Applying Divergence Principles to French to English Machine Translation

Violaine PRINCE
University Montpellier 2 / LIRMM-CNRS
161 Ada Street / 34392 Montpellier
cedex 5 FRANCE
prince@lirmm.fr

Jacques CHAUCHE
University Montpellier 2 / LIRMM-CNRS
161 Ada Street / 34392 Montpellier
cedex 5 FRANCE
chauche@lirmm.fr

Abstract

This article presents an implemented prototype of a French-to-English translation system which goal is to ensure a syntactic quality of translation. The method employed relies on a French parser, named SYGFRAN, which in turn is based on an engine that uses Markov’s algorithms applied to tree transducers. SYGFRAN produces a syntactic tree structure as output, with non terminal nodes tagged with part-of-speech constituents and dependencies. In this presentation, we show that divergence is the basis upon which we try to model the transformation path between an analyzed French sentence and its corresponding generated English sentence. We present a set of examples of divergence rules with different scopes, and a first evaluation of the prototype on two corpora: A random corpus, and a chosen corpus of 'syntactically difficult' French sentences.

Introduction

Syntactic quality in translation could be defined as a generated output satisfying the grammatical requirements of the target language (TL). It is generally poor in word-to-word translators, even those equipped with idiomatic phrase equivalencies ¹ or surface transformation rules ². Syntactic quality needs a more complex transfer (Lindop and Tsujii, 1991) since surface analysis is often unsufficient: Complete syntactic structures in source language (SL) must sometimes be thoroughly transformed to obtain a satisfactory result in TL. However, the majority of transfer-based systems relies on an architecture composed of two grammars (of both SL and TL) and a set of transfer rules between them (Arnold, 1993). This drives developers to write both a parser and a generator, and represents a heavy task. It ensures a possible bidirectional translation but, often, the system efficiency is impeded by the complexity of its components. Some authors have relied upon symmetrical features between parsing and generating sentences to alleviate the difficulty of the task (Dymetman et al., 1996), and this idea also runs with texts alignment as a lexical transfer device. In both cases, transformation, as a process, is not invoked, and replaced by concordance.

Transformation, as we understand it, is non symmetrical: The SL is thoroughly analyzed, thus an SL parser is required, whereas generation is much lighter. No grammar of TL is required, only a set of transformation rules modifying the intermediate structure resulting from SL parsing. This set generates the proper intermediate structure which, in turn, goes into the TL generating component, reduced to lexical and morphological resources (in TL). In this presentation, we try to show how we deal with issues tackled by complex transfer components without being burdened by the requirements of transfer-based translators. Due to intrinsic properties of the underlying theory, translation could be considered as a particular step of a (formal) rewriting system.

The other important trend in existing translation systems relies on the principle of interlin-
gu. These systems generally use a representation of the source text independently from SL and TL. Written in the interlingua (a pivotal neutral language) the representation is conceptual and mainly oriented toward semantics. In our system, there is a pivotal component, but it is not semantic and cannot qualify as a language. Our interpretation of an interlingua structure is a set of non terminal tree structures (intermediary states of an automat) marking out the transformation process, close in spirit to (Meyers et al., 2000) approach, but which mathematics are mainly defined in (Mendelson, 1987).

The computational linguistics theory that supports our approach is the theory of divergence. A fundamental article by Dorr (Dorr, 1994) has emphasized the fact that a TL structures its lexical compounds within a discourse, differently from what the SL does. If a taxonomy of divergence operations could be set, and then formalized into transformation rules, then divergence principles could ensure an appropriate TL generation, at least on the syntactic level.

In the following section we define tree transformation rules and their use for the sentence syntactic tree production by a French language parser called SYGFRAN. We show that the same mechanism applies either to produce a tree or to transform it, and thus produce the TL correct syntactic tree (previously to lexical transfer) (Fong, 1991).

In section 2 we describe divergence and particularly focus on divergence scope that could be embedded in tree transformation rules. We focus on Dorr’s solution for divergence classification, and offer a correspondence system between her taxonomy and our transformation rules. Section three briefly describes our prototype SYGToFoE, (meaning French to English with the SYG* systems) and the process of transforming a French sentence into an English one. Section four discusses a first set of results and conclusion sets up a framework for the prototype extension.

1 Divergence and Transformation Rules

1.1 Defining Tree Transformation Rules

A tree transformation rule is a premise-conclusion structure, where the premise contains a recognition pattern and the conclusion contains a transformed pattern.

Example: Let \( A( B(C, D)) \) be the recognition pattern of a rule. This pattern is that of a sub-

![Figure 1: A structured tree where the pattern is recognized](image1)

![Figure 2: The result of the rule application](image2)
a tree transducer system transforming a structure given in a language $L_1$ into a structure of a language $L_2$, providing that tree transformation rules are extensively specified. The underlying philosophy of this environment is that rewriting could be a translation if all transformation paths are completely determined (Chauche and Rolf, 1986).

The same author has developed a French language parser SYGFRAN, using SYGMART as a transformation engine. SYGFRAN contains 12,000 transformation rules and produces a syntactic tree of the French sentence, both in constituents and dependencies.

### 1.2.1 Parsing Transformation Rules in SYGFRAN

The initial sentence is directly transformed into a tree with a root node, given the tag "PH" (meaning sentence) and leaves being the words of the sentence.

**Example**  Let us take the sentence: *Arthur gives John a book*. The French corresponding sentence is: *Arthur donne un livre à Jean*. A first tree with a depth 1 is created, it is represented in figure 3. Through a complete set of transformation rules (each rule written according to a pattern similar to that described in subsection 1.1) named a recursive grammar, SYGFRAN finally produces a tree with some depth where:

1. Terminal nodes (leaves) contain the recognized lexical entries
2. Intermediate non terminal nodes (subtree roots) are numbered and assorted with a set of morphological and part-of-speech tags
3. The root node contains the fundamental root node "PH".

The parsed sentence is provided in figure 4. This tree is then considered as the input structure upon which divergence transformation rules will perform to produce the English syntactic tree. Next section explains how divergence helps to write transformation rules for the English sentence generation.

### 2 Divergence and Tree Transformation Rules for TL Syntactic Tree Generation

#### 2.1 Divergence Scope

Divergence is characterized by its scope. We propose the following definition. The syntactic divergence scope is:

(i) The number of nodes of a given subtree in the syntactic tree of $SL$ that are touched by one or many transformation rules, in order to obtain the correct corresponding subtree in $TL$. This defines the constituent scope of divergence.

(ii) The number of nodes of different subtrees in the syntactic tree of $SL$ that are touched by one or many transformation rules, in order to obtain the correct corresponding tree in $TL$. This defines a dependency scope of divergence.

(iii) If the syntactic tree of $SL$ needs a direct transformation at many depth levels, and if lexical divergence is important, then we consider that divergence scope is maximal. This mainly applies to

---

**Figure 3:** The original sentence before applying transformation rules

**Figure 4:** The parsed sentence as an output tagged tree structure
idiomatic or quasi-idiomatic expressions needing a heavy transformation.

2.2 'Divergence Operations' in Tree Transformation Rules

The different operations that are possible in tree transformation are the following:

(i) Insertion: The SL tree structure is modified by inserting one or several nodes to obtain the TL tree structure. Example: two insertions are necessary to translate the following French sentence Il veut sortir. into is English equivalent He wants to go out.

(ii) Deletion: Nodes existing in the SL structure are deleted in the TL output. Example: The French plural determinant for the subject noun phrase 'Des animaux' has to be deleted in English in the following sentences: (French) Des animaux dorment dans le pré. (English) Animals are sleeping in the meadow.

(iii) Permutation: Nodes (or complete subtrees) of the SL structure have to be switched in order to produce the correct TL structure. Example: the French pronoun 'lui' (him) acting as an object complement, is before the verb, and has to be put after the verb in English. (French) Il lui a donné un livre. (English) He gave him a book.

In many cases, different operations might be involved in the same sentence, and within the same divergence scope. For instance, our example: Arthur donne un livre à Jean, presents a permutation and a deletion to provide the English sentence Arthur gives John a book.

2.3 Dorr's Theory of Divergence

In her article, Dorr suggests to classify divergence according to its nature. She determined the following types of divergence:

1. Thematic: The theme is realized by different constituents in SL and in TL. This divergence might be observed in translating clitic pronouns or reflexive French verbs into a non clitic sentence in English. Example: je me lave les mains (word to word: I wash myself the hands) translates into I wash my hands, whereas je me coupe les cheveux (word to word: I cut myself the hair) translates into I cut my own hair.

2. Promotional: The second sentence, I cut my own hair, realizes a promotional divergence. The clitic pronoun (indirect object) is transformed into an adjective within the direct object constituent. Promotional divergence switches heads.

3. Demotional: another head switching divergence that lessen the importance of the constituent (symmetrical to the preceding one).

4. Structural: The SL word or constituent is transformed into a TL set of words. Example: a French verb like sortir is translated by the English phrasal verb go out.

5. Conflational: The SL constituent conflates into a single TL syntagm. Example: In French, the prepositional phrase Moulin à vent conflates into the English Windmill.

6. Categorial: When the part-of-speech tag changes when going from SL to TL. 7. Lexical: When the default equivalency (natural or apparently morphologically close translation) must not be used. Example: Je demande une réponse, is to be translated by I request an answer, and not by I demand an answer.

The author suggests different patterns for each type of divergence, named CLS (conceptual lexical structures), associated with possible operations on a syntactic structure, among which one can easily recognize our insertion, deletion and permutation operations. Moreover, the categorial divergence introduces a fourth operator we might call catmod, that does not modify the tree structure but operates on the tags decorating its nodes. From our point of view, Dorr’s types of divergence could be well represented in tree transformation rules provided that:

(i) Syntactic tree not only bears syntactic information (constituents) but also dependencies (semantic functions)

(ii) Semantic tags could be present to complete information about nodes (e.g.: Information such as 'location', 'temporal', 'manner', etc…).

Therefore, if the same intermediate representation (the tree structure) is complete enough, then it may both merge syntactic and semantic information. Thus tree transformation rules could deal with both a divergence pattern and a syntactic organization. Moreover, our tree transformation rules may tackle several types of divergence in the same sentence, if they co-exist. Table 1 shows a correspondence between Dorr’s divergence types, and divergence scope and operations.
Dorr’s Divergence Type | Divergence Scope | Transformation Operation
---|---|---
**Thematic** | Dependency | All types
**Promotional** | Constituent Head Switching | Permutation
**Demotional** | Constituent Head Switching | Permutation
**Structural** | Constituent | All types
**Conflational** | Constituent syntactic/part-of-speech | Insertion
**Categorial** | Constituent local operation:Catmod | 
**Lexical** | Constituent /not relevant | All types

Table 1: Divergence types, scope and operations.

3 Tree Transformation From French to English: the SYGFtoE Prototype

3.1 Launching Translation

Translation is launched through a call to SYGFtoE, our prototype developed in the SYGMART environment with a rule editor, and is written in a proprietary language adapted to Markov’s rewriting algorithms. Some rules are given in the next section that illustrate elements of this language. The translation method is the following:

1. The text (as long as the writer wishes) is sent first to the SYGFRAN French parser. A single tree is to be produced. Every sentence, recognized as such by punctuation marks, is a PH substree of the text. For instance, a text of 120 sentences has been sent to the SYGFRAN parser which produced around 2000 nodes.

2. The SYGFRAN tree structures coming as an output, are decorated with English words coming from the Lexical Transfer Module. Thus the tree is ready to enter the Tree Transformation System.

3. Tree transformation first works on the structure. It relies on recognizing patterns and conditions on nodes tags. If parsing has entirely succeeded, the tree transformation scope covers all the applicable rules. If parsing only produced partial structures, tree transformation is only enabled for very local rules with scarce recognition patterns and loose conditions.

4. Last, the system calls the Morphological Flexion Module (also not to be described here, for a lack of place) which properly conjugates verbs and provides gender and number agreement.

5. The writer receives his/her text translated into English.

The general architecture is provided in figure 5. In next subsection, we will focus on the prototype core: the Tree Transformation System.

3.2 Types of Tree Transformation Rules in the Prototype: Some Examples

The prototype embeds several types of transformation rules of different scopes. We only give a few examples of different scopes. SYGFtoE presently contains more than 100 transformation rules, which is very small compared to the parser SYGFRAN number of rules. However this small quantity of rules is particularly efficient, because it relies on an already existing translation, that is the word-to-word translation tree provided by the first step.

3.2.1 Rules of Constituent scope

**phrasal scope:** Noun phrases with the pattern: [DET] [NOUN1] [PREP] [DET] [NOUN2] that have to be transformed into:

(i) The pattern: [DET] [NOUN2]’s [NOUN1]: NOUN2 has the semantic tag ANIMATE. **Example:** la main de la fille, translates into: The girl(daughter)’s hand. Three operations are required: Deletion (of the preposition and the second determinant) permutation and insertion. The form of the rule is:

RGENANIMATE:
0(1,2, 3(4,5,6)) / 0: CAT=NOUNPHR; 1: CAT=DET; 2: CAT=NOUN; 3=CAT = PREPPHR; 6 = CAT = NOUN and (SE- MANT=ANIMATE) → 0(5,6, X, 2)/ X: LEM="s'

First, the rule states its recognition pattern with conditions on some nodes. The root node (here 0) indicates the divergence scope. Since the tested category is NOUNPHR (noun phrase) and only this, then divergence seems to be of phrasal scope, with a few local changes. This might map with some cases of structural divergence observed by Dorr.

(ii) The pattern [DET] [NOUN2] [NOUN1]: if NOUN2 is INANIMATE. Example: la portée de la divergence, translates into: The divergence scope.

RGENINANIMATE: 0(1,2, 3(4,5,6)) / 0: CAT=NOUNPHR; 1: CAT=DET; 2: CAT=NOUN; 3=CAT = PREPPHR; 6 = CAT = NOUN and (SE- MANT!=ANIMATE) → 0(5,6, 2)

(iii) Other noun phrase patterns with adjectives such as: [DET] [NOUN] [ADJ] where the adjective has to switch places with the noun.

syntactic scope: Rules that permutate constituents. The French pattern [VERB] [NOUNPHR1] [PREP][NOUNPHR2] where the verb is transitive and accepts an indirect object, has to be transformed in English into [VERB] [NOUNPHR2][NOUNPHR1] . There is a sort of head switching. In French, the direct object is a 'head' when compared to the indirect object. In English, it goes the other way round. Our example Arthur donne un livre Jean, states that livre is the direct object, and Jean the indirect object. The English translation transform it into: Arthur gives John a book.

RPERMDOOBJINDOOBJ: 0(1,2,3(4,5)) / 0: CAT=VERBPHR and SUB- CAT= transitive; 2: CAT = NOUNPHR ; 3: CAT=PREPNOUNPHR; 4:CAT=PREP → 0(1,5,2)

Verb phrases as roots of a divergence rule generally tend to widen the scope to important subconstituents with dependency functions.

3.2.2 Rules of Dependency scope

Thematic divergence is here considered but also every change that needs heavy dependency information. For instance, the French sentence: Il est mort le poète (word to word: he is dead the poet), has to be translated into the poet died, showing an important change from passive to active form.

The nature of the root node 0 being a sentence (PH) gives an idea of the scope of divergence: it has to be either dependency or maximal scope. In the recognition pattern, 1 must be a pronoun of the second person, and so forth. The rule produces the final pattern (0(3,4(4(X,7)))) where many nodes have been deleted, and a node X inserted, and with lexical changes on node 4. Since the Englishdict (E...) is followed by numbers, then it is a phrasal entry.

3.2.3 Rules of Maximal scope

Total transformation of the sentence for idiomatic expressions. For instance: tu reprendras bien un gâteau (word to word: you will take again a cake) has to be translated into: have another cake. Changes are lexical (reprendre means take again), thematic (the action is replaced by a passive attitude on behalf of the subject of the sentence), structural (insertion of a verb suffix in English, insertion of another, deletion of bien and the determinant). The form of the rule is the following: RIDIOM:

0(*,1,*,3(4,5(6)),7(*,8%9)) / 0: CAT=PH; 1:CAT=NOUNPHR and SUBCAT=PROP and PERS=2; 3: CAT = VERBPHR; etc... → 0(3,(4(X,7))) / 4: Englishdict(E(4,6,8)); X: Englishdict (E(4,6,8),2)).

The nature of the root node 0 a being a sentence (PH) gives an idea of the scope of divergence: it has to be either dependency or maximal scope. In the recognition pattern, 1 must be a pronoun of the second person, and so forth. The rule produces the final pattern (0(3,4(4(X,7)))) where many nodes have been deleted, and a node X inserted, and with lexical changes on node 4. Since the Englishdict (E...) is followed by numbers, then it is a phrasal entry.
4 Results

To test its abilities in ensuring syntactic quality, we have set an experiment with two corpora:

(i) One corpus randomly provided from a set of newspaper articles we had in storage for automatic classification purposes. It contained 700 sentences with an average of 14 words per sentence, which amounts to 9800 words.

(ii) Another corpus extracted from the Grevisse Grammar Handbook, providing sentence samples of grammar rules and exceptions (the equivalent of Collins in English). This corpus is not equivalent to the other in number of words, but is much more discriminant, when syntactic complexity is at stake. It contains 90 sentences of an average of 12 words (a total of 1080 words).

Contrary to statistical approaches, the prototype precision, and its divergence theory implementation are not dependent on the number of words processed, but on the number of transformation rules applied. The first corpus has been translated in 22 seconds, and the second in 45 seconds (total run time for SYFRAN + SYGFtoE).

The first corpus was not discriminant: It contained some ill-formed sentences, partially analyzed by SYGFRAN. Since SYGFRAN precision on a randomly chosen corpus is 34%, we had only a third of Corpus 1 sentences that entered SYGFtoE completely analyzed. This does not mean that all sentences were not translated. Lexical transfer was performed for all sentences, and partially analyzed structures were transformed with rules of a limited scope. But the rationale is that only a third of the output could be used as discriminant for judging SYGFtoE output. Whereas in the second corpus, since sentences were correctly, even 'skillfully', built, contained no misspelled words and no wrong phrases, the local precision of SYGFRAN analysis jumped up to 85%. We then had the ability to observe precision for SYGFtoE on an interesting set of syntactically difficult sentences, where several divergence rules were invoked for final transformation.

Table 2 summarizes the results obtained on both corpora. Recall is here a meaningless measure, since all sentences have been translated, at least with lexical transfer. Precision is much more significant. Two types of precision have been calculated:

- **Local precision**: on the number of completely analyzed sentences called \( cas \), with the following formula: \( Lp = \frac{wts}{cas} \)

where \( wts \) is the number of well translated sentences;

- **General precision**: on the total number of sentences in the corpus, called "tns", with the following formula: \( Gp = \frac{wts}{tns} \)

The results show that 126 sentences of the 238 well analyzed sentences of Corpus 1 and 66 out of the 76 well analyzed sentences in Corpus 2 have been correctly translated. Overall, 175 sentences of Corpus 1 have been correctly translated. The 49 well-translated, but ill-analyzed, sentences were distributed as following:

(i) Word-to-word translation was sufficient (34 sentences were in this case)

(ii) Partial trees were provided by SYGFRAN, and the constituent scope divergence rules applied by the prototype were sufficient (15 sentences).

The remaining 525 sentences of Corpus 1 were all translated: However, translation was a word-to-word or with some transformation rules applied, but translation ranged from 'slightly bad' to 'utterly dreadful'. Sometimes, transformation rules produced a wrong output. For instance, the ambiguous nature of the French determinant *les* led to incorrect translations. The sentence: *Les enfants chantent bien*, could be translated either by: *Children sing well* (if one is talking about children in general) or *The children sing well* (if one is watching children singing and makes such a remark).

This type of decision is difficult to make and needs a human translator, since it involves pragmatics. Considering Corpus 2, 72 sentences were correctly translated: the difference is only on six sentences where word-to-word translation was good enough. In the remaining 18 sentences, bad translations came out of misplaced translation. This led us to redesign transformation by specializing it further (limiting the scope of each transformation).

The interesting element is the average number of transformation rules: In a randomly chosen corpus, with plain sentences, an average of 4 per sentence has been observed, whereas the Grevisse sentences concerned up to 17 transformation rules. In this we do not count lexical transfer rules, only those for syntax, and for conjugation. This is what impedes the prototype runtime. At the same time, Corpus 2 is, for us, the most interesting. An 85% precision on a small corpus of difficult sentences better demonstrates SYGFtoE ability to support syntactic quality than any important number of
Table 2: First prototype results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Corpus1</th>
<th>Corpus2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences number</td>
<td>700</td>
<td>90</td>
</tr>
<tr>
<td>SYGFRAN precision</td>
<td>67%</td>
<td>85%</td>
</tr>
<tr>
<td>Number of complete SYGFtoE input trees</td>
<td>469</td>
<td>76</td>
</tr>
<tr>
<td>Runtime</td>
<td>22s</td>
<td>45s</td>
</tr>
<tr>
<td>Transf. rules avg. numb.per sentence</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Constituent scope divergence rules</td>
<td>75%</td>
<td>60%</td>
</tr>
<tr>
<td>Dependency scope divergence rules</td>
<td>21%</td>
<td>28%</td>
</tr>
<tr>
<td>Maximal scope divergence rules</td>
<td>4%</td>
<td>12%</td>
</tr>
<tr>
<td>SYGFtoE local Precision</td>
<td>69%</td>
<td>87%</td>
</tr>
<tr>
<td>SYGFtoE general Precision</td>
<td>61%</td>
<td>80%</td>
</tr>
</tbody>
</table>

plain and translated sentences. The used rules are mostly of a constituent scope. Let us notice however that Corpus2 requires more dependency and maximal scope rules.

Conclusion

For the moment, the prototype is running all the time. New sentences are provided by anonymous users through a Web window. This means that other experiments and new developments are on the worktable. The remarkable facts to be mentioned as a momentary conclusion are the following: (i) Using a system such as SYGFtoE does not enhance style: If discourse in SL is syntactically and stylistically poor, then it is not better in TL. On the other hand, SYGFtoE main goal is not to impede style and to maintain, as far as possible, the output syntactic quality. (ii) Translating through divergence seems a sound and economical way to process. Some authors (Dorr, Lindop, Tsuji, etc.) have used and modelled divergence. Dorr used it within a conceptual and semantic approach (CLS). We used it within a completely 'formal language'-oriented approach, and showed that Dorr’s classification could be reinterpreted within our framework. In both cases, divergence looked sound: It procrastinated its use until necessity imposed it. Although our corpora are not significant in volume, they gave clue to the fact that 10 to 20% of the French sentences could be translated word-to-word in English. This looks sensible: Divergence between French and English exists, but resemblances also. (iii) As a ‘con’, an approach as ours is very parser-dependent. SYGFRAN precision is a threshold the translator can hardly trespass. However, SYGFRAN improvement is daily made, and we are awaiting a 50% precision in a few months. At this level, we think that SYGFtoE will be able to provide an overall suitable translation, that might be compared with that of heavy systems such as SYSTRAN for instance. (iv) Last, to lessen the impact of the aforementioned limitation, it is important to say that all other tracks in translation, in particular lexical-based tracks, have neglected syntactic validity: The latter is important since translation errors are highly dependent of sentence construction. Our system is improvable in all its modules, but its originality dwells in its concern with syntactic quality.

References


