

**Teaching and translating synonyms: The case of *almost* and *nearly***

Belén Labrador de la Cruz

Universidad de León

El propósito de este artículo es presentar una serie de inequivalencias interlingüísticas y analizar específicamente una de ellas, lo cual nos lleva a revisar uno de los fenómenos intralingüísticos de mayor relevancia en lo que concierne a los estudios contrastivos: la sinonimia. El estudio que aducimos para sustentar nuestra hipótesis se ha realizado mediante una metodología empírica basada en corpus y los resultados prueban que incluso palabras tan parecidas semánticamente entre sí como *almost* y *nearly* difieren en gran medida en lo que respecta al uso del lenguaje. Sus significados son semejantes por lo que pueden ser equivalentes a un único término en otras lenguas, como es el caso del castellano (*casi*); sin embargo, no coinciden en su comportamiento lingüístico, el cual engloba índices de frecuencia, medio, contexto y patrones colocacionales.

**Abstract:** The purpose of this paper is to present a series of interlinguistic inequivalences and to specifically analyse one of them; this leads us to revise one of the most important intralinguistic phenomena with relevance in contrastive studies, synonymy, by means of an empirical methodology based on corpora. What is advocated here is the fact that even two such similar words as *almost* and *nearly* may be dissimilar in the realm of language use; they are very close in meaning and they usually correspond to a single item in other languages, for example in Spanish, but their behaviour differs in many respects.

A great deal has been said about the fact that translators, teachers and students of foreign languages sometimes cannot avoid thinking in terms of words, a tendency that is very much disapproved of and very much fought against by themselves. One of the reasons might be the influence of the written medium, where all words are set apart from one another by means of a blank space and another could be the way our mind works, trying to structure all knowledge. Whatever the reason, it is a risky but daily habit to give one-word translations for words in another language without paying attention to the context in which they occur.

This paper aims at raising the issue of intralinguistic synonymy and its relevance in the field of Contrastive Studies, Second Language Teaching and Translation. We advocate the idea that synonymy and other semantic and linguistic phenomena must be observed in the light of language use, within a functional approach and through empirical and descriptive research. The first part will consist of a theoretical exposition of the two most relevant facts concerning this subject, which are interlinguistic and intralinguistic relations of linguistic items, that is, their links with similar words in other languages on the one hand and their connection to similar words in the same language on the other. And the second part will provide some empirical evidence that shows that two so-called synonyms are not so to such a great extent as has always been considered and should therefore be taught differently.

## **PART I. THEORETICAL PRESENTATION**

### ***1. Interlinguistic relationships***

By way of example, in the present paper, we shall consider the case of *almost* and *nearly*. They are most often than not translated by *casi* in Spanish, consequently being thought of as interchangeable by Spanish students of English. One word in a language corresponds to two in another. This is quite a common phenomenon of which a large number of instances can be traced. The explanation that the Europeans Herder and von Humboldt first gave, which was later developed by the American structural linguists, especially Sapir and Whorf, is the cultural and linguistic variety, linguistic determinism - language determines thought and linguistic relativity - there is no limit to the structural diversity of languages.

This conveys the fact that vocabularies in different languages are not completely parallel: "It is well known that the vocabularies of languages tend to be, to a greater or lesser degree, non isomorphic. To the extent that this is so, some things will be more highly codable in one language than they are in another" (Lyons 1981: 306). These theories are usually supported with evidence like the huge range of word forms to refer to snow in Eskimo or to sand in most Australian languages, as compared with the limited number of words for these concepts in other languages. Another well-known example is the different division of the same reality, the colour

spectrum; the different sets of vocabulary of colour in different languages have been studied in connection with the thesis of linguistic relativity.

Is it possible then to establish correspondences between the *parts* (belonging to different languages) in which the *whole* of the meaning (common to both) is divided? It is, indeed, if we consider that the elements of such a comparison are different but they share a common factor, which is called *tertium comparationis*<sup>1</sup>, a kind of “third element of the comparison” that acts as an axis or link to make two elements comparable (in this case two sets of different words or expressions in two languages), elements which are neither totally alike nor totally distinct.

In our opinion the common factor that unites these comparable elements is a single semantic core which is materialized by a series of different formal resources in the languages. This is the reason why there exist elements which are semantically and/ or pragmatically similar and whose differences lie in the phonological, morphosyntactic and lexical levels. They are different materials carrying out the same function, and this function, taken as the *tertium comparationis* of the comparison allows the linguist to discover the similarities and differences in the resources that native speakers of two different languages employ in order to communicate similar or “communicatively comparable” messages.

Let us now move on to the classification of the problems that arise in translation due to the phenomenon of linguistic relativity mentioned above. Each word is a tool referring to a specific area of reality. And the difference between words in different languages does not only lie in the differences between their arbitrary significants, but, especially, in the scope of their meanings, because, not only the relationship between *le signifié* and *le signifiant* is arbitrary (Saussure 1916) but the creation of a *signifié* out of the division of the cognitive *whole* is also arbitrary. That is why we frequently find examples of inequivalence:

First there are conceptual or referential voids. Rabadán presents them as an extralinguistic limit to translation. She says that the inequivalences that derive from the linguistic representations of areas of experience not shared by two different cultures are most possibly the most difficult to surpass in a translation that is governed by the norms in the

target polysystem, and she adds, as an example, that it is not surprising the fact that a Spanish reader ignores the referent of the word *porridge* (Rabadán 1991: 164).

There are also cases of terminological voids or lexical gaps; what the culture lacks is a term to refer to a particular concept but the concept is known. Then a description is used until a borrowing or a loan process takes place. The borrowed word is usually adapted phonetically to the target language, e.g. *fútbol* (in Spanish from *football*) and two forms may coexist for a time before one of them prevails, e.g. *fútbol-balompié* (from ball = *balón* and foot = *pié*).

A third example is that of the so-called “false-friends”. They are usually cognates, that is, forms with the same root or origin that have led different lives and have therefore come to have different meanings. A typical example is that of *actually* - *actualmente*; the outlook of these two words is similar but *actualmente* means *at present* or *currently*.

Finally, another source of problems in translation is the area of asymmetrical correspondence. Both the concept and the term are known by the members of the two cultures but the way of dividing, classifying and naming the meaning is different. It is the case of having a single item in a language but two in another, which causes problems especially when translating into the second language. There are two situations: one in which the two words corresponding to a single term in the other language are never thought to be interchangeable but closely related, for instance, the *evening* is not the same as the *afternoon*, the former begins when the latter ends. However in Spanish a single term comprises the whole of this period of time. This can be named “exclusive asymmetrical correspondence” because the choice of an item excludes the possibility of the other as their meanings are different. The other situation is that of the “inclusive asymmetrical correspondence”, where the two items in a language are thought to be interchangeable, thus not exclusive, for instance *almost* and *nearly*, and correspond to one item in the other language - *casi*. A summary of these nets of interlinguistic relationships is shown in the following chart:

Table 1. Types of interlinguistic equivalences and inequivalences.

|             |  |  |
|-------------|--|--|
| a = b       | symmetrical correspondance               | e.g. <i>milk</i> = <i>leche</i>                          |
| a = 0       | referential voids                        | e.g. <i>fabada</i> , <i>porridge</i>                     |
| a = "a"     | lexical gaps                             | e.g. <i>wind-surf</i> , <i>paella</i>                    |
| a = b and c | exclusive<br>asymmetrical correspondance | e.g. <i>tarde</i> = <i>evening</i> &<br><i>afternoon</i> |
| a = b or c  | inclusive<br>asymmetrical correspondance | e.g. <i>casi</i> = <i>almost</i> / <i>nearly</i>         |

In their turn, these elements can be related to others in their own cultures, for instance, *tarde* as well as meaning the period of time that extends from around 2 p.m. to 10 p.m., also means *late*, the opposite to *pronto* (*early or soon*), so its relationships become more complex. Still, it is possible that some symmetric referential meanings have a different pragmatic function in two languages. Lado raises this matter with the following situation: He says that drinking milk is a usual habit in the United States and it has no connotation of social class, nationalistic or religious group, age, sex, etc., whereas in France, this habit is unusual and it has the meaning of special drink, special occasion or special group of people. And it is just the opposite with wine. (Lado 1957: 126). These apparently equivalent words might not be so in some particular context.

## 2. Intralinguistic relationships - Synonymy.

After these brief considerations on the matter of cross-cultural relations, what follows is an analysis of such a case as that of the words that have been selected for the purpose of this study. They have been classified here as terms having an *inclusive asymmetrical correspondance* with their equivalent in Spanish. They have always been taught as synonyms and most students regard them as completely interchangeable but are they really? Do real synonyms exist at all? It has often been said that if different words exist, it is because different meanings need to be conveyed, sometimes slightly different nuances, but then, more than one

word is necessary, thus, it is possible to assert that synonymy does not exist.

Or else, synonymy can be understood somehow differently. Palmer says that in his terminology, synonymy is a “sense relation”, not a “meaning relation”, and that this implies that there is synonymy when, in a given utterance, two or more items can be interchanged without any alteration in the purport of the utterance. He adds that, in that case, all the interchangeable items would make the same contribution and so would have “the same sense”, as in “the new porter will start (begin) work next week”, but he further explains that *start* and *begin* are not synonymous in all their senses, e.g. “The horse started back”, “she started when I came into the room suddenly”. In neither of these can *start* be replaced for *begin*. He then concludes that *begin* and *start* differ in their *meanings* even though they share some *senses*. To put it another way, synonymy is context-bound (Palmer 1972: 203).

We would like to go a step further and add another restriction to synonymy. We put into question that even “sense synonymy” exists or at least that there are as many cases as have always been considered as such. Our hypothesis is that *almost* and *nearly* may share all their senses and still not be completely synonymous because each one of them is commonly used in different patterns, beside different words. Consequently, they differ neither in their meanings, nor in their senses, but in their uses, in the preferences that native speakers have in order to choose one or the other to combine with other words.

Context is greatly involved in a two-fold way: situational context, that is, extralinguistic factors that set the conditions whereby an utterance is spoken and textual context or *co-text*, that is, the linguistic environment of the word or expression that is being studied, the collocational patterns. As Roos says: “Particular difficulties arise if the learner is confronted with a choice of *synonymous* lexical items which vary in usage. There are no easy answers available: problems of meaning can be solved by referring to a dictionary, problems of syntax by referring to a grammar. If the problems lie in between, both fail. This is the case with collocations” (Roos 1976: 65).

The meaning and the grammar of these as many other words are by no means problematic but sometimes doubts arise about how to use or translate them. We have chosen two apparently most simple and similar words in order to exemplify this idea. Some more research of this kind has been carried out so far, for example the comparison of *between* and *through* by Kennedy (1991) and that of *strong* and *powerful* by Church et al., (1991).

## **PART II. DESCRIPTIVE STUDY**

The methodology employed to analyse the textual context of *almost* and *nearly* is corpus linguistics, the study of language use by means of a computerised “collection of samples of running text” (Aarts 1991: 44). And the corpus we have used is the Bank of English of Cobuild, with the 50 million words of it that are available via Internet. It is composed of 11 subcorpora that correspond to a number of different text types, mainly, specialised magazines, newspapers, radio, oral English, books and ephemeral texts<sup>2</sup>. The variety of the corpus is also shown by the three diatopic variants it contains, although the proportion is not balanced: they are British English, American English and Australian English, ordered from a higher to a lower representation in the corpus. The results of the analysis are summarised as follows:

### ***Frequency and distribution***

In the first place, *almost* is much more frequent than *nearly*. The former has 14,162 occurrences in the 50 million word corpus whereas the latter occurs 6,000 times. As for their distribution, *almost* occurs most often both in British and American books, and most rarely in British and American ephemeral texts and in the oral subcorpus; the subcorpora of the press and radio are in between. *Nearly* differs in its distribution from *almost* in two senses: the medium and the variety. The British subcorpora are at the bottom of the list, in a lower position than the American except for the BBC World Service and Today newspaper, which are the second and third respectively in the ranking. And also *nearly* is most frequent in the subcorpora of the radio: American National Radio Public and British BBC World Service and least frequent in the British magazines, books, oral and ephemera subcorpora.

From all these data, an assertion can be made that there is a tendency for *almost* to appear in the written and more formal medium than in the oral and especially colloquial English, regardless of the diatopic variety. However, *nearly* seems to occur more often in the varieties of American English than in the other two and it is especially frequent in a kind of semiformal oral register.

### *Collocations to the left of the key word*

As far as the collocates are concerned, one of the most frequent to the left of *nearly* is *not*, which is, by contrast, very unusual next to *almost*. There are only 4 cases of *not almost* in the corpus and they are of a different kind altogether to the 112 cases of *not nearly*. As is shown in example (1), *not* is not modifying *almost*; they belong to different phrases, in this case even different clauses as well. The cooccurrence in cases such as this one is merely due to coincidence and it is not significant for the purpose of the study because there is not a strong link between the members of the combination analysed.

(1) Whether they're serious or not almost doesn't matter.

On the other hand, the occurrences of *not nearly* are significant and may cause problems when translating into Spanish or when encountered by Spanish learners of English as a second language. 56% of these occurrences are immediately followed by *as*, example (2), 20% are followed by *enough* as in (3), 1% is followed by a similar word, *sufficient*, as in (4), 11% occur next to *so* as in (5), 2% next to *such* as in (6) and another 1% is also followed by a particle with a comparative function, which is *that* as in (7).

(2) Learning to sail is not nearly as wet as surfing and windsurfing

(3) I could start practising not nearly enough for a Grand Slam

(4) ...for which a simple 'sorry' is not nearly sufficient

(5) She is such a little madcap not nearly so docile as her sister



(6) ... an off-white “Autumn Joy” but not nearly such a strong, reliable plant

(7) No, we’re not nearly that far ahead with blackflies

Finally, the remaining 9% is followed by a miscellanea of words: examples (8) and (9) Except for this remaining 9%, all of the other occurrences are examples of comparisons, a very definite context.

(8) No, not done, not nearly done

(9) I’m in my prime. I feel 22 not nearly 32.

It is also significant that there are no occurrences whatsoever of *very almost* whereas there are 82 occurrences of *very nearly*, followed either by a verb or by an adjective, as in (10) and (11):

(10) She very nearly lost her life

(11) The nineteen-year-old is very nearly old enough.

There is also a tendency for the auxiliary verbs to prefer the company of *almost* rather than *nearly*. There are only 10 occurrences of *will nearly*, 6 of them followed by *always*, as in (12), 2 followed by the verb *double* and 2 by the verb *triple*, as in (13) and there are 176 occurrences of *will almost*, 84 of which are followed by *certainly*, 6 by *always*, 2 by *immediately* and *double* and 1 by *instantly*, *surely*, *invariably*, *automatically*, *certain*, *yearn*, *be* and *a year ago* as in examples (14) and (15):

(12) This technique will nearly always identify food sensitivity

(13) The milk will nearly triple in volume

(14) The fuel tax will almost certainly be forgotten

(15) The new leader will almost immediately have to start working

Something similar is the case with *would*, which occurs 7 times next to *nearly* but 151 next to *almost*, 75% of which this combination is followed by *certainly* as in (16). In general, *almost* is more common after verbs than *nearly*. Some significant figures are the following: 130 occurrences of *is nearly* against 765 of *is almost*, 100 occurrences of *-’s nearly* against 481 of *-’s almost*, 60 occurrences of *are nearly* against 283 of *are almost*, 192

occurrences of *was nearly* against 806 of *was almost*, 67 occurrences of *were nearly* against 171 of *were almost*.

The combination *the nearly* occurs 71 times in the corpus but as many as 62% of the times it is followed by a number as in (17) and although *the almost* occurs more often, 157 times, it is followed by a number only 7% of the times (18); this combination usually precedes adjectives (19).

(16) such a resolution would almost certainly be defeated

(17) ... what will happen to the nearly 20,000 refugees along the border

(18) The authorities had warned the almost one-and-a-half million people living in this area to take precautions

(19) In Rwanda I saw the almost miraculous transformation of hundreds of listless, traumatised and hungry children into noisy, energetic youngsters

### ***Collocations to the right of the key word***

Some very frequent collocates to the right of *almost* are adverbs ending with *-ly*, some of them are on the 50 top collocations: there are 576 occurrences with *certainly*, 185 with *immediately*, 141 with *entirely*, 100 with *exclusively*, 58 with *daily*, 56 with *exactly* and 43 with *invariably*, examples (20) and (21). However, there are very few cases next to *nearly*: only 4 occurrences with *instantly*, 2 with *fully*, and some single instances of other *-ly adverbs* like *continuously*, *simultaneously* and *indefinitely*, examples (22) and (23).

(20) the reprisals the Iraqis would almost certainly take against the Kuwaities...

(21) Record contracts and world tours happened almost immediately.

(22) I forgot it nearly instantly

(23) a woman who fully or nearly fully breast-feeds her infant...

Regarding numbers, *nearly* seems to be more frequent before numbers (2,422 occurrences) than *almost* (1,648 occurrences) in all contexts and these figures become especially significant if we rate them against the total frequency of these two terms. 11% of the occurrences of *almost* precede a number whereas as many as 40% of the occurrences with *nearly* precede a number. This seems to be another piece of evidence that shows the fact that *almost* and *nearly* have not absolutely parallel behaviours.

Another difference regarding the collocations to the right of the key word is the largenumber of adjectives next to *almost*, 2,569, as compared with the 348 adjectives that occur behind *nearly*. At least one of them, *certain*, occurs only next to *almost*, 177 times, and never behind *nearly*. In both cases negative adjectives, both in form, like *impossible*, *illiterate*, *meaningless*, *unrecognisable*, *destitute*, *invisible*, *irrelevant* and in connotative meaning, like *empty*, *vacant*, *extinct*, *sick*, *wrong*, *naked*, *fatal*, *blind* are very common.

Surprisingly enough, the negative particles usually follow *almost* as opposed to the previous situation mentioned above in which it was *nearly* that was preceded by *not*. Whereas we have found 123 occurrences of *almost no*, as in (24) and (25), only 2 have been registered of *nearly no*: (26) and (27). Something similar occurs with *almost nothing*, example (28), with 74 occurrences and *nearly nothing*, with none.

(24) Nobody liked him and almost no one knew who Nike was

(25) and throughout that time we have had no government support

(26) But you will have nearly no chance

(27) there were nearly no fish

(28) We prepared it but in fact we got almost nothing in that area

## Conclusions

1. A close observation of the behaviour of two apparently equivalent words has shown that there “may be more to it than meets intuition” as refers the traditional establishment of sets of synonyms. By means of inductive and empirical tools such as corpora it is possible to reduce the so-called phenomenon of synonymy to lexical fields of very related words discarding the idea of absolute equal words.

These lexical fields would be arranged in concentric circles and the leap to an outer layer would show a lesser relationship, for example, *almost* and *nearly* would belong in the inner circle, and other resources to express a similar function would occupy the rest of the outer circles, like *a little less than*, *hardly*, *barely*, *approximately*, etc.

2. In a similar way as Palmer stated that two words may be synonymous in one or some of their senses but not in others, we consider that two words can be synonymous in one or some of their uses but not in others. In this paper we have outlined the uses in which *almost* and *nearly* differ, leaving aside those in which they behave similarly, as we have made the assumption that they are partly synonymous, but only in the contexts they share.

3. As some of their contexts seem to be exclusive to one of them, i.e. *not nearly*, *very nearly*, *almost no*, *almost nothing*, *almost certain* and in other cases, there are strong tendencies of co-occurrence, i.e. *almost* + *-ly* adverb, *the nearly* + a number, *almost* + adjective, auxiliary verb + *almost*. they cannot be interchangeable or at least, if they are said to be, all these contexts should be specify as restrictions.

4. Moving on to the field of applied linguistics, it is of great importance to foresee the implications that descriptive studies of this kind may and should trigger: when teaching these items both in English as a Second Language or in Translation, all the differences in use between them should be stated, thus providing the student with the necessary knowledge to see their similarity in meaning as well as being able to recognise when, where and why one has been privileged instead of the other.

Let us consider the case of a language with only one *prototypical equivalent* (terminology employed by Kalisz, 1988) for these two word-forms, which is Spanish. We have always been taught that *almost* and *nearly* are tantamount to *casi*, so one would presume that most of the Spanish people use them at random, which means approximately 50% of the times each. However, as the figures in the corpus show, in the native speakers' production *almost* seems to prevail over *nearly*.

As for the *co-text*, there are two cases of the series of them analysed in this paper which are ungrammatical or very odd in Spanish and they are the use of *not nearly* with comparisons and the collocation with *very*. In the first situation, all the sentence would usually be translated except for *nearly*, or else other resources to convey the same function might be employed in some cases, and as refers *very / muy*, again, other resources are used that substitute either *muy* or *casi*.

Regarding the rest of the collocational patterns, problems may arise not when translating into Spanish but into English, as the grammatical and collocational patterns coincide in the two languages but there is a two-sided choice. If only they are given a few guidelines that describe rather than prescribe the real and natural use of these words by native speakers, students will use them more appropriately and confidently, as a result of understanding that they are similar words but that their behaviour differs in some occasions.

5. Corpus linguistics is revealing itself as a suitable methodology for research in Descriptive Linguistics, including areas like Contrastive and Translation Studies, and the contribution of this kind of research is patent as well in Foreign Language Teaching. The shift in focus, from language system to language use that is said to have taken place some decades ago has caused the spread of corpus-building but it should also be supported by more studies that be born from corpora.

6. The observation of intra and interlinguistic nets of relationships among related word-forms allows for a better comprehension of the nature and functions of *assumed equivalent terms*, both as synonyms in the same language and as parallel items in two different languages.

All in all, the importance as well as the difficulty of this matter, teaching and translating synonyms, lies both in the links that join those synonyms together and in the links that join them to their *prototypical equivalents* in the target language (in translation) or the native mother tongue of the students. A contrastive study must take a global perspective that includes both nets of relationships.

## NOTES

1 For a historical review of the choice of *tertium comparationis* in Contrastive Linguistics see Rabadán (1991:43-45)

2 Ephemeral texts are brief texts that have a practical communicative function and are not usually published or recorded (e.g. tickets, brief notes, letters, bills, etc.)

## BIBLIOGRAPHY

Aarts, J. 1991. "Intuitions-based and observation-based grammars" in K. Aijmer and B. Altenberg eds. *English Corpus Linguistics*. London and New York: Longman. 95-111

Church, K., W. Gale, P. Hanks and D. Hindle. 1991. "Using statistics in lexical analysis" in U. Zernik ed. *Lexical Acquisition*. Englewood Cliff, NJ: Erlbaum. 115-64

Kalisz, R. 1988. "The problem of directionality in contrastive studies based on cognitive linguistics" in J. Fisiak ed. *Papers and Studies in Contrastive Linguistics* vol. 23. Poznan: Adam Mickiewicz University. 37-45

Kennedy, G. 1991. "*Between and through*: The company they keep and the functions they serve" in K. Aijmer and B. Altenberg (eds). *English Corpus Linguistics*. London and New York: Longman. 95-111

Lado, R. 1957. *Linguística contrastiva, lenguas y culturas*. Madrid: Ediciones Alcalá (translated into Spanish from *Linguistics across Cultures* por Joseph. A. Fernández)

Lyons, J. 1981. *Language and Linguistics*. Cambridge: Cambridge University Press

Palmer, L.R. 1972. *Descriptive and Comparative Linguistics. A Critical Introduction*. London: Faber & Faber

Rabadán, R. 1991. *Equivalencia y Traducción*. León: Universidad de León

Roos, E. 1976. "Contrastive Collocational Analysis" in J. Fisiak ed. *Papers and Studies in Contrastive Linguistics*. vol 5. Poznan: Adam Mickiewicz University. 65-76

Saussure, F. 1916. *Cours de linguistique générale*. Paris: Payot