existence of the source and the target compound, and not in the link between them.

Traditionally, we would say that these entries should be replaced by transfer rules giving the context for translating: something like "absorbed (modifying voltage)→tension". However, writing such a rule is not easy and suppresses the intrinsic "lexical" property of the source and target compound: "absolute voltage level" is translated by "niveau absolu de tension" and not by "niveau de tension absolu" which is not reflected by the rule. Moreover, building transfer rules is a complex task when accumulating a huge number of them. The combination of these rules is indeed complex to handle.

Alignment dictionary

The basis of the system is the capacity of aligning source and target descriptions. This alignment is performed with the aid of an "alignment" dictionary and by using alignment patterns. In the Toledo project, this alignment dictionary has been built incrementally according to the alignment needs. More generally, an alignment dictionary is roughly a transfer dictionary where all selection constraints (lexical, contextual or domain) have been relaxed, and where all implicit sub-token alignment patterns have been extracted.

For example:

coming from transfer rules	results
high→élevé (medicine domain)	high→élevé
high→haut	high→haut
high modify(COST)→ important	high→important
high order→ordre supérieur *and* order→ordre *and* $A_1N_2\rightarrow N_1A_2$	high→supérieur

Building such a resource can not be limited to an automatic projection. Let us take a sample from the English-French alignment dictionary built on the classic Systran dictionary. We have the following entries :

blood	sanguin,sang
high(A)	grand(A), supérieur(A), haut(A), fort(A), élevé(A), noble(A), extrême(A), large(A), important(A)
credit card(N)	carte bleue(N), carte de crédit(N)

Where the relation between first and second column is "N_s could be translated by N_t in certain context". And we do not have the following entries:

EN	FR	That could have come from :
blood	artériel	blood pressure→pression artérielle *and* pressure→ pression *and* NN→NA
credit	bleu	credit card→carte bleue *and* card→ carte *and* NN→ NA

This example proves that we must set a frame for this alignment dictionary. Accepting "blood→artériel" has probably some semantic relevance and may be productive but we have chosen to avoid such alignment rule to avoid subjectivity in extraction.

Finally, the constraints given by the transfer of the syntactic pattern is very flexible as we can see in the following examples:

(random access) device	unité à (accès aléatoire)	**$N_1N_2\rightarrow N_2$ à N_1**
(random access) device	unité à (accès aléatoire)	**$N_1N_2\rightarrow N_2$ à N_1**
sea view	vue sur mer	**$N_1N_2\rightarrow N_2$ sur N_1**
(sea view) room	chambre avec (vue sur mer)	**$N_1N_2\rightarrow N_2$ avec N_1**
act of legislation	acte législatif	**N_1 of $N_2\rightarrow N_1A_2$**
advance on salary	avance sur salaire	**N_1 on $N_2\rightarrow N_1$ sur N_2**
advance on expense	avance des frais	**N_1 on $N_2\rightarrow N_1$ DET N_2**
cash on delivery	Paiement à la livraison	**N_1 on $N_2\rightarrow N_1$ à DET A_2**
advance in technology	Avance technologique	**N_1 in $N_2\rightarrow N_1A_2$**
agreement in principle	accord de principe	**N_1 in $N_2\rightarrow N_1$ de N_2**
asset in kind	apport en nature	**N_1 in $N_2\rightarrow N_1$ en N_2**
traveler's check	chèque de voyage	**N_1's $N_2\rightarrow N_2$ de N_1?**
teller's check	chèque au porteur	**N_1's $N_2\rightarrow N_2$ au N_1**
earth's crust	Croûte terrestre	**N_1's $N_2\rightarrow N_2A_1$**

For that reason, we maintain parallel to the alignment dictionary, an open list of syntactic patterns. This list seems to be rather poor, but we keep at least the order of the tokens. Another approach with which we have experimented was to consider that all "function words" were not taken into account during alignment. This gave us very poor results because it produced frequent ambiguity, for instance between structures like **N_1 de N_2**, and N2 de N1 (tension de sauvegarde, sauvegarde de la tension...)

Alignment algorithm

Based on the previous resources and when given a source and target description, the alignment resource is very simple:
For each term **A** in the source description (could be a path of a graph, a dictionary entry, or even an extracted term), we translate this term by **B** if:

- **B** is an element of target resource (could be a path of a graph, a dictionary entry, or even an extracted text)
- we can decompose **A** in a syntactic pattern **S_A =N_1...N_i** and we can decompose **B** in **S_B=M_1...M_i** such that the patterns **$S_A\rightarrow S_B$** and each of the sub-expressions **(j,k)** described by the alignment of the patterns : **$N_j\rightarrow M_k$** (this alignments being either based on the alignment dictionary (in that case on lemma form for **N_j** et **M_k**) or based a recursive alignment of a whole expression).

For example if we have the source expression:
advance on professional expenses

We have the decomposition
advance on (professional expenses)

and the pattern **N_1 on N_2** aligned with (for instance) **N_2 sur N_1**
We find advance(N)→avance(N)
and for professional expenses: **$A_1N_2\rightarrow N_2A_1$**
and professional(A)→professionel(A)
and expenses(A)→frais(A)
frais professionels is in the target description, thus we have professional expenses→frais professionels
Finally, avance sur frais professionnels is in

the target description, thus we get:

```
advance on professional expenses→
                 avance sur frais professionnels
```

Enriching resources and extending them

We have presented the alignment resources and the alignment process. In order to come back to our first concern, which was to maintain and enrich a large multilingual database, we have now the capacity of re-aligning all the entries together (step 2 in Figure 1).

Moreover, with this alignment dictionary, we have brought more information to the resources from which we can now benefit for the following applications:

- Factorization of the entries based on the headword. This factorization (using for example the graph description) will combine some semantically connected entries and thus contribute to the organization of the linguistic information.
- Alignment with new resources: if we increase independently the source and target dictionaries, the alignment procedure will be able to detect if some of the new entries can be aligned. We have experimented with such a method to increase the bilingual dictionary of a new language pair: English-Hungarian. In this case, we only needed to extract new Hungarian compounds (this is easy because of internal compounding) and try to align them with the English reference dictionary. The idea to build bilingual glossaries with alignment is not new (Gale and Church, 1991). However, here the reference is not a bilingual corpus but “a priori” non-parallel resources.
- In the same way, the alignment resources can suggest, for any given term without an equivalent, a list of potential translations, and, by using a full search on a huge corpus, try to validate (i.e. to locate) one of the propositions. In technical field, this corpus validation is often a sufficient criterion for human revision
- Description of new linguistic phenomena. For instance support verb description is a very complex phenomenon to deal with in an explicit transfer way. With the implicit transfer module, we can have a link between each word in the monolingual description and its set of support verbs and align these support verbs during translation according to their modality.

Conclusion

The linguistic reorganization of huge existing resources seems to be a very promising way to extend the current pragmatic limits in dictionary enrichment. Moreover, the technology developed can directly be applied to a new description of specialized languages. In this case, we have shown the benefit in comparison with a standard customization.

In the two applications presented, a very important issue is to understand the nature of the information present in bilingual resources and to understand what is and what is not new information. This then allows us *to focus the description on the real linguistic information.*

References

Catizone R., Russel G., and S. Warwick *(1989)* Deriving Translation Data from Bilingual Texts, in Proceedings of the First International Acquisition Workshop, Detroit

Church K. and Gale W. (*199)1* Concordances for parallel texts. Proceedings of the 7th Annual Conference for the NOEDictionary and Text Research

Gross M. (*1997).* The construction of local grammars. In Finite State Language processing (Roches & Schabes eds.)

Gross M. (*1998).* A Bootstrap Method for constructing Local Grammars, in Contemporary Mathematics symposium proceedings (Neda Bokan Ed.)

Roche E., Schabes Y. (*1997).* Finite State Language Processing - Introduction (Roches & Schabes eds.)

Senellart J. (*1998).* Locating noun phrases with finite state transducers, in COLING-ACL'98 proceedings

Cédrick F., Senellart J. (*1999).* Réflexions sur la localisation, l'étiquetage, la reconnaissance et la traduction d'expressions linguistiques complexes, in TALN 99 proceedings

Senellart J. (*1999).* Description d'expressions linguistiques complexes par transducteurs linguistiques, thèse de doctorat 1999 - Paris

Acknowledgements

We would like to thank our colleagues Elke Lange and Jin Yang at Systran La Jolla for enriching and editing the paper.

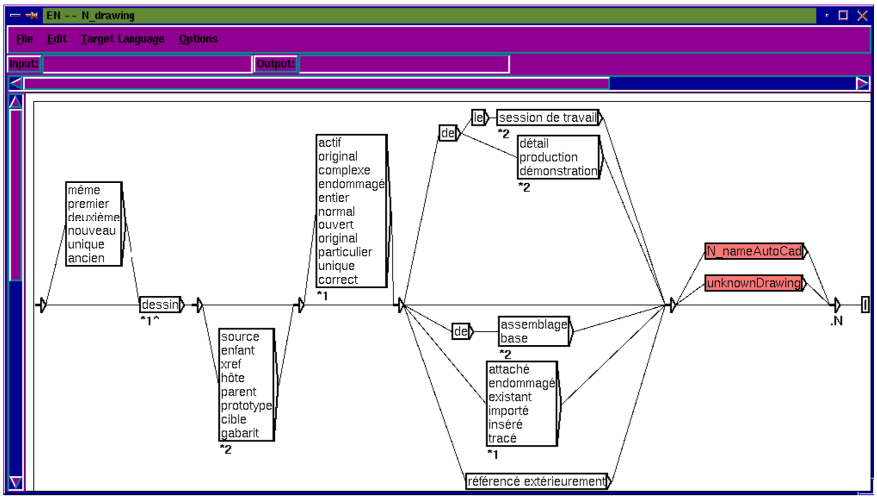

Figure 5. French translation of the grammar of "drawing" .