Cohesion in Human and Machine Translation

Mohamed Tohamy
Faculty of Arts & Humanities,
Suez Canal University

Abstract

Machine translation technology has consistently improved to the extent that it is frequently compared with human translation performance. This study reports a simple comparison of two translations of an Arabic literary text: a short story written by the Egyptian writer Yousef Idris. One translation is human made and the other is a machine translation using Google Translate. The study focuses on the use of some cohesive devices in the two translated texts. Two types of grammatical cohesion were selected for analysis, namely, reference and conjunction. The findings of the study show that cohesive properties vary according to the translation method involved, wither it is human or machine translation. The results obtained are indicative of the limitations of machine translation to produce a reliable translation of a literary work. Unlike the human translator whose use of cohesive devices is balanced and fairly distributed to the text, Google Translate uses such devices excessively, which makes the machine translated text incoherent, loose and ambiguous.

Key words: Machine translation, human translation, Google Translate, Cohesive devices

التماسك النصي في الترجمة البشرية والترجمة الآلية

المستخلص

ادي التحسن المستمر في تكنولوجيا الترجمة الآلية الى عقد المقارانات بين الترجمة الآلية والإداء البشري في الترجمة . تقوم هذه الدراسة على مقارنة بين ترجمتين لنص أدبي عربي وهو قصة قصيرة للكاتب المصري يوسف إدريس. إحداى الترجمات المستخدمة في المقارنة هي ترجمة بشرية والأخرى ترجمة آلية باستخدام ترجمة جوجل. تركز الدراسة على استخدام بعض ادوات التماسك في النصين المترجمين. وتوصلت نتائج الدراسة إلى القدرة المحدودة للترجمة الآلية للحصول على ترجمة موثوقة للعمل الأدبي. وتأتى هذه النتيجة مغايرة لما توصلت اليه الدراسة بخصوص الترجمة البشرية من حيث قدرة المترجم البشرى على استخدام أدوات التماسك بشكل متوازن مما يجعل النص متماسكا في كل اجزاؤه. توصلت الدراسة ايضا الى أن ترجمة جوجل تستخدم هذه الأدوات بشكل مفرط، مما يجعل النص المترجم غير متماسك وغير مترابط بل وغامض في كثير من الأحيان.

الكلمات المفتاحية: الترجمة الآلية، الترجمة البشرية، ترجمة جوجل، أدوات التماسك

1. Introduction

Using computers to translate natural languages with or without human intervention has become established since the idea was first proposed in the 1940s and started to be seriously investigated in the 1950s (Hutchins, 2001). Computer-based translation systems are now included in categorizations of translation modes according to the degree of human or machine involvement in the process of translation (e.g. Hutchins, 1986; Precup-Stiegelbauer & Laura-Rebeca, 2013). For instance, Hutchins and Somers (1992, p. 150) suggest a three-point categorization of translation modes. Totally human translation occupies one end of this categorization where translation is paper-and-pen based. At the other end comes the fully automatic or machine translation (MT) where no human intervention is expected. The middle position is occupied by the machine-aided translation (MAT) in which the output has to be submitted to revision and editing by a human translator.

Attitudes towards the quality of MT have been diverse. As early as 1951, it was regarded as "atrocious and fullers of howlers" (Holmström, 1951). In 1966, the US National Academy of Sciences charged the Automatic Language Processing Advisory Committee (ALPC) with evaluating the progress of MT research. The ALPC reported that automatic systems failed to provide good quality translations. The situation has not greatly changed ever since and it is now believed that MT output must be revised and edited by a human translator, if we want to achieve "a publishable quality" of this output (Hutchins, 2001, p.8). Kadiu (2019) argues that despite the advances in in translation software, MT still requires "human decision-making" (p. 92).

This study investigates variations in frequencies and distribution of cohesive devices in human and machine translation of Yousef Idris's short story [لغة الآي اي], translated as "*The Language of Pain*". Cohesion in this study is measured, according to a frame suggested by Lapshinova-Koltunski (2017), on the basis of three criteria: text language (original versus translated), translation method (human versus machine), and cohesion contrasts between the two languages according to Halliday and Hasan (1976). The model adopted for the study is limited to the analysis of two types of grammatical cohesive devices, namely reference and conjunction.

2. Cohesion

Halliday and Hasan (1976, p. 7) define cohesion as "non-structural text-forming relations of meaning within a text". Such relations are intended by the writer to guide the reader to grasp meaning relations of different parts of a text. Cohesion enhances the realization of continuity

and semantic unity of a text. Cohesive relations are distinct from other types of semantic relations in that the lexico-grammatical cohesive resources transcend clause boundaries and apply to the text as a whole (ibid, p. 297). Halliday and Hasan's model comprises two broad categories of cohesion: grammatical and lexical. Halliday (1973) and Halliday and Hasan (1976) propose an inventory of grammatical cohesive resources including reference, ellipsis, substitution, conjunction, in addition to lexical cohesion which comprises lexical devices such as nouns, adjectives and verbs. This study concentrates on two major types of grammatical cohesive devices: reference and conjunction, which are considered by Halliday and Hasan (1976) and Halliday and Matthiessen (2014) as explicit linguistic devices that identify and relate conceptual relations to other linguistic elements in other parts of the same text.

2.1. Reference

Reference is referring to an entity whose identity can be recovered from the context. Halliday and Hassan ((1976) stated that

> …reference is the specific nature of information that is signalled for retrieval. …[T]he information to be retrieved is the referential meaning, the identity of the particular thing or class of things that is being referred to; and the cohesion lies in the continuity of reference, whereby the same thing enters into a discourse a second time. (p. 31)

Reference can be either endophoric (textual) which creates cohesive ties inside the text itself or exophoric (situational) "which directs the receiver out of the text and into an assumed shared world" (McCarthy, 1991, p.41). Endophoric referencing is considered the core of cohesion theory (Crane, 2008, p. 133). Moreover, it can be either anaphoric (referring back to something said previously) or cataphoric (referring forward to a following element). Reference resources include personals, demonstratives, and comparatives (Halliday and Hassan, 1976, p. 33). Personal reference is tracking speech situations using pronouns such as *he*, *him*, *she*, *her* and possessive determiners such as *my*, *your*, *his, her* and possessive pronouns such as *mine*, *yours*, *his*, and *hers*.

Demonstrative reference is examined "by means of location, on a scale of proximity" (ibid, p. 37) using the adverbial demonstratives *here*, *there*, *now* and *then* to refer to the location in terms of space and time, and the nominal demonstratives *this*, *that*, *these*, *those* to refer to the location of a person or an object.

Comparative reference is reference of identity and similarity which are recognized through adjectives (like *same, equal, similar, different)* and adverbs (like *so, such, similarly, otherwise)*. Comparative reference is further sub-categorized into general comparison and particular comparison. Whereas the former is recognized in terms of "likeness and unlikeness without respect to any particular property" the latter is identified "in respect of quantity and quality" (p. 77). This is illustrated in the following examples:

Example 1:

General comparison: likeness and unlikeness

1.a. Most people have the *same* breakfast everyday.

1.b. The candidates gave three *similar* answers.

1.c. It's a *different* cat from the one we saw yesterday.

(Halliday & Hasan, 1976, pp. 78-90)

Example 2:

Particular comparison: quantity and quality

2.a. Take some more tea.

2.b. We are demanding higher living standard.

(Halliday & Hasan, 1976, p. 81)

2.2. Conjunction

The other grammatical cohesion device that is adopted in this study is conjunction. It is the use of conjunctives to connect parts of a text together. Semantically, conjunction is realized as "a specification of the way in which what is to follow is systematically connected to what has gone before" (Halliday & Hasan, 9176, p. 227). Halliday & Hasan Suggest an inventory of four types of conjunctions: additive (typified by the conjunct '*and'*), adversative (typified by the conjunct '*yet'*), causal (typified by the conjunct '*so'*) and temporal (typified by the conjunct '*then'*).

Two main grammatical categories of conjunctive expressions contribute to the cohesive structure of a text: adverbs and prepositions (Halliday & Hasan, 1976, p. 242-243). Adverbs are either simple such as *but, so, then* and *next*, or compound such as *-ly* adverbs (e.g. *consequently*), or adverbs in *there-* and *where-* such as *therefore* and *whereat*. Other compound adverbs include such words as *furthermore, nevertheless, anyway, instead* and *besides*. Prepositions functioning as conjunctives can be heads of prepositional phrases as in *on the contrary* and *in addition*, or prepositional expressions with *that* as *in instead of (that), as a result of (that)*, and *because of (that)*.

3. Cohesion in Arabic

Since this study investigates cohesion in human and machine translation from Arabic into English, it is appropriate to review cohesive

devices in Arabic. Although cohesive systems in English and Arabic are different, Halliday and Hasan's (1976) classification of cohesive devices is appropriate for the current study since their model of cohesion is mainly concerned with the unity of the text rather than sentence-specific features (De Beaugrande & Dressler, 1981).

A few studies have adopted Halliday and Hasan's model to investigate cohesion in Arabic texts (e.g. Chaalal, 2017; Williams, 1989; Al-Jabr, 1987; and Koch, 1981). These studies are based on a classification of cohesive devices in Arabic including reference, substitution, ellipsis, conjunction and lexical cohesion. As this study is concerned with reference and conjunction, this section is a review of these two devices.

3.1. Reference in Arabic

Beeston (1970) and Al-jabr (1987) present a classification of reference in Arabic including personals, demonstratives, and comparatives. Personal reference is realized in Arabic by explicit pronouns (الضمائر البارزة), which are visible in the clause, and implicit pronouns (الضمائر المستترة), which have no visible form but understood from context. Explicit pronouns are further sub-classified into independent (منفصلة) or enclitic (متصلة). Independent pronouns like أنا، هو, هي, انت, نحن (lit. *I, he, she, you, we*) are used to emphasize the role played by the referent in speech situation (Al-Shurafa, 1994, p.19). Enclitic pronouns never occur as detached components; they have to be attached to other words in the sentence. The examples in (3) below illustrate the two types of explicit pronouns (Arabic examples and their translations by Nagib (1990) are from my data):

Example 3:

3.a.

هل أنت حي أم ميت؟ فهمي رغم كل شيء حي أما أنا فلم أحي.

Are you alive or dead? In spite of everything Fahmy was alive, and he had lived. But I? No. I have not lived.

3.b

الأكل عندهم أن يحل موعد الطعام ويلتفون حوله في ترحيب ويتعازمون ويهزرون ويحسون أنهم يقومون باحتفال إنساني صغير.

Eating was the coming, the arrival of meal time … the eager gathering around the food inviting one another to eat, the joking and above all, the feeling that they were having a small celebration.

Unlike the enclitic pronouns in 3.b (ـهم, ـه, ـون) which are attached to nouns and verbs (عندهم – حوله – يتعازمون) , the independent pronouns in

3.a. (أنا , أنت) are detached from other words in the clause as this category of pronouns can stand alone as separate words in Arabic discourse.

Arabic personal pronouns are different from the English ones in that the former category reflects number and gender. Unlike English which has either singular or plural pronouns, Arabic has a special set of pronouns which are used to refer to two entities (dual). In example (4), the translator uses the coordinate structure "both husband and wife" as a translation of the Arabic "dual alif" (ألف الإثنين) in (أصبحا):

Example 4:

أصبحا بشيء من التحدي ينتظران الصرخة الثالثة.

Both husband and wife waited, almost defiantly, for a third scream.

One more difference between Arabic and English personal pronouns is that the singular non-human English pronoun "it" has no Arabic Equivalent (so it is translated as هو "he" or هي "she"). Gender differences are explicitly shown in the كتب/كتبت (he wrote/ she wrote) example where the feminine "t" (تاء التأنيث) indicates that the referent is a female entity.

Demonstrative reference in Arabic is similar to that of English in that both are classified in terms of proximity: *near* versus *far*. Still, Arabic demonstratives are inflected for number (i.e. singular, dual, and plural) and gender (masculine and feminine). Table 1 below contains a comparison of demonstratives in Arabic and English:

Table 1. Demonstratives in Arabic and English

	Near		Far	
	Arabic	English	Arabic	English
Singular	Masc. هذا	This	Masc. ذلك	That
	Fem. هذه		Fem. تلك	
Dual	Masc. هذان	Null	Masc. ذنك	Null
	Fem. هاتان		Fem. تانك	
Plural	Masc. هؤلاء	These	Masc. أولئك	Those
	Fem. هؤلاء		Fem. أولئك	

As for comparative reference, Al-jabr (1987, p. 80) and Chaalal (2017, p. 195) confirm that particular comparison is the only form of comparison that is found in Arabic. This type of comparison can be achieved by using a form that is "derived from a dynamic verb" (Al-jabr, p.80) and takes the form of "أ-------- من" , which can be illustrated in

examples such as "أكبر من" and "افضل من" (bigger than, and better than, respectively).

3.2. Conjunction in Arabic

Like English, conjunctions are used in Arabic to link together text parts. Conjunctions in Arabic are referred to as " أحرف العطف " or conjunctive particles. Yaquob (1971) and Abdullatif (1982) state that there are eight conjunctive particles in Arabic, expressing addition, sequence, grading, purpose, alternative, equation, negation and contrast. Table 2 shows that Arabic discourse uses the eight conjunction particles recognized by Arab grammarians in addition to other expressions which render equivalent meanings. Although this classification of conjunctions does not correspond to Halliday and Hasan's taxonomy, it is adequate for the current study as it covers the major categories of conjunctions recognized by Arab Grammarians.

Table 2: Arabic conjunction particles

Category	Conjunctive particle	Other expressions
Addition	و	علاوة على ذلك, ايضا
Sequence	فـ , ثم	من ناحية ثانية
Grading	فـ	بعد ذلك
Purpose	حتى	لهذا السبب, لذلك
Alternative	أو	بدلا من
Equation	أم	في أي الأحوال
Negation	لكن	على العكس
Contrast	بل , لا	في المقابل, ومع ذلك, بالرغم من ذلك

4. Literature Review

Several studies have been concerned with comparing HT and MT from the perspective of translation quality. It has been observed that many previous studies report the limited abilities of MT systems due to "their relative ignorance of linguistic structures" (Doherty, 2016, p. 953). For example, Ahrenberg (2017) suggests a scheme for comparing a human translation and a Google translation of an opinion article, published in Financial Times, translated from English into Swedish. Ahrenberg's study reveals differences between HT and MT in translation procedures. HT is reported to be similar to ST in text length, flow of information, and textual structure. However, MT proves to be unacceptable without human intervention as each sentence requires at least three edits to make it readable.

Christensen and Schjoldager (2016) collect data via a questionnaire to evaluate MT quality form the perspective of translation service providers. They focus on how the use of MT tools has impacted the translation

industry in Denmark. They report that translation software, particularly Computer-Aided Translation tools (CAT), have facilitated and improved the productivity and consistency of translation outputs, but they also report that these tools sometimes decrease the output quality.

Another study that evaluates translation quality is Bojar's (2011) where MT quality is analyzed on the basis of four categories of errors, namely, bad punctuation, missing word, word order and incorrect word. Using these error categories to analyze machine-translated texts, it is found that some errors, particularly lexical ones, are easier to correct than other types of errors. Bojar concludes that a fluent MT "output is not [necessarily] a good translation of the source text" (p. 74). Following the same line of analyzing errors made by MT systems, Abu-Ayyash (2017) investigates errors and non-errors made by three MT systems in rendering gender-bound structures from English into Arabic. The MT systems he used are Systran's Pure Neural Machine Translation, Google Translate, and Microsoft Bing. Abu-Ayyash's findings are in line with Doherty (2016), Christensen and Schjoldager (2016) and Costa, Ling, Luis, Correia & Coheur (2015) that there are problematic structures that MT systems cannot identify. Among these are pronoun-antecedent agreement and adjective-noun agreement.

Omar and Gomaa (2020) evaluate the versatility of two MT systems in the translation of literary texts to discover how a literary input affects the reliability of MT systems. They use two online translation systems, namely, Google Translate and Q Translate, to get Arabic translations of two English prose fictions. Comparing the MTs they get to a human made translation, they find that MTs contain lexical, structural and pragmatic errors, which negatively impact the reliability of MT systems in translating literature.

Trying to incorporate cohesive features into MT evaluation, few studies have demonstrated the differences between HT and MT in terms of cohesion. For example, Lapshinova-Koltunski (2015 & 2017) compare cohesive features and their distribution in English-to-German HT and MT. She found no significant differences between HT and MT justifying that variation in cohesive features is likely to be influenced by register rather than by translation method. Lapshinova-Koltunski (2017) admits that it is not an easy task to "discover considerable differences" (p.106) between HT and MT in terms of cohesion. However, she adds that the quality of MT cannot be compared to that of HT.

Chaalal (2017) studies the use of cohesive features in English/Arabic corpus based on the United Nations documents. She found that Arabic and English "have more similarities than differences in terms of cohesive devices" (p. iii). The similarities are preserved in the MT as the ST (and legal texts in general) is characterized by accuracy, transparency and formality.

The literature does not adequately cover the area of contrasting the use of cohesive devices in human made and machine translations. Thus, the main objective of the current study is to compare the use of some cohesive devices in human and machine translations of a literary text, namely, Yousef Idris's لغة الآي آي (1977), translated as *The Language of Pain*.

5. Research Questions

This study addresses a couple of questions related to Arabic/English human and machine translations. These questions cover cohesiveness and variation of cohesive links in human and machine translations:

1- To what extent are machine-translated texts cohesive compared to human-translated ones?
2- Is there any variation in the use of cohesive devices in human and machine translations?

6. Data

The data contain an Arabic source text (ST) which is translated twice: once by a human translator and another by an online translation tool. The ST is a well-known short story written by the prominent Egyptian author Yousef Idris. لغة الآي آي(1977) is a story that is contained in a volume of stories carrying the title of the story itself. The volume was translated into English in 1990 under the title *The Language of Pain* by Nawal Nagib and published by the General Egyptian Book Organization as an issue of the Contemporary Arabic Literature Series.

As for the machine translation (MT) version, a Microsoft Word Format of the ST was input into the Google Translation (GT) online tool to be translated from Arabic into English. Although there are many free online translation tools, GT is selected because it is the one that is most widely used worldwide. It has been reported that in 2016, GT translated texts from and into 103 languages, and more than one hundred billion words are translated daily (Turovsky, 2016). It has also been reported by many researchers (e.g. Hadla, Al-kabi and Hailat, (2014); Kadhim et al

(2013); Hampshire and Salvia (2010) that GT ranked higher than other free online translation tools in terms of accuracy, clarity and style.

7. Method

It has been mentioned above that this study concentrates on extracting and analyzing two major types of cohesive devices, namely reference and conjunction. For the Arabic component of the data, the researcher runs morphological segmentation, which is a process for analyzing a word into its constituent tokens/morphemes. For example, the word وسندخلهم is made up of the conjunction و , the future particle س , the verb ندخل and the object pronoun هم. The verb itself is made up of two units: the prefix ن indicating the first personal plural and the verb stem دخل . The whole word translates into the English: and *we shall enter them* and is thus a whole sentence. To perform the segmentation, the ArabicSOS package is used [https://dl.acm.org/doi/10.1145/3322905.3322927]. This is a software in which the authors use the "Gradient Boosting algorithm" to create a morphological segmentation system for Modern Standard Arabic. Mohamed and Sayyed (2019) state that the segmentation accuracy rate of SOS reaches a percentage of 98.47%. For English, the researcher performs Part of Speech tagging, which is assigning grammatical labels to words in sentences. The Spacy package is used for this purpose [https://spacy.io/api/annotation#pos-tagging]. Spacy is a software designed for natural language processing. It is used in this study to obtain part of speech tagging (See Appendix for screenshots of data analysis).

Reference in this study is analyzed in terms personals, demonstratives, and comparatives, following the model suggested by Halliday and Hasan (1976). Table 3 shows the items that are traced in the Arabic and English data.

Table3: Reference items traced in data analysis

Cohesive Device	Item Types	Realization in the Text
Reference	Personals	I – we- you – he – she – it – they my – our - your – his – her – its – their
		انا – نحن – **انتَ – انتِ –انتما – انتم – انتن** – هو – هي – **هم –هما – هن** يـ - نا – **كَ – كِ – كما – كم – كن** – هـ - ها – **هما – هم**
	Demonstratives	this – that – these – those
		هذا – هذه – هذان – هاتان – هؤلاء – ذلك – تلك – اولئك
	Comparatives	--er than
		أ --- من

ISSN 1110-2721

Occasional Papers
Vol. 76: October (2021)

Notes: Due to differences between the pronoun systems in Arabic and English, it is to be considered that انتن – انتم – انتما -انتِ – "انِتَ" are rendered in English as "you", " هن – هما -هم" as "they", " كم – كما – كِ – كَ" كن – as your, and " هم – هما" as their.

Conjunction is represented by four major types of items: additive, adversative, causal, and temporal. Table 4 shows that the Arabic connectives are mapped onto Halliday and Hassan's (1976; pp. 424-3) taxonomy so that the ST and TT can be appropriately analyzed and compared.

Table 4: Arabic and English connectives traced in data analysis

Cohesive Device	Item Types	Realization in the Text
Conjunction	Additive	and, and also, further, furthermore, moreover, also, what is more, besides, additionally, in addition to this, not only that but
		و – وايضا – علاوة على ذلك – ايضا – بالمناسبة – أى أن – بمعنى آخر – على سبيل المثال – فوق ذلك – في المقابل
	Adversative	yet, though, only, but, however, nevertheless, despite this, instead, rather, on the contrary
		لكن – مع ذلك – على اية حال – بالرغم من ذلك – في الواقع – في نفس الوقت – بالأحرى – على العكس – على الأقل – على اية حال – مها يكن الأمر – في أي الأحوال – بأي طريقة كانت
	Causal	so, thus, hence, therefore, consequently, accordingly, as a result, because of that
		لذلك – إذن – بسبب ذلك – لهذا السبب – نتيجة لذلك – آخذا بعين الإعتبار – بسبب – لأن – على هذا الأساس – في مثل هذا الظرف.
	Temporal	then, and then, next, subsequently, after that, presently, later
		عنئذ – ثم – بعد ذلك – في نفس الوقت – سابقا – آنفا – أخيرا – في الحال – في المرة التالية – في مناسبة أخرى – في اليوم التالي – حتى ذلك الحين – في هذه اللحظة – من الأن فصاعدا – بعد قليل.

8. Analysis and Results

The analysis relies on investigating the frequencies and distribution of cohesive devices in the three texts representing the data. Illustrative examples are provided to show how cohesive devices are translated in HT and MT.

8.1. Reference in Human and Machine Translations:

When applying the Hallidayan model of reference to Arabic/English original and translated texts, it has to be taken into consideration the fact that there are morphological and syntactic differences between English and Arabic. That is to say that cohesive devices used in English are different in some respects from Arabic. However, Mehamsadji (1988) argues that Halliday and Hasan's concept of cohesion "can be used to shed light on the different ways used in Arabic to achieve textual cohesion" (p. 26).

8.1.1. Personals

Table 5: Frequencies of personals in ST

Personals																				
Item	**أنا**	**نحن**	**انت**	**انت fem**	**انتما**	**انتم**	**انتن**	**هو**	**هي**	**هما**	**هن**	**يـ**	**نا**	**ك**	**كما**	**كن**	**ها**	**هما**	**هم**	**total**
Freq.	18	0	6	5	0	1	0	40	7	10	0	0	3	0	11	9	208	10	50	369

Table 6: Frequencies of personals in HT and MT

	Item	Human Translation	Machine Translation
Personals	I /me	85	92
	We	3	2
	You	49	54
	He /him	74	210
	She /her	32	51
	It	53	127
	They /them	61	58
	My	10	47
	Our	0	2
	Your	5	6
	His	31	151
	Her	8	16
	Its	17	25
	Their	7	23
Total		435	864

Tables 5 and 6 show that the overall frequencies of the personals category in the two English TTs exceeds their counterpart Arabic ST (Arabic = 369, HT = 435, MT = 864). Unlike English personals which demonstrate little variation regarding gender and, to some extent, number, each Arabic personal pronoun carries information about gender and number (See Table 7).

ISSN 1110-2721

Occasional Papers
Vol. 76: October (2021)

Table 7: Gender and number variation in Arabic personal pronouns

Item in Arabic	English Equivalent
a. masculine singular أنتَ	you
b. feminine singular أنتِ	you
c. masculine/ feminine dual أنتما	you
d. masculine plural أنتم	you
e. feminine plural أنتن	you

This entails that the analysis traces five items in Arabic أنتَ – **أنتِ** – أنتما **(**أنتم - أنتن –**)** and only one equivalent item in English (you). So whereas the frequencies of the personal pronoun *you* in HT and MT are 49 and 54 respectively, the Arabic equivalent items achieve a lower level of frequency (masculine singular = 6, feminine singular = 5, masculine/ feminine dual = 0, masculine plural = 1, and feminine plural = 0). This is due, to a large extent, to the general tendency of English to achieve clarity and explicitness of text by reiteration of pronominal referencing. The results represented in tables (5) and (6) are in line with Guitwinski's (1967, p.80) regarding the fact that reiteration is significant for textual cohesion as it helps the hearer/reader to recall specific items and associate them with others within the same text. Another reason for the higher frequency level of personals in HT and MT than ST is that Arabic tends to rely more on lexical cohesion (i.e. repetition of lexical items such as nouns, verbs, adjectives and adverbials) rather than grammatical cohesion (i.e. reference, and conjunctions) (Betti & AlFartoosy, 2019).

Comparing the frequencies of personals in HT and MT is indicative and sheds light on the question of whether or not MT can produce a cohesive literary text. As Table (6) shows, the overall frequency of personals in MT (f = 864) is almost as double as that of HT (f = 435). This excessive occurrence of personals in MT yields a text which is incoherent and loose. The following examples show how the overuse of personals may yield an inaccurate and incoherent text. In (5b), the insertion of the pronoun *him* makes the translated text confusing as it has no nominal reference, leaving the reader uncertain whether the sentence is a description of the inner feelings of one person trying to suppress his thought (which is the meaning intended by the author), or whether someone wants someone else to stop thinking (the meaning rendered in machine-translated text). In (6b), the excessive use of the pronominals *she*/*her* makes the whole extract ambiguous and loose as the reader is not sure about the reference of these pronominals. Unlike MT which is

characterized by redundant and ambiguous use of personals, HT exhibits a balanced and rational frequencies of such cohesive devices.

Example 5:

ولم يشأ أن يفكر أكثر ..

a) He did not want to pursue the thought further .. (HT)
b) He did not want him to think more.. MT

Example 6:

فتحت الزوجة فمها تصرخ في هوس من تأكد قولها، وانتظرت أن تنتهي الصرخة لتطلق صرختها هي. ولكن انتظارها طال وبدأت رغما عنها تسمع، ومن الذهول استمر فمها مفتوحا وأذناها بأمر قوة قاهرة تصغيان.

a) His wife opened her mouth to scream hysterically at the proof of her conviction and waited for the scream to end so that she could start hers. But she had to wait, and wait, and in spite of herself, she found herself listening, her mouth still open, her ears forced to listen. (HT)
(she/her = 6 occurrences)

b) The wife opened her mouth, screaming obsessively when she said that, and she waited for the screaming to end so that she would scream. But her wait was long and she began to hear in spite of her, and out of amazement her mouth continued open and her ears at the command of force majeure were listening. (MT)
(she/her = 9 occurrences)

8.1.2. Demonstratives

Table 8: Frequencies of demonstratives in ST

Item	اولئك	تلك	ذلك	هؤلاء	هاتان	هذان	هذه	هذا	Total
Freq.	0	0	4	1	0	0	11	23	39
Per.	0 %	0 %	10.3 %	2.5 %	0 %	0 %	28.2 %	59%	100%

Table 9: Frequencies of demonstratives in HT and MT

	Item	Human Translation		Machine Translation	
Demonstratives		Freq.	Per.	Freq.	Per.
	This	34	20.9 %	41	20.5 %
	That	117	71.8 %	156	78 %
	These	7	4.3 %	2	1 %
	Those	5	3 %	1	.5 %
Total		163	100 %	200	100 %

The disparity in the frequencies and percentages of demonstratives between the Arabic and the English data, on the one hand, and between HT and MT, on the other, is significant as it sheds light on the differences

between Arabic and English demonstrative systems. The data presented in tables 8 and 9 and figure .1 show that the frequencies of demonstratives in both HT and MT (163 and 200 respectively) highly exceed those of the Arabic text (only 39).

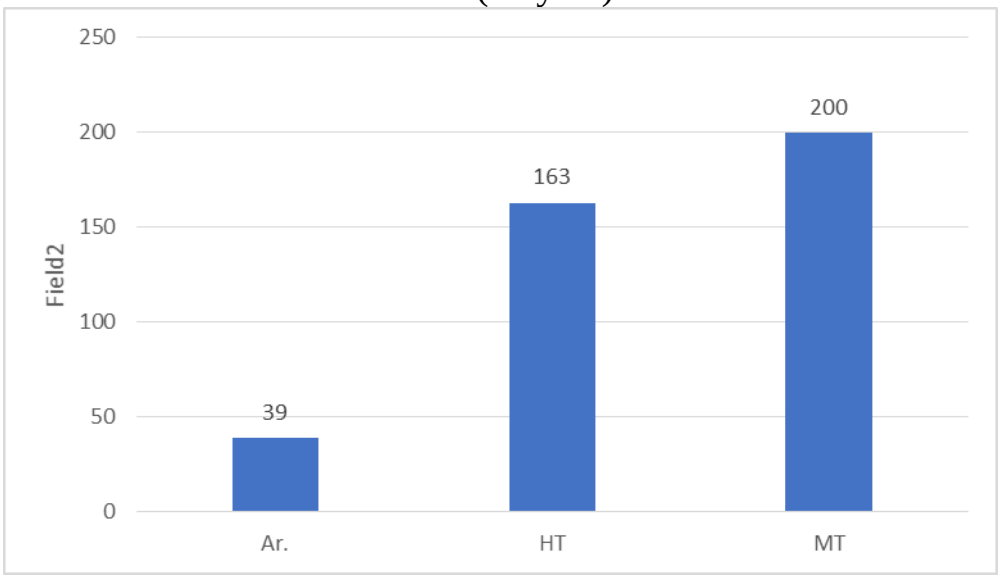

Fig. 1: Frequencies of Demonstratives in ST, HT, MT

This finding agrees with Biber et al. (1999) and Gray and Cortes (2011) who report that demonstratives are highly frequent in various types of English discourse as they potentially signify "immediate textual reference" (Biber et al. 1999: 349). Table (9) shows that the demonstrative *that* (HT =117 and MT = 156) is highly frequent in the ST data compared to other demonstratives. This is because *that* and (to a less extent) *those* have a pronominal function in English, whereas the others are more frequently used as determiners rather than as pronouns.

With regard to evaluating demonstratives as cohesive devices in HT and MT, it is noted that while HT uses demonstratives in a balanced way keeping the smooth flow of the text, MT overuses demonstratives (and other types of cohesive devices), which results in a rather loose text that is abundant in redundancy. Example (7) shows how the over use of the demonstrative *that* affects the flow of the text:

Example 7:

ST

شعور غامض يربطه بالصوت ويؤكد **أن** الصلة بينهما من صنعه ومسئوليته، و**أن** عليه وحده يقع التحمل للنهاية. وبالغريزة التفت .. كانت زوجته لا تزال على وضعها، فقط في اللحظة التي التفتت فيها ماءت مواء طال بعض الشيئ، ثم بإرادة نائمة انتقلت الى جنبها الأيسر وقربت ساقيها، ربما كان هذا هو الأثر الوحيد الذي أحدثه الصوت في جسدها المستسلم لأول مراحل النوم. وارتاح واطمأن بعض الشيئ وهو يواجه الأمر وحده، فقد كان ظهورها على المسرح ليلتها كفيلا بزيادة ارتباكه.

HT

A vague feeling linking him to the sound, proving that the relation between them was of his making and his responsibility and that it was he alone who must bear it to the end. He turned instinctively … his wife was still in her original position. When he turned, she mewed a little then shifted onto her left side and pulled in her legs: probably the only effect of the sound on her body, drifting into the first stage of sleep. He was reassured somewhat at having to face the situation alone, for her appearance on the scene at that moment would only have disturbed him more.

MT

A mysterious feeling that connects it with the sound and confirms that the link between them is his creation and responsibility, and that he alone should bear the end. Instinctively he turned… His wife was still on her condition, just the moment she turned, the water of the meow lasted for a bit, then with a sleeping will she moved to her left side and brought her legs closer, perhaps this was the only effect that the sound had on her body that surrendered to the first stages of sleep. He was relaxed and somewhat reassured while he was facing the matter alone, as her appearance on the stage that night was sufficient to increase his confusion.

While the Google MT uses the demonstrative *that* six times to render this passage in English, the human translator uses it only three times. It seems that the MT overuse of this demonstrative is at the expense of other cohesive devices that are used by the MT as shown in table 10:

Table 10: Overuse of "that" demonstrative in MT

	HT	MT	Comment
1.	A vague feeling linking him to the sound	A mysterious feeling that connects it with the sound	Ellipsis: in HT the pronominal connector "that" is omitted and the adjective clause is reduced to an adjective phrase.
2.	probably the only effect of the sound on her body	perhaps this was the only effect that the sound had on her body	HT uses the of-structure making the text more cohesive than the MT which uses the pronominal "that" to introduce a whole clause
3.	on her body, drifting into the first stage of sleep	her body that surrendered to the first stages of sleep.	Ellipsis: in HT the pronominal connector "that" is omitted and the adjective clause is reduced to an adjective phrase.

8.1.3. Comparatives

Table 11: Frequencies of comparatives in ST

Comparatives in ST											
Item	أكثر	أسرع	أعلى	أقوى	أقرب	أروع	أرقى	أذكى	أدق	أدنى	Total
Freq.	10	4	3	2	2	1	1	1	1	1	26

Table 12: Frequencies of comparatives in HT and MT

	Item	Human Translation	Machine Translation
Comparatives	---er/more than /	32	24

Tables (11) and (12) show that comparatives occur at lower frequencies than personals and demonstratives in the three texts under investigation (ST =26, HT = 32, MT =24). Thus it can be said that cohesion in the three texts depends more on pronominals and demonstratives than on comparatives. It is also observed that the frequencies of comparatives in the three texts do not show any significant differences.

Examples (8) and (9) show that whereas the common type of comparatives used in Arabic is the adjectival comparatives that has "a head + complement" structure, the recurrent pattern in English is the "--er/ more than" pattern.

Example 8:

وكنها لم تستجب وبفحيح **أكثر انخفاضا والحاحا** سألته

a)

b) But she did not stop. In an urgent, persistent hiss, she asked … (HT)

c) and with a **lower and more urgent** hiss, she asked him (MT)

Example 9:

كمن يفتش في كومة من قش قديم **عن ابرز ملامح** لطفل صديق كان **أعز عليه من نفسه**.

b) as though for a needle in a haystack, for the features of the childhood friend whom he had loved **more dearly** than himself (HT)

c) as if searching in a pile of old straw from the **tiniest features** of a friend who **was dearer to him than himself** (MT)

Example (8) shows that the Arabic comparative "head + noun" [أكثر انخفاضا وإلحاحا], which is disregarded in the HT, is translated literally in MT as "lower and mor urgent." In (9), the comparative أبرز ملامح is

ignored in HT, but mistakenly translated as "the tiniest features" in MT. The comparative structure "أعز عليه" is rendered by the two translators using almost the same wording, "more dearly" by HT versus "dearer" by MT.

8.2. Conjunction

Table (12): Frequencies and Percentages of Conjunctions in ST, HT, MT

	ST		HT		MT	
Conjunction Types	F.	%	F.	%	F.	%
Additives	605	85.7 %	384	68.4 %	489	73.6 %
Adversatives	43	6.1 %	83	14.8 %	60	9 %
Causals	16	2.3 %	63	11.3 %	87	13 %
Temporals	42	5.9 %	31	5.5 %	28	4.4 %
Total	706	100 %	561	100 %	664	100 %

Four groups of cohesive conjunctions are included in the study: additives, adversatives, causals and temporals. As shown in table (12), frequencies and percentages show differences in the occurrence of conjunctions in the three texts. It is interesting that the ST achieves a higher level of conjunction frequencies than the other two texts (ST = 706, HT= 561, MT = 664). However, the results show that there are slight differences between ST and MT. This might be explained in the light of the fact that MT almost renders literal translation of ST, which means that almost all the conjunctions occurring in the ST are literally rendered in MT. These slight differences are due to the inability of the Machine to understand the context in which these conjunctions occur. The table also shows that the three texts contain a higher number of additives (ST = 605, HT =384, MT = 489) than adversatives (ST = 43, HT = 83, MT = 60), causals (ST = 16. HT = 63, MT= 87), and temporals (ST = 42, HT = 31, MT = 28). This finding provides further evidence to Bystrova-McIntyre's (2012) and Chaalal's (2017) regarding the overwhelming occurrence of additives in original, human-translated and machine-translated texts. The following sections contain a detailed account of the results obtained for each type of cohesive conjunctions.

8.2.1. Additives

Table (13): Frequencies and percentages of additives in ST, HT and MT

Arabic (ST)			English (TT)				
Item	F	%	Item	HT		MT	
				F	%	F	%
و وأيضا	602	99.5 %	and and also	383	99.7 %	474	97 %

ISSN 1110-2721

Occasional Papers
Vol. 76: October (2021)

			as well as				
أيضا	3	.5 %	what is more also	1	.3 %	15	3 %
علاوة على ذلك أي أن بالمناسبة بمعنى آخر على سبيل المثال فوق ذلك في المقابل	0		moreover besides additionally in addition to this not only but	0		0	
	605	100 %		384	100 %	489	100 %

The results presented in table (13) show discrepancy in the use of additives between the Arabic ST on the one hand, and the two TTs on the other. Whereas the Arabic ST yields a high frequency (N = 605), the frequencies of the TT texts are much lower; it is 384 in HT, and 489 in MT respectively. Surprisingly, additives in the three texts under investigation are restricted to a limited number of items: in ST the recurrent items are "و" , "وأيضا" and "أيضا"; in TTs the four additive items which occur are "and", "and also", "what is more", and "also". The other additives that are listed in the table have zero frequency.

The results also show that the Arabic "و" and the English "and" do not have one-to-one relationship. The frequencies obtained are 602 for "و"/ "and" in the Arabic ST, 383 for HT and 474 for MT. In English, "and" can be said to perform two major functions: coordination of ideas and continuity of actions (Schiffrin, 1987). The Arabic "و", however, is the most frequently occurring connective as it is used to connect verbs and sentences. Bahloul (2008, pp. 89-91) reports that "و" can be used to perform as many as five functions in Arabic. First it can be used in sentence initial position to introduce a chunk of information. This is illustrated in example 10 where "و" introduces information about the state of silence in Al-Hadidi's house where the incidents of *The Language of Pain* take place. It is interesting to observe that "و" is absent in the translated texts as English rarely uses "and" in sentence initial position.

Example 10:

وكان الصمت الذي حل تاما ساحرا كالدواء الشافي المعجز.

The silence that followed was magic, like a miraculous healing balm. (HT)

The complete silence was as magical as a miraculous cure. (MT)

Secondly, "و" is also used to signal additive relations between phrases and sentences. In (11), it is used to connect two verb phrases, and in (12),

it connects a series of noun phrases. However, in (12), the HT uses the connective "as well as" as a synonym of "and" to avoid redundant repetition, while the MT uses it once before the last item in the coordinate series.

Example 11:

خصال جعلت الجميع **يدهشون ويفجعون** لاقدامة على سرقة العنزة،

It was because he had these qualities that people were **astonished and disappointed** when he stole the goat. (HT)
qualities that made everyone **astonished and mourned** for his attempt to steal the goat (MT)

Example 12:

وأنهم لفوا وتعبوا على جميع **"حكما" المركز ومستوصفاته ومستشفياته وحلاقي صحته والعرب** الذين يكوون بالنار، "ويخرمون" بالمسلة،

they had made the rounds of **all doctors, clinics and hospital** in their district, **as well as the local barbers, and even the Arabs** who specialized in cauterization and acupuncture (HT)
that they wrapped up and exhausted **all the "rulers" of the center, its clinics, hospitals, health barbers, and the Arabs** who made fire and "punched" with the obelisk (MT)

A third function of "و", according to Bahloul (2008), is that it expresses a successive relationship as it is used to connect episodes in a narrative. In the following example, the action of trembling is succeeded by the action of approaching. This is illustrated in example 13.

Example 13:

بدأت **ترجف وتقترب** من زوجها

Then she started to **tremble, and drew** closer to her husband (HT)
she began to **shake and approached** her husband (MT)

One more function of "و" in Arabic is that it can be used to connect two simultaneous actions as illustrated in example 14 below:

Example 14:

a) ولكنها بدت لأول وهلة جمادية ذات صليل كعظام البشر **تتكسر و تتهشم**

b) At first it was like the sound of matter the iron of a giant of incredible strength, with the merciless intent of **shattering the bones**. (HT)
c) … but at first it seemed to be a collective, tingling, like human bones being **broken and shattered** (MT)

In (14 a), the Arabic "**تتكسر و تتهشم**" [to be broken and shattered] refers to two actions taking place at the same time. However, it is rendered as "**shattering the bones"** in HT and "**broken and shattered**" in MT. It is interesting to note that these verbs are synonyms in Arabic as they both refer to the action of "smashing or destroying." Unlike MT, HT exhibits more freedom in the translation of the example to avoid repetition of synonymous items.

8.2.2. Adversatives

Table (14): Frequencies and percentages of adversatives in ST, HT and MT

Arabic (ST)			English (TT)				
Item	F	%	Item	HT		MT	
				F	%	F	%
لكن	31	72 %	but	50	60.2 %	49	81.8 %
بالرغم من (ذلك)	7	16 %	though / in spite	23	27.7 %	2	3.3 %
في نفس الوقت	3	7.4 %	yet	6	7.3 %	0	0 %
مهما يكن	2	4.6 %	despite	0	0 %	5	8.3 %
مع ذلك على أية حال في الواقع بالأحرى على الأقل على العكس في الواقع بأي طريقة	0		however	2	2.4 %	0	0 %
			rather	1	1.2 %	2	3.3 %
			instead	1	1.2 %	2	3.3 %
			nevertheless on the contrary	0		0	
	43	100 %		83	100 %	60	100 %

The results in Table (14) show differences in the use of adversatives in ST and TTs. The frequencies of this type of conjunction in TTs (HT = 83, MT = 60) are significantly higher than those of ST (=43). It is interesting to note that the adversative conjunction "but" is more frequently used than other adversatives in the three texts under investigation. This may be due to the fact that in both Arabic and English "but" [لكن] occurs in different positions in the sentence to contrast ideas and information. The Arabic "بالرغم من " (f = 7) and the English "though" (HT = 23, MT =2) rank next to "but" in the distribution of adversative conjunctions. Table (14) also shows that other types of adversatives are of low frequencies in TTs (Yet = 6 in HT, despite = 5 in TT).

An interesting point to make is that the discrepancies in the frequencies between the Arabic "لكن" and the English "but" is attributed to the fact that some Arabic items other than "لكن" are translated as "but"; such items include "إلا" , "أنما" and "و"as illustrated in examples 15 , 16 and 17 below. This is the reason why the frequencies of "but" in HT and MT are almost the same, 50 and 49 respectively.

Example 15:

في لحظة مر بخياله الف احتمال **الا** الإحتمال الوحيد الذي كان يخاف مروره.

In a second a thousand possibilities flashed through his mind , all **but** the one possibility he dreaded (HT)

In a moment, he passed through his imagination a thousand possibilities, **but** the only possibility that he was afraid of passing. (MT)

Example 16:

ليس خوفا منه **وإنما** من الشيئ المجهول المروع الذي يختفي لابد وراءه ويحدثه

not for fear of the sound itself, **but** because of that unknown terrible thing which inevitably lurked behind it (HT)

not for fear of him, **but** from the horrific unknown thing that disappeared behind him and caused it (MT)

Example 17:

ومن غير دعاء كان قد قرر أن يتكفل بالأمر

But their pleas were unnecessary, for he had already decided to help. (HT)

Without a prayer, he had decided to take care of the matter. (MT)

8.2.3. Causals

Table (15): Frequencies and Percentages of Causals in ST, TT and MT

Arabic (ST)			English (TT)				
Item			Item	HT		MT	
	F	%		F	%	F	%
لذلك	4	25 %	so	38	69.8 %	15	89.7 %
لأن	7	43.75 %	because (of that)	10	15.9 %	7	8 %
اذن	2	12.5 %	Thus, hence	5	7.9 %	2	2.3 %
بسبب ذلك	2	12.5 %	therefore	2	3.2 %	0	0
على أساس	1	6.25 %	As a result	2	3.2 %	0	0
نتيجة لذلك آخذا بعين الإعتبار في مثل هذا الظرف	0		Consequently Accordingly	0	0	0	0
	16	100 %		57	100 %	24	100 %

A comparison between the frequencies and distribution of causals in the data shows that the frequencies of causals in the TTs are higher than them in the Arabic ST (ST =16, HT = 57, MT = 24). As indicated in Table (15), the most obvious difference appears in occurrence of "لذلك"/ "so" (ST = 4, HT =38, MT = 15). The low frequency of causals in Arabic data is attributed to the phenomenon that Arabic additives are sometimes used with causal meaning as shown in the examples 18 and 19 below where "و" are translated as "so" by the human translator. MT ignores the causal use of the additive "و", which makes the MT text less cohesive than the HT rendering.

Example 18:

وأمر بدخولهم

So he had them come in. (HT)
He ordered their entry. (MT)

Example 19:

وأجال بصره محاولا أن يعثر على من يصلح ليكون أبا لفهمي أو عمه

And **so** he gazed at them, trying to find one who might be Fahmy's father or uncle (HT)
He looked back, trying to find someone who could be the father of my understanding or his uncle (MT)

8.2.4. Temporals

Table (16): Frequencies and percentages of temporals in ST, HT, and MT

Arabic (ST)			English (TT)				
Item			Item	HT		MT	
	F	%		F	%	F	%
ثم عند /حين	16	36.6	then	18	58	8	28.6
ذلك(بعد -قليل)	14	34	After that	11	35.6	12	42.9
ثم	4	9.8	Next	1	3.2	7	25
اخيرا	4	9.8	Lately	1	3.2	1	3.5
سابقا	2	4.9					
في الحال	2	4.9					
Total	42	100		31	100	28	100

Table (16) shows slight difference in terms of frequencies and distribution of temporals in the three texts under investigation (ST = 41, HT = 31, MT = 28). The results obtained in this context provide further evidence to support Grisot and Blochowiak (2021) regarding the similarity of the rates of temporal relations in original and translated texts (P. 415). The slight differences are indicative of the phenomenon that the

conjunction "و " and its English equivalent "and" are sometimes used with temporal meanings as shown in example (20) where the Arabic "و " , which is used to indicate a temporal sequence relation, is translated as "then" in HT and as "and" in TT. The example also shows that the human translator uses "then" as a translation of the conjunction "و" because she is more aware of this meaning variation than MT. Example (21) provides further support as the Arabic "ثم " is translated as "and" by HT and as "the" by MT.

Example 20:

... كشريط اللمبة حين يحمر من تلقاء نفسه **ويقصر ويحترق** لدى فراغ الكيروسين؟

…like a flame of a gas lamp when the kerosene runs low, fuming red, **then** shrinking, **then** burning itself out? (HT)

… like a lamp strip when it turns red on its own **and** shortens **and** burns when the kerosene is empty? (MT)

Example 21:

وأن الذي باستطاعته ان يتفوق كطفل لابد باستطاعته أن يتفوق كشاب **ثم** كرجل.

and that he who could shine as a child would surely shine as a youth, **and** as a man. (HT)

and that he who could excel as a child must be able to excel as a young man and **then** as a man (MT)

9. Conclusion

This study traces some cohesive devices in human made translation and machine translation of a literary text, namely, Yousef Idris's short story [لغة الآي آي], translated as "The Language of Pain." The analysis is limited to reference and conjunction as two crucial features of grammatical cohesion. The research questions posed in section 5 inquire about the cohesiveness of MT compared to HT, and whether or not there is variation in the use of cohesive devices in the two TTs that are examined in the study.

The results show that cohesive properties vary according to the translation method involved: HT or MT. Both human and machine translations show differences in the frequencies of cohesive devices from those found in the original Arabic ST. This is due to differences in some aspects of cohesion in both English and Arabic. The results also show variation regarding the frequencies and distribution of cohesive devices in HT and MT.

Regarding reference, the results show significant differences in the occurrence of personals and demonstratives in both HT and MT. On the one hand, personals in MT are almost as double as those occurring in HT. On the other hand, the occurrence of demonstratives in MT highly exceeds it in HT. However, this does not mean that the machine produces a more cohesive text than the human translator. Th excessive occurrence of cohesive reference makes the text loose and incoherent. As for conjunction, the results also show that MT achieves higher frequencies of conjunctions in general, and of additives and causals in particular, than HT. Like reference, the redundant use of conjunctions on the part of MT yields a text which is loose, ambiguous, and confusing to the reader. It is recommended that MT product be submitted to human revision and editing in terms of text cohesion so as the grantee that the final product meets the requirements of textual cohesion in the target language.

References

Abdullatif, M. H. (2003) (in Arabic). *binaa aljumla al'arabeyya*. [بناء الجملة العربية]. Cairo: Dar Ghareeb

Abu-Ayyash, E. A. (2017). Errors and non-errors in English-Arabic machine translation of gender-bound constructs in technical texts. *Procedia Computer Science 117:73-80*. DOI:10.1016/j.procs.2017.10.095

Ahrenberg, L. (2017). Comparing machine and human translation: A case study. Conference Paper. *RANLP 2017 - Workshop on Human-Informed Translation and Interpreting Technology*. DOI:10.26615/978-954-452-042-7_003

Al-Jabr, A.M. (1987). *Cohesion in text differentiation: A study of English and Arabic*. (Unpublished Doctoral Dissertation, University of Aston, Birmingham). Retrieved from http://eprints.aston.ac.uk/10288/

Al-Shurafa, N.S.D. (1994). Text linguistics and cohesion in written Arabic. *JKAU' Arts and Humanities,* 17, 17-30.

Bahloul, M. (2008). *Structure and Function of the Arabic Verb*. London: Routledge.

Beeston, A.F.L. (1970). *The Arabic Language of Today*. London: Hutchinson.

Betti, M. & AlFartoosy, M. (2019). Ellipsis and reiteration in English and Arabic: A contrastive study. *English Language and Literature Studies*, (9) 1. ISSN 1925-4768 E-ISSN 1925-4776

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson.

Bojar O (2011) Analysing error types in English-Czech machine translation. *The Prague Bulletin of Mathematical Linguistics* No. 95, 2011, pp. 63–76.

Bystrova-McIntyre, P. (2012). Cohesion in Translation: A corpus Study of Human-Translated, Machine-Translated, and Non-Translated Texts (Russian into English). Unpublished PhD Dissertation, Kent State University. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=kent1353451112&disposition=inline

Chaalal, I. (2017). Arabic/English Translation of Cohesive Devices in the United Nations Texts: A Corpus-based Study. (Unpublished PhD Dissertation, University of Mentouri – Constantine 1) Retrieved from https://bu.umc.edu.dz/theses/anglais/CHA1482.pdf

Christensen, T. & Schjoldager, A. (2016). Computer-aided translation tools – the uptake and use by Danish translation service providers. *The Journal of Specialised Translation (25).* https://www.researchgate.net/publication/292150749

Costa, A., Ling, W., Luis, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for Machine Translation error analysis. *Mach Translat* (2015) 29:127-161. DOI 10.1007/s10590-015-9169-0

Crane, P., (2008). Texture in text: A discourse analysis of a news article using Halliday and Hasan's model of cohesion. *Nagoya University of Foreign Studies Research Journal*, 30, 131-156.

De Beaugrande, R. & Dressler, W. (1981). *Introduction to text linguistics*. London: Longman.

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication 10(1)*: 947-969. doi:10.2478/v10108-011-0005-2

Gray, B. and Cortes, V. (2011). Perception vs. evidence: An analysis of *this* and *these* in academic prose. *English for Specific Purposes* 30, 31-43. https://doi.org/10.1016/j.esp.2010.06.004

Grisot, C. & Blochowiak, J. (2021). Temporal relations at the sentence and text genre level: The role of linguistic cueing and non-linguistic biases- An annotation study of a bilingual corpus. *Corpus Pragmatics (2021) 5*. Pp. 379–419. https://doi.org/10.1007/s41701-021-00104-5

Hadla, L., Hailat, T., & Al-Kabi, M. (2014). Evaluating Arabic to English machine translation. *International Journal of Advanced Computer Science and Applications*, 5 (11). Pp. 68-73

Halliday, M. A. K. (1973). *Explorations in the Functions of Language* London: Edward Arnold

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. New York: Longman.

Halliday, M. K. & Matthiessen, C.(2014). *Halliday's Introduction to Functional Grammar*. London: Routledge.

Hampshire, S. & Salvia. C. (2010). Translation and the Internet: Evaluating the quality of free online machine translators. *Quaderns. Rev. trad.* 17. Pp. 197-209

Holmström, J.E. (1951): *Report on interlingual scientific and technical dictionaries*. Paris: Unesco.

Hutchins, J., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic Press Limited.

Hutchins, J. (1986). *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood and New York: Halsted Press.

Hutchins, W.J. (2001). Machine translation and human translation: in competition or complementation? *International Journal of Translation*, 13 (1-2). pp.5-20.

c) Idris, Y. (1977). (In Arabic) *لغة الآي آي [The language of Pain]*. Cairo: Misr Printing House

d)

e) Idris, Y. (1990). *The Language of Pain and Other Stories* (Nagib, N. Trans.). Cairo: GEBO. (Original work published in 1977).

f)

Kadhim, K., Habeeb, L., Sapar, A., Hussin, Z., & Abdullah, M. (2013). An evaluation of online machine translation of Arabic into English news deadlines: Implications on students' learning purposes. *The Turkish Online Journal of Educational Technology*. 12(2). Pp. 39-50.

g) Kadiu, S. (2019). *Reflexive Translation Studies: Translation as a Critical Reflection*. UCL Press. http://www.jstor.com/stable/j.ctv6q5315.9

h)

Koch, B.J. (1981). Repetition in Discourse: Cohesion and Persuasion in Arabic Argumentative Prose. (Unpublished PhD Dissertation, University of Michigan). Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/127596

Lapshinova-Koltunski, E. (2017). Cohesion and translation variation: Corpus-based analysis of translation varieties. In Katrin Menzel, Ekaterina Lapshinova Koltunski & Kerstin Kunz (eds.), *New perspectives on cohesion and coherence*, 95–118. Berlin: Language Science Press. DOI:10.5281/zenodo.814468

Lapshinova-Koltunski, Ekaterina. (2015). Exploration of inter- and intralingual variation of discourse phenomenon. In *Proceedings of EMNLP 2015 Workshop on Discourse in Machine Translation*. Lisbon.

McCarthy, M. (1991) . *Discourse analysis for language teachers*. Cambridge: CUP.

i) Mehamsadji, M. (1988). Cohesion and Text Development in Writing Arabic. Unpublished PhD dissertation: University of Stanford. Retrieved August 2021. URL: http://usir.salford.ac.uk/id/epring/2200

j)

k) Mohamed, E. & Sayyed, Z. A. (2019). Arabic-SOS: Segmentation, stemming, and orthography standardization for classical and pre-modern standard Arabic. In *DATeCH2019: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. NY, USA: Association for Computing Machinery.

l)

Omar, A. & Gomaa, Y. (2020). The machine translation of literature: Implications for translation pedagogy. *International Journal of Emerging Technologies in Learning (iJET) 15(11)*. Pp. 228- 234. https://doi.org/10.3991/ijet.v15i11.13275

m) Precup-Stiegelbauer, & Laura-Rebeca. (2013). Automatic translations versus human translations in nowadays world. *Procedia - Social and Behavioral Sciences*, 70(0), 1768-1777. DOI: http://dx.doi.org/10.1016/j.sbspro.2013.01.252

n)

o) Schiffrin, D. (1987). *Discourse Markers*. New York: Cambridge University Press

p)

Turovsky, B. (April 28, 2016). "Ten years of Google Translate". Google Translate Blog. Retrieved from https://www.blog.google/products/translate/ten-years-of-google-translate/

Williams, M.P. (1989). *A Comparison of the textual structures of Arabic and English written texts: A study in the comparative orality of Arabic*. (Unpublished PhD Dissertation, University of Leeds). Retrieved from http://etheses.whiterose.ac.uk/469/

Yaqoub, E. B. (1971) (in Arabic) *Encyclopedia of Arabic Language Sciences* [موسوعة علوم اللغة العربية]. Beirut: Dar Alhekma

Appendix

Fig. 1: Desktop screen shot of data processing

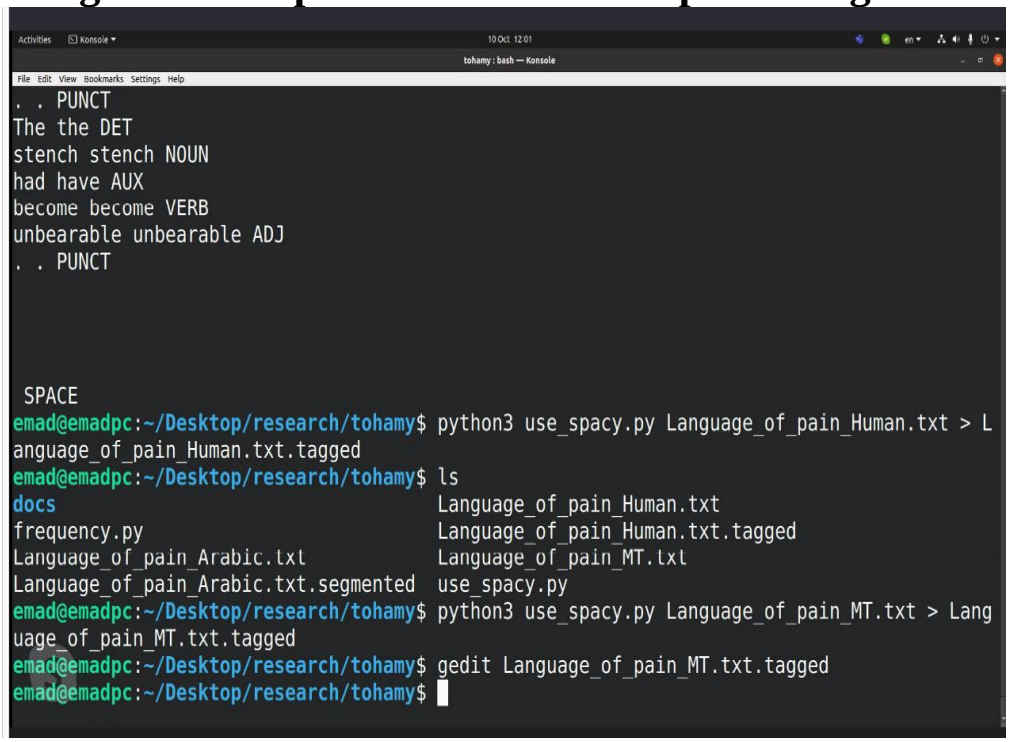

Fig. Screenshot of Morphological segmentation of the Arabic Data

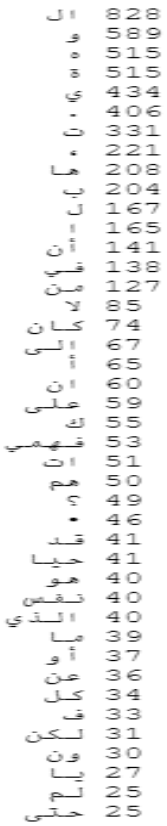

Fig. A screenshot of part of speech tagging of the English data

A detailed screenshot of part of speech tagging of the English data

SPACE

It -PRON- PRON

was be AUX

not not PART

exactly exactly ADV

a a DET

cry cry NOUN

,, PUNCT

but but CCONJ

it -PRON- PRON

was be AUX

the the DET

first first ADJ

time time NOUN

... ... PUNCT