Test suites as a means for the evaluation of machine translation systems

Yves Lepage

CNRS-ELSAP & EDF-DER

Abstract

In this paper, we present a method for writing test suites for the evaluation of machine translation systems. The items of the test are sentences extracted from a corpus and modified for our purpose. This test addresses a specific linguistic problem and is divided into two parts: the selection test suite and the linguistic problem test suite. The design of the first suite is based on the study of a source corpus; the design of the second one is also based on a small enquiry among native speakers of the target language.

1 Introduction

We present the work currently done at ELSAP (CNRS) and DER (EDF) in the writing of a test suite for the evaluation of machine translation systems. At DER, the interest in machine translation is related to its possible application in an information retrieval system for documentation. In this system, the automatic indexation of plain texts is mainly based on the analysis of noun phrases. So, it would be envisaged to index texts from other languages by only translating the noun phrases they contain. To carry out the work on evaluation of MT systems relevant for this framework, we chose to assess the translations into English of the French preposition *de* in noun phrases.

2 Extracting test suites from a corpus

Our approach in building the test suite has been to extract the items for the test suite from a real corpus. We hope this allows us to better meet the real phenomena which can be encountered in texts of interest for DER.

3   Work on the corpus

3.1   The French corpus

Each occurrence of the form *de* has been automatically marked in a corpus. The collided or elided forms of *de* have been taken into account: *du = de + le, des = de + les* or indefinite article in plural, *d' = de* before a vowel. The number of such occurrences in the corpus is about 650.

Example: ...paramètres * de contrôle...

3.2   The English corpus

As we had SYSTRAN raw translations for the corpus at our disposal, we manually marked the translation equivalents of *de* in the English texts.

Example: ...operating * parameters...

3.3   Associated structures

A list of the different structures corresponding to the form has been drawn up, in French as in English. Each mark is related to one of these structures. For French, the structures are:

structure: gn (det(*), N)
scheme: [det] N
example: * des études

structure: gn (prep(*), det, N)
scheme: de la (du) N
example: varie * de 60 à 280° C

structure: ga (a, gnp (prep (*), det, N))
scheme: A de N
example: accompagné * d'une documentation

structure: phinf (sub (*), v)
scheme: V de V
example: permet * de traiter tout problème

structure: locprep
scheme: prep X de
example: à partir * d'un langage, au sein * des utilisateurs

structure: locadv
scheme: de X
example: * de même, * d'ailleurs

structure: gn (det, N, gnp (prep (*), det, N))
scheme: N1 de N2
example: analyse * des structures

For English, the structures are:

structure: gn (det (*), N)
scheme: [det] N
example: * study groups

structure: ga (a, gnp (prep (*), det, N))
scheme: A prep N
example: accompanied * by an important documentation

structure: phinf (sub (*), v)
scheme: V to V
example: allows * to deal with any problem

structure: locprep
scheme: prep X de
example: * from the beginning of

structure; gn (det, N gnp (prep (*), det, N))
scheme: N1 de N2
example: numerical simulation * of materials behavior

structure: gn (gn (N), *, N)
scheme: N2 N1
example: project * group

structure: gn (a, *, N)
scheme: adj2 N1
example: the Aster mechanical * code

structure: gv (det, gnp (prep (*), N))
scheme: V [prep] N
example: by sequencing * the various routines

3.4   Cross references

Given one French structure and one English structure, all the occurrences of the French *de* corresponding to this structure and whose English translations correspond to this English structure may be listed. For example, here is an excerpt from the list of the translations of the French preposition *de* in noun phrases, translated by an English adjective:

structure: gn (det, n gnp (prep (*), det, n))
schéma: N1 de N2
exemple: analyse * des structures

structure: gn (a, *, n)
schéma: A N
exemple: the ASTER mechanical * code

2 ; R : UN CODE D'Analyse * des Structures et DE
2 ; IS CODE FOR STRUCTURAL * AND THERMOMECHANICAL

3 ; lyse des Structures et * DE Thermo-mécanique pour
3 ; L AND THERMOMECHANICAL * STUDY AND RESEARCH

16 ; iels, en élaborant le code * de mécanique ASTER. La
16 ; ting the ASTER mechanical * code. The development a

24 ; ; - des études dites * de "pointe" avec ajustem
23 ; 1 studies, - advanced * studies with adjustment

62 ; posées par les paramètres * d'exploitation de
62 ; e imposed by the operating * parameters of the m

134 ; isées et des efforts * d'apprentissage
134 ; used and a learning * effort by the co-

142 ; ' interfaçage avec le code * d'élasto-acoustique VA
142 ; the VARIA elasto-acoustic * code (see other "Outst

145 ; tation d'un élément * de thermique qui exploi
145 ; llation of a thermal * element which uses the

163 ; coude Pour une tuyauterie * d'épaisseur 21,44 mm
163 ; elbow For a 21.44 mm thick * pipe with a 385

174 ; ns une zone de 20 mm * d'épaisseur située au
173 ; °C in a 20 mm thick * region located be

4   The "selection" test suite

4.1   Usefulness of the "selection" test suite

Not all French structure types are relevant for our problem. We listed down five structure types, the *de* preposition in noun phrases being only one of them. We then separated the structure types into those relevant for the very problem we had chosen, and the others. These last ones were used to design what we called the "selection" test suite: a test suite ensuring that the machine translation system recognizes some *de* occurrences as not being the preposition in noun phrases. Apart from certain locutions or idioms, the translations of the items in this test suite must not be noun phrase structures.

1. Article:
   *...utilisant comme système de sauvegarde * **des** systèmes tels que bandes,.,.*

2. Subordination:
   *le but est * **de** mémoriser l'expérience*

3. Preposition introduced by a verb or an adjective:
   *accompagné * **d'**une documentation importante*

4. Idioms, prepositional or adverbial locutions:
   *sous forme * **de** scénario*

5. Introducing a noun complement:
   *les paramètres * **d'**exploitation*

4.2   Design of the "selection" test suite

Each item of this test is a French sentence involving one *de* corresponding to one of the four non noun phrase structures. The goal of this test is to "trap" the system in its recognition of *de* not appearing in noun phrases. We extend the test in two directions. First we try to refine the description of each structure, second we complicate the items of the test suite.

4.3   Simple items

These items are obtained by simplifying sentences from the corpus into very simple sentences containing only *de.*

Example:

*Le groupe de projet, auquel participent **des** groupes d'études de cinq services, a été initié fin 1988 pour fédérer les développements de la DER dans le domaine de la simulation numérique **des** comportements d'origine mécanique, thermique ou thermo-mécanique **des** matériels, en élaborant le code de mécanique ASTER.*

Des *groupes participent au projet.*

Below is an excerpt from the "selection" test suite showing simple items for many structure types,

Des *groupes participent au projet.*
*Il nécessite **des** efforts,*
*II faut implémenter **des** fonctionnalités.*
*L'analyse pour **des** études.*
*Le code est développé pour **des** études industrielles.*
*La majorité sont **des** applications à caractère scientifique.*
*Le paramétrage permet **de** traiter tout problème.*
*Le code est accompagné **d'**une documentation importante.*
*Nous disposons **d'**un ensemble d'outils.*
*Au cours **de** développement, on réalise l'accès aux opérateurs validés.*

4.4   Refined items

For each structure, we try to refine the description so as to cover all the possibilities of *de* appearing in this structure. Here is an example of the refinement we have for *des* as indefinite article in plural.

subject: **Des** *groupes participent au projet.*
direct object: *Il nécessite **des** efforts.*
indirect object: *Le code est développé pour **des** études industrielles.*
predicative adjective: *La majorité sont **des** applications à caractère scientifique.*

4.5   Complex items

The complex items are obtained by introducing what we know to be difficulties for machine translation systems. We introduce coordination, broken constructions, we change the position of the occurrence in the sentence and create artificial ambiguities solved by a human being as not being noun phrase preposition.

Des *groupes participent au projet.*
Des *groupes et* **des** *départements participent au projet.*
Des *groupes, qui dépendent de différents départements, participent au projet.*
Des *groupes d'études participent au groupe de projet.*
Le groupe de projet, auquel participent **des** *groupes d'études, a été initié fin 1988.*
Le groupe de projet, auquel participent en premier lieu **des** *groupes d'études, a été initié fin 1988.*

4.6   Reading of the selection test suite

The "selection" test suite written comprises about 100 items. It has just been translated by SYSTRAN and the raw translations obtained are currently being analysed.

Apart from the idioms and locutions, the reading of the translations obtained is as follows: as the occurrences of the *de* form do not correspond to a noun phrase preposition, consequently, the raw translation must not use a rendering of a noun phrase. In examining whether this is the case or not, we are not concerned with the fact that the raw translation be a good translation or not. The role of the "selection" test is to select the very problem we are concerned with by elimination.

The case of idioms and locutions is a more tricky one, as the translation of some locutions may imply the use of noun phrases. These items constitute a fuzzy border between the "core selection" test and the "real problem" test. Here, the appreciation of the "rightness" of the translation requires us to know whether the translation is a good one or not.

5   Concentrating on the linguistic problem

We are concerned with the structure *N1 de N2* noted down in the corpus. The number of occurrences for this structure is high: about 450 among 650 total occurrences.

The previous characterisation of the structures was done in terms of string schemes and syntactic structures. For the French texts, a further study is done which involves logical relations descriptions and semantics. For the English side, an enquiry is made for gathering the different possible translations and confirming our assumptions on them.

5.1 Working on the French side

5.1.1 Logical relations

Some occurrences of *N1 de N2* in the French corpus may be analysed in predicate-arguments terms, depending on wether *N1 or N2* can be considered as a predicate.

We note pred*(N)* the predicate extracted from *N (*pred*(introduction) = introduire),* qual*(N)* the quality obtained from *N*(qual *(possibilité) = possible)*. We number the arguments according to Eurotra usage. Here is a list of the different structures we got:

N1 = arg1(pred(*N2*)) *procedure d'introduction*

N2 = arg1(pred(*N1*)) *participation d'EDF, comportements des matériels*

N2 = arg2(pred(*N1*)) *simulation des comportements*

N2 = arg1(pred(*N1*)) or *N2* = arg2(qual(*N1*)) *disponibilité des réseaux, ouverture du service, possibilité de connexion*

N1 = arg1(pred(*N1*) and *N2* = arg2(pred(*N1*)) *Direction de la Production*

5.1.2 Semantics

Some occurrences of N1 or N2 may also be characterised by their meaning. We grouped some of them under different classes:

N1 = collection: *base de données, bibliothèques de programmes, catalogues de données*

N1 = form: *groupe de travail, document de spécification*

N1 = means, instrument: *système d'exploitation, réseau de transport, application de gestion*

N1 = place: *centre de calcul*

N1 = owner: *ordinateur du (centre de calcul), (centre de calcul) de la DER*

5.2 Working on the English side

5.2.1 Translation possibilities

The raw translations we had at our disposal have been analysed and the structures occurring for the translation of *N1 de N2* have been listed.

We got five possible translations:

N1 of N2: the capacities of the computers ← les possibilités des ordinateurs, parametrisation of the (data structures) ← paramétrages des (structures de données)

N1 other prep N2: group from 5 departments ← groupe de 5 départements

N2 N1: computer center ← centre de calcul, data base ← base de données

adj2 N1: the mechanical code ← le code de mécanique, future prospects ← perspectives d'avenir

Ving2 N1: the document describing the formulation ← le document de description des formulations

where *adj2* stands for an adjective "derived" from *N2, Ving2* a progressive form of a verb "derived" from *N2.*

A common translation of *N1 de N2* in English, say the *N2's N1* structure, does not appear in the corpus. We decided to add this structure in our list as it seems reasonable.

5.2.2 Setting up an enquiry

The translations we have are raw translations from SYSTRAN. In order to work on more reliable data on the English side, we set up an enquiry. Native English-speakers will be asked to tick in arrays to indicate the possible translations for a given French noun phrase (double tick for the best translation). A short truncated context is given around the noun phrase.

Here is an extract of an answer sheet:

Initié fin 1988 pour fédérer les développements de la DER dans le domaine
* de *la simulation numérique des comportements d'origine mécanique.*

N1 of N2	N1 other prep N2	N2 N1	N2's N1	adj2 N1	Ving2 N1

other translation: ....

5.2.3   Making assumptions

We make assumptions on the usual translations of the different classes on the logical relations and the semantic levels. We intend to ask native speakers of English and English professors their opinion about these assumptions.

For example, we assume that the translation of *N1 de N2,* where *N1* is a collection, is *N2 N1* as deduced from the occurrences we gathered from the corpus:

> *base de données→ data base*

> *bibliothèque de programmes → program library*

Moreover, the answers of the enquiry will provide us with new data against which we will check our assumptions.

6   The "linguistic problem" test suite

6.1   Design of the test suite

The last step is the writing of the test suite concerning the linguistic problem. Among all the assumptions we made, some will be verified. The related sentences may constitute the items of a test suite. The others will be classified according to the best translations proposed in the answer sheets.

6.2 Reading of the test suite

After having the test suite translated by a machine translation system, the results will be analysed with the help of the answer sheets from the enquiry. If the best translation is given by the system, the better; a possible translation is accepted as a good one; if the translation does not cope with the answer sheets, we will consider it has to be rejected.

7 Conclusion

We described a general way of designing a test suite to assess the behaviour of a machine translation system on a particular linguistic problem.

We decided mainly to extract the items of the test suite from a corpus of real texts to better cope with the interests of the demander.

A test suite consists of two parts: a "selection" test suite and a "linguistic problem" test suite. The first one draws the boundaries of the selected problem, the second one, based on a linguistic study of the corpus and on an enquiry among native speakers, is used to assess the quality of the translation of the linguistic problem examined.