Translating *Le Monde* and *El País* English Business Loan Terms into Arabic: A Corpus-based Analysis

Mohamed M. Mostafa
Gulf University for Science and Technology, West Mishref, Kuwait

**Abstract**—A major trend in corpus linguistics has been the use of highly specialized corpora to achieve several different purposes, including teaching technical writing, language learning, and teaching translation techniques. This study aims to design and use corpora in order to aid translators master the translation of English business terms used routinely in *Le Monde* and *El País* into Arabic. The *Le Monde* corpus included 456,122 words, while the *El País* corpus included 874,687. Corpora satisfied several rigorous criteria such as authorship, topic, genre and register. Since translators need a variety of reference sources in order to come up with quality translations, concordance, collocation, and word cluster analyses were performed to reveal how different language patterns are used in their usual contexts. We found that corpora might help translators in navigating across various possibilities of terminologies and expressions, which have to be first decoded prior to being re-coded into the target language. Corpora examples also make the target text more successful in cross-cultural communication contexts.

**Index Terms**—corpus linguistics, translation studies, business terms, concordance, collocation, word clusters

I. INTRODUCTION

Computer technology plays a central role in the study of vocabulary, lexicography and translation. In fact as long ago as 1980s, a leading linguist noted that computers ability to store and scan very long texts “provides a close-to-objective basis on which language patterns can be observed” (Sinclair, 1984, p. 3). He also noted that computational linguistics evidence provided by concordances, for example, is quite superior to any other traditional method. Thus, as soon as translators make full use of this evidence, “it will be impossible to go back to a reliance on pre-computational techniques” (Sinclair, 1985, p. 87). In a similar vein, Chapelle (2001, p. 38) argued that computer technology has resulted in what one can describe as a “corpus revolution.” By using computers in extracting and analyzing meaningful data from general and specialized corpora, corpus linguistics can contribute to the field of translation by performing important tasks such as determining collocations, word clusters, sub-categorizations, and validating linguistic hypotheses (Miangah, 2012). Corpus-based discourse analysis can also raise new research questions, remove bias, and identify linguistic norms and outliers (Baker, 2006).

A major trend in corpus linguistics has been the use of highly specialized corpora to achieve several different purposes, including teaching technical writing (Noguchi, 2004), language learning (Miangah, 2012), and teaching translation techniques (Frankenberg-Garcia, 2012). Thus, by using modern computer technology in translation, semantic and linguistic biases may be corrected. For example, if we examine the two verbs “quake” and “quiver”, which belong to the same semantic class “shake” in two widely-used dictionaries such as the COBUILD and the Longman, we find that both dictionaries claim that the two verbs are intransitive. However, the entry for “quake” in the Oxford dictionary is intransitive, while the same dictionary lists “quiver” as a transitive verb. To solve this dilemma, Atkins and Levin (1995) used a very large English corpus comprising more than 50 million words and were able to detect possible uses of both verbs in transitive forms such as “it quaked her bowels” and “…quivering its wings.” Thus, by using a large enough corpus it seems possible to correct for traditional dictionaries’ errors and to provide evidence for accurate entries (McEnery and Wilson, 1996).

Computer-assisted technology may also be used also to detect nuances or shades of meaning. For example, Nelson (2005) used a specialized English corpus to investigate the usage of two terms often used interchangeably: “global” and “international”. The author found that there are clear differences in usage as the word “global” often collocated with terms such as “business activities”, while the word “international” collocated usually with terms such as “companies and institutions.” This contrasted with the word “local” which collocated with a large number of non-business words. The author concluded that while the three words share the same semantic class, the word “local” collocated significantly with a distinctly non-business activity. Probably one of the major contributions of computer technologies to translation studies has been the creation of what is called “open-ended monitor corpus”, which allows translators to keep track of new words entering the language, words changing their usual meaning across time, or the balance of word usage based on factors such as formality, genre, etc. Thus the ability to use computer software to detect collocation and
word clusters/bundles rather than the word at the individual level means that it is now possible to systematically investigate how phrases and words collocate and how such collocations change over time (Castejon, 2012).

This study aims to design and use corpora in order to aid translators master the translation of English business terms used routinely in *Le Monde* and *El País* into Arabic. Although there are several specialized dictionaries in the field of business, the concordance produced through the analysis of an electronic corpus shows the translator how a term or an expression is actually used within a particular context. Hultsjn (1992) argued that, compared to consulting a dictionary, a corpus typically calls for deeper processing which, in turn, enhances the learning process. We believe that by doing this, the corpus designed can be used as a complementary resource that allows translators to make judicious use of certain business terms. Thus, translators might use this corpus to verify or reject the use of particular word meanings given in a dictionary.

II. LITERATURE REVIEW

Corpora and translation studies started to emerge in the 1980s. It is generally acknowledged that Gellerstam’s (1986) study is the first to use a corpus-based translation approach (Zanettin, 2012, p. 12). In Gellerstam’s (1986) seminal work original words distribution of English adverbials were compared to their corresponding translations in Swedish. Laviosa (2002) noted that since the publication of this work, corpus-based translation studies have flourished as a distinct body of research spanning both theoretical and descriptive levels. In a similar vein, Olohan (2004) argued that the use of corpora as a research tool in translation studies has created a new stream of research into the field of translation. In fact several translation aspects have been investigated using corpus-based translation approaches. For example, the translation universals hypothesis, which argues that there are invariant features that characterize all translated texts irrespective of source language and translation direction, has been tested using corpora approaches (e.g., Becher, 2010; Jantunen, 2002). Corpus-based studies have also examined translators’ styles and words choice motivations (Saldanha, 2011). Corpus-based interpretation studies have investigated areas such as hesitation and disfluencies (Straniero-Sergio & Falbo, 2012).

The simplification hypothesis, the assumption that linguistic patterns in a corpus of translations are much simpler than a corpus of comparable texts in the same language, was also extensively investigated. For example, Laviosa (1997, 1998) found that the range of vocabulary or lexical variety and the information load or lexical density are both lower in translations corpora. This study has also been replicated in several languages, including Spanish (Corpas-Pastor, 2008) and Chinese (Xiao, He, & Yue, 2010). The normalization/conventionalization hypothesis, which argues that original corpora include more creative linguistic patterns compared to comparable translated corpora, has also been investigated extensively in the literature. For example, several authors found that translated texts in a corpus tend to conform to target language rather than source language patterns and norms. This results in producing more conventional rather than creative target strings (e.g., Kenny, 2000; Puurtinen, 2003). However, the underrepresentation hypothesis, which argues that unique elements in language tended to be underrepresented in translation, was less investigated. For example, Tirkkonen-Condit (2005) studied the use of the particle “kin” in a corpus of texts translated into Finnish against a comparable corpus of texts written in Finnish across five different genres. The author found that the frequency of the particle “kin” is significantly lower in Finnish translations compared to original texts.

Several studies have recently attempted to detect “translators’ fingerprints” based on corpus analysis. For example, Wang and Li (2012) investigated translator’s style based on a corpus including Chinese translations of Joyce’s *Ulysses*. The authors argued that translators normally leave some traces of lexical idiosyncrasies that may be detected by analyzing translation corpora. In a similar vein, Li, Zhang and Liu (2011) studied different corpora representing English translations of *HongLouMeng*—one of the greatest classical Chinese novels. The authors were able to detect stylistic differences among different translators of the same novel. These differences were attributed to differences in translators’ socio-political, cultural and ideological perspectives. Rybicki and Heydel (2013) used stylometry and stylistics methods based on most-frequent-word frequencies to identify different translators of the same text. The authors used a corpus of Polish translations of English novels to determine the point (the chapter) in which one translator took over from the other. Based on a corpus including 144 texts of 133, 159 words representing letters written by Vincent van Gogh and by his brother Theo, Forsyth and Lam (Forthcoming) investigated whether authorial discriminability is preserved by translators when translating the letters from the original French into English. The authors found that much of the stylistic authenticity between the two brothers was preserved in the English translations. A similar study has recently investigated author discriminability between the Holy Quran and prophet’s *hadith*/statements. A computer-assisted corpus analysis revealed that the two texts cannot be written/translated by the same “author” (Sayoud, 2012).

From this brief review we find that although numerous studies have attempted to assess the role played by corpora in translation studies, virtually no studies have focused on examining corpus-based translation into Arabic. Notable exceptions include Alharbi and Swales (2011) who used corpora to investigate English and Arabic translations of abstracts in bilingual academic journals and Merakshi and Rogers (2013) who used a corpus of Arabic and English texts of Scientific American to investigate the translation of culturally-bound metaphors into Arabic. In this study we aim to fill this literature gap by empirically designing corpora that might help Arabic translators master the translation of English business terms used routinely in *Le Monde* and *El País*. © 2015 ACADEMY PUBLICATION
III. METHODOLOGY

A. Corpus

Since the major task of any translator is to develop a true mapping of words, a balanced corpus is nowadays essential in translation studies (Cermak, 2010). Following Le Poder’s (2012) study, we investigated a corpus of English business terms appearing both in the French newspaper Le Monde and the Spanish newspaper El País from January 2007 to December 2010. We selected a random sample of 20 terms listed in the author’s study and built our corpus around them. Table 1 shows a summary of corpora used in this study. It should be noted that corpora created for the purpose of investigating the translation process, generally known in the literature as “translation-driven corpora, can be monolingual or bilingual as well as comparable or parallel (Zanettin, 2012).

The vast amount of electronic texts available on the newspapers’ web sites facilitated the creation of our corpora in a relatively short time. The Le Monde corpus included 456, 122 words, while the El País corpus included 874, 687. Corpora satisfied several rigorous criteria such as authorship, topic, genre and register (Biber, 1993). Thematically, the articles included in the corpora included areas as diverse as markets, enterprises, stock exchanges, consumption, employment, public accounts, etc. It should be noted that both the French newspaper Le Monde and the Spanish newspaper El País are characterized by frequent contributions from economists who write in a sophisticated, yet accessible language for non-specialists. We used a newspaper-based corpus since it is argued that, compared to other translation modalities, journalistic translation has changed the most by the advent of the Internet (Bielsa and Bassnett, 2009). In fact such corpora have been used in several studies, including Lobanova, Kleij and Spenader (2010) who investigated the definition of antonyms using three Dutch daily newspapers and Jones (2002) who used the British newspaper The Independent to build a corpus of 55,411 sentences to investigate the functions of synonyms and antonyms in context. Tse (2003) used a corpus including 514, 691 words based on three British newspapers, namely The Independent, the Guardian and The Daily Telegraph, to investigate theoretical factors determining the use or omission of the definite article preceding multi-word organization names. Similarly, Baroni and Bernardini (2006) used a two-million-word Italian newspapers’ corpus to study the language properties of the “translationese”.

B. Procedures

Having compiled the corpora in French and Spanish, we turned to preparing the texts for analysis. The texts original html format was changed into plain text format compatible with the text analysis software packages used. Following Le Poder (2012), the Chip Cache Plus software was initially used to eliminate photos, graphs and tables. This is a necessary step that must be conducted before using the text analysis packages. To conduct the quantitative analysis part of this study we used AntConc 3.3.5 (Anthony, 2012), QDA Miner 4.1.3 (Provalis Research, 2013), the Signature Stylometric System (Millican, 2003), and the R Stylo package 0.5.2 (Eder, Rybicki, & Kestemont, 2013). These software packages were used for coding and analyzing textual data. The packages were selected because of their extensive tools that can be used to identify hidden patterns in textual data because they include powerful concordance and frequency generators. Cluster of words and lexical bunde analyses can also be performed using these packages.

IV. RESULTS

A. Quantitative Data Analysis

Statistical analyses for business English terms in corpora were conducted using both R and Excel Data Analysis software packages. Table 2 presents descriptive statistics results for both the Le Monde and El País corpora. From this table we see that the mean for English business terms used in the El País corpus is much higher than the mean in the Le Monde corpus. This is to be expected given well-known French tendency to preserve the French language. However, a t-test reveals no significant differences between corpora. In fact we found a strong positive correlation between both corpora (r = 0.88, p < 0.001). Finally, we conducted both a sentence length and a word length analysis for both corpora. For example, from figure 1 we see that the majority of sentences in the Le Monde corpus range from 10 to 20 words.
Table 2. Descriptive statistics of business English terms in corpora

<table>
<thead>
<tr>
<th></th>
<th>Le Monde</th>
<th>El País</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>3.00</td>
<td>1.00</td>
</tr>
<tr>
<td>First Quartile</td>
<td>7.75</td>
<td>7.25</td>
</tr>
<tr>
<td>Median</td>
<td>17.00</td>
<td>43.50</td>
</tr>
<tr>
<td>Mean</td>
<td>76.35</td>
<td>144.10</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>41.75</td>
<td>81.00</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>154.17</td>
<td>260.36</td>
</tr>
<tr>
<td>Maximum</td>
<td>530.00</td>
<td>974.00</td>
</tr>
</tbody>
</table>

Figure 1. The Le Monde Corpus sentence lengths. X-axis represents the number of words in a sentence and Y-axis represents frequency. The figure shows more or less a Gaussian distribution with some oscillations.

B. Frequency and Alphabetical Lists

A useful first step in analyzing a corpus is to generate a word list. Analyzing frequency of appearance or simply the incremental count of appearance of particular words or phrases might provide insights into a particular topic. In fact O’Leary (2011) argues that despite the simplicity of such approach, it can be used to predict characteristics of the topic analyzed. In a similar vein, Barlow (2004, p. 207) argued that generating word lists represent “the most radical transformation of a text used in linguistic analysis.” Word lists generated are very useful since they highlight which words occur frequently in the corpus and, therefore, merit further investigation.

Three English business terms dominate the distribution of the twenty terms selected [Subprime(s), Hedge fund(s), and Swap(s)]. Such terms might be characterized by different meanings depending on the context in which they are used. An implication here for translators might be to consult several examples to clarify the different patterns of the use. It should be noted that word frequency lists are generally much more useful than alphabetical order lists since the latter results generally in producing a list of function words (such as le, l’a, les) that do not really tell us much about the essence of the corpus. However, a closer look at such function words may lead in some cases to the discovery of some interesting patterns (Romer & Wulff, 2010). To gain more insights into the word lists, the generated list should be compared to a general reference corpus that is usually much larger and more general. Such comparison results in determining “keyness” values. A high keyness value is given if a word occurs significantly more in a selected corpus than it would be expected to occur in the reference corpus. We did not do this comparison since we focus only on the English terms used within French and Spanish specialized corpora. From the word lists a type to token ratio may be computed in order to determine the lexical variety in the corpus against a general corpus. However, Kenny (2001) argued that this index is extremely sensitive to text and corpus length.

C. Concordances

A concordance represents accurately and explicitly the different language patterns used in usual contexts. Barnbrook (1996, p. 65) argued that the major aim of a concordance is to “place each word back in its original context, so that the details of its use and behavior can be properly examined”. KWIK, or key word in context, is usually the format used to present a specific term. Figure 2 represents three examples of concordances produced for the terms “stress test”, “double dib” and “outsourcing”. As can be seen from this figure, the search word is displayed in the middle of the screen, while the context is displayed left and right of it. The use of computer software saves a lot of time going back and forth across the corpus to determine the relevant context of a particular term. The search word is also known as the “node” and the concordance is read vertically not horizontally. The phraseological patterns and the contextualized meaning(s) of a particular word can be determined by looking at the context words to the left and to the right of the node. From this figure we see that the term “subprime” is usually associated with the word “mortgage” and translated into French as “crédit hypothécaire” or “credit immobilier” since these two words are the most frequently associated with the node “subprime(s)”. The term may be translated into Arabic as “الرهن العقاري”. Thus, such a piece of information is in fact very important since it provides us with a great insight into (1) the syntactic contexts in which the
node occurs, (2) the semantic properties of the node’s syntactic companions, and (3) the membership of the node in classes of semantically similar words (Atkins, Fillmore, & Johnson, 2003). Figure 6 shows concordance for three more terms. Similar conclusions may be obtained from studying examples in this figure.

<table>
<thead>
<tr>
<th>Node</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress Test KWIK</td>
<td>65</td>
</tr>
<tr>
<td>Double dip KWIK</td>
<td></td>
</tr>
<tr>
<td>Outsourcing KWIK</td>
<td></td>
</tr>
</tbody>
</table>

Concordance analysis reveals not only what words combine with other words but also where they occur in a text and how they are distributed across different texts in a specific corpus. This type of analysis reveals whether a particular term has a preference to occur at the beginning, at the end, or with no specific preference in a specific text. This analysis is conducted via a concordance plot. Since concordances are normalized, each bar in a certain barcode represents an occurrence of a term in a text. For example, the term “stress test” in this figure seems not to be evenly distributed as there is high concentration around the middle and no occurrences towards the end.

Thus a concordance plot may be a useful tool to show not only “how often a search term appears in a corpus of data, but also where and in what distribution” (Anthony, 2004).
Extracting concordances leads to the next logical step which is analyzing co-occurred words to the left and to the right of the node or the target word. This helps in identifying collocations or word associations. This is an important step since the translator is essentially a text producer and needs to know how words are used and how words associate with other words (Tagnin & Teixeira, 2012). Das (2012, p. 325) defines a collocation as “a set of words occurring together more often than by chance in a corpus.” Although this term is defined in the literature in a variety of ways, the common denominator refers to some sort of “syntagmatic relationship between words” (Walker, 2011). The contextual span for the search or the window of words both sides of the node is determined by the researcher. However, based on a paper by Sinclair and Jones (1974), a span of four words on either side is commonly used in the literature (e.g., O’Halloran, 2007). Frequency or statistical significance approaches may be used to rank collocations for different positions within a cut-off span. We found from the mutual information/t-score analysis that words such as “test”, “tests”, and “fâmeux” collocate significantly with the term “stress test”. Any collocate with a t-score > 1.96 is statistically significant at the alpha level 0.05. Thus the combination of such significant terms with the node or target word is not the result of chance. An alternative way to detect collocations is known as the n-grams, where n usually varies between two and five words (Anthony, 2004). This method is also known as the “lexical bundles” method (Biber, Conrad, & Cortes, 2004). These bundles represent frequently juxtaposed items “which do not usually constitute semantic units in themselves” (Peters, Jones, Smith, Winchester-Seeto, Middledorp, & Petocz, 2006). We found that the most frequently occurring three-word lexical bundles are “milliards de dollars”, “milliard d’euros”, and “aux Etats Unis.” Collocations and lexical bundles may also be visualized graphically using proximity plots. Figures 8 and 9 show proximity plots for the terms “subprime” and “swap.swaps”, respectively.

For example, We found that the English business term “subprime” is associated with French terms such as “credits”, “risqué”, “hypothécaire”, and “immobilier”. Thus, this term may be translated as “crédit hypothécaire” or “credit immobilier” or as “la ren se ar set” in Arabic.

We used multidimensional scaling (MDS) to investigate the relationships in our corpora. MDS is a statistical technique used in order to visually detect complex patterns in high-dimensional datasets (Borg & Groenen, 1997). We found that the sub-corpora representing the terms “High Frequency Trading” and “bluechips” lie far away from the bulk of the data and might be regarded as outliers. This might be the result of very technical terms used in such sub-corpora since it is well-known that the more specialized the text, the greater the difference it exhibits in terms of terminological sources (Tagnin & Teixeira, 2012). For such texts the meaning might be inferred from the observation of several authentic examples in the corpus since in such cases the mere use of correct terminology does not guarantee in itself a good translation (Byrne, 2006).

Finally, to cluster sub-corpora, we used the Ward’s method which produces a set clusters based on proximity of sub-corpora. This allowed us to detect which sub-corpora were the most similar. Finally, we conducted a Delta-normalized bootstrapped cluster analysis (Hoover, 2004) to produce a consensus tree showing distances between texts included in the Le Monde sub-corpora. A bootstrap consensus tree was constructed based on the similarity between sequences of most-frequent-word frequencies (MFW) in each sub-corpus. Bootstrapping has shown to alleviate several problems attributed to the original Delta-normalized method (Smith & Aldridge, 2011; Burrows, 2002). Following Rybicki and Heydel (2013), MFW varied between 100-1000, culling set at 100% (only words appearing in each sub-corpus are used in the analysis), and personal pronouns are excluded in order to avoid possible false attributions. We found that texts such as “subprime” and “stress test” are grouped together. Similarly, texts such as “hedge funds” and “swap” are grouped together. These texts might include similar terminologies and thus the translator might use the same translation strategy in translating them.

V. CORPUS-BASED TRANSLATION EXAMPLES

One of the most important characteristics of corpus-based studies is the large number of authentic examples generated. Several authors have found that corpus-based examples help with language comprehension and translation (e.g., Laufer, 1993; Frankenberg-Garcia, 2012). This is because corpus-based examples generally provide added value to the core meaning expressed by traditional dictionaries (Al-Ajmi, 2008). In this section we present some examples extracted from the Le Monde and El País corpora.

A. Le Monde Corpus

« Et si le scénario cauchemardesque d’un “double dip”, une double récession, aux Etats-Unis semble écarté par les investisseurs, depuis un peu plus d’une semaine, ces derniers ont trouvé d’autres motifs de stress. »

« Jean-Claude Trichet, président de la Banque centrale européenne (BCE), a révisé à la hausse, jeudi 2 septembre, les prévisions de croissance de la zone euro en 2010 et 2011, rejetant le spectre d’un “double dip”, d’une double récession. »

© 2015 ACADEMY PUBLICATION
« De fait, les grands dangers qui planaient sur l'économie mondiale ne se sont pas concrétisés : il n'y a pas eu de "double dip" aux États-Unis - c'est-à-dire de double récession - , pas (encore) d'« éclatement de la zone euro » ni d'« atterrissage violent de l'économie chinoise ».

« Le Wall Street Journal affirme, mardi 7 septembre, que les résultats des tests de résistance "stress tests" - des banques européennes "ont sous-estimé les montants de titres de dette publique potentiellement risqués détenus par certains établissements".

« La Commission de régulation bancaire chinoise a ordonné aux établissements de crédit du pays de réaliser des "stress tests" pour évaluer leur résistance à une diminution de moitié des prix de l'immobilier de logement dans les plus grandes villes. »

« La crise financière provoquée cet été par la chute du crédit hypothécaire à risque américain (subprime) n'est pas finie. »

« L'entreprise reste une étonnante poule aux œufs d'or : au deuxième trimestre, elle a dégagé un cash-flow (flux de trésorerie) opérationnel de près de 9 milliards de dollars. »


« En général, les fonds souverains choisissent de placer leur argent dans les "blue chips" - valeurs de grandes entreprises jugées peu risquées - . »

« A l'inverse des fonds d'investissement traditionnels, les feeder funds ne gèrent pas de portefeuille diversifié. Il s'agit d'intermédiaires qui déléguent la gestion des fonds collectés auprès des grosses fortunes et d'institutions à des conseillers en investissements. »

« Notamment à cause d'un fort recours à l'"outsourcing", en d'autres termes l'externalisation d'une partie de l'industrie vers l'Europe de l'Est et les pays émergents. »

B. El País corpus

“La inquietud sobre el excesivo poder de los fondos especulativos, conocidos como hedge funds, fue uno de los temas estrella de la reunión de los ministros de Economía de los Veintisiete celebrada ayer en Berlín.” (30 April 2007).
“Entonces descubrimos la existencia de un segundo sistema financiero que obtiene beneficios de miles de millones de dólares para los directivos de los hedge funds y también para los grandes bancos.” (26 September 2010).

“El éxito de los «hedge funds» (fondos de alto riesgo) se basa en gran parte en la desregulación de que gozan, subrayó el vicepresidente, quien indicó que, «para no matar la gallina de los huevos de oro»…” (21 April 2007).

“Y sigue creciendo, la presión por mantener la estrategia de superar los limites de la crisis, porque todos sabemos que hay una ley que dice que si no se supera la crisis, se desvanece.” (26 September 2010).

“El Banco de España hace meses que está realizando exámenes exhaustivos a las tripas financieras de las cajas, no muy alejados de los llamados stress test, que la Reserva Federal realiza a los mayores bancos de Estados Unidos.” (18 September 2009).

“Para romper esta situación, el Gobierno español pidió a la UE que acelerase la publicación de las pruebas de resistencia o stress test de la banca.” (4 July 2010).

“Fueron incapaces de detectar la enfermedad de las subprime y otorgaron la máxima calificación a entidades que se desplomaron con el huracán financiero.” (18 September 2010).

“... también quiere realizar una auditoría limitada o due diligence que duraría unas tres semanas desde la puesta a su disposición de la documentación requerida.” (18 March 2007).

“Juan Ignacio Crespo, de Thomson Reuters, asegura que eso no supondrá una recaída en la recesión -la famosa double dip- al menos por ahora.” (29 August 2010).

VI. CONCLUSIONS AND IMPLICATIONS

Computer technology plays a central role in the study of vocabulary and translation. By using computers in extracting and analyzing meaningful data from general and specialized corpora, corpus linguistics can contribute to the field of translation by performing important tasks such as determining collocations, word clusters, sub-categorizations, and validating linguistic hypotheses. In fact corpora play a major role in every aspect of translation studies, including theoretical, descriptive, contrastive, and teaching (Laviosa, 2011). In this study we used French and Spanish corpora of Le Monde and El País daily newspapers to investigate English business terms translations. Our corpora satisfied several rigorous criteria such as authorship, topic, genre and register. We argue that using these corpora might help translators understand the context in which English business terminology is used. Specifically, we claim that our corpus-based analysis might help in (1) building better dictionary entries regarding the English business terms used in Le Monde and El País daily newspapers, (2) compensate for the poor representation of such terms in a traditional dictionary, and (3) obtain context-specific information about syntactic and semantic usage of the terms used in Le Monde and El País corpora.

In fact the corpora used in this study might be exploited by translators in several ways. For example, translators can refer to the target corpus for finding information related to collocations. In fact adjectives that collocate with nouns have been proven to be very useful in understanding the context (e.g. Miangah, 2012). Translators also can use corpora in order to verify selection of a particular word or phrase based on other tools. Thus, corpora can be used to find out the
most appropriate equivalent of specific terms in target language for which other translation tools such as dictionaries suggest unsuitable translations. For example, Boulton (2012) presents an excellent example showing the limitations of traditional dictionaries compared to corpora. The French sentence « Je suis paralysé entre le brulot et la chanson d’amour. » The author found that with careless usage of dictionary, “le brulot” may be translated as “fire ship”, “pamphlet”, or “gnat”. However, based on a large corpus, the author shows that the most appropriate translation is “revolutionary”, “rebel”, or “protest”. This is a type of a song for which the French singer Renaud is famous for. All these possible translations are found in the corpus used.

Thus, since the translator is basically a text producer, he or she should be able to determine how words are used and how they associate with other words in a specific context. In fact Wright (1993, p. 70) emphasizes this fact by arguing that “documents must speak ‘the language’ of the target audience and should resemble other texts produced within that particular language community and subject domain. These considerations frequently require that translators move beyond merely correct strategies in terms of lexical and grammatical content in order to account for stylistically appropriate solutions.”

REFERENCES

Mohamed M. Mostafa has received a PhD in Business and Marketing from the University of Manchester, UK and a PhD in International Relations from the University of St. Gallen, Switzerland. He has also earned a MS in Applied Statistics from the University of Northern Colorado, USA, a MA in Social Science Data Analysis from Essex University, UK, a MSc in Functional Neuroimaging from Brunel University, UK, a MA in Translation Studies from the University of the Port Said University, Egypt. He was employed at universities in the USA, Portugal, Egypt, Cyprus, Turkey, France, Jordan, United Arab Emirates, Bahrain and Kuwait.