DOMENEC MENDEZ
LiveScripts Association
Terrassa – Barcelona – Spain
domenec.mendeth@gmail.com

Language and Identity

1000 Families' is one of my favorite art books. A compilation of typical family pictures from all over the world in which the German photographer captured with great success the essence of every culture and country. From my modest point of view, it would have been equally interesting to reflect a great deal of that diversity without leaving the boundaries of any European country to show the fact that migratory movements have spread the population across every corner of the planet since the very beginning of time. Take for instance, Spain, it is obvious that the blood flowing through our veins once beat up in the hearts of Iberians, Celts, Greeks, Romans, Germans, Jews and Moors -not to mention Altamira or Atapuerca. There is nothing more beautiful than the blending of cultures and races and for that reason, I could not be prouder of my birthplace and origins.

These previous circumstances -emigration, mixing of races- together with trade and war justify the existence of vocabulary -even noun phrases and syntactic structures- common to some or all Western languages. But exactly, how many words between two different languages still keep a somewhat similar spelling and meaning? How many are alike but have another meaning -false friends? Or how many entries have a translation whose morphology is completely different? And this is actually the scope of this paper:

1. Setting up a criteria to compare the lexicon of two languages based on the morphological and semantic similarity between one term from one language and the most common translation/s for that term in another language. 2. Applying such criteria to compare the vocabulary of two languages in order to quantify the results. 3. Suggesting a practical use for this research, mainly in *applied linguistics* and *bilingual lexicography*. 4. And finally, commenting on the expectations and consequences that these new resources in *foreign language teaching* could have in students as well as the marketing impact of this implementation.

Criteria of Classification

1. First Group or true cognates.

1. 1. Scientific and Technical terms -*mathematics, telephone, monopoly*…- loan terms -*potato, orange, anorak, croissant*…- and other cognate terms in which there is an obvious likeness between an entry and its most common translation. 1.2. Words like *colleague, pharmacy, syringe, phenomenon, have, various*… have some spelling differences with their translation but they share their main meaning or moneme.

1. 3. Indo-European cognates like *who, how, what, head, heart, first, fist, hill, helmet, horse, cow, day, night, mother, father, wind*… which share their main moneme with their equivalent translations and usually have a very high use frequency. In general, we can suspect that there is some sort of etymological relationship between two terms when the sounds of their graphemes in either their first syllable or stressed position share the same or a proximate point of articulation -mainly bilabial phonemes /p/ /b/, alveolar-dental phonemes /t/ /d/ and velar phonemes /k/ /g/- and experience variations in the manner of articulation, like voicedness, frication or nasality.

2. Second Group or partial cognates.

2. 1. Both terms compared keep a very similar orthography but they do not share their main moneme, they share instead a second less common meaning, *college, professor, crystal*, *compass*, for instance. 2.2. Those entries in L1 whose direct translations in L2 are not morphologically alike but are nevertheless related etymologically to another term in L2 of the same family. For instance, *close* has not much in common with *cerrar* but it keeps a certain resemblance with *clausura*, and this similarity can be useful for FLT purposes. 2.3. Compound words like *corkscrew* whose separated parts abide by any of the rules above. 2.4. Onomatopoeic words like *tick, click, boom*, bark,

3. Third Group of false friends

3.1 false cognates like *actual, sensitive, sensible, mañana*, escalator, carpet…

4. Fourth Group of non-cognate close set elements

4. 1. Articles, pronouns, prepositions, auxiliary and modal verbs, even days of the week or months of the year are classified according to the previous rules. However, if they do not show any likeness with their translation, they are put into this group because they normally have a very high frequency of use and therefore, they should be potentially easy to learn.

5. Fifth Group of non-cognate pairs of terms.

5. 1. That is, those words whose translation is so different that the student has any morphological reference to identify or relate both terms. Often they are terms with very low use frequency which increases the difficulty to recognize and learn them.

Objectives

This criteria should allow us to accomplish the following three objectives: First, identifying and quantifying the amount of true and partial cognates between both languages (columns named Group 1 and Group 2 on the following charts), so that they can be furthermore classified into semantic fields for a more appropriate use in FLT. This research could also be used as the basis for other investigations searching for cognate Noun Phrases, and Verb patterns or structures with a similar word collocation. Second, quantifying the amount of false friends

(column named Group 3 on the following charts) not to prevent students against their wrong use, because there are already specialized wordbooks of this kind, but in order to make a call to the common sense of publishers as to integrate this list of false cognates with another much more numerous list of true and partial cognates. Third, quantifying specially those completely dissimilar pairs of terms between L1 and L2 (column named Group 5 on the following charts) because it was precisely this percentage (indicated on the next column) the one needed to determine the viability of a new tool in Applied Linguistics which I started to conceive, a series of ***Cognate Bilingual Dictionaries*** among main Western languages whose main features and contents will be explained later.

Results

English-Spanish Morpho-Semantic Comparison

INITIAL	G1	G2	G3	G4	G5	% (G5)	Entries Compared
A	264	26	27	26	46	12	390
B	89	32	33	5	203	56	362
C	378	51	50	2	107	18	588
D	190	20	32	1	87	26	330
E	190	11	25	15	28	10	269
F	111	53	27	10	92	31	291
G	67	22	11	-	67	40	167
H	57	41	37	6	103	42	240
I	159	6	15	11	15	8	206
J	19	12	4	-	11	24	46
K	7	8	-	-	15	50	30
L	39	31	30	3	89	45	190
M	180	32	26	1	52	18	291
N	55	9	14	2	29	25	107
O	56	7	-	7	36	34	106
P	327	53	34	-	69	14	483
Q	7	4	1	-	12	50	24
R	160	36	27	-	76	25	299
S	220	73	49	18	356	50	716
T	110	43	38	17	110	34	318
U	36	5	8	4	46	46	99
V	94	9	6	-	4	4	113
W	7	21	21	14	112	67	166
X	1	-	-	-	-	-	1
Y	2	4	-	5	6	35	17
Z	5	1	-	-	1	14	7
TOTAL	**2829**	**610**	**506**	**147**	**1768**	**30**	**5860**

Conclusions

Applied linguistics is the branch of linguistics that addresses real-world problems like **first language education** and **second language acquisition**. *Historical* or *diachronic linguistics*, on the other hand, deals with language change, and *comparative linguistics*, which is one of their subfields, focuses on comparing languages in order to establish their historical relatedness. There are other branches of *linguistics* which involve the comparison of languages and which are not, nevertheless, part of *comparative linguistics: Linguistic typology*, for instance, compares languages as to classify them by their features and *contact linguistics*, examines the linguistic result derived from the contact of speakers of different languages.

Well, we could say that this is a *typological* research on *semantics* which occasionally relies on the *comparative method* to classify some dubious morphemes as cognates. Or, in other words, a *synchronic* approach to moderns *semantics* with an implied *diachronic* perspective and with the strictest objective of being implemented in *applied linguistics* and *bilingual lexicography*.

On these regards, the somewhat surprising results on both charts more than measuring the high relatedness of both languages, the data mirrors the tremendous interrelationship of worldwide cultures, the leading roles of two countries throughout history and the close geographical proximity of their communities of speakers.

Moreover, in spite of the corpus' oral language preference, the sample can not compensate for the fact that real communication, whether written or spoken, follows other patterns of production and word selection, which are totally unrelated to the trends of semantic similarity described on these charts.

Furthermore, if rigorous criteria of spoken intelligibility or *phonetics* had been considered, dissimilarities would have increased dramatically, even in the first category of identically spelled bound morphemes. For instance, the pronunciation of *literature* and *comfortable* have almost nothing in common with the pronunciation of those terms in another language, although their spelling hardly changes.

Therefore, it is extremely important to point out that the value of these reckonings are entirely restricted to the confirmation of our premises and objectives regarding their use in *applied linguistics*, particularly *bilingual lexicography,* which mainly uses written resources to classify and display their findings, and on this account, the sample should be perfectly representative of other dictionaries with an equal coverage as well as the vocabularies of some other languages. For example, it seems obvious that still higher rates of cognate terms will be found if, unlike this study, the languages compared belong to the same Indo-European branch like Spanish/French, English/Dutch or English/German.

However, the most important conclusion confirms the existence of, on one hand, so many true and partial cognates between both languages and, on the other hand, such a low percentage of dissimilar pairs of entry/translation that, instead of producing specific listings, it made better sense, from the point of view of both pedagogy and marketing, to compound a special bilingual dictionary. This brand-new type of bilingual dictionary, besides being consulted as such, would give students easy access to three types of related information:

First, a comprehensive list of true and partial cognates between both languages, including affixes, noun groups, idioms and syntactic structures conveniently displayed by means of an adequate layout and graphic design; second, a list of false friends marked in red if necessary; and third, another list of those entries -and syntactic structures as well- potentially most difficult to learn because of either the morphological differences with their translations or low frequency of use. As we said before, this new kind of teaching resources are called ***Cognate Bilingual Dictionaries*** among main Western languages.

In this sort of publication, entries are headed by superposed pairs of cognate terms. That is, first the term in L1 with its phonetic transcription and immediately under -instead of next to it- the cognate translation and pronunciation in L2 whenever it might be possible, like this:

Communication (n) /kc,mjuni'keiSn/

Comunicación (n) /komunika0i'on/

Next, the rest of the entry will display monemes, synonyms, examples, etc… the reason for such graphic design could not be more simple: <u>there is no easiest way to show similarities and differences of spelling and pronunciation as to impress the reader's mind and help him associate both terms</u>. In any case, these associations will eventually take place of course, but our particular graphic layout should better contribute to speed up this association process by reducing the passive stage of mere word recognition and sooner incorporate that word into the speaker's active vocabulary.

Nowadays, the design of conventional bilingual dictionaries shows each lexicon as linguistic realities totally independent one from the other, without any absolute connection of any kind now or ever. In my opinion, that traditional design tends to confuse students and block them because it interferes with their *linguistic competence*, that is, their growing intuition to know the right use of a term, its collocation and synonymous expressions. However, the graphic design of cognate pairs of terms heading the entries of CBD gives a more real and truthful lookout on the interrelations and interdependence of modern languages. This new lookout should encourage the students' assimilation as well as their production whether written or spoken particularly in the lower levels.

Another particular feature of CBDs consists of listing all terms for Groups 1, 2, 3 and 4 mentioned before in one single part for each initial and immediately after the rest of words from Group 5 of dissimilar terms. Obviously, dividing the corpus of every initial into two parts -to isolate those potentially more difficult terms to learn- can cause the disadvantage of having to search twice for the same word. Yet, if we look carefully into the data of Group 5 on both charts, we will realize that in more than half of the initials, the amount of dissimilar terms is so small as to be listed in one or two pages which eases the search.

Besides, there are many occasions in which students consult a bilingual dictionary not to look up a term for the first time, but to search for other meanings, or synonymous expressions, or spelling, or pronunciation and therefore, they know beforehand which section to choose. Moreover, splitting up the vocabulary of each initial allows for the following advantages:

a. It influences students in favor of taking an active role when consulting a CBD, challenging them to an activity of word recognition in which they must rely on the linguistic competence of their own language to decide whether looking up a term in the cognate or non-cognate section. b. Searching for a term in the non-cognate section of a CBD also influences good students in favor, even unconsciously, of paying more attention to assimilate that kind of dissimilar term. c. Listing entries in two different sections and one after the other makes it easier for students to consult and complete the self-assessment activities designed to focus on each kind of vocabulary. d. Of course, the existence of two different sections poses no trouble at all when using the digital version of the CBD in CD.

To sum up, CBD is a new attempt to increase the variety and offer of the many bilingual dictionaries existing in the market today, like visual dictionaries, phrasal verbs wordbooks, false friends listings, etc… Publishers, textbooks and methodologies can change from one course year to the other. However, a *Cognate Bilingual Dictionary* can be used to unify criteria of vocabulary and structures required for each level; it can also be used for students to measure their personal progress in lexicon throughout the years, so he can have the certainty of being able to understand a great deal of language and to express any idea with either simple words, or the common complex words that he already knows from his own language. *Cognate Bilingual Dictionaries* stand half way

between a reference book and a workbook, they are for personal use only, and therefore increasing consumption.

Needless to say, it will be a wider practice of speaking-writing output together with the automation of speech habits that will lead the individual to thoroughly assimilate the new code, but exposure to more appropriate pedagogical resources will significantly increase the user's *linguistic competence* and accelerate their wide language acquisition process.

After quite a few years of teaching modern languages, it is obvious that even good students are, on one hand, unable to use basic cognate vocabulary and, on the other hand, they make mistakes of false friends and partial cognates which do not fit in with the several years studying and being exposed to the foreign language. For sure, we teachers have prevented them, remarking similarities and differences of meaning and collocation, but of course this task should be reinforced by a reference book.

And here lies without any doubt the most significant concept to highlight in this article: from my modest point of view there is a key difference between, on one hand, a conventional bilingual dictionary of about 6000 entries per language, like the one used for this research, and on the other hand, the rest of reference books with a higher coverage, whether bilingual or monolingual. And this fundamental difference has been almost overlooked by most publishers who use a very similar design for all of them and put them in the same sack of reference books. This key factor consists in the fact that whether nobody would ever try to memorize a volume, or even a page, of the *British Encyclopeadia*, for instance; this circumstance is totally different with regards to a conventional bilingual dictionary. In this case, if a given student goes through the complete foreign language learning process, by the end of the fifth or six years, he should be able to recognize about 90 % of the entries in the dictionary and use actively 60 or 70 % of those entries. Therefore, it seems reasonable to display all graphic resources, like those explained before, to ease that assimilation process.

This article is actually an invitation to colleges, universities and institutions all over the world to complete the statistics of cognate rates and listings with the rest of Western languages according to the criteria mentioned before. It is also an invitation to keep researching on related topics, for instance, a comprehensive listing of cognate noun phrases, idioms and verb structures, or cross-cognates among all Western languages, or the possible existence of cognate terms in non-European languages. And of course, this is also an invitation for a publisher to pioneer the compilation of one *Cognate Bilingual Dictionary.*

The prospects of success could not be more promising. Just one CBD English-Spanish, Spa-Eng, would have a potential market of about ten million consumers in Spain, UK, USA and the rest of Latin-American countries; even more if we consider that Spanish is frequently the second language option learnt by speakers around the world. Now, keeping in mind that most of those speakers already know English, our Cognate Bilingual Dictionary would be perfect for them because it would help those students assimilate Spanish vocabulary at the same time that they are reviewing English vocabulary.

And this is only one volume, according to the data collected on this research, at least ten more CBDs could be completed, that is, English-French, English-Italian, English-Portuguese, English-German, English-Dutch, but also Spanish-French, Spanish-Italian, Spanish-Portuguese, French-Italian, French-Portuguese, etc… all of them with millions of potential sales and sustained! For let us remember that these publications should be designed as a new tool half way between a reference book and a workbook, and thus personal material not to be shared like a regular dictionary.

What consequences can we expect from the implementation of such a new tool in *applied Linguistics*?

In the short term, it is likely that this kind of publication could draw a special interest and motivation from the general public for linguistic topics, particularly tourists and sales representatives who, instead of limiting themselves to the knowledge of one single lingua franca in all their international trips, might dare learn some basics from Italian, Portuguese, French or Spanish. Mainly, if that acquisition is based on the knowledge of their own language or others that he might be familiar with and if that wordbook comes with a useful phrases, crosswords, etc… Perhaps some of those occasional readers might well go on satisfying their curiosity with basic language courses.

In Language class, cognates already represent a very motivating factor for students, but they frequently refrain from using them -or do it in a wrong way- because of the lack of adequate technical support, and because they are perfectly aware that many expressions are partial cognates and false friends. If these students have access to a didactic resource like a CBD that systematizes and exemplifies the proper use cognate terms in their wide range of morpho-semantic similarities and differences, they might well improve their academic results in all skills, particularly written and oral output.

It is obvious that the most spoken languages -English, Spanish, French…- do not represent a single country any more and have turned into international communication means. *Cognate Bilingual Dictionaries*, with its openly merging or blending design, give every language the character of a natural Esperanto and legitimates, more than ever, their international use not only as a trade and scientific tool, also culturally. In other words, in the long term, the continued use of these new tools in *applied linguistics* should provide a better appreciation and understanding for other Western languages and foreign cultures.

To sum up, CBDs are only the consequence of the tremendous scientific and technical advances that lead our society for the last two centuries and blend absolutely all frontiers among any field of human knowledge to produce a single ***oneness*** within our apparent variety. CBDs are also the echo from the delicious cultural mixing of the arts, the music and dancing. Denying them is to deny us.

Works Consulted

- Al-Kasimi, Ali M. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill, 1977.
- *The American Heritage Word Frequency Book*. New York: American Heritage Publishing Co., Inc., 1971.
- *The Barnhart Dictionary of Etymology*. N.p.: The H.W. Wilson Company, 1988.
- Bergman, Peter M. *The Concise Dictionary of twenty-six Languages in Simultaneous Translation*. New York: Bergman Publishers, 1968.
- *The Cambridge Encyclopedia of Language*. London: Cambridge University Press, 1987.
- *Collins Pocket Diccionario Español-Inglés English-Spanish* 8th ed. Barcelona: William Collins Sons & Co. Ltd. Ediciones Grijalbo, SA, 1986.
- *Ethnologue*. Dallas: Summer Institute of Linguistics, Inc., 1988.
- Hood Roberts, A. *A Statistical Analysis of American English*. London: Mounton & Co., 1965.
- Knorre, Dorwick, Vanpatten and Villarreal. *Puntos de partida*. 3rd. Ed. New York: Random House, 1989.