Translation, The Biggest Challenge For Computers

By Mrs. Abadou Fadila
Phd student at the University Frères Mentouri- Constantine
Supervisor: Mr. Khadich Salah- University of Khanchela

Abstract:

The large number of users of computers and information technology understand the relationship between computer and language as the computer provides great potential to study, analyze and simplify human languages and also to facilitate their education. But the relationship is not so, it involves a lot of complexities and ambiguities. The natural language processing is a difficult and an incomprehensible expression/term because of the challenges that it poses. The automatic processing of language is primarily for the development, the creation and the formulation of machine translation, whether written or spoken, because translation is the only means by which nations can cope with the explosion of knowledge and information; no society or nation can live independently of this knowledge revolution in front of the hegemony of the English language in the world. The purpose of this research is to know the challenges posed by translation in this field and to what extent the computer succeeded in providing a translation equal to the human translation.

Key words: *Natural language processing, machine translation, computer*

Résumé:

Le grand nombre d'utilisateurs d'ordinateurs et de technologies de l'information comprend la relation entre l'ordinateur et la langue comme l'ordinateur offre un grand potentiel pour étudier, analyser et simplifier les langues humaines et même pour faciliter leur éducation. Mais la relation est plus profonde, elle implique beaucoup de complexités et d'ambiguïtés. Le traitement du langage naturel est une expression / un terme difficile et incompréhensible à cause des défis qu'il pose. Le traitement automatique du langage est principalement destiné au développement, à la création et à la formulation des systèmes de traduction automatique, écrite ou orale, car la traduction est le seul moyen par lequel les nations peuvent faire face à l'explosion des connaissances et des informations; aucune société ou nation ne peut vivre indépendamment de cette révolution de la connaissance face à l'hégémonie de la langue anglaise dans le monde.

Mots clés: traitement automatique du langage naturel, traduction automatique, ordinateur.

Introduction:

The technical revolution has led to the development of all fields of knowledge. And the human beings use of machines is increasing day after day because of the accuracy of the machine in addition to the availability and the speed of delivery in the process of its production. In fact the machine did not find a great challenge to enter any kind of factory and to achieve success in its performance.

One of the devices that push technology is the computer that is the highest product of human being till now. Actually, the area that challenges this device is language because it is the peculiarity of man; to speak about human languages is to deal with "natural languages", because they are, in a way, spontaneous collective creations to which we cannot attribute a precise date of birth. Natural languages are thus opposed mainly to "artificial languages" or "formal languages" such as computer programming languages or mathematical logic.

The object of linguistics, or of the language sciences, is the scientific study of natural languages. Behind the apparent diversity of human languages, linguists try to track common operations, shared structures, and universals not to control or limit their natural evolution but to observe them as they speak and write to each other. For this study, computer science plays increasingly an important role, via the field of Natural Language Processing (NLP). Computer science is a recent scientific discipline, which should not be reduced to the simple use of computers and programs; its name refers to it as the science of "automatic processing of information".

Under the heading of automatic natural language processing (NLP), we group together all research and development aimed at modelling and reproducing, with the help of machines, the human capacity to produce and understand linguistic utterances for communication purposes. This will be a question of human language, hence the adjective 'natural', and not formal language, because the formal languages are precisely designed and optimized in the optics of algorithmic manipulation. It is quite different for the natural language, whose automatic processing poses major difficulties especially in translation because of the number of languages involved in the translation process.

Although man is still dependent on his own ability to perform translations between different languages, the technical revolution and the development of the computer led to serious thinking in the introduction of computing into translation, known as Machine Translation (MT) or Computer Aided Translation (CAT). This depends on several factors, the first is to develop the device to the extent that it can deal with the languages in this area, and the second is to prepare languages in a way that allows the computer to deal with them. Translation is the biggest challenge for computers in the field of human languages, for the simple reason that dealing with human language depends on the mental capacity of humans, especially in the field of translation, it is more complex because of the number of languages included in the translation process; and this is not an automatic work as in other things, such as manufacturing and operation of vehicles of various types.

1. The difficulties of NLP: ambiguity and implicit

The difficulties of the NLP are mainly of two kinds, and emerge either from the ambiguity of the language, or from the quantity of implicit contained in the natural communications.

1.1. Ambiguity

Natural language is ambiguous. This ambiguity, far from being marginal, is one of its characteristic features. We can see the result of an inevitable compromise between, on the one hand, an almost unlimited capacity of expression, and on the other, the constraints related to the limitation of the physiological resources implemented (size of the memory to long and short-term, density of phonetic space, articulatory constraints, etc.). This ambiguity manifests itself in the multitude of possible interpretations for each linguistic entity relevant to a level of treatment, as shown by the following examples:

- Ambiguity of graphemes (letters) in the process of orthographic encoding: compare the pronunciation of 'i' in 'lit, poire, maison';
- Ambiguity of the endings in the process of conjugation: thus a / s / final marks at the same time the plural of the nouns, the adjectives (in French), and the

second/the third (sometimes also the first) person of the singular of the verbal forms;

- Ambiguity of the grammatical function of groups of words, illustrated by the sentence:

He pursues the girl on a bike.

In this example, 'on a bike' is either a complement of way to continue (and it is 'he' who pedals), or a complement of the noun 'girl' (and it is 'she' who mills);

- Ambiguity of the scope of quantifiers, conjunctions, prepositions. So, in: All my friends have a drink.

We can assume that everyone had a different drink/cup

But in: All the witnesses heard a cry.

It is likely that it was the same cry for all witnesses.

- Ambiguity about the interpretation to be given in context to an utterance. Compare the "meaning" of no in the two following exchanges:

(a) If I go to class tomorrow? No (negation)

(a) You're going to class tomorrow! No ! (I cannot believe)

In accordance with what have been mentioned above, we have especially emphasized the ambiguities of recognition, but the problems naturally arise in a symmetrical way with regard to generation: how to choose the sentences produced in such a way as to limit ambiguities for the receiver? How to select from a set of synonyms and among a set of paraphrases?

1.2. Implicitness

Language activity is always part of a context of interaction between two humans, sensibly endowed with a knowledge of the world and its functioning such as the immense majority of the elements of context necessary for the disambiguation but also with the understanding of 'a natural statement that can remain implicit. The situation

changes completely as soon as a machine tries to fit into a process of natural communication with a human: the machine does not have this background knowledge, which makes the complete comprehension of the majority difficult, if not impossible, if we do not have additional knowledge bases, giving access to both knowledge about the world (or domain) in general (static knowledge) and the context of enunciation (dynamic knowledge). [1]

1 Kouassi R. R. , 2009, *La problématique de la traduction automatique*, Revue n. 4 du Laboratoire des théories et des modèles linguistiques LTML, Université du Cocody, ISSN 1997 4256, parution décembre 2009, pp. 1-30.

Fortunately, there are many applications for which these difficulties can be, to a large extent, circumscribed. Therefore, since the framework of the analyzed texts is restricted to a particular sub-domain (legal texts, scientific texts, information server specialized in sports information, etc.), it becomes possible, on the one hand, to ignore a large number of ambiguities, in particular semantics; and on the other hand, to formally represent a large number of knowledge necessary for understanding the statements of the field considered. In fact, certain activity domains or specific interaction contexts seem to drastically restrict all possible (or acceptable) statements, considerably simplifying the processing of these real sub-languages by a machine.

2. Language Dynamism and Machine Translation:

Translation is a dynamic process that consists of Transferring into another language (terminating language - LT) the meaning of a text [written in another language (original language LO)] in the sense intended by the author"

> 'verter a otra lengua (lengua terminal - LT) el significado de un texto [escrito en una otra lengua (lengua original LO)] en el sentido pretendido por el autor.'[2]

Automatic translation poses the problem of the universals of language and that of the principles and inherent parameters of internal functioning to each natural language. Two essential facts are to be noted here: the dynamism of the natural metalanguage and the dynamism of the enunciation context.

2.1. The dynamism of the natural metalanguage

The natural metalanguage is the speech of the language on itself. It expresses the abstract microphenomena of work in the natural language. The language thus folds back on itself to account for its own functioning. The natural metalanguage is the foundation of any use of the language in context. The capture of this metalanguage makes it possible to explain the expressions or effects of meaning.

---

2 Newmark, P., 2004, *Manual de traducción*, Madrid, Cátedra, P.19

Machine translation is fundamentally based on approaches that are organized around the fact of language as a phenomenon, that is, as a realized fact or product. Through establishing correspondences between words and expressions and writing computer processing algorithms to program these correspondences, the computer will produce the expected response when the stimulus associated with is engaged. This back and forth between stimuli and responses is the result of programming rules, probabilities and matches. It is an important work that leads to more or less satisfactory results. Using statistics, the frequencies of better translations from increasingly rich and varied corpora will be modelled. However, the fact that the analyses are focused on data or corpora drawn from the linear chain of statements poses the problem of the adequacy of rules and programs. What is to be modelled is the transition from the signified of power to the signified of effect, the operations of setting into speech.[3]

2.2. The dynamism of the enunciative context

A text is the result of an enunciation. This enunciation is possible by dynamic mechanisms architected by a subject enunciating in a situation or a context of enunciation. Any statement produced is therefore new when the spatio-temporal parameters are modified at each instance, which underlies the fact that the computation base of the computer model is constantly obsolete. This implies permanent updates for efficiency always preserved or improved. Systran has set up this type of self-regulating module for continuous improvement of translated texts. But this module always works from corpora of sentences and expressions stripped of their contexts of enunciation. The model will hardly interpret the senses.

We have tested (two online) translation software by submitting the same texts to translate. They are *Google Translation* and *Systran* because they position themselves as the market leaders in our days due to the hybrid engine they use in addition to the syntactic-semantic analysis and the statistical approach that is based on aligned bilingual corpus. Indeed, a link is created between each part of the text of the source language and the corresponding part in the target language. This link is usually created at the sentence level.

3 Kouassi R. R. (2009), ibid.

The statistical analysis uses the redundancies existing in this corpus to estimate the parameters of the translation process. "Statistical translation is possible because ad hoc models are coupled with dynamic programming algorithms that maximize a translation function from a source sentence to a target sentence." [4]

The results are shown below:

Example one:

Text	**Systran**	**Google Translation**
L'ONU s'efforçait hier d'obteni r en dernière minute le soutien	UNO endeavoured yesterda y at the last minute to obtain	The United Nations was trying yesterday to obtain last

des Grecs et des Turcs pour son plan de réunification de Chypre. Le pr ojet final présenté par Kofi Annan au terme d'une semaine d'âpres négociations en Suisse doit être soumis par référendum aux Chypriotes, le 20 avril. (Le Figaro, 1.04.04)	the support of the Greeks and the Turks for his plan of reunification of Cyprus. The final project presented by Kofi Annan at the end one week according to negotiations in Switzerland must be subjected by referendum with Cypriot, on April 20th. (Le Figaro, 1.04.04	minute support from the Greeks and Turks for its plan to reunify Cyprus. The final draft presented by Kofi Annan after a week of tough negotiations in Switzerland must be submitted by referendum to the Cypriots on 20 April. (Le Figaro, 1.04.04)

Src.: L'ONU s'efforçait hier d'obtenir **en dernière minute** le soutien des Grecs et des Turcs…

Gt.: The United Nations was trying yesterday to obtain **last minute** support from the Greeks and Turks...

Ref.: the problem here is of sentence structure; The United Nations organisation was trying yesterday **at the last minute** to obtain the support of the Greeks and the Turks...

- The problem here is of syntax structure; '*En dernière minute*' is a discourse connective that mentions the time, but it is ambiguous in the MT output, yet, it is vital for the correct understanding of the text.

---

4 Lavecchia, Caroline, Kamel Smaili et David Langlois, « Une alternative aux modèles de traduction statistique d'IBM: Les triggers inter-langues », *LORIA/Speech Group*, Campus scientifique, Juin 2008.

Src.: ... pour **son** plan...

Sys.: ...for **his** plan...

Ref.: ...for **its** plan …

- The problem here is of referencing anaphora; '*Son'* is a personal pronoun. It refers to the UNO. It is wrongly rendered as 'his' because The context of the preceding sentences is absent, meaning that the reference is undetermined.

Src.: Le projet final présenté par Kofi Annan **au terme d'une semaine**…

Sys. : The final project presented by Kofi Annan **at the end one week**...

Ref.: The final draft presented by Kofi Annan **after a week...**

- The problem here is of syntax structure; 'Au terme d'une semaine' is a temporal discourse connective, it is not clear in the translation of systran; the natural logic of the sentence *' at the end one week'* is distorted, a functional word(preposition) is missing.

Example two:

Text	Systran	Google translate
Droite comme un i dans sa cou rte robe de velours noir, Marie -Laure a peur. C'est pour la première fois qu'elle monte sur scène. Elle joue l'ouvreuse dans la pièce *Chantier interdit au public*. (À côté d'elle) Il y a un homme au canotier et au nœud papillo n et un autre vêtu de blanc, à la mourstache remontante et un autre, courbé sur sa canne, au complet de velour s vert. [Le Figaro, 20.10.03]	Right-hand side like an I in its short black velvet dress, Marie-Laure is afraid. It is for the first time that it goes up on scene. She plays the usherette in the part Building site prohibited with the public. (At side of it) There is a man with the rower and the noeu butterfly and another dressed in white, wit h the moustache going up and another, curved on its cane, complete of green velvet .	Straight as an i in her short black velvet dress, Marie-Laure is afraid. It is for the first time that she goes on stage. She plays the opener in the room Chantier prohibited to the public. (Beside her) There is a man with a canoe and a butterfly, and another dressed in white, with a rising mourstache, and another, bent over his cane, in a suit of green velvet. [Le Figaro, 20.10.03]

Src.: Droite comme un i …

Sys. : Right-hand side like an I ...

Gt.: Straight as an i ...

- There is a problem of lexical cohesion, *'Right-hand side like an I'* is not the appropriate translation for *'Droite comme un i'* that means *'Who stands very straight, rigid'*. The wrong word renders the sentence incoherent.

Src. : … dans **sa** courte robe de velours noir, Marie-Laure a peur.

Sys. : ... in **its** short black velvet dress, Marie-Laure is afraid.

Ref.: ... in **her** short black velvet dress, Marie-Laure is afraid.

- The problem here is of referencing anaphora; '*Sa'* is a possessive adjective pronoun. It refers to Marie-Laure. It is wrongly rendered as 'its' because the context of the preceding sentences is absent, meaning that the reference is undetermined which makes the sentence incoherent.

Src. : C'est pour la première fois **qu'elle** monte sur scène.

Sys. : It is for the first time that **it** goes up on scene.

Ref.: It is for the first time that **she** goes on stage.

- The first problem in this sentence is of referencing anaphora; '*elle'* is a personal pronoun. It refers to Marie-Laure. It is wrongly rendered as 'it' because the context of the preceding sentences is absent, meaning that the reference is undetermined which makes the sentence incoherent

Src.: Elle joue **l'ouvreuse** dans **la pièce** *Chantier interdit au public…*

Sys. : She plays **the usherette** in **the part** Building site prohibited with the public.

Gt.: She plays **the opener** in **the room** Chantier prohibited to the public...

Ref.: she plays the usherette in the piece *'chantier prohibited to public'*

- there is no lexical cohesion in this sentence, the words are 'opener/ part room building site' are not appropriate for context that is missing because of the sentence-by-sentence system of translation. For the word 'la piece' , it is a piece of theatre; for chantier it can be put as it is because it is a title of the piece of theatre.

Src.: À côté d'elle…

Sys. : At side of it....

Gt.: Beside her...

Ref.: beside her

- First, there is a problem of referencing anaphora, it is not appropriate to refer to *'elle'*, the correct word is the possessive adjective *'her'*. Second, there is a problem of sentence/syntax structure in *'At side of it*, it is a discourse connective expressing the place'; normally we should say *'at her side'* or *'by her side' or 'beside her'*

Example three:

Text Source	Google Traduction
Le succès des cafés où l'on discute Paris, 18h 30. Au premier étage d'un café du XI ème arrondissement, un petit groupe de gens s'installe autour des tables. Certaines personnes se connaissent, d'autres non, mais tous se disent bonjour chaleureusement. Elles sont de tous les âges, de tous les milieux : trois femmes élégantes, deux jeunes, des hommes de cinquante ans et quelques retraités… ce soir comme chaque semaine, le bistrot est réservé aux participants d'une réunion un peu particulière. Au début, il y a déjà dix ans, personne n'y croyait vraiment. Accueillir des gens dans un café, pour parler de littérature, de politique, de philosophie ou simplement des petits soucis de la vie quotidienne, l'idée semblait tellement bizarre! Et portant, aujourd'hui, ces rendez-vous sont devenus célèbres et de plus en plus de Parisiens participent à ses réunions. Autour d'un verre, chaque personne présente ses idées et écoute celles des autres. Un animateur est choisi parmi tous et c'est lui qui organise la discussion. « à quoi sert la culture? », « qu'est –ce qu'éduquer? », « comment mieux vivre une période de chômage? », etc. tous ces thèmes attirent une clientèle nombreuse. Devant le succès de ces café où l'on discute, certains patrons ont eu l'idée de créer d'autres lieux, d'un style un peu différent, mais en gardant la même ambiance de convivialité. Samedi prochain, par exemple, le café Zanzibar devrait ouvrir ses portes dans le XX ème arrondissement : chacun pourra y parler de ses problèmes en demandant conseil à un médecin. Moins sérieux mais aussi outil, le Bricolo Café vous permet d'apprendre à bricoler. Deux fois par jour, à 12h30 et 16h00,	**نجاح المقاهي حيث مناقشة** باريس، 30. 18 H في الطابق الأول من القهوة الحادي عشر عشر منطقة، وهي مجموعة صغيرة من الناس استقرت حول الجداول. بعض الناس يعرفون، وليس بعض، ولكن كل التحية بحرارة. وهم من جميع الأعمار، من جميع الخلفيات: ثلاث نساء أنيقة، شابين من خمسين عاما وبعض المتقاعدين ... هذه الليلة كما كل أسبوع، محجوز حانة صغيرة للمشاركين في اجتماع خاص نوعا ما. في البداية، يكون هناك بالفعل عشر سنوات، لا أحد يعتقد حقا. تحية الناس في مقهى للحديث عن الأدب والسياسة والفلسفة أو مشاكل صغيرة فقط من الحياة اليومية، يبدو أن فكرة غريبة جدا! ومع اليوم، أصبحت هذه الاجتماعات الشهير والمزيد والمزيد من الباريسيين حضور اجتماعاتها. مشروب، كل شخص لديه أفكاره والاستماع إلى الآخرين. يتم اختيار ميسر من جميع وانه هو الذي نظم مناقشة ما هو الغرض من الثقافة؟ ما هو التعليم؟ "، كيف يمكن أن يعيش بشكل أفضل فترة من البطالة؟ »، الخ. كل هذه المواضيع جذب عملاء كبيرة. ونظرا لنجاح هذه المقاهي أو يجري مناقشتها، وكان بعض زعماء فكرة إنشاء أماكن أخرى، وهو أسلوب مختلف قليلا، ولكن الحفاظ على نفس أجواء العيش المشترك. يوم السبت المقبل، على سبيل المثال، مقهى زنجبار المقرر افتتاحه في حي العشرين: الجميع سوف يكون الحديث عن مشاكله مع المشورة من الطبيب. أقل خطورة ولكن أيضا أداة Bricolo مقهى

un vendeur très aimable vient vous montrer comment changer les fenêtres de votre appartement, refaire les peintures, etc. grâce à ces petits cours, le bricolage n'aura plus de secret pour vous! Et de semaine en semaine, de café en café, les gens se cherchent et se retrouvent. Le café, lieu de rencontres et de discussion, devient alors un lieu de vie indispensable. **Infos Matin, 27 mars 2003**	يسمح لك لتعلم العبث. مرتين في اليوم، في تمام الساعة 12:30 و16:00، بائع ودي جدا فقط تظهر لك كيفية تغيير النوافذ في شقتك، الطلاء الجديد، الخ ذلك بفضل هذه الدروس الصغيرة، دي لن يكون لديك أسرار بالنسبة لك! ومن أسبوع لآخر، من مقهى إلى مقهى، والناس تسعى وتجد. المقهى مكان الاجتماع والمناقشة، يصبح مكانا للحياة الحيوية. مورنينغ نيوز، 27 مارس 2003

Fr. : Au premier étage **d'un café** du **XI ème arrondissement**

Ar.:من **القهوة الحادي عشر عشر منطقة**

There is no **lexical cohesion** in the translation of this sentence; the word *'un cafe'* does not mean the drink that we take but the place where this service is offered.

We also find a problem in the order of words, meaning that the sentence is not properly constructed in the target language. 'IX' can be considered as an adjective of 'arrondissement'. In Arabic the noun is always followed by the adjective that qualifies it, so it is more correct to say: "في الطابق الأول لمقهى في الحي 11 "

Fr. : un petit groupe de gens s'installe autour **des tables**

Ar.: وهي مجموعة صغيرة من الناس استقرت حول.

In this sentence we find the word *"des tables "* translated by "**الجداول** " but it does not carry this meaning, as they refer to "table to sit down to eat or drink"

Fr. : Certaines personnes se connaissent, **d'autres non**, mais **tous se disent bonjour chaleureusement.**

Ar.: بعض الناس يعرفون، **وليس بعض**، ولكن **كل التحية بحرارة**.

The translation is not well structured; there is no logical link between its items. For example, "بعض الناس يعرفون" the meaning is incomplete. The translator does not really

understand that "se connaissent" is a reflexive/pronominal verb, meaning that the doer and the receiver of the action are one person. And the expression " كل التحية بحرارة " perhaps the reader can understand the meaning but they may be surprised by the linguistic structure of the sentence, it is more correct to translate the verb in the original text "se disent" by a verb in the target language to get the meaning of the sentence and its structure also, we say:

"بعض الأشخاص يعرفون بعضهم، أما آخرون فلا، لكن الكل يلقي السلام بحرارة

Fr. : ce soir comme chaque semaine, **le bistrot est réservé** aux participants d'une réunion un peu particulière.

Ar.: هذه الليلة كما كل أسبوع، **محجوز حانة صغيرة** للمشاركين في اجتماع خاص نوعا

"محجوز حانة صغيرة" is the controversial expression in the translation, since the adjective precedes the noun (it is not in Arabic), it is better to say " حانة صغيرة محجوزة ", the error here concerns the syntax structure of the sentence and the words order.

Fr. : Au début, il y a déjà dix ans, **personne n'y croyait vraiment**.

Ar.: في البداية، يكون هناك بالفعل عشر سنوات، **لا أحد يعتقد حقا**.

In this sentence, there is an omission; the word "y", which refers to *'les cafés'*, is not translated. It is more correct to say: "في البداية، قبل عشر سنوات، لم يكن أحد يؤمن بهذه المقاهي"

Fr. : ces rendez-vous sont devenus célèbres et de plus en plus de **Parisiens participent à ses réunions**

Ar.: أصبحت هذه الاجتماعات الشهير والمزيد والمزيد من **الباريسيين حضور اجتماعاتها**.

"والمزيد من الباريسيين حضور اجتماعاته" , This sentence is incorrect, no verb is found, although the original sentence contains it, to correct the syntax of the sentence we

say: " أصبحت هذه الاجتماعات شهيرتا و المزيد المزيد من الباريسيين يحضرونها " or وقد تضاعف عدد الباريسيين

" فيها "

Fr. : chaque personne **présente ses idées et écoute celles des autres**

Ar.: كل شخص **لديه أفكاره والاستماع إلى الآخرين**

The same mistake is found in this sentence, the use of "الاستماع" to translate the verb "écoute", it is best to translate the verb by a verb in this sentence to give a correct sentence.

Fr. : « **à quoi sert la culture?** », « qu'est –ce qu'éduquer? », « **comment mieux vivre** une période de chômage? », etc. tous ces thèmes **attirent** une clientèle nombreuse.

Ar.: ما هو الغرض من الثقافة؟ ما هو التعليم؟ "، كيف يمكن أن يعيش بشكل أفضل فترة من البطالة؟ » الخ.

كل هذه المواضيع جذب عملاء كبيرة.

The same problem is found here: *tous ces thèmes attirent une clientèle nombreuse*

Fr. : Devant le succès de ces cafés **où** l'on discute,

Ar. : ونظرا لنجاح هذه المقاهي **أو** يجري مناقشتها،

'**Où**' expresses the place and not the choice as stated in the translation.

Fr. : grâce à ces petits cours, **le bricolage n'aura plus de secret pour vous!**

Ar. : ذلك بفضل هذه الدروس الصغيرة، **دي لن يكون لديك أسرار بالنسبة لك**

دي لن يكون لديك أسرار بالنسبة لك, we should put a question mark on this phrase. what means "De", it can be translated by " إصلاح/ تصليح ". Concerning " لن يكون لك أسرار بالنسبة لك " , we should say " لن يكون له " i mean " تصليح " because the writer addresses the participant in these cafés and not the act of repairing. We can provide a better translation:

"بفضل هذه الدروس لن يكون لعملية التصليح إسرار بالنسبة لك"

Conclusion:

The examples mentioned above highlight some of the issues that MT approaches deal poorly with.

Lexical cohesion

Often, lexical cohesion occurs not simply between pairs of words but over a succession of a number of nearby related words spanning a topical unit of the text. These sequences of related words will be called lexical chains that consist of a sequence of related words that contribute to the continuity of meaning based on word repetition, synonymy and similarity. There is a distance relation between each word in the chain, and the words co-occur within a given span. Lexical chains do not stop at sentence boundaries. They can connect a pair of adjacent words or range over an entire text. Lexical chains tend to delineate portions of text that have a strong unity of meaning. [5]

Somasundaran et al. (2014) consider how **lexical chains,** affect discourse coherence. They use lexical chaining features such as length, density, and link strength to detect textual continuity, elaboration, lexical variety and organisation, all vital aspects of coherent texts. They claim that the interaction between lexical chains and discourse cues can also show whether cohesive devices are organised in a coherent fashion. [6]

There are two major reasons why lexical cohesion is important for computational text understanding systems:

1. Lexical chains provide an easy-to-determine context to aid in the resolution of ambiguity and in the narrowing to a specific meaning of a word.
2. Lexical chains provide a clue for the determination of coherence and discourse structure, and hence the larger meaning of the text.[7]

5 Morris, J. , Hirstt. G., 1991, *Lexical Cohesion Computed by The causal Relations as an Indicator of the Structure of Text*, Association for Computational Linguistics.

6 Somasundaran et al., 2014, *Lexical Chaining for Measuring Discourse Coherence Quality in Test-taker Essays*, Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 950–961, Dublin, Ireland, August 23-29 2014.

7 Morris Jane, Hirstt. G., ibid

MT has been shown to be consistent in its use of terminology [8], which can be an advantage for narrow texts domains with significant training data. However, MT systems may output direct translations of source text items that may be inappropriate in the target context. Moreover, while a specific target text word may correctly translate a source text word in one context, it may require a totally different word in another context.

Referencing Anaphora resolution is a very challenging issue in current MT [9] approaches because they translate one sentence at a time that makes the context of the preceding sentences absent, meaning that the reference is undetermined. Even once it is correctly resolved, reference resolution is directly impacted by linguistic differences, for example, the target language may have multiple genders for nouns while the source only has one. The result is that references can be missing or wrong.

Discourse connectives, those which can be temporal or causal in nature, are vital for the correct understanding of discourse. Yet in MT systems they can be incorrect or missing the fact that distorts the meaning of the text. In particular, where discourse connectives are ambiguous or implicit, the MT system may choose the wrong connective translation because it cannot detect it while the human translator can.

Syntax structure: Different languages have different syntactic structures; each language contains specific rules for properly connecting syntactic items to form a sentence. In MT system, the syntax of the target language may get distorted, when it is too close to the syntax of the source language that leads to an incoherent sentence formation because it violates its syntactic rules and gives a syntactically (semantically)

In addition to the classical phenomena, such as ambiguity, references and ellipses, found in the Latin languages, and that can make it difficult for the computer to deal with either comprehension, re-generation or translation, there are other problems specified to Arabic and some Semitic languages, as: The absence of verbal movements (vowels), the

absence of punctuation marks and the problem of agglutination, and other characteristics that would also limit the performance of the computer since the Arabic language is still experiencing a significant delay in this area compared to other languages, mainly French and English, due to the stagnation witnessed (in several fields) in the arabe world.

As a result, the translation activity is under the full domination of Man. It is a truly (socio-) cognitive, linguistic and stylistic activity that the machine will have a lot of troubles to integrate it totally.

> 'Le système de traduction complètement automatique qui produira les textes idiomatiques comparables aux traductions humaines n'est qu'un rêve pour ceux qui n'en ont pas l'expérience. Toutefois, dans les conditions propices, ces systèmes, loin d'être parfaits, peuvent être utilisés avec profit et succès.' [10]

8 Carpuat. M., Simard., 2012, *The trouble with smt consistency*, In Proceedings of WMT, pages 442–449, Montreal, Canada.

9 Nov´ak. M., 2011, *Utilization of anaphora in machine translation*, In WDS Week of Doctoral Students, June.

10 Hutchins, J., , *Vers une nouvelle époque en traduction automatique*, Troisièmes Journées Scientifiques LTT, Montréal, 30 septembre 1993

Bibliography:

Carpuat. M., Simard., 2012, *The trouble with smt consistency*, In Proceedings of WMT, pages 442–449, Montreal, Canada.

Hutchins, J., 1993, *Vers une nouvelle époque en traduction automatique*, Troisièmes Journées Scientifiques LTT, Montréal, 30 septembre 1993

Kouassi R. R., 2009, *La problématique de la traduction automatique*, Revue n. 4 du Laboratoire des théories et des modèles linguistiques LTML, Université du Cocody, ISSN 1997 4256, parution décembre 2009, pp. 1-30.

Lavecchia, C., et al., 2008, *Une alternative aux modèles de traduction statistique d'IBM: Les triggers inter-langues*, LORIA/Speech Group, Campus scientifique, Juin 2008.

Morris, J. , Hirstt. G., 1991, *Lexical Cohesion Computed by The causal Relations as an Indicator of the Structure of Text*, Association for Computational Linguistics.

Nov´ak. M., 2011, *Utilization of anaphora in machine translation*, In WDS Week of Doctoral Students, June.

Newmark, P., 2004, *Manual de traducción*, Madrid, Cátedra.

Potet. M., Esperanc. Besacier. E. L. and Blanchon. H., 2012, Collection of a large database of french-english smt output corrections.

Somasundaran et al., 2014, *Lexical Chaining for Measuring Discourse Coherence Quality in Test-taker Essays*, Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 950–961, Dublin, Ireland, August 23-29 2014.