Revue de Traduction et Langues Volume 18 Numéro 02/2019, pp. 138-153
Journal of Translation and Languages **مجلة الترجمة واللغات**
ISSN (Print): 1112-3974 **EISSN (Online): 2600-6235**

Coherence in Machine Translation Output

Abadou Fadila
University of Constantine1 Frères Mentouri -Algeria
abadou.fadila.07@gmail.com
0000- 0002-3212-1697
Pr. Khadich Saleh
University of khanchela Abbès Laghrour-Algeria
lissaniat@yahoo.fr
0000-0002-8959-4217

To cite this paper:
Abadou, F. & Khadich, S. (2019). Coherence in Machine Translation Output. *Revue Traduction et Langues 18(2),* 138-153.

Received: 09/09/2018; **Accepted**:29 /12/2019, **Published:** 31/12/2019

Abstract: *Coherence is a cognitive process. It plays a key role in argumentation and thematic progression. To be characterised by appropriate coherence relations and structured in a logical manner, coherent discourse/text should have a context and a focus. However, it receives little attention in Machine translation systems that considers the sentence the largest translation unit to deal with, the fact that excludes the context that helps in interpreting the meaning (either by human or automatic translator). In addition to that, Current MT systems suffer from a lack of linguistic information at various stages (modelling, decoding, pruning) causing the lack of coherence in the output. The present research aims at, first, capturing the different aspects of coherence, and second, introducing this notion in texts generated by machine translation based on sentence-by-sentence basis, in order to see and discuss the several phenomena that can lead to incoherent document translations with different language pairs.*
Keywords: *Coherence, (Lexical) Cohesion, Discourse Connectives, Machine Translation, Referencing Anaphora*.
Résumé : *La cohérence est un processus cognitif. Elle joue un rôle clé dans l'argumentation et la progression thématique. Le discours/texte cohérent doit avoir un contexte et un objectif caractérisés par des relations de cohérence appropriées et structurés de manière logique. Cependant, elle a reçu peu d'attention dans la traduction automatique (MT) car la plupart des systèmes traduisent le texte phrase par phrase, indépendamment du contexte. De plus, les approches actuelles de la traduction automatique souffrent d'un manque d'informations linguistiques á différentes étapes (modélisation, décodage, transformation), ce qui entraîne à des textes incohérents. La présente étude s'intéresse d'abord à la description des différents aspects de la cohérence, et ensuite à l'introduction de cette notion dans les textes générés par la traduction automatique pour voir et discuter les différents phénomènes pouvant conduire à des traductions incohérentes avec des différentes paires de langues.*
Mots clés : *Anaphores, Cohérence, Cohésion (Lexicale), Connecteurs de Discours, Traduction Automatique.*

Corresponding author : Abadou Fadila

1. Introduction

Machine Translation or TA is a sub-domain of computational linguistics that works on the theory and practice of the use of computers for the translation of written and oral texts from one natural language to another. It is also called computer-assisted translation (CAT) or software translation. In fact, the best translation software provides, at the level of isolated sentences from different human languages, correct results, but when we try to translate series of sentences the results are rather disappointing if not disastrous. This is because the fact that the different systems of translation can analyse relatively well the syntactic-semantic relations that exist within the sentence but they are hardly able to grasp the relations established between sentences. Thetransphrastic relations without which a sequence of sentences could never form what is commonly called a text whatever the genre to which it belongs. While it has been extensively commented upon, coherence is not taken into account in machine translation systems because the models that integrate and exploit knowledge or language resources do not rely on a global vision of the text and the themes discussed. They are based on the notion of saliency of a textual unit, a sentence or a paragraph, and this salience is calculated independently on the thematic structure of source texts.

2. Related Works

The Trouble with Machine Translation Coherence (Sim Smith et al., 2016) consists of an Analysis of adapted coherence models in an MT setting. It shows that assessing coherence in SMT is a far harder task for existing models than trying to reorder shuffled texts.

Sim Smith et al. investigate local coherence models for a different scenario, where texts are automatically translated from a given language by systems of various overall levels of quality. Coherence in this scenario is much more nuanced, as elements of coherence are often present in the translations to some degree, and their absence may be connected to numerous types of translation errors at different linguistic levels. There are undeniably grammatical issues, but arguably a proportion of these do indirectly affect coherence.

A Coherence Corpus in Machine Translation (Sim Smith et al., 2015) includes corpus analysis, and examines the types of coherence errors that frequently occur in SMT. It finds that different language pairs result varying types of coherence errors. (Sim Smith, 2017)

Topic-based coherence modelling for statistical machine translation (Xiong, D. et al. 2015) proposes topic-based coherence models to produce coherence for document translation. these models are based on the continuity of sentence topics in a text through extracting, automatically, a coherence chain for each source text to be translated, adopting a maximum entropy classifier to predict the target coherence chain that defines a linear topic structure for the target document.

The experiments of this research show that the proposed coherence models achieve substantial improvements over both the baseline and models that are built on either document topics or sentence topics obtained under the assumption of direct topic correspondence between the source and target side. Additionally, the target translations

generated by these models are more coherent and similar to reference translations than those generated by the baseline.

Modelling lexical cohesion for document-level machine translation (Xiong, D. et al. 2013) proposes three different models to capture lexical cohesion for document-level machine translation. They integrate the three models into hierarchical phrase-based machine translation and evaluate their effectiveness on the NIST Chinese-English translation tasks with large-scale training data. The Experiments show that all three models can achieve substantial improvements over the baseline and that the mutual information trigger model performs better than the others.

The Review of Discourse-Based Machine Translation Evaluation (Zhang, Y. 2018)[7]consists of making the difference between the two methods of the evaluation metrics of machine translation: the one based on discourse structure and the other on discourse features. It shows the advantages and the advantages of each category.

Sim smith, K. (2017) proposed ways to automatically assess the coherence of machine translation output. He evaluates existing monolingual coherence models on this new task, identifying issues and challenges that are specific to the machine translation setting. The researcher also proposed a new coherence model through exploring the cross lingual transfer of discourse relations in machine translation and measuring the correctness of the discourse relation in comparison with to the source text rather than to a reference translation.

The research shows how the new and adapted model correlates with human judgements of translation quality. It also suggests that improvements in general evaluation within machine translation would benefit from having a coherence component that evaluated the translation output with respect to the source text.

3. Global Coherence

Any text is composed of clearly identifiable units that are linked together and articulate with one another. For Jean-Michel Adam, a text is made up of five types of links: connections (connectors and textual organizers), implications (ellipses, presuppositions, and implicit), links of the signifier (repetition of phonemes, syllables, lexemes, morphosyntactic groups), and links of the signified (anaphors).

For Adam, the text exists if these types of link are actually updated: "each of these operations is a factor of textuality, but none is sufficient to make a text a coherent unity". (Adam, 2005: 85)

It is also necessary that these operations should be organized in identifiable configurations: textual sequences. They are more or less typed and have as characteristics the correspondence to categories of macro-semantic relations memorized by socio-cultural impregnation.

The text is; therefore, a verbal material constructed within the framework of a language formation referring to a genre whose internal organization, based on discrete and localizable units, denotes its coherence.

In fact, coherence manifests itself at the global level of the text; it concerns its general meaning. For a text, to fulfil the conditions of textual coherence, it must obey four rules: a progression of information, a close relationship between passages and ideas, a lexical field and a non-contradiction.

> Pour qu'un texte soit cohérent, « il faut qu'il comporte dans son développement linéaire des éléments à récurrence stricte » (métarègle de répétition) et « que son développement s'accompagne d'un apport sémantique constamment renouvelé » (métarègle de progression). Par ailleurs, pour qu'un texte soit cohérent, « il faut que les faits qu'il dénote dans le monde représenté soient reliés » (métarègle de relation), ce qui signifie qu'il doit exister des relations qui permettent de lier les éventualités décrites et de structurer le discours : ces relations sont appelées relations rhétoriques ou discursives. (Charolles, 1978 : 42)

3.1. Progression of Information

All texts must follow the principle that corresponds to the rule of progression of information. It is important that a text presents new information to have a communicative interest. The interest of the text will be weak, if it consists in repeating in different ways the same information one will say that it tramples.

The progression of information assumes that each new sentence must bring new information that is logically related to the previous information. There are different ways to advance information as using textual organizers by following the rules of the paragraph as well as the sequences of the explanation.

3.2.Relationship Between Passages

In a very coherent text, the passage from one idea to another must be clear. This logical passage is important so that the reader does not have the impression that the author passes without transition or reason from one subject to another. In general, this passage is between paragraphs. It can be found in some writings that the end of each paragraph announces the opening or the beginning of the next paragraph that should bring a new or a complementary idea to the main one.

3.3.Non-contradiction

It is important to avoid any contradiction between sentences and paragraphs, because non-contradiction ensures the credibility of the text through avoiding opposing information, said or implied. There are, in fact, two types of contradiction:

> The enunciative contradiction: Avoid abrupt changes, such as change of time (from simple past to past tense), or change of person (from pronoun "he" to pronoun "I")
> The contradiction in referential plan: This contradiction is relative to the coherence of the referential plan. For example, the main point in the text or in the paragraph revolves around the Internet, and the author begins to talk about sport. This change of subject, also called an inappropriate digression, can hinder the textual coherence. (Alkhatib, 2012 : 54)

4. Local Coherence (Cohesion)

Le mot cohésion désigne […] l'ensemble des moyens linguistiques qui assurent les liens intra- et inters phrastiques permettant à un énoncé oral ou écrit d'apparaître comme un texte. (Charaudeau& Maingueneau, 2002)

There are three main components of cohesion: temporal and spatial connectors, anaphora and lexical field.

4.1.Connectors

Connectors are phrases, groups of words or words that indicate the organization of a text. They announce a new passage, summarize, mark a transition, and conclude ... They are often placed at the beginning or at the end of a paragraph. They can indicate that, in the same textual sequence, one changes place, time, aspect treated, argument, etc. They explicitly emphasize that we change the subject (concerning ..., as for ...) and they indicate the end of the passage (finally, in conclusion ...). The grammarian S. Chartrand calls the connectors "text organizers". In this sense, they play a discursive role different from the relationship markers; they intervene on the passages that are presented as coherent units.

4.2. Anaphora

Ensuring coherence of a text is, among other things, checking whether certain elements of meaning are common from one sentence to another. Thus, the resumption of information between sentences certifies that there is a link between them. From one sentence to another, the information is taken over by substitutes whose pronouns constitute an important category.

4.3. Lexical Field

The term *'lexical field'* refers to the vocabulary that constitutes a text. This vocabulary must correspond to the type of writing and the subject treated. In other words, in literary texts, it is necessary to use words of high language (literary language supported); in the scientific text, it is necessary to resort to specialized scientific words, etc. Improper use of vocabulary can affect the coherence of texts and therefore their comprehension.

5. Machine Translation

Machine translation (MT) refers to the computerized systems that can produce translations with or without human assistance. The challenge in MT is how to program a computer to understand a text as a man does and also to create a new text in a target language as it would be written by a human. (Thi-Ngoc-Diep DO. 2012.: 12)

We distinguish TA from computer assisted translation TAO (machine aided human translation) where the goal is to help a human to perform a translation task using online electronic dictionaries, databases terminology, translation memories, etc.

هناك طريقتان معروفتان في الترجمة الآلية. فالأولى، أي برامج الترجمة بمساعدة الحاسوب Computer Aided Translation ما هي إلا امتداد لقدرات المترجم. أي أن المترجم يقوم بالترجمة ويستعين بهذه البرامج لزيادة سرعة الإنتاج وتحسينه. أما الثانية، أي برامج الترجمة الآلية Machine Translation، فهي برامج قائمة بذاتها، لا يكون للمترجم أي دور في الترجمة سوى أن يلقم البرنامج بالنص الأصلي ليقوم البرنامج بترجمته بصورة كاملة. عندئذ يقوم المترجم بتحرير الترجمة وتدقيقها. وبالنسبة لهذه البرامج تكون الترجمة ركيكة ويكون من الأفضل إعادة الترجمة بدلا من تدقيقها على الأقل بالنسبة للترجمة من وإلى اللغة العربية. (كبير زهيرة، 2018: 13-14)

In general, the process of MT of a text consists of three fundamental steps: 1) **analysis**: to analyze a source text in intermediate representations in a source language, 2) **transfer**: to

transfer these intermediate representations to intermediate representations in a target language, and 3) **generation**: to generate a new text in a target language from intermediate representations in this language.

> **Machine-aided translation** (*also* automatic translation, computer translation, machine translation)**:** Transmission of a natural-language text into an equivalent text of another natural language with the aid of a computer program. Such programs have (with varying specializations and success) lexical, grammatical, and, in part, encyclopaedic knowledge bases. Machine-aided translation consists of three components: **(a)** analysis of the source language by means of **parsing; (b)** transfer: the transmission of information from the source language into the target language; **(c)** synthesis: the generation of the target language. (Hadumod, B. 1996: 172)

Machine translation is composed of two essential parts: software automatic translation and automatic translation on line.

5.1.Software Automatic Translation

Software automatic translation (Machine translation or offline translation) is a process based on translation software installed on a computer. This software is increasingly sophisticated; it offers opportunities appreciated by users as academics, business professionals, students, programmers, web designers, etc. Translate Pro, Systran Pro, Power Translator Pro, Babylon are an example of software automatic translation.

Systran operates a hybrid translation engine that integrates statistical analysis with the traditional semantic-syntactic analysis of the source text. This approach allows the software to choose the most frequent solution between two propositions of the semantic-syntactic engine. In addition, it integrates a continuous improvement module.

This hybrid engine allows Systran to position itself as the market leader in our days. Previously, the method used by the software was based on a semantic-syntactic analysis system. The engine analyzed each source sentence and created the syntactic tree to represent its components and the relationships that unite them. Then, each expression was translated using a dictionary. Once the tree was fully translated, the software returned the target sentence. The dictionary, then, constitutes a central element: the more complete it is, the better is the result. Yet, even with highly supplied dictionaries, it is almost impossible to produce a completely correct target sentence as far as the dictionary, which is a collection of lexical data, will find it difficult to account for contextualized or new words and expressions. (Kouassi, 2009: 7)

5.2 *Automatic Translation*

Automatic translation is a service of translation of texts on the Internet. It works in the same way as offline software translation but it requires an internet connection. In fact, the software is either not installed on the computer or installed very minimally (through, for example, the gadgets developed by Microsoft Windows Vista). In recent years, the web has seen a flourishing of tools allowing users to translate instantly texts when they do researches. The most used online translators in our days are: Systran Net, Google

Translate, Promt, Reverse, Yahoo Babel Fish, Babylon, Bing Translator. ((Kouassi, 2009: 7)

Google translate adopts a new method:

> On the one hand, we introduce into the computer billions of words coming from monolingual texts in the target language; on the other hand, we add texts that parallel the two languages. These are created from samples of translations made by professional translators. Then, we apply statistical learning techniques to create a translation model. We have had excellent results in this domain. (http://www. google.fr/intl/fr/ help/ faq_translation.html#statmt)

In response to the increasing inadequacies of the rule method due to the complexity of natural language, some laboratories and research groups move towards statistical methods. With computers, it becomes more and more powerful; it is now possible to tap into the vast corpus of computerized databases to reuse fragments of sentences already translated by professional translators. This is the birth of the statistical translation method.

This method reached its peak in the 2000s with Google which retrieves all the translations existing on the Internet to build the architecture of its translation tool '*Google Translate'*. This statistical approach 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. A statistical analysis uses the redundancies existing in this corpus to estimate the parameters of the translation process.

6. Issues of incoherence in MT systems

Measuring coherence in MT is important because the translations that are generated by standard MT systems (a sentence-by-sentence basis) can lead to incoherent document translations because these systems apply a syntactic- semantic analysis which is not enough to ensure coherence that covers the whole text in addition to the context that surrounds it.

Example one:

Text	Systran	Google Translation
L'ONU s'efforçait hier d'obtenir en dernière minute le soutien des Grecs et des Turcs pour son plan de réunification de Chypre. Le projet 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)	UNO endeavoured yesterday at the last minute to obtain 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	The United Nations was trying yesterday to obtain last 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.

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

Sys. : ...forhis 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 Annanau 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.

Src. : Droite comme un i …

Example two:

Text	Systran	Google translate
Droite comme un i dans sa court e robe de velours noir, Marie-Laure a peur. C'est pour la premi ère fois qu'elle monte sur scène. Elle joue l'ouvreuse dans la pièce *Chantier interdit a u public*. (À côté d'elle) Il y a un homme au canotier et au nœud papillon et un autre vêtu de blanc, à la mourstache remontante et un autre, courbé sur sa canne, au complet de velours 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 noeud butterfly and another dressed in white, with th e 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]

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

Gt.: Straight as an i ...

Ref.:

- 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 côtéd'elle…

Sys.: At side of it....

Example three:

Text	Systran	Google translate
Bertrand Delanoë est revenu lun di au Conseil de Paris, 43 jours a près la grave blessure dont il a ét é victime, en présidant, longtem ps et activement, la séance plénière de cette assemblée municipale où la politique a repr is sa place comme avant."C'est u n Conseil de Paris comme les au tre.Tout de suite, l'ambiance a été à la contradiction. ça ne m'a pas s urpris et je ne demande rien d'autre", a commenté le maire PS. Ouverte sur un accueil chaleureux des élus parisiens qui l'ont applaudi, la session s'est to ut de suite poursuivie par un incide nt de séance et a été marquée p ar une grande pugnacité de l'opposition de droite sur le dossier des "frais de bouche" du couple Chirac, a u lendemain du congrès fondate ur de l'Union pour un mouvement populaire (UMP). Le consensus a été de mise, mais brièvement, quand Bertrand Delanoë, apparemment ravi et e n très bonne forme, a fait son entrée dans l'hémicycle : poignées de mains aux élus et fo nctionnaires, brève allocution po ur dire son "grand plaisir" à se re trouver parmi eux. Resté éloigné de	Bertrand Delanoe was allocated Monday to the Council of Paris, 43 days after t he serious wound of which it was victim, while cha iring, a long time and actively, the plenary session of this municipal assembly where the policy took again its place like f ront. "It is the Council of Paris like the others. Immediately, environment was with contradiction. That doe s not have me surprised and I do not ask anything else", the mayor PS commented on. Opened on warm welcome of the Parisian elected officials who applauded it, the session continued immediately with an incident of meeting and was marked by a great pugnacity of the right-wing opposition on the file of the"ex penses of mouth" of the couple Chirac, the shortly after the congress foun der of the Union for a popular movement(UMP). The consensus was of setting, but briefly, when Bertrand Delanoë, apparently charmed and in the very fine shape, came out in the	Bertrand Delanoe was allocated Monday to the Council of Paris, 43 days after t he serious wound of which it was victim, while chai ring, a long time and actively, the plenary session of this municipal assembly where the policy took again its place like front. "It's the Council of Paris like the others. Immediatel y, environment was with contradiction. That does not have me surprised and I do not ask anything else", the mayor PS commented on. Opened on warm welcome of the Parisian elected officials who applauded it, the session continued immediately with an incident of meeting and was marked by a great pugnacity of the right-wing opposition on the file of the «expenses of mouth" of the cou ple Chirac, the shortly after the congress foun der of the Union for a popular movement (UMP). The consensus was of setting, bu t briefly, when Bertrand Delanoë, apparently charmed and in the very fine shape, came out in the hemicycle: handshakes to the e

Gt.: Beside her...

Ref.: beside her/ at her side

*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'*.

l'Hôtel de Ville après l'attentat à l'arme blanche perpétré contre lui le 6 octobre, le maire s'est félicité du "compor tement collectif" de toute la mun icipalité. "La boutique a tourné", s'est-il réjoui. Mais le chef de son opposition, Claude Goasguen, après lui avoir sou haité la bienvenue, a tout de suite laissé la politique "reprendr e ses droits". Il s'en est pris avec vigueur à une "dérive des travau x du Conseil" où se répandent "d es calomnies, des suspicions". Il a demandé une suspension de sé ance pour "rappeler" au conseil s a vocation : "S'occuper de la vie quotidienne des Parisiens et non se transform er en tribunal révolutionnaire du passé".	hemicycle: handshakes to the el ected officials and civils servant, short short speech to say its "great pleasure" to find itself among them. Remained far away from the Town hall after the attack with the knife perpetrated against him O ctober 6th, the mayor was pleased with the "collective behavior" of all the municipality. "Theshop turned", it was delighted. But the chief of its opposition, C laude Goasguen, after him to have welcomed, immediately let the policy "take again his rights". He was caught some with strength with a "drift of work of theCouncil" wh ere spread "calumnies, suspicion s". He asked an adjournment "to recall" to the council his vocation: "To deal with the daily life of the Parisians and not to transform themselves into revolutionary tribunal of the past".	lected officials and civils servant, short short speech to say its "great pleasure" to find itself among them. Remained far away from the Town hall after the attack with the knife perpetrate d against him on October 6th, the mayor was pleased with the "collective behavior" of all the municipality. "The shop turned", it was delighted. But the chief of <u>its</u> opposition, Claude Goasguen, after him to have welcomed, immediately let the policy "take again his rights". He was caught some with strength with a "drift of work of the Council" where spread "calumnies, suspicions". He asked an adjournment "to recall" to the council his vocation: "To deal with the daily life of the Parisians and not to transform themselves into revolutionary tribunal of the pa st".

Src. : … 43 jours après la grave blessure dont <u>il</u> a été victime,
Sys.: ... 43 days after the serious wound of which <u>it</u> was victim,

Gt.: ... 43 days after the serious wound of which <u>it</u> was victim,

Ref. :43 days after the serious wound of which <u>he</u> was victim,

- The problem in this sentence is of **referencing anaphora**; '*il'* is a personal pronoun. It refers to Bertrand Delanoë. 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. : … où la politique a repris sa place <u>comme avant.</u>

Sys.: ... where the policy took again its place like front.

Gt.: ... where the policy took again its place like front.

Ref.: ... where the policy took again its place as before.

- There is no **lexical cohesion** in this sentence, the word *'front'* is not appropriate for this context, *'commeavant'* is a discourse connective expressing time (the past), the correct equivalent here is *'before'*.

Src. : Ça ne m'a pas surpris …

Sys.: That does not have me surprised...

Gt.: That does not have me surprised ...

Ref.: That has not surprised me...

- There is a problem of **sentence word order** or **syntax structure**, 'does not have me surprised' is not correct grammatically, because languages differ from each other in many things, the placement of words according to the role they play within a sentence is an example. In this sentence, 'me' (for the object) should be put after the verb not before as in French.

Src. : Ouverte sur un accueil chaleureux des élus parisiens qui l'ont applaudi,

Sys.: Opened on warm welcome of the Parisian elected officials who applauded it, ...

Gt.: Opened on warm welcome of the Parisian elected officials who applauded it, ...

Ref.: Opened on warm welcome of the Parisian elected officials who applauded him, ...

- The problem in this sentence is of **referencing anaphora**; '*l'* is a personal pronoun replacing the object. It refers to 'le maire'. 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. To be correctly expressed we say 'who applauded him'.

Src. : … pour dire son "grand plaisir" à se retrouver parmi eux.

Sys.: ... to say its "great pleasure" to find itself among them.

Gt.: to say its "great pleasure" to find itself among them.

Ref.: ... to say his "great pleasure" to find himself among them.

- The problem in this sentence is of **referencing anaphora**; '*son'* is a possessive adjective pronoun. It refers to '*Delanoë'*. 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. To be correctly expressed we say *'his "great pleasure'*.

Src. : Mais le chef de **son** opposition…

Sys.: But the chief of **its** opposition...

Gt.: But the chief of **its** opposition...

Ref.: But the chief of **his** opposition...

- The problem in this sentence is also of **referencing anaphora**; '*son'* is a possessive adjective pronoun. It refers to '*Delanoë'*. 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. To be correctly expressed we say *'his "great pleasure'*.

Src. : après lui avoir souhaité la bienvenue…

Sys.: ...after him to have welcomed, ...

Gt.: ... after him to have welcomed, ...

Ref.: after welcoming him/ after having welcomed him...

- There is a problem of **word order**, or **syntax structure**; the pronoun referring to the object is put before the verb in French '*lui avoir souhaité'* but this is not the case in English language saying *'him to have welcomed'* makes the sentence ungrammatical and the text incoherent, the correct form is to put the pronoun after the verb. Here the natural logic of the sentence is distorted, with the subject coming after the verb, directly affecting the coherence.

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

- **Lexical cohesion**

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 and 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. (Morris& Hirstt, 1991)

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.

MT has been shown to be consistent in its use of terminology (Carpuat& Simard, 2012), 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 different word in another context.

Referencing Anaphora resolution is a challenging issue in current MT approaches (Novak, 2011) 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 has only 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 (Potet et al. 2012). 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) ill-formed utterance.

Consequently, we believe that a coherent discourse should have a context and a focus, be characterised by appropriate coherence relations, and structured in a logical manner.

7. Conclusion

Coherence is undeniably a cognitive process guided by linguistic elements discernible in the discourse that does include cohesion. it describes how a text becomes semantically meaningful overall. However, it has received little attention in Machine Translation (MT) because most decoders work on a sentence-by-sentence basis isolated from context due to both modelling and computational complexity. Moreover, Current MT approaches suffer from a lack of linguistic information at various stages (modelling, decoding, pruning) causing the lack of coherence in the output.

In other words, in machine translation, it exists the analysis step in which the machine takes into account the sentence (the largest grammatical unit in syntax) to analyse without referencing to the preceding and the following sentences; the fact that makes the sentence in isolation without context (pragmatic coherence) that we need to interpret the meaning of the word within the sentence and within the text as a whole. Moreover, the syntactic semantic analysis of the source text neglects the theme and the rheme that guarantee the sequence of ideas (the thematic coherence).

Therefore, to improve the performance of translation systems, ensure textual coherence and give satisfactory translations, we should add to the phase ‘semantic-syntactic analyses. the thematic analysis can be done through segmenting the text into linguistic elements that should have a clear relationship to each other (cohesive segments) and into textual units that refer to the same theme; and a pragmatic analysis that concerns the aspects related to the context of enunciation (extra-linguistic reality).

References

[1] Adam, J.-M. (2005). *La linguistique textuelle, Introduction à l'analyse des discours*. Paris: Armand Colin.

[2] Alkhatib, M. (2012). La cohérence et la cohésion textuelles : problème linguistique ou pédagogique ?'Didáctica. *Lengua y Literatura. vol. 24.* P : 50-54.

[3] Bussmann. H. (1996). Routledge dictionary of language and linguistics. Translated and edited by Gregory Trauth and Kerstin Kazzazi. London and New York: Routledge.

[4] Carpuat. M. Simard. (2012). The trouble with smt consistency. In *Proceedings of WMT. 442–449. Montreal,* Canada.

[5] Charaudeau, P., Maingueneau, D. (2002). *Dictionnaire d'analyse de discours*. Paris: Seuil.

[6] Charolles, M. (1978). Introduction aux problèmes de la cohérence des textes. In: *Langue française*.

[7] DeyiXiong, Guosheng Ben, Min Zhang, YajuanLuandQun Liu. (2013). Modeling lexical cohesion for document-level machine translation. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. 2183-2189.

[8] DeyiXiong, Min Zhang, Member IEEE, and Xing Wang (2015). Topic-based coherence modeling for statistical machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.23, N°3.*

[9] DO. Thi-Ngoc-Diep. (2012). *Extraction de corpus parallèle pour la traduction automatique*. Universite de Grenoble.

[10] Halliday, M.A.K., Hasan, R. (1976). *Cohesion in English.* London: Longman.

[11] Kouassi R.R. (2009). *La problématique de la traduction automatique.* Abidjan: Université de Cocody.

[12] Maingueneau, D. (1996). *Les termes clés de l'analyse du discours.* Paris: Seuil.

[13] Marion, P. et al. (2012). Collection of a large database of French-English SMT Output Corrections.

[14] Morris, J., Hirstt. G. (1991). *Lexical cohesion computed by the causal relations as an indicator of the structure of text.* Association for Computational Linguistics.

[15] Novak. M. (2011). Utilization of anaphora in machine translation. In WDS Week of Doctoral Students.

[16] Sim Smith, K. (2017). Coherence in machine translation. University of Sheffield.

[17] Sim Smith, K., Aziz, W., and Specia, L. (2015). A proposal for a coherence corpus in machine translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation,* pages 52-58, Lisbon, P*ortugal. Association for Computational Linguistics.*

[18] Sim Smith, K., Aziz, W., Specia, L. (2016). The Trouble with machine translation coherence. *Baltic J. Modern Computing, Vol.4, N°2,* pp. 178–189.

[19] Sim Smith, K. and Specia, L. (2017). Examining lexical coherence in a multilingual setting. Translation and Multilingual Natural Language Processing. Language Science Press, Berlin.

[20] SolmazS. (2014). Syntactic and lexical simplification: the impact on efl listening comprehension at low and high language proficiency levels. *Journal of Language Teaching and Research, Vol. 5, No3,* pp. 566-571.

[21] Swapna, Somasundaran, Jill, Burstein. Martin, Chodorow, (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.

[22] Wong, B. T.-M. and Kit, C. (2012). Extending machine translation evaluation metrics with lexical cohesion to document level'. In *Proceedings of EMNLPCoNLL*, pages 1060-1068

[23] Yunzhen Zhang, En Guo, Hui Yu, WeizhiXu. (2018). Thereview of discourse-based machine translation evaluation. *Advances in Intelligent Systems Research, vol 147*. International Conference on Network, Communication, Computer Engineering.

[24] Bebir, Z. (2018). Al-tarjama Al-'Āliyya Al-Waqi' wa l Āfāq. *Revue Traduction et Langues 17 (2)*.