Revue SCIENCES HumaineS n°45, Juin2016, Vol A, pp63-89

Cohesive Devices in the United Nations Arabic Texts and their English Translations: A Corpus-based Study

Abstract:

This paper examines one essential aspect of equivalence in translation: textual cohesion. It aims to contrast Arabic and English cohesive devices in some United Nations texts, find out how they are interpreted as equivalents and examine the regularity in shifting cohesion. A parallel corpus of texts of 15 Arabic texts and their English translation counterparts are selected and the distribution of their cohesive devices is analysed. Using computer software, the use of cohesive devices is compared. Findings reveal that Arabic and English have more similarities than differences in terms of the cohesive devices used though there are obvious differences in the frequency of their occurrence. The similarities are significantly preserved for the purpose of accurateness, transparency and formality that characterise the language of legal texts.

Imen CHAALAL
Faculty of Letters and Languages
Department of Letters and English
University of Mentouri
Constantine

ملخص:

يدرس هذا البحث أحد الجوانب الهامة في مسألة المعادلة في الترجمة و المتمثلة في التماسك النصي. وتهدف هذه الدراسة إلى مقارنة أدوات الربط في اللّغتين العربية و الإنجليزية المستعملة في بعض نصوص الأمم المتحدة و تحديد إلى أي مدى يمكن اعتبارها متكافئة، بالإضافة إلى دراسة أهمية التغيرات في تحقيق المعادلة بين النصوص. من خلال بناء مدونات متوازية لقد تم وضع 15 نصاً عربياً في وثائق الأمم المتحدة إلى جانب 15 نصاً مقابلاً لها باللّغة الإنجليزية، بالإضافة إلى توزيع أنواع هذه الأدوات. باستعمال برامج معلوماتية تتم المقارنة بين استعمال أدوات الربط في المدونتين باللغتين العربية و الإنجليزية. وتكشف النتائج بأن ثمة أوجه تشابه أكثر من أوجه الاختلاف بين اللّغتين فيما يخص أدوات الربط المستعملة في حين أن هناك أكثر اختلافات وبشكل واضح على مستوى وتيرة ورود هذه الأدوات. في الواقع، لقد تم حفظ أوجه التشابه في الترجمة بهدف تحقيق الدقة و الشفافية و الصورية المميزة للغة النصوص القانونية .

Introduction :

Nowadays, the accelerated phenomenon of globalization has significantly affected the fields of translation, contrastive studies and corpora, and has allowed them to gain popularity more than ever. Scholars working in these fields, especially Baker [1], are credited for pushing these areas towards empirical, interdisciplinary and multilingual investigations. Hence, translation and parallel corpora are conjoined together in order to undergo a contrastive study of Arabic and English cohesive devices used in some United Nations Texts.

In the history of contrastive studies, the focus had always been on the micro-linguistic approach to language, i.e. the analysis of grammar, phonology and lexis. However, with the development of linguistic studies, contrastive studies became more concerned with macro-linguistic analyses, i.e. the study of discourse analysis and text linguistics. As a result, the shift of interest from words in isolations to whole texts found also its way in translation studies. Henceforth, the topic of cohesion was introduced because it plays a significant role in the organisation of unified texts, and became one major topic of text translation. Newmark [2] as cited in Baker [3:180] pointed out that "the topic of cohesion… has always appeared to me the most useful constituent of discourse analysis or text linguistics applicable to translation."

During the process of translation, it is inevitably important to add or to omit parts of text in order to maintain its cohesiveness; such changes are generally known as shifts of cohesion which help translators to create accurate and natural translation products. Based on Halliday and Hasan's model of cohesion [4], the study attempts to observe variation in the way each language makes use of cohesive devices to signal cohesion relations and to identify the shift of cohesion, at the textual level, in the English translations of some Arabic United Nations texts.

The questions asked are whether the diverse grammatical and lexical structures of Arabic and English exhibit different types of cohesive devices in translation, and whether there are any significant differences in the frequency of occurrence of these devices across the two languages. The focus, then, will be on the distribution of grammatical and lexical cohesion in the translated text and the examination of how these devices can render the source language texts and the target ones into both cohesive and coherent products. Thus, by adopting the notion of translation equivalence as the best basis for contrastive analysis, the study aims basically (1) to scrutinise the behaviour of cohesive devices across the two languages and note to what extent they are interpreted as equivalents and (2) to examine, occasionally, the occurring patterns of shifts of cohesion in the translation product. Moreover, based on the assumption that the contribution of corpus-based studies in contrastive studies may be used for practical purposes in translation, and grounded on the very significant role of parallel corpora in the examination of the differences and similarities between languages, this research makes use of a unidirectional parallel corpus of Arabic and English United Nations texts.

Textual Cohesion

The last few decades have witnessed a growing interest in the application of discourse analysis and text-linguistics theories on translation studies. This

increased propensity fostered linguists to scrutinise aspects of cohesion in translation. While first eminent scholars such as de Beaugrande and Dressler [5] and Halliday and Hasan [4] defined cohesion in one single language, English, some others like Baker [3], Hatim and Mason [6] and Blum-kulka [7] defined cohesion in translation studies.

The concept of cohesion as suggested by Halliday and Hasan [4] and later on developed by Halliday [8] has played an important role in discourse analysis and text-linguistics. These two fields emphasise on one major argument: that any piece of text or discourse is produced for a specific function, either to inform readers or to persuade them; what is important is that writers use the texts' structures and semantic wholeness to achieve one particular goal. Following this view and in their influential work, Halliday and Hasan [4] proposed the concept of *texture* referring to the "property of being a text" [4: 2]. According to them, every text has its own features of organisation that help people distinguish between a text and a random collection of sentences. That is to say, by means of cohesive ties, people do relate sentences in a particular sequence in order to convey their intended meaning. Cohesion, thus, is an indispensable element to provide texture; it is a semantic relation which is classified in details into five types: reference, conjunction, ellipsis, substitution and lexical cohesion. Halliday [8: 311] argued that cohesion is one aspect of the study of texture and maintained that:

> It is important to be able to think of text dynamically, as an ongoing process of meaning; and of textual cohesion as an aspect of this process, whereby the flow of meaning is channelled into a traceable current of discourse instead of spilling out formlessly in every possible direction.

Similarly, de Beaugrande and Dressler [5] proposed that for a piece of writing to be considered a text, it must meet seven standards of textuality; otherwise the text cannot be considered communicative. Cohesion, thus, is made as one of seven standards of textuality: cohesion, coherence, intentionality, acceptability, informativity, situationality, and intertextuality. They explained that cohesion:

> Concerns the ways in which the components of the SURFACE TEXT, i.e. the actual words we hear or see, are mutually connected within a sequence. The surface components depend upon each other according to grammatical forms and conventions, such that cohesion

> rests upon GRAMMATICAL DEPENDENCIES…. All of the functions which can be used to signal relations among surface elements are included under our notion of COHESION. (9: 30)

Therefore, being a necessary element in the creation of texts, cohesion strongly influences the quality of the translation product. The significance of this concept in the framework of translation studies has been discussed by many researchers from different perspectives, for example: Baker [3], Blum-Kulka [7], and Hatim and Mason [6].

Baker [3], for example, reiterated that the main goal of translators is to guarantee a degree of equivalence at the textual level, rather than at the word or sentence level. According to her, translators intend to produce natural translated texts that are fluent and accurate without appearing foreign versions. That is why translators are requested to adjust some of the features existing in the source text in order to fit the organisation of the target texts.

Hatim and Mason [6], as well, maintained that translators should consider the texture features existing in each language while interpreting the communicative meaning. According to them, cohesive devices are language specific because of the languages' different origins; that is why, they may pose great challenges for translators. They expressed this as follows:

> The various activities of translation criticism, translation assessment and revision all run the risk of concentrating on features of texture without relating them to the communicative process which engendered them. Texture needs to be seen an integral part of what one is doing with one's language. (6: 194)

In her discussion on shifts in cohesion, Blum-kulka [7:17] defined cohesion as "an overt relationship holding between parts of the text, expressed by language specific markers." She postulated one main argument adopted from the idea that the process of translation necessarily involves shifts in textual and discoursal relationships. Her argument is grounded on the perception that translation is viewed as an act of communication, and, therefore, all differences connected to both linguistic and cultural aspects holding between the two languages must be taken into consideration.

According to her, shifts of cohesion are evident in translation and fall into two types: *shifts in levels of explicitness* and *shifts in meaning*. In the first type, she explained that cohesive explicitness refers to shifts in the type of cohesive devices used. These shifts are achieved through the substitution of cohesive

devices with some others which do increase the level of cohesion in the target text. In the second type, i.e. *shifts in meaning*, she explained that the translation product demonstrates a change in the explicitness and implicitness of the meaning of the source text. Factors that explain these phenomena are explication, stylistic preferences or culture-bound translation norms.

Blum-kulka [7: 19] indicated that:

> On textual level, shifts in levels of explicitness through translation have been claimed to be linked to differences in stylistic preference for types of cohesive markers in the two languages involved in translation.

And because in any language transfer there is a tendency to explicate, this strategy has been postulated as a universal strategy used by both novice and professional translators. Blum-kulka [ibid: 21] noted that "explication is a universal strategy inherent in the process of language mediation, as practiced by language learners, non-professional translators and professional translators alike."

All in all, it might be said that the aforementioned studies revealed the significance of the concept of cohesion within the framework of translation studies. They showed elaborately how the significant role of cohesion in organising the linguistic elements into unified texts became very pertinent to the study of translation.

Contrastive Analysis, Translation and Corpus-based Studies

It is a common fact that any contrastive analysis emphasises on the so called *tertium comparationis* (TC). This means that any two languages must have some common measure by which they can be compared in order to display the similarities and differences, otherwise such a task will not be possible. On this basis, James [10] reached the conclusion that translation equivalence is the best TC for contrastive analysis. He asserted that "For two sentences from different languages to be translationally equivalent they must convey the same ideational *and* interpersonal *and* textual meanings" [James, 10: 178]. Therefore, by investigating textual equivalence in corpus based translation studies, this study will uncover clear differences and similarities between the cohesive ties across the two languages in terms of the categories used and their occurrence frequencies.

In fact, the development of computer technology and the emergence of corpus linguistic studies have brought many advantages to the study of linguistics; in the sense that they have succeeded to provide significant help in drawing attention to the study of language in use rather than the study of linguistic systems in the abstract. According to Teubert [11], real language data displayed in the form of larger corpora were strongly needed to undergo adequate investigations. He [12: 1] argued that "Today, the corpus is considered the default resource for almost anyone working in linguistics. No introspection can claim credence without verification through real language data. Corpus research has become a key element of almost all language study."

Likewise, Baker [1] anticipated that the wide accessibility of large corpora of original and translated texts, in addition to the advances of corpus-driven methodology would facilitate for translators to reveal "the nature of translated text as a mediated communicative event." [1: 243]. Following this prediction, many researchers, to name but a few, Ebeling [13], Johansson [14], and Mauranen [15] started to view the approach of corpus-based studies as a practical and a new successful way for both translation studies and contrastive linguistics. They approved, as well, the significance of parallel corpora for the study of specific language patterns cross-linguistically, and helped translators to simplify the understanding of source language texts and improve their production skills.

Although the application of corpus-based linguistics in translation studies received high attention in the last two decades; the type of translation corpora, however, was seriously questioned by some researchers. According to Teubert [16], translated texts do corrupt the language, and therefore, cannot create a consistent foundation for contrastive analysis. He stated:

> Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing the language. That is why translations have no place in reference corpora. (16: 250)

Nevertheless, the significance of translation corpora was defended by many others. Zanettin [17:21], for example, argued that:

> A decision to exclude translations on the assumption that the language of translation "corrupts" the standard norm of reference does not seem to be justified by theoretical considerations. On the other hand, corpora designed to

> investigate regularities of translation usually include a translational subcorpus. This subcorpus can be compared with different types of subcorpora in the same language, whose design will depend on the purpose of the investigation.

In a similar vein, Mauranen [15: 161] stressed that, translation corpora "constitute a valuable source of evidence for contrastive research" as they provide "language that has been used in its normal communicative contexts by a large number of users" [Mauranen, 15: 161]. Therefore, it might be said that the use of translation corpora, also known as parallel corpora, offers practical and effective advantages for contrastive studies. In other words, they reveal clearly the differences and similarities between the two languages and display a solid foundation of comparison for the particular features involved in any two languages. Johansson [14: 4] stated that "[parallel corpora] give new insights into the languages compared -insights that are likely to be unnoticed in monolingual corpora". Also, Mauranen [15: 161] indicated that "a parallel corpus can capture relations of sense as well as form, which would be very hard to capture without such data."

Finally, it is worth mentioning that many researchers have presented different classifications of corpora .Types of corpora include monolingual, parallel, multilingual and comparable corpora. See: Zanettin [18], Laviosa [19], Fernandes [20], Johannson [21], and Biel [22]. The corpus under study is called a *parallel corpus* (Baker, [23] & McEnery et al., [24]), and a *translation corpus* (Granger, [25] & Johansson, [14]); or a *translational corpus* (Lauridsen, [26]). It is, in fact, the most common version, where only two languages are involved; one sub-corpus consists of original texts, for example in *Arabic*, while the other of translated texts in a different language, *English*.

Methodology

As pointed out above, although parallel corpora play a vital role in finding equivalents, some researchers, however, believed they may not be the most convenient source of data for translators. The main reason is that the difficulties encountered in the process of translation may distort the ultimate originality of the translation product and would, eventually, make the target texts less typical and lead readers to question their quality. Therefore, in order to avoid such a complication and reduce any shortcomings, we relied on authentic texts derived from the United Nations' documentation, and produced by reasonably competent professional translators. In this view, the samples

given in Arabic and English will be derived from a unidirectional aligned Parallel Corpus of the United Nations Texts (PCUNTs). It is worth mentioning that presenting a set of aligned parallel texts, as Barlow [27] reiterated, is very remarkable because it helps users to see every sentence with its corresponding translation, and therefore, to compare the translated texts with their originals.

The Parallel Corpus of United Nations Texts (PCUNTs)

Selection of Texts for the Parallel Corpus (PCUNTs)

The texts were selected according to their reputation and availability. They were exclusively derived from UN resolutions issued by the Security Council and the General Assembly; the two main organs of UN organisations and institutions. Based on the assumption that the translations of UN bodies are of great proficiency and of a high degree of reliability, the sample texts were selected. In terms of availability, the texts are available online and downloadable via the official website of UN Documentation(1) (The ODS).

This parallel corpus (PCUNTs) consists of *nine* UN General Assembly Resolutions and *six* Security Council Resolutions published over a period of three years (2011-2013), and related to the most relevant events in the Middle East and North Africa, tackling mainly issues related to the crises in Syria and Libya. The contents of the parallel corpus are summarised in **Table1**. The parallel corpus encompasses a total of 30 texts, organised in an aligned paragraph pattern where the Arabic sub-corpus is established along with its translational counterpart in English. That is to say, 15 Arabic original texts and 15 English parallel translations are covered. The wording of texts ranges from 200 to 2000 running words per text. The whole corpus contains around (36,903) word tokens; the Arabic sub-corpus has slightly fewer word tokens (16,635) in comparison to the English one, (20,268) word tokens. Because cohesive devices tend to be more frequent in the corpus, it might be said that such a relatively small corpus is sufficient and seems to be adequate for the purpose of analysis.

Furthermore, since the present study is based on the model of Halliday and Hasan [4], it is important to emphasise that cohesive devices between sentences are “the only source of texture” [4: 9], and, that “it is the inter-sentence cohesion that is significant, because that represents the variable aspect of cohesion, distinguishing one text from another” [Halliday & Hasan, 4: 9]. However, in this paper, the analysis of cohesion is not merely restricted to inter-sentential ties for one main reason: that the punctuation system in Arabic and English is very flexible, and the notion of sentence boundaries is not

specific. For example, one whole paragraph in Arabic may contain one single sentence, whereas its translational counterpart in English definitely differs. Accordingly, the analysis of cohesive devices within sentences, i.e. at the intra-sentential level, is to be covered as well.

UN Texts	N° of Texts	Topics	Word Tokens in AUNTs	Word Tokens in EUNTs	Year of Publications
Security Council Resolutions	6	Identical letters from the Permanent Representative of the Syrian Arab Republic to the United Nations addressed to the Secretary-General and the President of the Security Council. Letters from the Permanent Representative of Libya to the United Nations addressed to the President of the Security Council.	9,589	12,037	2011-2013
General Assembly Resolutions	9	Reports of the Secretary-General concerning the situation in the Middle East: Replies received from the Syrian Arab Republic. Letters from the	7,046	8,220	2011-2013

		Permanent Representative of the Syrian Arab Republic to the United Nations addressed to the Secretary-General. Annexes to letters from the Permanent Representative of the Syrian Arab Republic to the United Nations addressed to the Secretary-General.		
Total	15		16,635	20,268

Table 1: Summary of the Corpus of United Nations Texts

Tools and Procedure of Analysis

For the purpose of analysis, we used two main software tools developed by Laurence Anthony ([28] & [29]). First, the AntConc (Anthony Concordancer) software (Anthony, [28]) is used; the word list function of the AntConc allows us to create the list of the most frequent cohesive devices in the parallel corpus (PCUNTs), and to compare the two sub-corpora in terms of the types of cohesive devices used. Second, the AntPconc (Anthony Parallel Concordancer) software (Anthony, [29]) is used in order to examine the differences and similarities between Arabic and English cohesive devices. Through the parallel concordance function of the AntPconc, the samples allow us to distinguish the differences in real context, and therefore, to detect any shift of cohesive patterns in the translated corpus.

Therefore, with the help of these two software tools (Anthony, [28] & [29]), all the cohesive devices in Arabic and English texts are computed and the types of cohesive devices in the two sub-corpora are identified according to the taxonomy of Halliday and Hasan [4], as in **Table 2**. Then, the total number of all types of cohesive devices in Arabic United Nations Texts (AUNTs) and English United Nations Texts (EUNTs) are compared, and run in SPSS 22 software (Statistical Package for the Social Sciences).

Types	Subcategories of Cohesive Devices
Reference	R.1 Pronominals, R.2 Demonstratives and Definite Article 'the', R.3 Comparatives
Substitution	S.1 Nominal Substitute, S.2 Verbal Substitute, S.3 Clausal Substitute
Ellipsis	E.1 Nominal Ellipsis, E.2 Verbal Ellipsis, E.3 Clausal Ellipsis
Conjunctions	C.1 Additive, C.2 Adversative, C.3 Causal, C.4 Temporal, C.5 Other Types
Lexical Cohesion	LC.1 Reiteration: Repetition, Synonymy or Near Synonymy, Superordinate or Hypernyms, General Term. LC.2 Collocation: Relation of Antonomy, Relation of Complementarity, Relation of Part to whole or Meronymy, Relation of part to part. Relation of Co-hyponymy, Words from the same ordered series.

Table 2: Summary of Cohesion and Coding Scheme (Halliday & Hasan, [4])

Results and Discussion

Distribution and Comparison of Cohesive Devices

The overall frequency and distribution of the identified cohesive devices across the two sub-corpora (AUNTs) and (EUNTs) are sorted out in Table 3. The figures listed in the table below showed a total of 7963 of the five types of cohesive devices in AUNTs, whereas EUNTs showed a total of 6456.

Corpus Cohesive Devices	AUN Texts N° of Occurrences	AUN Texts Percentage %	EUN Texts N° of Occurrences	EUN Texts Percentage %
Reference	2222	27.90	957	14.82
Substitution	141	01.77	156	02.41
Ellipsis	16	00.20	50	00.77
Conjunction	1703	21.38	1359	21.05
Lexical Cohesion	3881	48.73	3934	60.93
Total	**7963**	**100%**	**6456**	**100%**

Table 3: Frequencies and Percentages of Cohesive Devices

The word frequency analysis revealed that the two sub-corpora followed different decreasing orders in terms of their frequencies. In the (AUNTs), lexical cohesion was the most frequently used device (48.73%) followed by reference (27.90%) and conjunction (21.38%). Both substitution (01.77%) and Ellipsis (0.20%) appeared with very low frequencies. Unlike the (AUNTs), the mostly used devices in (EUNTs) were lexical cohesion (60.93%), followed by conjunction (21.05%) and then reference (14.82%).Similar to (AUNTs), both substitution (02.41%) and ellipsis (0.77%) were of low frequencies. Thus, it can be said that the three mostly used cohesive devices were lexical cohesion, reference and conjunctions. The distribution of these devices in the two sub-corpora displayed many similarities in the choice of the types of cohesive devices, though the differences occurred in the second most frequent type of cohesive devices. AUNTs exhibited a preference for reference while EUNTs showed preference for conjunctions. However, both substitution and ellipsis were of little use and were considered as marginal phenomena in the two sub-corpora.

Because United Nations texts are descriptive and argumentative in nature aiming at presenting facts and persuading readers, lexical cohesion plays a vital role in the organisation of information. Besides, like all legal texts, UN texts are very formal and always require a lot of transparency, precision and accurateness; these characteristics explain overtly the scarce use of substitution and ellipsis which may cause misunderstanding and ambiguity.

Distribution of the Subcategories of Conjunctions

Table 4 below provides a summary of conjunctive cohesion in the two sub-corpora. The table displays the five types of conjunctions, their number of occurrences and percentages in AUNTs and EUNTs.

Table 4: Frequencies and Percentages of the Subcategories of Conjunctions

Corpus	**AUN Texts**		**EUN Texts**	
Conjunctive Devices	N° of Occurrences	Percentage %	N° of Occurrences	Percentage %
Additive	1408	82.67	1062	78.14
Adversative	37	02.17	25	01.83
Causal	97	05.69	90	06.62
Temporal	82	04.81	72	05.29
Others	79	04.63	110	08.09
Total	**1703**	**100%**	**1359**	**100%**

A glance at the table reveals that additives (82.67% in AUNTs and 78.14% in EUNTs) are by far the most frequent devices of conjunctions, followed by causals (05.69% in AUNTs and 06.62% in EUNTS) and temporals (04.81% in AUNTs and 05.29% in EUNTs). Adversatives (02.17% in AUNTs and 01.83% in EUNTs), however, were found to be the least frequently used in the PCUNTs. Other types of conjunctions i.e. continuative conjunctions (04.63% in AUNTs and 08.09% in EUNTs) were of very low frequencies in AUNTs, whereas in EUNTs they were found in the third rank.

The nature of UN documents (resolutions), which is typically descriptive and rigid, explains the remarkable predominance of additives in the PCUNTs. Moreover, the importance given to causals and temporals in the second and third ranking respectively in AUNTs is due to, first, the argumentative characteristics of UN texts that aim to persuade readers, and, second, to their narrative function of successions of facts, events or precise reports that have taken place. The highly use of continuatives in EUNTs justifies the importance of logical relations in English United Nations texts. That is to say, English highlights the explicit means to show semantic relations at the inter-sentential level, that is why continuatives are highly employed. Continuatives like *regarding* and *with reference to* are maintained in English in order to show clearly the semantic relations of sequence.

On the whole, conjunctive devices were utilised in AUNTs as frequently as in EUNTs. No significant difference was found in the first four subcategories in terms of their occurrence ($p>0.05$), but there was a significant difference in the statistical result of continuatives ($p=0.002<0.05$). The paired samples t-test results are presented in Table 5 below.

Table 5: T-test Results for of the Subcategories of Conjunctions

Conjunctive Devices	Mean	Std. Deviation	t	df	Sig.(2-tailed)
Additive	23.06667	23.14386	3.860	14	0.066
Adversative	0.80000	3.36367	0.921	14	0.373
Causal	0.46667	3.35659	0.538	14	0.599
Temporal	0.86667	6.33434	0.530	14	0.604
Continuatives	-2.06667	4.00832	-1.997	14	0.002

The wordlist function of the AntConc Software tool (Anthony, [28]) revealed that the most remarkable difference between the two sub-corpora lies in the frequency of additives و *wa* (1333 word tokens) and *and* (908 word tokens) which are significantly different. In fact, Arabic و *wa* as shown in **Figure 1** is the most frequent of all conjunctions in Arabic and ranks first in all Arabic words. Almost all linguists who approached Arabic were struck by its high frequency. Dudley-Evans and Swales ([30], cited in Al-Jaber [31]), for example, examined the redundancy of و *wa* in Arabic and explained that a number of factors contribute to its high frequency. Al-Jaber [31] summarised them as follows: First, the lengthy sentences that characterise Arabic generate the abundant use of و *wa*. Also, the trend of Arabic towards using coordination as a favoured structural device enriched the use of conjunction و *wa*. Moreover, the occurrence of و *wa* with other conjunctions such as the additive أيضاً *aj.dan* and adversative لكن *lakin*, which is very usual, in Arabic enhances its frequency. Equally the conjunction *and* occurs with other conjunctions in English, for example *and also*, but does not occur with *but*. Therefore, this extensive usage of و *wa*, in addition to its multifunctional nature working as causals, temporals or adversatives increases its frequency; this is also valid to the conjunction *and* but with lower frequencies.

Word List Results 3

Word Types: 2676 Word Tokens: 20268

Rank	Freq	Word	Lemmas
1	1937	the	
2	1024	of	
3	908	and	
4	840	to	
5	524	in	
6	353	arab	
7	312	that	
8	216	a	
9	208	for	
10	187	syrian	
11	185	by	
12	182	on	
13	174	x	
14	173	with	
15	131	its	
16	128	syria	
17	124	is	
18	118	states	
19	115	council	

Word List Results 4

Word Types: 4185 Word Tokens: 16635

Rank	Freq	Word	Lemmas
1	1333	و	
2	578	في	
3	385	من	
4	284	على	
5	238	العربية	
6	174	إلى	
7	133	سوريا	
8	130	التي	
9	123	أن	
10	106	السورية	
11	103	الدول	
12	91	مجلس	
13	88	العربي	
14	74	مع	
15	72	الأمن	
16	72	عن	
17	71	هذه	
18	70	لبنان	
19	66	تنفيذ	

Figure 1: Most Frequent Words in the Two Sub-Corpora

Distribution of the Subcategories of Reference

As gleaned from Table 6, AUNTs exhibited a total of 2222 of four types of reference devices, whereas EUNTs showed a total of 957.

Table 6: Frequencies and Percentages of the Subcategories of Reference

Corpus	**AUN Texts**		**EUN Texts**	
Reference	N° of Occurrences	Percentage %	N° of Occurrences	Percentage %
Pronominals	1934	87.03	586	61.23
Demonstratives	203	09.13	226	23.61
Comparatives	75	03.37	113	11.80
Others	10	00.45	32	03.34
Total	**2222**	**100%**	**957**	**100%**

Detailed analysis showed that pronominals or personal reference (87.03% in AUNTs and 61.23% in EUNTs) was the most prevalent subcategory of reference in the parallel corpus, followed by demonstratives (09.13% in AUNTs and 23.61% in EUNTs) and then comparatives (03.37% in AUNTs and 11.80% in EUNTS). Other types of reference including *blend words* and

numbering references[(2)], which characterise the language of legal texts, came out with very low frequencies (0.45% in AUNTs and, 03.34% in EUNT). While AUNTs depend heavily on pronominals more than EUNTs do, demonstratives and comparatives are found in EUNTs more frequently than in AUNTs. Thus, it can be said that AUNTs are more explicitly cohesive than their English counterparts through the use of pronominals. EUNTs, however, exhibit more emphasis on repetition for the aim of accuracy instead of reference, as we will see later.

Despite the same distribution of the subcategories of reference, significant differences exist in the occurrence frequencies of two types of reference, *Pronominals* and *others*, between AUNTs and EUNTs ($p=0.048<0.05$ and $p=0.037<0.05$ respectively) as shown in table 7 below.

Table 7: T-test Results for the Subcategories of Reference

Reference	Mean	Std. Deviation	t	df	Sig.(2-tailed)
Pronominals	116.53333	208.31908	2.167	14	0.048
Demonstratives	-0.66667	21.06679	-0.123	14	0.904
Comparatives	-2.86667	11.06388	-1.003	14	0.333
Others	-2.06667	3.47371	-2.304	14	0.037

In fact, Arabic has three types of pronouns: independent, enclitic and implicit, in addition to special pronouns for the category of dual. English, however, does not have such a variety of pronouns. That is why pronominals are found considerably more in the Arabic texts. Williams [32] maintained that the high frequency of pronominals in Arabic texts is attributed to the nature of Arabic verbs which contain an implicit pronoun, in addition to their fully inflected nature for number and gender. Pronominals in Arabic, thus, obtain a greater referential significance more than their English counterparts. In this view, it can be said that the abundant use of pronominals in Arabic UN texts is a consequence of the legal texts' deep concern to preserve maximum levels of precision and reduce misunderstanding.

It is important to mention that the use of blend words, such as: *hereby*, *herebelow*, and *herebefore* are very essential to ensure the accuracy of legal texts. (Karakira, [33]). This kind of reference is widely utilised in UN texts. It is through a specific reference to the whole text or to any of its parts that textual cohesion is maintained in UN texts. An extract from the PCUNTs presenting the reference device *herewith* and its concordance list are displayed below:

> I have the honour to transmit **herewith** a letter from H. E. Mr. WalidAl-Moualem, Minister for Foreign Affairs of the Syrian Arab Republic dated14 March 2011, addressed to the Secretary-General and the President of the Security Council, regarding the capture and imprisonment of the two Syrian citizens, Mr. Majed Al Chaer and Mr. Fidaa Chaer from the Occupied Syrian Golan, by the Israeli occupation authorities last year and the recent issuing of unjust judgments against them **(see annex).**

Concordance Hits 5

Hit	KWIC	File
1	I have the honour to transmit herewith a letter from H. E. Mr. '	Nouveau document texte.txt 0 1
2	t, I have the honour to transmit herewith the position of the Syr	Nouveau document texte.txt 0 2
3	t, I have the honour to transmit herewith a letter that represents	Nouveau document texte.txt 0 3
4	e I have the honour to transmit herewith a letter dated 22 Janua	Nouveau document texte.txt 0 4
5	1 I have the honour to transmit herewith a letter dated 2 Februa	Nouveau document texte.txt 0 5

Figure 2: Concordance Sample from PCUNTs for the Reference *herewith*

Another cohesive device that is significant in United Nations texts is the feature of articulation and numbering. Karakira [33] asserted that because legal texts follow rigid methods of dividing documents into parts or paragraphs, references are clearly made throughout the text. Thus, it can be said that this elaborate referencing system within the one text is a strong cohesive feature shared by all legal texts. Accordingly, this type of cohesive reference is significantly found in this parallel corpus. These examples are extracted from the PCUNTs:

> **In resolution 65/18**, the Assembly declared that the Israeli decision of 14 December 1981 to impose its laws, jurisdiction and administration on the occupied Syrian Golan was null and void and had no validity whatsoever, as confirmed by **the Security Council in its resolution 497 (1981)**, and called upon Israel to rescind the decision.

Another example is:

> Arab Republic has also declared in all international forums its full commitment to the relevant international resolutions and has called for their implementation, in particular **Security Council resolutions 242 (1967), 338 (1973) and 497 (1981).**

These two passages explicitly display the use of articulation and numbering in the PCUNTs. The numbers of resolutions are significantly referring to specific and prior resolutions established by the United Nations authorities.

Distribution of Substitution and Ellipsis

Table 8: Frequencies and Percentages of Substitution and Ellipsis

Corpus	**AUN Texts**		**EUN Texts**	
Cohesive Devices	N° of Occurrences	Percentage %	N° of Occurrences	Percentage %
Substitution	141	01.77	156	02.41
Ellipsis	16	00.20	50	00.77

The table showed that substitution (01.77% in AUNTs and 02.41% in EUNTs) and ellipsis (0.20% in AUNTs and 0.77% in EUNTs) are quite infrequent in the two sub-corpora.

The scarcity of these devices in written discourse has been confirmed by many researchers. Williams [32], for example, pointed out that Arabic tends to avoid ellipsis. In this parallel corpus, both substitution and ellipsis are considered to be marginal phenomena in both Arabic and English United Nations documents; however, English tends to use them more frequently than Arabic. Such scarcity of use of these devices is due to the nature of the United Nations texts, which are particularly precise, aiming at achieving exactness of meaning and reducing any possible ambiguity that may affect the information.

Distribution of Lexical Cohesion

Table 9: Frequencies and Percentages of Lexical Cohesion

Corpus	**AUN Texts**		**EUN Texts**	
Lexical Cohesion	N° of Occurrences	Percentage %	N° of Occurrences	Percentage %
Reiteration	3437	88.55	3177	80.75
Collocation	444	11.44	757	19.24
Total	**3881**	**100%**	**3934**	**100%**

As shown in Table 9, reiteration (88.55% in AUNTs and 80.75% in EUNTs) was the dominant device of lexical cohesion in AUNTs and EUNTs in terms of occurrence frequencies, while collocation (11.44% in AUNTs and 19.24% in EUNTs) showed a very low frequency. Significant differences were found in the occurrence frequencies of the two subcategories of lexical cohesion between AUNTs and EUNTs ($p=0.006<0.05$).

Table 10: T-test Results for Lexical Cohesion

Cohesive Devices	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Lexical Cohesion	-26.50000	405.17219	-0.092	1	0.006

The data contained in table 9 above show that lexical reiteration is much more frequent in AUNTs, whereas lexical collocation is employed more frequently in EUNTs. Lexical reiteration contributes significantly to the cohesion of United Nations texts, in the sense that patterns of repetition, synonymy, or general words are included for the sake of emphasis and the guarantee of a consistent flow of ideas. As far as collocation is concerned, though it is much less used than reiteration in the two sub-corpora, it still plays a vital role at the inter-sentential level; it exceeds the boundaries of sentences and even paragraphs in order to guarantee the required accurateness and clarity of texts. Therefore, it can be said that lexical cohesion is achieved partly through reiteration and partly through collocation which are related to the nature of legal texts.

In what follows, repetition, the most prevailing type of lexical reiteration, will be discussed.

Corpus	AUN Texts		EUN Texts	
Lexical Cohesion	N° of Occurrences	Percentage %	N° of Occurrences	Percentage %

Repetition	2931	85.27	2649	83.38

Table 11: Frequencies and Percentages of Repetition

From the table above, it is evident that the repetition patterns are the prevailing lexical cohesive devices used in the parallel corpus. The figures listed in the table show a total of 2931 of repetition in AUNTs, whereas EUNTs show a total of 2649. In other words, AUNTs (85.27%) tend to use repetition more than their English counterparts (83.38%). Significant differences exist in the occurrence frequencies of repetition between AUNTs and EUNTs ($p= 0.035<0.05$).

Table 12: T-test Results for Repetition

Lexical Cohesion	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Repetition	25.46667	42.27777	2.333	14	0.035

Thus, it can be said that lexical repetition is of a high frequency and inevitable in the organisation of UN texts across the two languages. As far as English is concerned, Wright and Hope [34] asserted that lexical explicitness, the most remarkable feature of texts, is contingent on lexical cohesion, particularly repetition rather than reference such as pronominals. Similarly, Arabic tends to favour lexical repetition; Williams [32:126] maintained that "in Arabic 'the same theme' is repeated in 'successive clauses....more frequently than English does, even when it is grammatically possible to omit it'" (quoted in Al-Jaber, [31]). Moreover, Al-Jaber [31] explained that the tendency of Arabic texts to use lexical repetition very frequently is said to be due to the abundant word root which can generate many derivatives. For example, the words[3] تعليم *ta.cli:m*, عالم *.calim*, يعلم *ju.calim* and معلم *mu.calim*, are all derivatives of the word *.cilm* علم. Furthermore, since there is a strong trend towards description and argumentation in United Nations texts, the use of lexical repetition which is ascribed to rhetorical devices, such as assertion and exaggeration aims at persuading readers. Koch [35] examined repetition in argumentative discourse. She wrote: "repetition of form and content yields much cohesion to Arabic texts. Repetition is a means of persuasion in Arabic argumentative discourse." (Koch, [35], quoted in Al-Jaber, [31:167])

To clarify the significance of repetition, an extract from the PCUNTs below shows the term *Syria* reappearing in every sentence in order to put emphasis on 'Syria the Republic'.

إن ما ذكره التقرير حول الجهود **السورية** لتنفيذ أحكام قرار مجلس الأمن ١٥٥٩- ٢٠٠٤ هو اعتراف صريح بأن **سوريا** قد قامت بتنفيذ ما يخصها في هذا القرار، ومن غير المقبول الاستمرار بزج اسم **سوريا** في تقرير الأمين العام حول تنفيذ القرار١٥٥٩- ٢٠٠٤ بالرغم من أن **سوريا** قامت بتنفيذ ما يخصها من أحكامه . فقد أشاد التقرير ليس فقط بإجراء انتخابات رئاسية ونيابية حرة ونزيهة في لبنان أي من غير تدخل أو نفوذ أجنبي، وبجهود **سوريا** حول سحبها لقوّاتها ومعدّاتها العسكرية من لبنان، وإنما أشاد أيضا بإقامة علاقات دبلوماسية كاملة بين **سوريا** ولبنان.

Therefore, it can be said that the semantic stability of lexical repetition, established through the connection of lexical items, is maintained throughout the text. This semantic stability relies on the precise repetition of previously mentioned items occurring at both the intra-sentential and inter-sentential levels. That is why, when repetition is abundant in both Arabic and English UN texts, the aesthetic and stylistic features of language are abandoned for the purpose of transparency and accurateness of meaning.

Summary of the Results

In the light of the above discussion, here is a summary of the obtained results:

1- The distribution of the types of cohesive devices and their subcategories is almost identical in the two languages. In terms of their frequencies, there are significant differences.
2- Considering the argumentative and descriptive nature of legal texts (United Nations texts), lexical cohesion seems to be the most frequently used cohesive device in the two languages.
3- Since legal texts require much specificity and transparency, substitution and ellipsis are considered to be marginal phenomena that may cause undesirable effects of ambiguity and misunderstanding.
4- As for conjunctive cohesion, Arabic texts are found to use additives more frequently than English. The high occurrence of these devices, especially additive و *wa*, are due to the trend of Arabic towards coordination.
5- As for reference, Arabic seems to use a higher proportion of pronominals than English. While Arabic depends heavily on pronominals, demonstratives and contrastive reference are more frequently used in English.
6- Lexical repetition is used more frequently in Arabic than in English. As (Al-Jaber [31]) affirmed, this device contributes largely to the surface connectivity and semantic continuity of Arabic texts which

derive much cohesion from the great redundancy resulting from lexical repetition.

Conclusion

A contrastive analysis of cohesive devices in Arabic and English United Nations texts showed that Arabic and English texts have more similarities than differences in terms of the types of cohesive devices used, whereas differences markedly occur in their frequencies. The similarities are significantly preserved for the purpose of accurateness, transparency and formality that characterise the language of legal texts. However, differences occasionally arise because of the diverse grammatical and lexical structures of Arabic and English; shifts of cohesion, therefore, occur in order to achieve the closest equivalent meaning in the translation product. This paper may help both students of translation and professionals working in law improve their discourse skills and produce more accurate translations. However, it is necessary to mention that this study attempted to shed light on some of the most prevailing patterns of cohesion; lexical collocation, for example, was not covered; hence, further research is required to address this area in more details.

References

1- Baker, M. (1993). Corpus linguistics and translation studies - implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Ed.), *Text and technology: In honour of John Sinclair*. (pp.233-250). Amsterdam: John Benjamins.

2- Newmark, P. (1987). *A textbook of translation*. London: Prentice-Hall International.

3- Baker, M. (1992). *In other words: A coursebook on translation*. London, England: Routledge.

4- Halliday, M.A.K., & R, Hasan. (1976). *Cohesion in English*. London: Longman.

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

6- Hatim, B., & Mason, I. (1990). *Discourse and the translator.* London: Longman.

7- Blum-kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-kulka (Eds.), *Interlingual and intercultural*

communication: Discourse and cognition in translation and second language acquisition studies. (pp.17-35). Tübingen: Gunter Narr Verlag.

8- Halliday, M. A. K. (1994). *An Introduction to functional grammar*. China: Edward Arnold Publisher.

9- De Beaugrande, R.A., & Dressler, W. (2002). *Introduction to text linguistics* .London: Longman. A Digital Printing. Retrieved from: http://beaugrande.com/introduction_to_text_linguistics.htm

10- James, C. (1980). *Contrastive analysis*. *London*: *Longman.*

11- Teubert, W. (2004). Language and corpus linguistics. In M.A.K. Halliday, W.Teubert, C. Yallop & A. Čermáková, *Lexicology and corpus linguistics*. (pp.73-112). London: Continuum.

12- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*. *10* (1), 1-13.

13- Ebeling, J. (1998). Contrastive linguistics, translation, and parallel corpora .*Meta:* Translators' Journal, *43*(4), 602-615.

14- Johansson, S. (1998). On the role of corpora in cross-linguistic research. In S. Johansson & S. Oksefjell (Ed.), *Corpora and cross-linguistic research: Theory, method, and case studies*. (pp.3-24).Amsterdam: Rodopi.

15- Mauranen, A. (1999). Will 'translationese' ruin a contrastive study?. *Languages in Contrast. 2* (2), 161-185.

16- Teubert,W. (1996). Comparable or parallel corpora?. *International Journal of Lexicography. 9* (3), 238-264.

17- Zanettin, F. (2011). Translation and corpus design. *SYNAPS- A Journal of Professional Communication. 26* /2011.

18- Zanettin, F. (2000). Parallel corpora in translation studies: Issues in corpus design and analysis. In M. Olohan (Ed.), *Intercultural faultlines: Research models in translation studies I textual and cognitive Aspects*. (pp.105-118).Manchester: St Jerome.

19- Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Amsterdam: Rodopi.

20- Fernandes, L. (2006). Corpora in translation studies: Revisiting Baker's typology. *Fragmentos. 30*, 087-095.

21- Johansson, S. (2003). Reflections on corpora and their uses in cross-linguistic research. In F., Zanettin, S., Bernardini, & D., Stewart. *Corpora in Translator Education.* (pp.135-144). New York: Routledge.

22- Biel, Ł. (2009). Corpus-based studies of legal language for translation purposes: Methodological and practical potential. In C. Heine & J. Engberg (Eds.) *Reconceptualizing LSP. Online Proceedings of the XVII European LSP Symposium 2009, Aarhus 2010*. (pp1-15).

23- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*. 7(2), 223–243. Amsterdam: John Benjamin's.

24- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

25- Granger,S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual corpora and learner corpora. In Ajmer et al., (Eds) *Languages in Contrast*, 37-51.

26- Lauridsen, K. (1996). Text corpora and contrastive linguistics: Which type of corpus for which type of analysis?. In K. Aijmer, B. Altenberg & M. Johannson (Eds.), *Languages in contrast. Papers from a Symposium on Text-based Cross Linguistic Studies*. (pp.63-72).Lund: Lund University Press.

27- Barlow, M. (1996). Analysing parallel texts with ParaConc. I*n Proceedings from ALLC- ACH, University of Bergen, Norway.* Quoted in P. Danielsson, (2003). Units of meaning in translation — how to make real use of corpus evidence. *Translating and the Computer.* London: Aslib.

28- Anthony, L. (2011). AntConc Build 3.2.4. [Free computer Software Tool]. Centre for English language education in science and engineering. Waseda University: Tokyo. Retrieved February 25, 2014 from http://www.antlab.sci.waseda.ac.jp/software.html

29- Anthony, L. (2013). AntConc Build 3.4.1. [Free Computer Software Tool]. Centre for English language education in science and engineering; Waseda University: Tokyo. Retrieved February 25, 2014 from http://www.antlab.sci.waseda.ac.jp/software.html

30- Dudley-Evans, R., & Swales, J. (1980). *Study modes and students from the Middle East.* ELT documents log. The British Council: London

31- Al-Jaber, A.M. (1987).*Cohesion in text differentiation: A study of English and Arabic*. Doctoral dissertation. University of Aston, Birmingham.

32- Williams, M.P. (1983). A problem of cohesion. In J. Swales & H. Mustafa (Ed.) *English for Specific Purposes*. (pp.118-128). University of Aston, Birmingham.

33- Karakira, S. (1997). *Lexis versus text: The case for translating English legal texts into Arabic*. Master dissertation. The University of Western Sydney, Australia.

34- Wright, L., & Hope, J. (2005). *Stylistics: A practical coursebook*. London: Routledge.

35- Koch, B.J. (1981). *Repetition in discourse: Cohesion and persuasion in Arabic argumentative prose*. Unpublished PhD dissertation. University of Michigan

36- Al-Qahtani, D.M. (2004). *Semantic Valence of Arabic Verbs*. Beirut, Lebanon: Libraire du Liban Publishers.

Appendix

Transliteration Tables

Transcription adopted for this paper of Arabic pronunciation as appears in Al-Qahtani's Semantic Valence of Arabic Verbs [36].

Arabic Alphabet	Symbols	Arabic Alphabet	Symbols
ء	?	ض	0d
ب	b	ط	0t
ت	t	ظ	D0
ث	T	ع	.c
ج	Z	غ	^g

ح	.h	ف	f
خ	X	ق	q
د	d	ك	k
ذ	D	ل	l
ر	r	م	m
ز	z	ن	n
س	s	ه	h
ش	S	و	w
ص	.s	ي	j

Table 1: Consonantal Symbols

Vowels		**Symbols**
Short	◌َ	a
	◌ُ	u
	◌ِ	i
Long	ا	a:
	و	u:
	ي	i:

Table 2: Vocalic Symbols

NOTES

1 - http://documents.un.org/
2- Blend words and numbering references do not appear in Halliday and Hasan's classification [4], but it is very important to mention their occurrence since they are attributed to the language of UN texts.
3 - For the transliteration tables see the appendix.