British Journal of Translation, Linguistics and Literature (BJTLL)
Homepage:https://journals.ukbrighthorizons.co.uk/index.php/bj/about

Bilingual Dictionaries: From Theory to Computerization

Sherif A.Sattar Okasha
Ph.D. candidate in translation studies, University of Birmingham, UK
Corresponding email: okashacat@gmail.com

ARTICLE DATA

Received: 30 Sept. 2021
Accepted: 18 Nov. 2021
Volume: 1
Issue: FALL
DOI:

KEYWORDS
lexicography, bilingual dictionaries, semantic concordance, parallel corpus

ABSTRACT

This paper suggests a computationally-enhanced model of an English –Arabic dictionary based on a systematically empirical linguistic analysis of the source language and target language systems in contradistinction to the introspective intuitions of bilingual lexicographers. In this model, computerized text corpora and bilingual semantic concordances play a key role in turning out a reliable bilingual dictionary that does not only serve the purposes of all types of Bilingual Dictionary users but will also be a robust bilingual repertoire in bilingual Natural Language Processing systems such as rule-based Machine Translation.

1. Introduction

Notwithstanding the great advances in the fields of lexical semantics and computational lexicology, bilingual lexicography (BL) is still a far cry from being a scientific discipline per se. Bilingual comparative analysis of the source language and the target language has not yet built itself into the toolkit of the bilingual lexicographer. Computerization as far as bilingual lexicography is concerned is still restricted to such surface-level automation as can be sufficient to transform a book dictionary into a computerized form. This attitude is definitely oblivious to what potentialities artificial intelligence and smart computation can have for updating the linguistic content of bilingual dictionaries beyond what mere CD –Rom churning can. On the other hand, linguistic theories on bilingual lexicography have been governed-somewhat unconsciously-by commercial considerations. Still in the literature on bilingual dictionaries we can read something about the "purpose" of the dictionary and whether it is targeted for production of the TL by SL users or comprehension of an SL by certain TL users, depending on the direction of the SL-TL pair. This view has always governed such critical issues as sense discrimination in both the source language and target language, rendering the need for semantic disambiguation in a bilingual dictionary (BD) subject to the pre-determined purpose of the dictionary. This paper tries to expose the shortcomings of this view, adopting a different theoretical position which sees the unity of purpose as the basis of building the architecture of the bilingual dictionary so that it becomes suited to the needs of all users, be they average users ,specialized ones ,language learners or translators and be they native speakers of the source language or the target language. At the same time, it would be fair to argue that that eclectic view of the bilingual dictionary can be attributed to the limited space made available in paper dictionaries. However, such an argument, one can contend, is no longer valid once we have adopted full-fledged computerization- with its immense potential for storage and retrieval of large chunks of data- as an irrecoverable substitute for the paper dictionaries. In this way, an integrated bilingual dictionary which unifies purpose and content may well come into existence.

1.1 Theoretical Framework

Bilingual contrastive analysis can be done in two stages which can also be regarded as two paradigms for this type of analysis. The first stage is the preparatory stage, which involves a thoroughgoing comparative and contrastive analysis of the two language systems and the relative position syntactic categories occupy in both of them before embarking on the compilation work. The second stage is the compilation stage in which contrastive analysis focuses on the lexical transfer part of the process. In this part, the lexicographer will select lexicographical equivalents for SL words from a repertoire of translational equivalents provided by bilingual text corpora. Most existing BDs reflect a level of contrastive analysis based on either of the two stages just mentioned. This is why BD theorists classify bilingual dictionaries correspondingly into two broad categories which reflect either one or the other of these two paradigmatic stages. These two categories are: the segmental BD and the idiomatic BD (Piotrowski: 1994.p.148). A segmental BD contains decontextualized lexemic equivalents which are supposed to be substitutional forms to be used by bilingually competent users such as translators. An idiomatic dictionary contains highly contextualized lexemic equivalents together with preconstructed expressions. Thus, it is best suited for production of SL texts by non-native speakers of the TL or for comprehension of L1 by L1 learners when they are native speakers of L2. It can be suited also for communication based on comprehension by, say ,tourists or businessmen, but not so much for translation. This is because translators need ready lexemic equivalents which they can substitute for the source words in the target text at hand rather than idiomatic paraphrases since they are supposed to be already aware of the semantic subtleties of both languages.

It seems, then, that segmental BDs are the most suitable for the purposes of translators. However, segmental BDs usually contain lexical equivalents which serve as contrastive lexical components in the TL system rather than "real" translational equivalents that can be substituted automatically for SL words. For example, all known French-English dictionaries supply the quantitative adjective some as a direct equivalent of de, despite the fact that a corpus-based statistical study conducted by Catford (1965) had found that the actual translational equivalent of de in English is (0), that is, it is not translated. Yet this equivalent was motivated by a belief that the two words occupy the same position in their respective language systems. At any rate, there is certainly a difference between using translation as a paradigm against which we model our BD and considering it the be-all and end-all target of the BD.

To use translation as a paradigm in BL is to consider it as a tertium comparation, that is a third model against which the other two approaches of the segmental and idiomatic BD are compared with a view to integrating them into a single approach. This segmental-idiomatic approach assumes that translational equivalents can be included in a BD as lexicographical equivalents if they follow a regular pattern of occurrence. The pattern should be so regular that translation equivalents can be reduced to a definite and at the same time variegated number of lexicographical equivalents which represent this pattern in a balanced manner. At the same time, they are to be excluded by the lexicographer when they are irreducibly irregular or infrequent and randomly dispersed in TL stretches of discourse. We cannot hope that the successful lexicographical equivalents will be fit for substituting SL words in all relevant contexts, but we can expect them to be so for the greatest number of contexts in which SW is likely to occur. It should be also noted that this substitutability presupposes an unchanged SW status on the morphological and syntactical levels and that any change at these levels may affect this substitutability so that the one remaining constant will be: meaning.

This integrational approach cannot be fully realized in a paper dictionary because in such a case translational equivalent will stand as segmental equivalents which, due to considerations of space, will not be accompanied by a representative variety of expressions in which they occur in TL texts and thus will serve only one purpose, that of translation. However, in a corpus-based bilingual dictionary, these expressions will serve two purposes: to show the validity of the translational equivalents as lexicographical equivalents, account for their diversity and to be explanatory examples for unsophisticated users. The electronic dictionary seems to be the optimum solution for implementing this view. It should be noted that this solution cannot come out in the form of an automatic acquisition of the lexicographical equivalence data provided by the lexicographer as an output from the compilation stage but rather in the form of this data linked to the natural contexts from which the equivalents were derived. This will require building bilingual semantic concordances, a possibility which will be discussed in section v.

1.2 Translation V.s Lexical Transfer

Before we start, a certain stumbling block has to be removed which has often stood in the way of compiling a bilingual dictionary based on a sound linguistic basis: that is lexicographers' inattention to the difference between lexical transfer involved in translation and that involved in bilingual dictionary-making. Bilingual dictionaries may go to extremes in stating what should remain implied, which results in an explanatory equivalent rather than a lexical one. Such a kind of equivalent will soon prove to be a fiasco once we encounter the SL word in a different context than that which the lexicographer had in mind while lexically transferring it into the TL. For example, the English noun abortionism is translated by Al-Nafees English-Arabic dictionary as تأييد حرية الإجهاض (literally: supporting the freedom of abortion). When this noun occurs in a sentence like: The US supports abortionism, it becomes easy to see how erratic such an equivalent is, due to the lexical tautology it causes when we use it in translating this sentence into Arabic. It transpires that the more terminological equivalent حرية الإجهاض "freedom of abortion" is the proper one, for it serves both the purpose of comprehension and that of production and would cover a wider spectrum of the contextual occurrence of the SL term than the explanatory equivalent.

2. Contrastive Semantic Analysis:

2.1 Polysemy in the source language

Perhaps the most important challenge for a bilingual dictionary user, be he a reader or a translator of a text written in the source language, is to figure out the meaning of the lexical unit for which he seeks a lexical equivalent from between the lines of the source language text. The next step is to spot the nearest equivalent to that meaning from the "map" of lexical equivalents listed by bilingual dictionaries for that lexical unit. If the reader or translator is already familiar with all the senses of the source word, he will not make a hard job of "recognizing" the proper TL equivalent as he goes through his bilingual checklist. Otherwise, the practiced user, say a translator or a specialized reader, will perhaps first resort to a SL monolingual dictionary, in order to compare the different meanings listed under the entry for the SL word with the contextualized lexical unit, as it occurs in the source text at hand, till he settles on a satisfactory sense mapping. Then he may consult a bilingual dictionary in search of an exact TL equivalent. As for the language learner or the general user, they may well dispense with the SL dictionary intermediation simply by browsing all the lexical equivalents catalogued by the bilingual dictionary for the source word. The browsing will continue till they find an approximation which they think is the closest thing to the meaning of the source word in the given text, which is an even harder task.

It is our contention that the bilingual dictionary should reduce these steps to a minimum and save its users all this trouble by stating the various meanings of the source language word. Most bilingual dictionary theorists argue that the bilingual dictionary should not state the different meanings of the SL polysemous word unless there is semantic ambiguity in both languages. That is, when there is a polysemous target word for each meaning of a polysemous source word.

The problem with such views is that they restrict comprehension and production to the limited area of temporary users such as language learners and general readers. What about advanced bilingual dictionary users like translators and academic writers? A translator, for example, would want to use the dictionary for comprehension and production at the same moment: comprehension of the SL and production of the TL. Therefore, he would like to have a well-defined SL meaning linked to an accurate TL equivalent, regardless of whether he is a native or non-native speaker of the source language, and to the elimination of SL dictionary intermediation.

There are two models of the monolingual lexicon which the bilingual lexicographer can choose from when he sets about the task of incorporating the SL meanings into his dictionary. These two models are: the sense enumerative lexicon and the generative lexicon. The former assumes that a multi-sense word has a definite number of meanings which may be unified under one sense spectrum, a phenomenon which lexical semanticists call polysemy, or they may not be unified by the same sense spectrum, a phenomenon traditionally known as homonymy. A prototypical example of polysemy is that of the noun bank, which could mean a 'financial institution' or the 'building' used by that institution. The same word can also provide us with a typical example of homonymy when it means 'side of a river', a meaning which has nothing to do with the previous ones. As for the generative lexicon, it rejects the idea of a word having a pre-determined set of meanings on grounds that word meaning is affected by the context, the linguistic and the non-linguistic one, and is constantly subject to change in such a way that the sense enumerative lexicon cannot track.

Thus meaning, according to this model, is generated from usage. Let's take the example of an adjective like fast. According to the sense enumerative lexicon, three sense spectrums can be tracked of this word within which any subsequent usage of it has to be understood. The word fast may indicate the speed of an event or an action as in fast trip, or it can indicate the speed of an object when it is the initiator of the speed as in fast runner and fast car. Finally, it can indicate the speed of an object when this object, which is expressed by the noun the adjective qualifies, is the product of the speed rather than the producer or initiator of it as in fast meal. When an expression like fast road occurs, it is automatically mapped, according to the sense enumerative lexicon, to the second meaning. This will be rejected by the generative lexicon model on the grounds that what is being described as "fast" here is not the road, but, rather, the cars speeding on it, which is a new meaning generated from the context and other meanings can be generated from other contexts if we have a reliable corpus.

In order for the generative lexicon model to be implemented in a bilingual dictionary, this will require computerized bilingual text corpora where SL meanings are generated from the contextual co-occurrences of SWs and then mapped to their TL equivalents. The computational paradigm can provide us with a means to integrate the two models of the generative and sense enumerative lexicons. This comes about by extending the repertoire of the sense enumerative lexicon beyond a finite list through comparing the already given meanings against corpus sense-in-text and generating new meanings to be constantly added to the list of meanings.

2.2 Lexical Equivalents in the Target Language

One can argue that bilingual dictionary theories focus mainly on word-to-word equivalence and sense–to-word equivalence and don't give due attention to meaning-to-meaning equivalence. Before carrying the discussion, a step further, I would first like to make clear what I mean by these three terms. Word-to-word equivalence is the simplest form of lexical equivalence; it exists when there is a monosemous source word mapped to a monosemous target word. Sense-to-word equivalence occurs when there is a polysemous source word for each meaning of which there is a separate lexical item in the TL lexicon, which does not intersect semantically with it except in respect of that meaning. In other words, the target word in such a case could be monosemous or polysemous. If it is monosemous, there will naturally be semantic equivalence between it and the particular SW meaning for which it was selected. If it is polysemous, the semantic equivalence will hold only between one of its meanings and the meaning of the SW for which it was selected, while other SW meanings will be covered by other, different TWs and so on.

Meaning-to-meaning equivalence, on the other hand, occurs when all the senses of a SW can be mapped to all the senses of a TW without need to go to different TWs to translate the different SW senses. From now onwards I will give a lexical equivalent resulting from meaning-to-meaning equivalence the term semantic equivalent while a lexical equivalent resulting from sense-to-word equivalence, or word-to-word equivalence will be assigned the term lexical-word equivalent.

2.2.1. Semantic Equivalents

A semantic equivalent in the sense just defined could be isomorphic or non-isomorphic, depending on the degree to which the meanings of both the source word and the target word are identical. An isomorphic semantic equivalent occurs when there is a source word which has a certain number of senses or semantic extensions, linked by the same semantic spectrum, and a corresponding target word, having the same number of senses and the same collocational range. Therefore, the TW is said to represent an isomorphic semantic equivalent of the SW if (1) the meanings of the TW are linked by the same semantic spectrum as that whereby the SW meanings are linked; (2) the TW is valid as a lexical equivalent of the SW in all of the latter's contextual co-occurrences (i.e. its immediate collocational range, which the lexicographer discovers through a thorough-going corpus investigation of the word). In such a case, the lexicographer, and often the translator as well, will not need, as we have noted, to go to a separate lexical item in the target language lexicon for each meaning of the source word and will use the same isomorphic TW for all meanings. For example, the English verb collapse has three meanings linked by the semantic spectrum of “falling down". This “falling down” could be literal, figurative or psychological, as illustrated below by 1. (a),(b) and (c) respectively:

1. (a) The building collapsed

(b) Negotiations collapsed

(c) The man collapsed

It is to be observed that the Arabic verb ينهار (collapse) has the same three meanings of the English verb and in this way, there will be no need to use a lexical-word equivalent pertaining to a different semantic spectrum or an explanatory equivalent which, in addition to being lexically clumsy, does not communicate the SW meaning precisely, as we find in Al-Mawrid English-Arabic dictionary. In this dictionary, we encounter the Arabic verb يخفق (fail), which means: to fail, as the equivalent of the second sense of collapse. For the third sense, the dictionary supplies a paraphrase: يصاب بضعف شديد (literally: to be affected by severe weakness). This means that the isomorphic semantic equivalent is the ideal lexical equivalent not only on account of its broad semantic coverage but also for its semantic exactitude. One can argue that behind this bilingual semantic isomorphism are macro-level universal principles underlying human cognition. To verify this claim no doubt requires detailed empirical research into many translational language pairs. It can be noticed that the second and third senses exemplified by 1(b) and (c) are a metaphoric extension of the first concrete sense exemplified by 1(a). The comparative corpus analysis of the Arabic translation of collapse in different texts where it occurs, in these three senses, reveals that translators favour the bilingual cognitive metaphor of falling down, lexically realized in the Arabic verb ينهار (collapse), over a lexical-word equivalent pertaining to a different semantic spectrum. This reveals that the semantic equivalent ينهار (collapse) is the absolute equivalent of the word due to its semantic comprehensiveness and the diversity of the SW contextual co-occurrences it covers (about 50 out of 50 occurrences found in one computerized bilingual corpus); it therefore qualifies as an isomorphic semantic equivalent.

By a non-isomorphic semantic equivalent is meant a polysemous target word semantically identical with a polysemous source word in respect of some senses only, or in respect of all senses, but not all contextual co-occurrences. According to this definition, a non-isomorphic semantic equivalent is produced in either of two cases:

(a) the source word and the target word are identical in respect of some of their senses, but not all of them. For example, the Arabic verb يكسر (yaksar) is fit as an equivalent of the English verb break in almost all its senses which are related by the sense spectrum of 'splitting in a harsh manner'; yet it is not a correct equivalent for one of these senses – that of 'cutting' as it occurs in a sentence like: The dog broke the girl's skin, in which case the proper TL equivalent is the Arabic verb يقطع (cut). (b) The source word and the target word are identical in respect of all their senses, yet the target word cannot cover all the collocational co-occurrences of the source word in one or more of these senses (in this case, it is sufficient for a target word to cover only one contextual co-occurrence of each sense of the source word in order to say that there is a non-isomorphic semantic equivalence between the source word and the target word). To illustrate this case, we can return to the example of the adjective fast we mentioned before with its three sense subspectra of event-speed, agent-speed and patient-speed in a sense enumerative lexicon as has been demonstrated before. We find that the English-Arabic lexicographer and/or translator will often use one Arabic word – سريع (fast) – to express the three broad meanings of the English fast. It so happens that the Arabic adjective سريع has these three major senses or, rather, sense subspectra: Arabic native speakers say: ولد سريع (a fast boy), جري سريع (fast run), قطار سريع (fast train).

Yet this Arabic semantic equivalent is still non-isomorphic because it does not cover all the contextual co-occurrences of the source word. For example, fast café will not be translated into standard Arabic as مقهى سريع (fast cafe), because سريع does not collocate with, مقهى , Standard Arabic for coffee shop, in this variety of the Arabic language. The translator or the lexicographer will therefore paraphrase the English NP rendering it as: مقهى للمشروبات السريعة (a café for fast drinks). Hits of the Arabic monolingual corpus for this Arabic adjective tell us that مقهى سريع (fast café) is mostly used informally to mean: a high-speed cybercafé!

2.2.2 Lexical-word Equivalents

The lexical-word equivalent is used in either of two cases: the first case occurs when the SW is polysemous; here it is used either to fill in inadequate coverage gaps left by a non-isomorphic semantic equivalent or as the sole type of equivalent when there is no semantic equivalent. The second case is encountered when the SW is monosemous, in which case the lexical-word equivalent is naturally the only choice available.

2.2.2.1 Lexical-word Equivalents When SW is Polysemous

When the SW is polysemous, the lexical-word equivalent is relevant only in either of two cases: a) when there is no semantic equivalent, isomorphic or non-isomorphic, for the source word. For example, the English adjective fat, has two senses related by the same sense spectrum, i.e., that of size. The first one falls within the semantic field of human body adjectives as in the nominal compound fat man, while the second one falls within the semantic field of adjectives that describe inanimate objects as in the nominal phrase: a fat book. In modern standard Arabic, there is no single adjective lexeme that combines these two senses precisely and so the lexicographer finds himself forced to resort to discrete lexical items as lexical-word equivalents in the target language: بدين, literally: large-bodied for the first sense and ضخم , (large-sized) for the second. b) There is only a non-isomorphic semantic equivalent for the source word and so either the semantic coverage gaps or the collocational coverage gaps have to be filled by lexical-word equivalents in the manner described before. For example, the Arabic verb يشق (split), could also be suggested as a possible lexical-word equivalent for that sense of *break* uncovered by the non-isomorphic equivalent يكسر (break), i.e. break in the sense of breaking the skin, as illustrated above in the discussion of the non-isomorphic semantic equivalent. As for gaps resulting from the inadequacy of collocational coverage by a non-isomorphic equivalent, such gaps are also filled by lexical-word equivalents, as exemplified earlier.

2.2.2.2 Lexical-word Equivalents When SW is monosemous:

When the source word is monosemous, the dichotomy of the semantic equivalent and lexical-word equivalent disappears and only the second pole of it survives – i.e., the lexical-word equivalent. Strikingly enough, the relationship between the two poles is not one of binary opposition but rather one of complementarity: The lexical-word equivalent, when properly employed, fills in gaps left by a non-isomorphic semantic equivalent. For a monosemous source word, the situation is different: there is no scope for such gaps since the source word has a single meaning and the lexical-word equivalent is the only lexical equivalent possible. There are three cases for the lexical-word equivalent when the source word is monosemous:

a) The lexical-word equivalent is monosemous and its meaning is identical to that of the source word. Examples of this phenomenon abound in all language pairs, and it is indeed one of the reasons why lexical transfer between languages is possible. It can be observed among abstract lexical items as well as concrete lexical items. Nouns indicating plants and animals in English, for example, are mostly monosemous words for which there are equally monosemous nouns in Arabic. A word like bravery in English has many synonymous lexical-word equivalents in Arabic, all of which are single-meaning words.

b) The lexical-word equivalent is monosemous yet its meaning is not identical to that of the source word. The result is that the source word meaning is acquired by the target word and added to its already existing single meaning. For example, the Arabic noun أصالة (‘aSāla), which originally meant antiquity or precedence of occurrence of something, came to acquire the meaning of ‘creative thinking’ when it was used as a translation of the English noun originality which means ‘creative thinking’ or ‘newness based on creative thinking’. What happened is that the English source word extended the Arabic sense spectrum of the Arabic word-equivalent so that it means also ‘precedence of thinking’, a sense unfamiliar to the word before this translation came into existence.

c) The lexical-word equivalent for the monosemous source word is polysemous. Here the polysemy problem is transferred from the source language to the target language and in this case, it ceases to be a comparative problem of lexical equivalence between the source language and the target language, but rather one of comprehension related only to the target language. To explain this point, let us pick an example. The English noun science has a single meaning – i.e., that of ‘experimental study of the natural world’. The Arabic target word علم (learning) has two meanings: the first one refers to knowledge in general and the noun in this sense behaves as a deverbal noun which inherits the argument structure of the verb from which it is derived – the Arabic verb يعلم (know). The second meaning refers to ‘experimental science’. Having selected this Arabic equivalent, it will then be the task of the lexicographer to select from its two meanings the one which can be mapped to the source word science – in this case the second meaning, of course – since the target language speaker certainly needs this mapping in order to “comprehend” the meaning of the source word.

The common mistake which bilingual lexicographers inadvertently make is that they usually fail to recognize the significance of differentiating between the semantic equivalent and the lexical-word equivalent. They tend to introduce lexical-word equivalents for the different meanings of the source word without making sure that there is one lexical equivalent which can be suitable as a TL semantic equivalent to all or most of these senses, which could be the first lexical equivalent introduced. In this way, they bar the target language from revealing its semantic richness on the one hand and a considerable part of its expressive force is lost in the translation on the other hand, as we have seen in the case of collapse.

2.3 Grammar and Meaning in a BD

There is a systematic relationship between meaning and grammar which affects the choice of lexical equivalents in a BD. We will restrict the concept of grammar in this section to that common sense found in traditional textbooks which focuses on basic syntactic and grammatical properties of words. Substitutability of a given TL equivalent is not a given. It depends on many factors. One of these factors is the variability of the syntactico -semantic properties of the Sl word. For example, the English noun suicide can be countable or uncountable. The conceptual lexical equivalent of this English noun is انتحار (suicide), which is lexically substitutable for the SL noun only when the latter is uncountable. When suicide behaves syntactically as a countable noun, this equivalent should be changed into حالة انتحار (suicide case).

The countable-uncountable alternations turn out to be responsible for many semantic alternations between an abstract concept and an abstract entity within the same lexical unit. As an example, there is the alternation between abortion (uncountable, abstract concept) and an abortion (particular event, countable). As we mentioned earlier, An English -Arabic dictionary has to provide two different equivalents for the two variants of the English noun, إجهاض (abortion) for the uncountable variant and عملية إجهاض (An abortion operation)for the countable one. It's only when such variations show a regular, systematic pattern that reflects on TW substitutability that they have to be tackled by a BD at all. One way to do this in a paper dictionary is to list them as subentries under their lemmatized forms and list the lexicographical equivalents in the opposite direction.

Shifting the focus to adjectives, we can say that, in some cases the syntactical position of the adjective either before or after the noun can have some bearing on its semantic interpretation in a way which affects the choice of lexical equivalents in Arabic. It should be noted first that we do not mean by the syntactic position of adjectives those cases in which the adjectives is grammatically fixed in one position only, either attributively or predicatively. This having been said, we can proceed. When a regular adjective is used attributively, its meaning may be slightly different than when it is used predicatively after a copulative verb. For example, in 2a and 2b below

2a He is a tense person

2b H is/looks tense

it is easy to notice that tense in 2a expresses a rather stable trait in the noun described by the adjective while in 2b it refers to a temporary state of affairs.

Generally speaking, lexical equivalents of adjectives will not be affected by their mobility. However, when the meaning alternation resulting from this mobility is not reflected by the corresponding position of the regular adjectival equivalent; the alternation has to be preserved in the target language with lexical means by introducing a semantically different adjective for each position. So, it seems that one Arabic equivalent for tense in both its syntactical positions is unlikely. The Arabic adjective mutawatir, supplied by three English-Arabic dictionaries, is a stative adjective and so will be fit to substitute for tense in the predicative position illustrated by 2b. For the attributive position exemplified by 2a, we suggest قلوق (restless), which is an inherent adjective in Arabic and is therefore more semantically felicitous in this position.

In order for the lexicographer to make precise predictions of this kind, he has to restrict his test criteria to two variables only: the syntactical position of the adjective and its meaning and neutralize any other variables that may influence his decision such as the communication situation in the texts he is examining. To achieve this end, test sentences of a simple structure like that of 11 and 12 above should be gleaned out of text and analyzed.

3. Contrastive Morphological Analysis

Arabic is often described as a non-concatenative language. This is because word formation in Arabic is based on the derivation of various morphological patterns from a single root rather than a concatenation of affixes to a stem. Each morphological pattern reflects a set of semantic patterns. But this does not mean that there is no affixation in Arabic morphology. In modern Arabic morphology, concatenation and affixation play a central role in word formation and coinage in order to cope with the terminological needs of the language in the different domains. However, progress in Arabic morphology has been very slow and random in terms of extending the semantic applicability of already existing morphological patterns.

Such slow and random progress has had negative influence on lexical transfer from foreign languages, especially English, into Arabic. This influence consists in using certain Arabic morphological patterns as equivalents to some derivational patterns in English without careful study based on contrastive analysis at the morpho-semantic level. For example, The Arabic nominal category known as artificial masdar (adjectival noun) is often used both in the translation of English “isms” and names of sciences which end with the suffix "ics".To give but a few examples, there is اشتراكية for socialism, معلوماتية and أسلوبية for informatics and stylistics, respectively.

A careful contrastive analysis of the Arabic artificial masdar and the equivalence patterns based on it reveals that it is not an accurate choice for translating science names which end in ics.The line of reasoning on which we base our argument is as follows. The artificial masdar in Arabic is semantically parallel to a relational adjective. A relational adjective is an adjective which indicates a relation to a noun and ascribes the attributes of this noun to the noun which it qualifies. It may be used as an inherent adjective as in معاملة إنسانية (human treatment) and هجمات وحشية (brutal attacks) wherein the attributes of a human and those of a (brute) are used to qualify the deverbal noun معاملة (treatment) and the plural noun هجمات(attacks),respectively. Or it may be used to indicate the mere existence of a relation as in اعتبارات سياسية (political considerations), that is, considerations related to politics. In this way this noun-related adjective in Arabic serves a twofold function: it can be used subjectively as an inherent adjective and objectively as a relational adjective. By analogy, the artificial masdar can be used to do these functions nominally.; For example, The nouns إنسانية (humaneness) ,وحشية (brutality) and همجية refer to subjective personal traits ,while ألوهية (divinity) refers to a relation as in the phrase ألوهية المصدر (divinity of origin).However,the latter,relationl use of the artificial masdar is very rare in Arabic.

In English, isms can also be used objectively as names of doctrines or subjectively to name individual intellectual attitudes. In this way there is semantic symmetricality between the Arabic artificial masdar and an English ism, which makes the former a suitable pattern for translating such isms. On the other hand, names of sciences are characterized by a neutral degree of objectivity since they refer to disciplines of knowledge which are concerned with objective realities. Therefore, their lexical equivalents have to be as neutrally objective, which the artificial masdar is not for all intents and purposes.

It is to be observed that using the artificial masdar in the translation of names of sciences, whether natural or human sciences, is a relatively new trend. The more established one is the use of a pluralized relational adjective on the grounds that the noun which it qualifies is elliptically slashed. On this assumption, a noun such as رياضيات (mathematics) is semantically a reduced form of أمور رياضية (mathematical matters) in such a way that the plural noun أمور(matters) is slashed and replaced by the plural morpheme .ات . What has been said of mathematics can also be said of linguistics, which is often translated as لسانيات.

We conclude thus far that the pluralized relational adjective is more appropriate, from the semantic point of view, for the translation of science names since it is elliptically generated from a semantically neutral nominal compound. The artificial masdar,on the contrary, is less appropriate due to the fact that it is often used to label personal traits or value-laden doctrines, which all runs counter to the objective nature of science. Shifting the focus again to the English-Arabic BD, we find that we cannot burden the bilingual lexicographer with finding solutions to such complicated problems in Arabic morphology. It is the role of Arabic-language academies to solve these problems. Then, lexicologists can receive the results of their research and use them in their arduous contrastive analysis which is essentially related to the preparatory stage. Later on, it will be the task of lexicographers to put such results into practical application in the compilation stage. Without parallel tagged text corpora, no such comparative morpho-semantic analysis of the lexical categories in both languages can be hoped for.

4. Contrastive Syntactic Analysis

In a corpus-linked bilingual dictionary syntax acquires a particular importance due to the interdependent relationship between syntax and semantics in general. There are already many theories which try to frame the relationship between syntax and semantics, the most important of which, in my view, as far as bilingual lexicography is concerned, is the valency grammar theory, which was developed by the French linguist Lucien Tesniere (1893-1954). The valency metaphor is derived from chemistry and refers to the tendency of an atom to acquire or lose a certain number of electrons while it forms a bond with the atom of another chemical element. In language, the atoms are the syntactic categories and electrons are the arguments which they acquire or lose in their interaction with other syntactic elements. Syntactic valencies represent the argument structures of the lexical items. The syntactic valencies of a verb are the subject, object or complement arguments and those of a noun or adjective are the phrasal complements which are attached to them and tied to their semantic representation.

Such quantitative specification of syntactic valencies suits the segmental nature of the lexicon and makes it easier for computers to deal with them as minimum coded units, such as V, which stands for a univalent (i.e. intransitive) verb, Vn which stands for a bivalent verb whose argument structure consists of a subject and a direct object, Vpr for a bivalent verb with a subject and a prepositional complement forming its argument structure and so on.

Semantic valencies represent the semantic content of the syntactic arguments in the form of semantic features and taxonomies, as we will see in the next section.

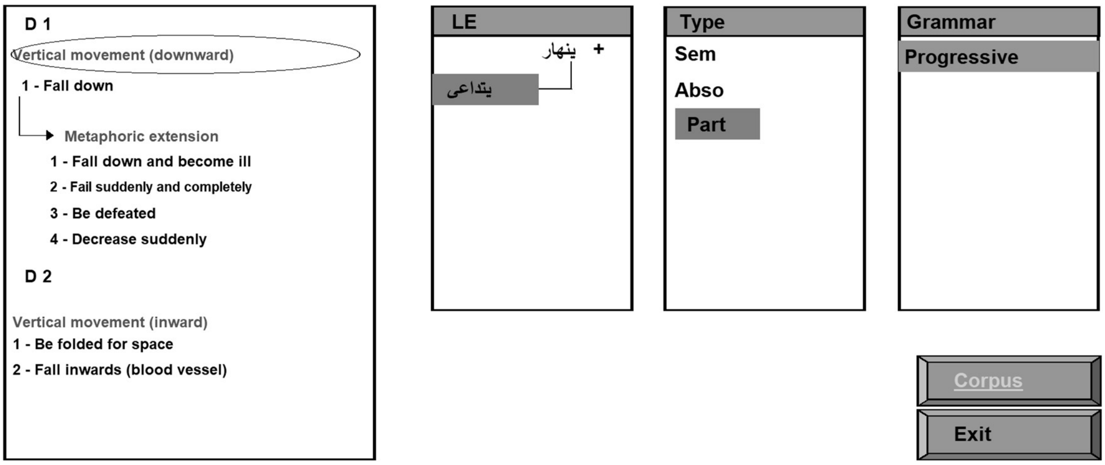
5. Implementation Mechanisms &The Role of Computers

In a semantically organized computerized English-Arabic dictionary, syntactic valency (SVL) is the 'blade' whereby a lexical entry is divided into lexemes and the conceptual content of each lexeme into lexical units. Each set of lexical units is unified by a semantic spectrum, which could be a semantic extension, a semantic field or a cognitive metaphor. Semantic extension is a method of relating senses of a polysemous word semantically rather than at a level of semantic organization. A set of senses unified by semantic extension of a core concept usually have a semantic equivalent in the target language. For example, love in the sense of ‘strong liking’ as in love of horses is a semantic extension of the primary sense of love as ‘warm affection’. In Arabic both senses will have the semantic equivalent حب (love). Semantic field is a broad term for taxonomy, a feature or a dimension. Senses grouped under a given syntactic valency can be divided into taxonomic subsets. For example, the noun bed has several senses that can be divided taxonomically. The first sense is assigned the taxonomy furniture while the other two senses are grouped by the taxonomy land surface (seabed, bed of roses, a bed of rock). Needless to say, it is sufficient to attach the taxonomy name to the first sense of the subset unified by the same taxonomy. However, when a semantic extension leads to a change of taxonomy the sense generated by extension should be assigned its own taxonomical label even if it happens to have the same semantic equivalent in the TL. As an example, the first sense of bed is semantically extended to mean ‘a state of sleep’, as in the sentence: she put the child to bed. The latter sense has to be assigned the taxonomy state. A semantic feature can be used to group senses in a manner which shows a certain contrastive value. For example, the semantic feature ‘inchoative’ (i.e., gradual) can be assigned to the first three senses of the verb decline (decrease gradually, go into a worse condition and slope downwards). For these three senses there is an inchoative verbal equivalent in Arabic, that is, the semantic equivalent ينحدر (literally: slope down).

It is important to note that these levels of semantic organization are not mutually exclusive in theory. A feature, in principle, can well be combined with a taxonomy (e.g., to narrow down its applicability). Semantic extension, far from being a level – as we have just noted – is a technique which can permeate all levels. The message is that we use the single semantic spectrum which is most suitable to highlight contrastive properties of the two languages in so much as they affect our choice of lexicographical equivalents, and not to show the semantic features of each language separately. The taxonomy ‘decrease verbs’, for example, does not bring into focus the contrastive inchoative feature of the English verb decline and the Arabic verb ينحدر since ‘decrease verbs’ in English and Arabic could be inchoative or non-inchoative. This is why we use the semantic feature inchoative+ on its own for grouping the above-mentioned senses of decline into one set, rather than the taxonomy.

Unlike a feature, a dimension represents a concept on a scale of continuous, graded properties rather than a set of binaries, discrete ones. For example, in the semantic representation of the verb collapse as a univalent verb V, the dimension of movement grades from vertical downward movement to vertical inward movement. Between these two-dimensional spectra stands the cognitive metaphor of falling down. A cognitive metaphor is a semantic extension of a dimension or a dimensional spectrum. Senses unified by a cognitive metaphor will mostly have one isomorphic semantic equivalent while senses unified by a dimension could be covered by a non-isomorphic semantic equivalent and lexical-word equivalents that fill the non-isomorphic gaps (see below). Thus, a dimension is conceptually more comprehensive than a cognitive metaphor. The latter generates from the concept several metaphoric senses on the same point of the dimensional scale.

In our would-be **English–Arabic Bilingual Dictionary** there is a separate screen for each syntactic valency. Figures 1 and 2 show a semantic representation of the English verb collapse as a univalent verb (V) together with its Arabic equivalents in a linguistically based, corpus-based and corpus-linked electronic English-Arabic dictionary. To simulate the mouse shifts in the original prototype, the V screen of collapse is split here into two screens.



D 1

Vertical movement (downward)

1 - Fall down

Metaphoric extension

1 - Fall down and become ill

2 - Fail suddenly and completely

3 - Be defeated

4 - Decrease suddenly

D 2

Vertical movement (inward)

1 - Be folded for space

2 - Fall inwards (blood vessel)

LE

ينهار +

يتداعى

Type

Sem

Abso

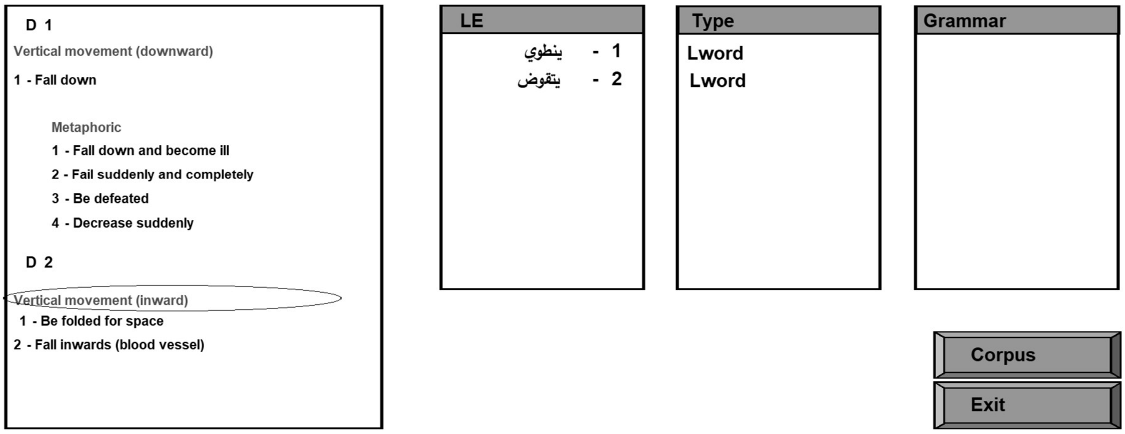
Part

Grammar

Progressive

Fig 1: Dimension 1 (D1) of collapse-V (encircled): downward vertical movement

The first text box to the left in Fig 1 shows the first dimension of the verb collapse, which covers the concrete concept of falling down and is metaphorically extended to cover four other related senses, all of which are linked in the data set to their relevant semantic equivalents as shown in the first list box to the left (where LE stands for Lexicographical Equivalent). The next list box shows the type of equivalent. Abso stands for absolute equivalent, i.e., an equivalent which covers a great number of contexts; Part is short for partial equivalent, i.e., an equivalent which covers a limited number of contexts. The partial equivalent يتداعى is linked to a special grammatical feature in the third list box which specifies that it can be used only as an equivalent of the source verb when the latter occurs in a progressive aspect. This is because يتداعى is an inchoative verb while collapse is a terminative verb and so it cannot be an equivalent for it when it occurs in the past or present simple tenses.



D 1
Vertical movement (downward)
1 - Fall down

Metaphoric
1 - Fall down and become ill
2 - Fail suddenly and completely
3 - Be defeated
4 - Decrease suddenly

D 2
Vertical movement (inward)
1 - Be folded for space
2 - Fall inwards (blood vessel)

LE	Type	Grammar
1 - ينطوي	Lword	
2 - يتقوض	Lword	

Fig 2: Dimension 2 (D2) of collapse V: Vertical inward movement.

Fig 2 shows the second dimensional spectrum D2 which relates to downward vertical movement. It covers two senses which are completely different in meaning and register yet are related by the same dimension. They are linked to two different lexical-word equivalents in the second list box. L-word in the type box stands for Lexical-word equivalent.

To link a bilingual corpus properly to the bilingual database we need to build a bilingual semantic concordance (BSC). A semantic concordance (SC) is defined by Miller et al (1993,303) as "a textual corpus and a lexicon so combined that every substantive word in the text is linked to its appropriate sense in the lexicon". A bilingual semantic concordance can then be defined as "a bilingual textual corpus and a lexical database so combined that every substantive word in the SL text is linked to its appropriate sense in the SL lexicon and its TL equivalents in the parallel corpus and the TL lexicon"

Building a BSC as such from scratch is both costly and time-consuming. Using commercially available tools will make our job much easier and more cost-effective. These tools are a bilingual machine-readable dictionary, a part-of-speech-tagged bilingual corpus and a grammatically annotated computerized English dictionary.

Syntactic categories and their valencies in the form of V, Vpr, Adj:pr, V.to.inf. etc can be extracted from an English electronic dictionary which has such tags for each lexical unit. Then they can be mapped manually to their lexicographical equivalents in the electronic Bilingual Dictionary. The syntactic tags of the part-of-speech tagger are also to be mapped to the part of speech tags of the English lexicon (V, adj, N etc). In this way we can build a crude English parser which we can use to do an automatic syntactic tagging of the corpus texts. Then human syntactic and semantic taggers will have to improve automation results by manual bootstrapping. This will involve correcting errors of automatic syntactic tagging by linking corpus lexemes to their correct syntactical valencies provided by the SL lexicon. It will involve also semantic tagging of corpus words by linking them to their proper senses of the SL lexicon. Thanks to the close relationship between semantics and syntax, we assume that most of the words that were correctly syntactically tagged by the parser are also semantically tagged in a correct way. Of course, if we had a semantically disambiguated parser, this would save a lot of manual tagging. Finally, the Arabic hits in the TL side of the bilingual corpus will appear with the SVL-linked lexicographical equivalents. Now that the bilingual corpus has been linked to a bilingual dictionary, the lexicographer becomes ready to embark on his arduous task of compiling his own linguistically based, corpus-based bilingual dictionary. Among the myriad tasks he will have to undertake is that of updating the lexicographical equivalents of the traditional Bilingual Dictionary, classifying them semantically and adding new ones based on extensive corpus research.

6. Conclusion

Lexicography needn't depend on only lexicology and lexical semantics for its methodology and metalanguage, and it has to develop its terminology and linguistic toolkit. This will inevitably lead to the birth of a new science of bilingual contrastive semantics as an applied subdiscipline of lexicography rather than as a branch of theoretical semantics.

The major points which we need to re-emphasize in conclusion are: First, the importance of selecting a computationally tractable model for a monolingual dictionary to be used as an input for the bilingual dictionary. Second, the need to focus on the semantic expansion of the Arabic lexicon not just its lexical word power so as to provide the lexicographer with a repertoire of word-senses that ultimately extend the applicability of already existing lexemes. This can be achieved through compiling an Arabic dictionary in which semantic generation is based on extensive corpus-based analysis not just on the intuitions of lexicographers. Third, the integrational approach to the BD suggested by the author cannot be achieved without a parallel computationally integrative approach. Such an approach certainly draws heavily on state-of-the-art techniques in Natural Language Processing and data mining as well as the traditional interface-oriented software mechanisms in revolutionizing the content and structure of the Electronic BD. In this way it exacts a radical change in the non-linguistically minded interface culture propagated by current computerized BDs.

References

[1] Alberton, D.J: .(2003) . http://www.linguistik.uni-erlangen.de/~msbethke/papers/Valency.pdf

[2] Al-Kasimi,M.Ali.(1977).Linguistics and bilingual dictionaries.Leiden.E.J.Brill.

[3] Bell, Roger.T (1995).Translation and Translating: Theory and Practice. Longman London and New York.

[4] Catford, John C. (1965). A linguistic Theory of Translation.London:Oxford University Press

[5] Crystal,David.(2003).A dictionary of linguistics and phonetics. Oxford

[6] Fellbaum, Christiane. (1998).A WorldNet Electronic Lexical Database.MIT .

[7] Larson, Mildred L. (1984).Meaning-based translation University Press of America,London

[8] Lyons, John (1982) .Language, Meaning and Context.London:Fontana.

[9] Miller,G.A.(1995).Wordnet:A lexical database for English. Communications of the ACM.

[10] Piotrowky,Tadeuz.(1994).Problems in Bilingual Lexicography.PHD,Wroclaw.

[11]Pustejovsky, J.James.(1998).The Generative Lexicon. Massachusetts Institute of Technology.

[12] Zgusta, Ladislav (1988)."Translational Equivalence in Bilingual Lexicography",Bahukosyam dictionaries.vol.9.

[13] Ba'labaky,Munir.(2003),Al-Mawrid English-Arabic Dictionary.Dar Al-I'lm lilmala<yin

[14] Wahba,Magdy(2003).Al-Nafees English-Arabic Dictionary. Beirut For Publishing and Distribution.

[15] Sakhr Online Dictionary, www.ajeeb.com