# Contrastive Lexical Analysis

PID_00249322

Ana Fernández Montraveta

Recommended minimum reading time: 3 hours

**Ana Fernández Montraveta**

PhD in Linguistics and
Communication (UAB), expert in
Lexicology, Semantics and Corpus
Linguistics.

Associate Professor. Departament de
Filologia Anglesa i de Germanística.

Coordinator of the Linguistic
Applications Inter-University
Research Group.

# Contents

# Introduction

**Contrastive Lexical Analysis** is a very active area of research within the field of Contrastive Analysis (CA) and, by extension, within the field of Applied Linguistics.

In order to define what this area is about let's define the words in the term. Lexical analysis belongs to the field of Lexicology, the branch of linguistics concerned with the study, at a theoretical level, of the lexical component of languages. Lexicology is closely related to lexicography, which is about dictionary creation: i.e. sense distinctions and definitions, and, at a more general level, with the macro and microstructure of a dictionary.

The other word of the term, *contrastive*, describes a methodological approach to the study of languages that, as you already know, is essentially based on the contrast of languages; that is, how similar or different languages are, at different levels of linguistic analysis. In this unit we are concerned with how languages can be compared if we consider them at the lexical level.

Thus, the main focus of Contrastive Lexicology is studying and comparing the lexical systems of two or more languages. The most common view of lexical systems (the vocabulary or lexicon of a language) is that of a list of words that have a specific form and meaning. Consequently, it is commonly thought that the way languages might differ at the lexical level is either formally (i.e. the way you write or pronounce words is different) or semantically (i.e. the senses a word expresses do not correspond).

Nevertheless, a commonly forgotten aspect of the lexicon is the referential value of words, that is, the fact that words are used to refer to reality. As we all know, words denote objects, ideas, actions, properties of objects, etc. that belong to the world we live in. Sometimes, reality, or, rather, «our perception of reality as a community of speakers» might differ. It is precisely for this reason that it is hard to talk about a total equivalence of the words in the different languages.

In this unit we will review several approaches to the analysis and contrast of the lexical systems of English, on the one hand, and Spanish and Catalan, on the other.

# Objectives

After having studied this unit, the student will be able to:

**1.** Contrast equivalent senses of lexical items in English, Spanish and Catalan using Lado's classification.

**2.** Become familiar with the terms *false friend*, *cognate* and *false cognate*.

**3.** Understand the concept of *lexical divergence*, and become aware of the mismatches between English and Spanish/Catalan at lexical level.

**4.** Become familiar with the concept of *Extended units of meaning*, be able to identify them, and find the best equivalent in the target language.

**5.** Understand the terms *neologism* and *borrowing* and the most common types according to their formation processes.

**6.** Get to know the most common translation strategies used in the translation of neologisms and borrowings.

# 1. Lado's classification

As we said above, Contrastive Lexical Analysis, and therefore Contrastive Linguistics (CL), is considered a sub-discipline of Applied Linguistics. Back in the 1950s and 1960s, this methodological approach was customary practice, particularly in fields related to language teaching and learning. This is probably the reason for the spread of CL and its popularity then. At the time, it was believed that most of the errors made by foreign language students could be explained through the differences in the structure (grammar) and the lexis (words) between the target language (TL), i.e. the language being learned (L2), and the language of the learner, i.e. their mother tongue (L1).

Robert Lado, a professor of English as a Foreign Language (EFL) and a scholar at the University of Michigan, is considered one of the founders of modern CL. Lado, together with Charles C. Fries, postulated in 1957 what is known as the strong version of the contrastive hypothesis, which defends the idea that difficulties in the learning of a second language can be predicted on the basis of contrastive analysis between the system of the learner's native language (L1) and the system of the TL they are studying (L2). By foreseeing the differences between the student's L1 and L2, more adequate materials can be designed, thus making the learning of a foreign language an easier task.

In his seminal book *Linguistics Across Cultures: Applied Linguistics for Language Teachers* (1957), Lado outlines the methods for comparing the systems of two languages at different levels. The book was intended as an introduction to the methodology used in contrastive analysis. He considered not only the linguistic levels: phonetic (comparison of sounds), structural (comparison of the grammatical structures) and lexical (the level we are concerned with in this unit), but also the cultural level and the writing system of languages.

In this unit, we are particularly interested in his proposal for the analysis of the lexical level. Lado's classification of contrasts between words accounts for the ways in which pairs of equivalent words relate in two languages, taking into consideration both their meaning and their form. It has been widely accepted and adopted in the world of language teaching and, in fact, the terminology he proposed in this work is still in use in the EFL classroom. Probably, you have seen lists of false friends and cognates in many books and teaching materials and are already familiar with these concepts.

In this section, we present an adaptation of Lado's (1957) original proposal. We will consider 4 different degrees of formal and semantic similarity between words and their equivalents: false friends, cognates, partial cognates / partial

false friends, and all the other possibilities: similar words with different forms, words which are different in both form and meaning and, finally, similar words with different connotations.

## 1.1. False friends

False friends are words that are very similar in form but completely different in meaning. Below we present a table (Table 1) with several very well-known false friends for the pairs of languages English-Spanish and English-Catalan (as you can very well imagine, and given the proximity of Spanish and Catalan, many false friends are common):

**See also**

For a more comprehensive list, visit the sites in the Section «Useful Online Resources».

**Note**

Keep in mind that when we talk about a word we always refer to a specific meaning of that word, i.e. one of its senses.

Table 1: List of false friends in English, Spanish and Catalan

| English | Spanish/Catalan | Spanish/Catalan | English |
|---------|-----------------|-----------------|---------|
| SL | TL | TL | SL |
| actual | real | actual | current, present |
| actually | en realidad / en realitat | actualmente/ actualment | nowadays |
| advice | consejo/consell | aviso/avís | notice, warning |
| agenda | orden del día / ordre del dia | agenda | diary |
| argument | discusión / discussió | argumento/argument | plot |
| assist | ayudar/ajudar | asistir/assistir | attend |
| carpet | moqueta, alfombra/catifa | carpeta | file |
| constipated | estreñido/restret | constipado/constipat | having a cold |
| conference | congrés/congreso | conferencia/ conferència | lecture |
| content | satisfet | contento/content | happy |
| deception | engaño/engany | decepción/decepció | disappointment |
| embarrassed | avergonzado/avergonyit | embarazada/ embarassada | pregnant |
| idiom | refrán/proverbi | idioma | language |
| library | biblioteca | librería/llibreria | bookstore |
| molest | abusar sexualment(e) | molestar | bother |
| ordinary | normal | ordinario/ordinari | vulgar |
| preservative | conservant(e) | preservativo/ preservatiu | condom |
| prove | demostrar | probar/provar | try (out) |
| realize | darse cuenta / adonar-se | realizar/realitzar | carry out |
| resume | reanudar/reprendre | resumir | summarize |

| English | Spanish/Catalan | Spanish/Catalan | English |
|---|---|---|---|
| SL | TL | TL | SL |
| sensible | sensato/assenyat | sensible | sensitive |
| sympathetic | comprensivo/comprensiu | simpático/simpàtic | friendly, nice |
| topic | tema | tópico/tòpic | cliché |
| violate | transgredir | violar | rape |

## 1.2. Cognates

So far, we have learnt the concept of false friend and seen some examples. On the other extreme we find cognates. A cognate is a word which is very similar in both form and meaning in both languages. There are several reasons that can explain why some languages share very similar or identical lexical items: the word may be a borrowing from the other language, or else, both languages borrowed it from a third language. Another possibility is that they can be traced to the same etymological source, that is, they share the same origin. Cognates are believed to help students learn a foreign language since being aware of them is an easy way to enlarge their L2 lexicon. For this reason, you can find many paper and Internet-based sources providing lists of cognates. Find below (Table 2) some examples:

**See also**

For a more comprehensive list, visit the sites in the Section «Useful Online Resources».

Table 2: List of cognates in English, Spanish and Catalan

| English | Spanish | Catalan |
|---|---|---|
| map | mapa | mapa |
| centre | centro | centre |
| galaxy | galaxia | galàxia |
| generic | genérico | genèric |
| obtain | obtener | obtenir |
| sacrifice | sacrificio | sacrifici |
| yoga | yoga | ioga |
| zebra | cebra | zebra |

The ending of some words can help us identify possible cognates. This is the case of words ending in *-al* and *-or*, which derive from Latin:

    *-al*: animal, hospital, manual, etc.

    *-or*: actor, director, etc.

New cognates are created in the three languages by means of adding the same, or a very similar suffix, to the same root. On some occasions, there maybe a small change in the spelling of the suffix:

| Suffix | English | Spanish/Catalan |
|---|---|---|
| **-ist/-ista** | idealist | idealista |
| | narcissist | narcisista/narcissista |
| **-ism/-ismo/-isme** | idealism | idealismo/idealisme |
| | communism | comunismo/comunisme |
| **-nce/-ancia/-ància** | abundance | abundancia/abundància |
| | intelligence | inteligencia/intel·ligència |
| **-ty/-dad/-tat** | curiosity | curiosidad/curiositat |
| | sincerity | sinceridad/sinceritat |
| **-tion/-ción/-ció** | association | asociación/associació |
| | inspiration | inspiración/inspiració |
| **-y/-ía/-ia/-io-i** | agony | agonía/agonia |
| | salary | salario/salari |

Just as we have seen above, there are other word categories, such as adjectives, that also present parallel word formation processes:

| Suffix | English | Spanish/Catalan |
|---|---|---|
| -al | accidental | accidental |
| | brutal | brutal |
| -ble | comfortable | confortable |
| -ive/-ivo-a/-iu-iva | active | activo-a/actiu-va |
| | abusive | abusivo-a/abusiu-va |
| -nt/-ante/-ant | abundant | abundante/abundant |
| -ous/-oso-a/-ós-osa | ambitious | ambicioso-a/ambiciós-osa |
| | delicious | delicioso-a/deliciós-osa |

## 1.3. Partial cognates / Partial false friends

Before starting this subsection, remember that words are pairs of form and meaning. However, one single form can have more than one meaning.

Lemmas can sometimes be polysemic, that is, they can have more than one sense (meaning) and, for this reason, it is possible to find that one of the senses of a lemma in one language (SL) is a cognate with a word in the other (TL) whereas another sense in the SL corresponds to a different word in the TL. This happens because the senses of a lemma do not necessarily have a complete correspondence. Therefore, we sometimes use the word *partial* when we talk about cognates and false friends.

Partial false friends and partial cognates are used as synonyms in this section. Below we present a brief list (Table 3) of partial false friends. As you can see, there are two directions in which these words can be analysed, namely from Spanish or Catalan into English and from English into Spanish or Catalan. We provide two translations (words), separated by a slash (/) when there is not a complete correspondence of the senses of a lemma and there are 2 possible equivalent words in the TL.

Table 3: List of partial false friends in English, Spanish and Catalan

| Partial false friends / Partial cognates | | |
|---|---|---|
| **English** | **Spanish** | **Catalan** |
| action/share | acción (L1) | acció (L1) |
| alternatively (L1) | alternativamente | alternativament |
|  | por otro lado | d'altra banda |
| blank/target/white | blanco (L1) | blanc (L1) |
| brave (L1) | valiente/bravo | valent/brau |
| demand/lawsuit | demanda (L1) | demanda (L1) |
| familiar/relative | familiar (L1) | familiar (L1) |
| figure (L1) | cifra/figura | xifra/figura |
| paper (L1) | papel/trabajo | paper/treball |
| table (L1) | mesa/table | taula |
| theatre (L1) | teatro/quirófano | teatre/quiròfan |
| story/history | historia (L1) | història (L1) |

**Note**

The base form of a word in lexicography is known as *lemma* and each meaning associated to this form is called a *sense*.

**Example**

The word *acción-acció* (taking Spanish and Catalan as the SL) has 2 senses and, depending on the sense we want to translate, we have two possible equivalent lexical items in English: *action* or *share*.

**See also**

For a more comprehensive list, visit the sites in the Section «Useful Online Resources».

**Note**

As can be observed in the list shown in Table 3, the lexical items presented can be considered either false friends, in one of the senses, or cognates, in the other. This is why they are called partial false friends / partial cognates.

## 1.4. Other possibilities

Obviously, the possibilities are more extensive than the categories we have just seen. Many lexical items in different languages refer to the same object or idea; for this reason, they can be considered to be similar in meaning but they can have a complete different form (in the languages we are describing this is due to a different etymology).

Again, this similarity in meaning can be either total or partial since words can have different senses. Besides, the words in the two languages can also have different connotations, in the sense that some figurative or vulgar uses do not always coincide. Below you can find some examples of total similarity (Table 4.1), partial similarity (Table 4.2) and different connotations (Table 4.3):

Table 4.1: Total similarity in meaning

| English | Spanish | Catalan |
|---------|---------|---------|
| skirt | falda | faldilla |
| homework | deberes | deures |

Table 4.2: Partial similarity in meaning

| English | Spanish | Catalan |
|---------|---------|---------|
| high/tall | alto | alt |
| short | corto/bajo | curt/baix |

Table 4.3: Different connotations in meaning

| English | Spanish | Catalan |
|---------|---------|---------|
| cock[1] (taboo) | gallo | gall |
| bloody[2] | sangriento | sangnant |

[1] *rooster* usually used.

[2] In English it is a commonly used swear word.

# 2. Lexical divergences

As we have already observed, the analysis of lexical items provided in the Section «Lado's classification» is mainly concerned with the way pairs of words relate, with the aim of using this contrastive analysis to deal with vocabulary teaching in an FL class. As has been already mentioned, this type of classification was key within the field of second language teaching (SLT) and is still important and valid. Nevertheless, when comparing the lexicons of two languages at a more theoretical level, or from the point of view of translation, for example, we encounter different types of problems not mentioned in Lado's classification.

As has been said, in this section we are mainly concerned with problematic cases. There are several reasons why there might not be a direct correspondence between a word in the TL and the SL, and most of them are due to language idiosyncrasies. Again, when such a lack of correspondence happens at the lexical level, it can either be total or partial, as we will see. This lack of direct correspondence is called **lexical discrepancy**. Janssen (2004) proposes a summary of the types most commonly described from a theoretical perspective:

- When we explore the syntactic properties of some lexical items (especially verbs) we find **syntactic divergences**. For example, the verb *buscar* (in both Spanish and Catalan) requires a DO, when we describe the object being searched for or a prepositional phrase with the preposition *en/a*, when we describe the location, the place where the search is conducted. In English, the equivalent verb *search* subcategorizes a prepositional phrase (PP) (for) when we mention the object being searched for and a noun phrase (NP) when we identify the place:

  Buscó la libreta / Va buscar la llibreta / He searched for his notebook.

  Buscó en la habitación / Va buscar a l'habitació / He searched the room.

- Differences in the lexicalization of meaning: both languages can express a specific meaning but they use a different linguistic mechanism in order to convey it (Fernández Montraveta, 2000). That is, the SL lexicalizes a meaning in just one word whereas the TL transmits the same meaning using more than just one word, (e.g. a phrase in which a more general term is used and then it is further specified by an adjunct). Sometimes, it can also be the case that the meaning or concept does not have a lexical item to express it in the TL and, therefore, an explanation is required.

  in front of / opposite / en frente de / davant de

  shallow / poco profundo / poc profund

> **Note**
>
> In this section we will present some of these translation problems. The focus of the contrastive analysis this time is on the SL so we will describe the possible problems when there is a lack of complete equivalence from one lexical system (SL) into the lexical system of the TL.

> **Bibliographical reference**
>
> **B. J. Dorr** (1993). *Machine Translation: A View from the Lexicon*. Cambridge, MA: MIT Press.

> **Note**
>
> As can be seen in the examples above, this is just a formal problem. We will describe this discrepancy in the Section «Lexical mismatches».

> **Note**
>
> This problem is described in the Section «Lexical gaps».

- Divergences in connotation happen when the nuances in meaning between the equivalent words in both languages target are not exactly the same. Remember the connotations of the words: *agonia/agonía* versus *agony*.

- Differences in denotation: the denotation of the word in the SL only partially overlaps the denotation of the «equivalent» word in the TL. For example, if we compare the words: *convento* and *convent*, the Spanish word refers to both the building and the religious community. Also, it can be used for both men and women. On the other hand, the English word *convent* only refers to the building in which a community of religious women lives (Section «Differences in denotation»).

## 2.1. Lexical gaps

**Lexical gaps** are also known as **lexical holes**. Gaps are found at several levels of linguistic analysis, but in this section, we are only concerned with those gaps found at lexical level. In this section, we will also consider lexical gaps at the phrase level, so idioms and restrictive collocations are also reviewed.

According to the definition provided in the dictionary of *Grammatical Terms in Linguistics* (Trask 1993:157), the term lexical gap refers to «the absence of a hypothetical word which would seem to fit naturally into the pattern exhibited by existing words. For example, English has sets of animal names like stallion, *mare*, *horse* and *buck*, *doe* and *deer*, but the set *bull*, *cow* lacks the obvious third term to complete the set».

From the point of view of contrastive lexicology, we consider that a lexical gap between 2 or more languages happens when a word that exists in the SL does not have an equivalent counterpart in the TL. The reason is probably because this latter language (TL) has not lexicalized that particular meaning. That is, the word is missing in one of the languages.

> From the point of view of translation, lexical gaps refer to lexical items for which there is no direct translation in the TL, but if you are a translator you might need to translate them nonetheless and, therefore, you need to find a solution.

A lexical gap, thus, happens when a concept is expressed in the SL using a word whereas in the TL we need to use a free combination of words in order to convey the same concept or else use a more generic term and it is underspecified. Let's see some examples:

| Spanish/Catalan | English |
|---|---|
| sobremesa/sobretaula | time spent at the table talking after a meal |
| merienda/berenar | (afternoon) snack |

As we have said, idioms can also pose this kind of translation problem. Let's consider the following Spanish idioms and expressions:

Tener un humor de perros

De perdidos al río

Más chulo que un ocho

As you can see, you could find alternative ways to convey the same, or a similar meaning, but the translation has to go from the global meaning to the form, which probably will be a paraphrase or an explanation of the source expression. For example, we could translate the previous idiomatic expressions as follows:

To be in a filthy mood

There's nothing to lose

To be a real show-off

As you can see, the meaning is an approximation to the meaning of the expression as a whole; these idiomatic expressions can never be translated literally.

## 2.2. Lexical mismatches

Unlike lexical gaps, **lexical mismatches** are related to how information is expressed. That is, in the SL a specific meaning is lexicalized in an item whereas in the TL this specific meaning component is expressed through a different linguistic mechanism.

There are several possible causes for this kind of mismatch. Sometimes the SL incorporates a meaning by means of a morphological process and this process is not possible in the TL. This would be the case of the examples below. As you can see a productive morphological rule is applied in Spanish and Catalan but the meaning contributed by the morphological affix is expressed in English (our TL in this example) through a different lexical item:

- Derivational suffix → two lexical items
  limon**ero**/llimon**er** → lemon **tree**
- Derivational prefix → two lexical items
  **mis**calculate → calcular mal / calcular **erròniament**

On other occasions, the SL resorts to a compounding process to create a new word and this kind of process is not possible in the TL, where two independent lexical items are needed:

**Note**

The classification of mismatches presented in this subsection is an adaptation of:

**A. Fernández Montraveta** (2000). *The Analysis Of Verbal Lexical Items And Translation Mismatches In English And Spanish: A Theoretical Framework*. PhD, Universitat Autònoma de Barcelona.

> Compounding → two lexical items
>
> wheelchair → silla de ruedas / cadira de rodes
>
> headache → dolor de cabeza / mal de cap

In the mismatches seen in the examples above, the problem is of a lexical semantic nature: meaning is expressed in just one lexical item in the SL whereas in the TL two items are required. In the examples below, we will see a different problem: one of the languages lexicalizes semantic components in a way that is not possible in the other and, thus, if we want to convey the same meaning we need to express the meaning in other constituents of the sentence. These kinds of mismatches have been discussed in Talmy and Pustejovsky works.

> Different semantic information (meaning components)
>
> To swim across → cruzar nadando / creuar nedant
>
> To run across → cruzar corriendo / creurar corrent

As you can see, the meaning component of manner, is lexicalized in the English verbs. However, in Spanish and Catalan this meaning component is usually expressed as an adjunct. According to Talmy, English, on the one hand, and Spanish or Catalan, on the other, belong to two different types of languages because of the way they lexicalize the meaning components of motion events. He calls the former type a satellite-framed language and the latter a verb-framed language.

Satellite-framed languages conflate manner of motion in the verb (***swim across***, ***run*** *across*, etc.) whereas verb-framed languages express this component, if they do, as an adjunct to the verb (*cruzar* **nadando** or *cruzar* ***corriendo***). Manner is hardly ever lexicalized in verbs in either Spanish or Catalan while in English it is quite common in motion events but also in other semantic fields like verbs of seeing (*peep, peer, blink, frown, glimpse, glance*, etc.) or verbs of saying (*say, yell, giggle, groan, grumble, growl*, etc.).

Another example of this kind of mismatch is the one provided by Pustejovsky (1991) for English. As can be noticed, it is related to the same linguistic phenomenon. The lexicalization (also called conflation) of some meaning components in English is impossible in Spanish or Catalan and, therefore, this conflated meaning is expressed in another lexical item, in this case, as an object:

> to dust → quitar el polvo / treure la pols
>
> to butter → poner mantequilla / posar mantega

Finally, the last type of lexical mismatch that we are going to review in this subsection refers to the grammatical information that is sometimes contributed by the lexical item. Occasionally, there is not a complete

**Bibliographical references**

**L. Talmy** (1985). «Lexicalization patterns: Semantic structure in lexical forms». In: T. Shopen (ed.). *Language typology and syntactic description, Vol.3: Grammatical categories and the lexicon* (p. 57-149). Cambridge: Cambridge University Press.

**J. Pustejovsky** (1991). «The Generative Lexicon». *Computational Linguistics* (vol. 17, num. 4).

equivalence in the kind of information provided by the word and then we need to use a different linguistic mechanism to express the same grammatical meaning:

- Different grammatical information: countable *versus* uncountable:
  (piece of) furniture → mueble/muebles - moble/mobles
  (piece of) news → noticia/noticias - notícia/notícies

- Different grammatical information: singular versus plural:
  People are very nice in this town → La gente es muy amable en este pueblo / La gent és molt amable en aquest poble
  The police are looking for him → La policía le está buscando / La policia l'està buscant.

As you can see, the examples above represent two different types of mismatch. In the first example, we have a word that is uncountable in English whereas in Spanish or Catalan it is countable, from the point of view of grammar. So the mechanism used to singularize it is different: in English we need to use a phrase (*a piece of*) whereas in Spanish and Catalan we use the productive zero ending for the singular and the *-s* suffix for the regular plural (and for uncountable use). The second kind of grammatical mismatch refers to the different information regarding number. Groups of people in English are usually considered plural but in Spanish and Catalan they are singular.

## 2.3.  Differences in denotation

As we already mentioned in the Introduction, the lexicon has traditionally been understood as a list of forms with a meaning associated, but it is more than just that. Lexical items, words, always denote an object from our world; it can be concrete or abstract; it can be an action or a relation. For this reason, it can be assumed that the differences between languages are not only formal or semantic; they can also be conceptual.

Three types of concepts are generally distinguished in the conceptual base of:

- **Universal concepts** are those that exist in all languages or cultures in the world; they are supposed to be innately understood and represent universally meaningful concepts. As a matter of fact, there are very few. According to some authors (Wierzbicka), in this type of universal concepts we find concepts such as *I*, *this*, *say*, *want*, *then*, *before* and *after*. Universal concepts would constitute a kind of metalanguage of meaning (Goddard & Wierzbicka).

**Bibliographical reference**

I. Kecskes, T. Papp (2000). *Foreign Language and Modern Tongue*. London: Lawrence Erlbaum Associates, Publishers.

Nevertheless, some scholars (Croft, among others) contest the existence of such universal concepts, especially from the field of typological analysis. Some data provided from this field contradicts the existence of such concepts; for example, in Japanese, there are two words for *water*, one for cold water, *mizu*, and one for hot water, *o-yu*. Another example can be found in the Yimas language of New Guinea where there's no word for *water* at all but just a word for *liquid*, *arm*, also used for other liquids such as petrol.

- **Language or culture specific concepts** are those particular to a given culture. Examples in this class are most typically concepts related to specific food or cultural traditions, e.g. *sardana*, *paella*, *Yorkshire pudding*, *Guy Fawkes*. However, there are others, such as *agobiado*, *atabalat*, *saudade* that are more related to the feelings (or view of the world) shared by a community of speakers.

- **Decompositional concepts** are those that share some features with the same concept in another language, but not all of them. They usually refer to complex entities such as social relations or social celebrations: e.g. *family*, *wedding*, *party*, etc. As you must be aware of, the way a wedding or a funeral is celebrated varies in different cultures, although there are parts of it that might be common. For this reason, the translation of these concepts is always an approximation to the source meaning.

**Bibliohgraphical references**

A. Wierzbicka (1996). *Semantics: Primes and Universals*. Oxford University Press.

C. Goddard, A. Wierzbicka (2007). «Semantic primes and cultural scripts in language learning and intercultural communication». In: G. Palmer & F. Sharifian (eds.). *Applied Cultural Linguistics: Implications for second language learning and intercultural communication* (p. 105-124). Amsterdam: John Benjamins.

W. Croft (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. London: Oxford University Press.

# 3. Current trends in Contrastive Lexicology

The position of the lexicon as a level of linguistic analysis has undergone changes over the years: from being considered a secondary level of linguistic analysis now it occupies a central position in several theoretical approaches. For example, within the framework of *Systemic Functional Grammar*, Halliday was one of the first linguists to defend its importance. From the postulates of this framework, the grammatical and lexical levels are intimately intertwined, being idioms and collocations the proof of this close relation.

Sinclair, from a different theoretical approach, also defended the centrality of the lexicon and its relation to grammar by postulating the existence of what he called **extended units of meaning**.

> Extended units of meaning are phrases that go beyond single words but possess just one referential meaning.

Many properties that had traditionally been described from the grammatical level started to be regarded as properties that could be best explained from the lexicon (Sinclair). From a pure lexicologist approach, some grammatical properties, such as subcategorization patterns, are believed to be projections of lexical properties (Levin). Following these linguistic trends, in the last few decades, contrastive lexicology has also shifted its focus of interest.

Regarding methodology, a really important change that happened in the 80s modified the methods used in contrastive linguistics, and, in fact, in linguistics in general. It is the framework known as the **empiricist movement**. This approach defends the necessity to study language from the point of view of the real performance of its speakers, its authenticity, and move away from introspection as the only research method in linguistic description. Thus, linguist introspection is abandoned, and the observation of real language becomes central. This new methodological approach is known as **Corpus Linguistics**.

Kučera and Francis' computational analysis of the Brown Corpus constitutes a fundamental work in the field of corpus linguistics. They analyzed American English from several perspectives such as language teaching, sociology or psychology, among others. *The American Heritage Dictionary of the English*

**Bibliographical reference**

**M. Halliday** (1961). «Categories of the Theory of Grammar». Word (vol 17, num 3, p. 241-292).

**Bibliographical reference**

**J. Sinclair** (1996). «The Search for Units of Meaning». *Textus* (num 9, p. 75-106).
**J. Sinclair** (1998). «The lexical item». In: E. Weigand (ed.) *Contrastive Lexical Semantics* (num. 1-24). Amsterdam: John Benjamins.

**Bibliographical references**

**J. Sinclair** (1987). «Grammar in the dictionary». In: J. Sinclair (ed.). *Looking up: An account of the COBUILD Project in Lexical Computing* (p. 104-115). London and Glasgow: Collins.
**B. Levin** (1993) *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.

**Bibliographical reference**

**J. Sinclair** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

*Language* (1969) combined both descriptive with prescriptive information, that is, how language should be used and how it is actually used. It was the first time a dictionary presented information extracted from corpora, and it represented a milestone in the field of lexicology and lexicography.

**Bibliographical reference**

H. Kučera, N. Francis (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

In parallel to the construction of the Brown Corpus of American English, in England, the Survey of English Usage (University College in London) was the first center to conduct research on corpora. It was founded in 1959 by Randolph Quirk. Quirk *et al.*'s famous *Comprehensive English Grammar* also followed the empiricist approach, and is based on information compiled in corpora. Also, in England, the COBUILD series, which included a dictionary and a grammar among several other resources, followed the methodology based on corpus linguistics.

**Bibliographical reference**

R. Quirk, S. Greenbaum, G. Leech, J. Svartvik (1985). *Comprehensive English Grammar*. London: Longman.

Nowadays, to browse multilingual corpora and extract lexical information has become extremely popular. The kind of lexical information that can be extracted from corpora ranges from the study of frequencies of use of a specific word to the identification of collocations or the study of idiomatic phrases (extended lexical units). Thus, searching corpora allows us to study words in context, that is, it permits us to study lexemes in their immediate context. The study of context allows us to find the semantic preferences of words and their most frequent collocates.

## 3.1. Extended units of meaning

The new model of what constitutes lexical meaning has supposed a revolution in current lexicography and, by extension, in contrastive lexicography. In traditional lexical semantics, the unit of analysis was the lexical item, the word defined as a string of letters placed between two blank spaces. In fact, Firth was the first scholar to put forward the idea that meaning is not only found isolated in words and, therefore, he proposed collocations as another unit of analysis. Sinclair's works represented the formalization of the idea and a turning point in the field. He proposed the concept of **extended lexical units** as the unit of analysis. Some recent studies based on corpus linguistics provide support to this idea that meaning is a phraseological phenomenon (see Korning, among many others).

Sinclair's model establishes extended lexical units as the primary unit of lexical analysis. He defined them as **abstract units** which include not only the meaning conveyed by the lexical items in isolation (lexical semantics), but also incorporate the meaning contributed by the semantic and pragmatic relations established among all the words of the extended units.

From the point of view of contrastive lexicography, proposals based on the analysis of just single words, such as Lado's classification, even though still valid and useful, have been exceeded by new proposals that incorporate a more global vision of languages and can provide us with a more comprehensive understanding of how the lexicons of languages relate to one another.

In what follows we are going to review the new units of lexical analysis proposed by Sinclair: idioms, collocations (lexical collocations and grammatical collocations - i.e. colligations) and proverbs. All of them have in common that, in fact they act as a single unit of meaning that has to be learned and processed as a chunk and, therefore, a speaker or a student of an L2, or even a translator, needs to be aware of this fact. Being able to master them will give a native-like quality to students' productions. This is why these concepts are so important in contrastive analysis, foreign language teaching and translation.

### 3.1.1. Idioms

The most common definition of the term **idiom** is that it is a fixed or frozen expression whose meaning is figurative, that is, it cannot be built or predicted compositionally from the meaning of each individual word. It is precisely for this reason that a lexical item which is part of an idiom cannot be replaced by a synonym or near-synonym; if it were, the idiomatic meaning would be lost.

**Bibliographical references**

**J. R. Firth** (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.

**J. Sinclair** (1996). «The Search for Units of Meaning». *Textus* (num 9, p. 75-106).

**J. Sinclair** (1998). «The lexical item». In: E. Weigand (ed.) *Contrastive Lexical Semantics* (num. 1-24). Amsterdam: John Benjamins.

**K. Korning Zethsen** (2008). «Corpus-based cognitive semantics: Extended units of meaning and their implications for Translation Studies». *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, [S.l.] (num. 7, Jun. 2013). ISSN 2295-5739. Available at: <https://lans.ua.ac.be/index.php/LANS-TTS/article/view/218>. Date accessed: 26 Jun. 2017.

**Example**

The literal translation for the expression *hot potato* would be in Spanish *patata caliente*. Nevertheless, when it is used idiomatically it means *tema candente*. Something similar happens when you listen to somebody saying that you have heard something from *the horse's mouth* meaning that they have heard something, in fact, from *una fuente autorizada*.

As you can see, idioms have to be learnt as a unit, and we always need to check a dictionary if we are not sure how to translate them into our own language.

You can find information about idioms in specialized dictionaries. Currently, this information is also found in general monolingual and bilingual dictionaries. For example, the *Cambridge Online Dictionary* includes idioms as entries and always marks them as such. There is a section at the end of the relevant entries called *idioms* where we can find information about idiomatic expressions:

In Figure 1 we partially reproduce the relevant information for this section (*Idioms*) in the entry «finger»:

Figure 1: Idioms included in the entry «finger»



In Figure 2, we can see part of the terms presented in the Thesaurus of terms related to the word «fighting». It can be noticed how idioms are presented and marked. This is the case of *tilt at windmills* and *to the last (man)*:

Figure 2: Thesaurus by *Cambridge dictionary*: words related to «fighting»

Finally, in Figure 3 we can see another way of presenting information in the *Cambridge dictionary*. In this case, we can see all the words that are alphabetically placed before and after «fighting». What is interesting for us is how the entry «fight your corner» has been marked as an idiom:

Figure 3: Words related to «fighting» in alphabetical order
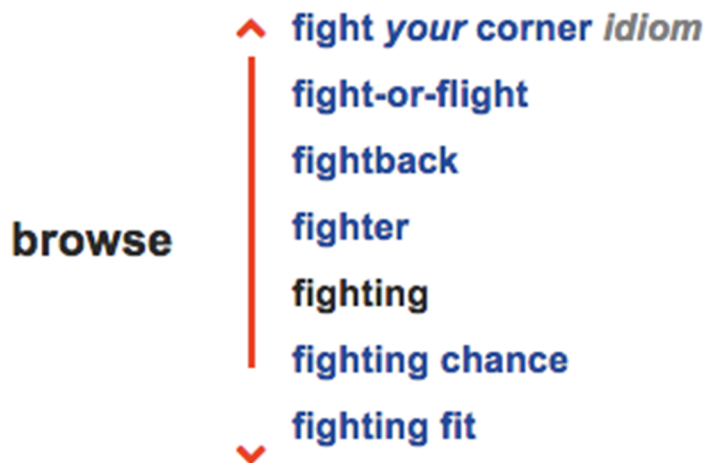


## 3.1.2. Collocations

**Collocations** are combinations of words that co-occur more frequently than would be expected by chance. Unlike idioms, the meaning of collocations is compositional and they could still be understood if one of the words was not the «right» one in the combination (the word that collocates). Nevertheless, we need to say that for a collocation to sound natural few possibilities are allowed. The correct use of collocations can be regarded as a 'proof' to native-like proficiency. Let's see some examples.

A word like *traffic* would collocate with the following adjectives, nouns or verbs, among many others:

Congested/heavy/thick **traffic**

Light **traffic**

Fast/fast-flowing **traffic**

Generate/increase **traffic**

**Traffic** jam/light

If we take in consideration the morpho-syntactic category of the lexical items in the collocation, there are six different types of possible combinations:

- Adjective + noun (usually noun + adjective in Spanish and Catalan)
  - Great joy
  - Alegría intensa
  - Crua realitat

- Noun + noun (usually noun + prepositional phrase in Spanish and Catalan)
  - Bed linen
  - Ropa de cama
  - Mal de panxa

- Verb + noun
  - Strip the beds
  - Paliar el aburrimiento
  - Esclatar una revolta

- Adverb + adjective
  - Jolly nice
  - Rematadamente mal
  - Mortalment ferit

- Verb + prepositional
  - Dream of
  - Pensar en
  - Somiar amb

- Verb + adverb
  - Eat properly
  - Vivir intensamente
  - Anar endavant

As can be observed in the examples above, knowledge about collocations is a kind of lexical knowledge that cannot be learned by rules. It is idiosyncratic, usually language dependent, and needs to be learned as regular vocabulary of the L2, even though collocations go beyond the natural boundaries of what has traditionally been considered a word.

The limits between idioms, restricted collocations and free combinations are not always clear. In many dictionaries, idioms and collocations are nowadays explicitly marked, particularly in modern dictionaries. A selection of such resources is presented at the end of this unit under the epithet.

**See also**

See the Section «Useful Online Resources».

### 3.1.3. Colligations

The term **colligation** was first coined by Firth in 1968. He defined colligation as the relation established between a word and the words in its immediate context, but, unlike other lexical relations, this relation defines the grammatical level, so it is a relation between a word and the grammatical categories that can appear next to it more frequently. They are also called **grammatical collocations**.

Hoey calls this kind of relation «lexical priming» and he defines it in the following terms: «a lexical item may be primed to occur in or with a particular grammatical function». In order to exemplify this concept, we can borrow Sinclair's example: the unit *true feelings* has a strong tendency to colligate with a possessive adjective whereas a unit such as *true story* with a determiner:

**my** true feelings

**his** true feelings

**their** true feelings

**a** true story

**the** true story

**that** true story

In Spanish and Catalan we can also find examples of colligations such as the following:

*en* (demonstrative) *sentido → en este/ese/aquel sentido*

*en* (demonstrative) *sentit → en aquest/aquell sentit*

### 3.1.4. Proverbs

Finally, we will briefly review **proverbs**. Formally, proverbs are usually complete sentences but, from the point of view of meaning, they are rather similar to idiomatic expressions. They usually express cultural or traditional truths which are shared by a community of speakers. They tend to refer to biblical stories and social conventions metaphorically and, occasionally, they present a moral that might be shared by the source and the target cultures. Speaking of English and Spanish or Catalan, this is usually the case. They belong to popular knowledge and are usually learned by repetition from older generations to newer.

Currently proverbs are included in regular dictionaries, usually at the very end of the lexical entry. In the dictionary we briefly discussed above, proverbs are included as entries and are, in fact, labelled as idioms (see Figure 4). Find below a partial reproduction of the thesaurus for the entry «engagement». We can notice the proverb *marry in haste, repent at leisure* with the word *idiom* right next to it.

Figure 4: Example of a proverb in the *Cambridge Online Dictionary*



Proverbs have sometimes a direct equivalent but sometimes, as was the case with idioms, they have to be translated from the global meaning, from the moral they convey:

A bird in the bush is worth two in the hand

Más vale pájaro en mano que ciento volando

Més val ocell al puny que una grua lluny

A stitch in time saves nine

Más vale prevenir que curar

Val més prevenir que curar

# 4. Anglicisms: neologisms and borrowings

**Anglicisms** are words or phrases that have been borrowed from English into another language. In our case, we are interested in those anglicisms that are part of the Spanish and Catalan lexicons. As you must be aware of, there are many words in these two languages that come from English, particularly in the fields of technology, economics, sports and media.

The reason for these **borrowings** is diverse. First of all, we need to consider the status of English as *lingua franca*. For example, many words from the worlds of sports and media have entered other languages, either because they did not have an equivalent in the TL or else for reasons of prestige. In addition, many new technological words and concepts have appeared recently, that is, the concept or object did not exist previously.

> **Example**
>
> This would be the case of lexical items such as *the Internet*, or *personal computer* (PC), and have later been adopted by other languages.

**Neologisms** are «newly coined lexical units or existing lexical units that acquire a new sense». (Newmark). Three stages have been described in the process of creation of a neologism: when they are first created and are, therefore, **unstable**; in the second phase they start being used, i.e. **diffused**, but not completely accepted and, finally, they are said to be **stable** when they are widely accepted by the speakers. The translation of neologisms poses a big problem for translators, especially when they are in the first 2 phases.

> **Bibliographical reference**
>
> **P. A. Newmark** (1988). *Textbook of Translation*. Prentice Hall International (p. 140).

According to Newmark there are 12 classes of neologisms. First, he distinguishes between those lexical items that already existed but start being used for a new sense to denote a new concept or object (neologism of sense) from those words that are completely new (neologism of form). He further subdivides the former group according to their form: simple (words) or complex (collocations). Let's see an example for each of the above categories:

> A word such as *entrada*, for example, has acquired a new sense when used in the domain of technology. *Entrada*, when used with this meaning, refers to the port used for connections in computers. An example of the second type, collocation, would be *tren de alta velocidad* (known in Spanish as AVE) used to describe the new kind of train (object) but with words that were not neologisms.

In the latter group of neologisms, new forms that appear in a language, he distinguishes classes according to the process of creation. Below, we review just the most important ones:

- **New coinages**. New words that appear as a result of coinage (a pure invention), examples of this kind are *zipper* or *quark*. Sometimes they are

brand names, when this is the case usually the original names are kept in other languages.

- **Derived words**. The great majority of neologisms belong in this group. Words such as *teleconference* or *animatronics* are the result of a derivation process. The translation of these words is usually rather literal and language-dependent. They will follow the equivalent derivational process in the TL, whenever it makes sense.

- **Abbreviations** are the next type in this classification with *RSS* or *TCP* as examples. Abbreviations are commonly borrowed as they are in the SL: *www* (*world wide web*), *IP* (*internet protocol*), *HTML* (*hypertext mark-up language*). The examples mentioned are all abbreviations commonly used in the technological field.

- New **collocations** are very common, particularly in technology and social sciences: *e-book*, *netbook*, etc. The tendency with this type of neologisms, especially in technology, is to borrow the word to the TL. This tendency is not observed in other fields. Sometimes they are translated as calques: *acid rain* was translated as *lluvia ácida / pluja àcida*.

- **Eponyms** are neologisms derived from a proper noun or brand, for example *beliebers* or *nylon*. They are usually well accepted and therefore maintained in the TL after they are known.

- **Phrasal noun**. In this category we find new nouns or adjectives that appear as the result of transforming a phrasal verb into a phrasal noun. This possibility is obviously restricted to English with examples such as *add-on* or *lock-out*. Neologisms of this kind are always translated into Spanish or Catalan by the appropriate target form, if it exists:

add-on → añadido/afegit

laid-back → relajada/relaxada

lock-out → cierre patronal / tancament patronal

- **Transferred** words are neologisms that are borrowed from another language, for example *ponzu* or *taekwondo*. From the point of view of translation, they are often kept as they are in the SL if they are popular in the target culture, as is the case of the word *taekwondo*. Otherwise, they are sometimes translated with either a functional-descriptive equivalent or with a more generic one.

ponzu → salsa ponzu

taekwondo → taekwondo

# Summary

In this unit we saw in the Section «Lado's classification» an adaptation of Lado's classification (1957), which represents the first serious attempt to lexical analysis from a contrastive perspective. This classification presents a typology of relations established between two equivalent word senses in terms of form and meaning. It supposed a major milestone in the area of Contrastive Analysis and it was devised with the purpose of creating appropriate teaching materials in the area of Foreign Language Teaching.

In the Section «Lexical Divergences», we moved on to review the concept of Lexical Divergences and we presented other classifications of lexical problems that were not accounted for in Lado's proposal since they are studied from the perspective of translation. Thus, in this section, we reviewed how lexical items in the SL are translated into the TL and the kind of problems that arise when there is not a complete equivalence between the concepts denoted by the lexical items in two languages or the grammatical information does not fully coincide.

In the Section «Current trends in Contrastive Lexicology» some current trends in cross-linguistic lexical studies were discussed. More precisely, we reviewed the concept of extended lexical units and highlighted the important role they play nowadays in lexicology.

Finally, in the Section «Anglicisms: neologisms and borrowings» we briefly analyzed the concepts of neologism and borrowing and very briefly commented on some translation strategies and problems related to them.

# Self-evaluation

**1)** In the following list you can find some words in different languages and how they are translated into English. Analyze the relations you see in terms of the semantic features expressed in the following English pairs: *leg/foot*, *arm/hand*, *finger/toe*. Think about the features which English uses to differentiate each member of the pair. Write a contrastive statement for each translation:

**a)** jalka (Finnish) *vs*. leg, foot
**b)** nogá (Russian) *vs*. leg, foot
**c)** käsi (Finnish) *vs*. arm, hand
**d)** ruká (Russian) *vs*. arm, hand
**e)** dedo (Spanish) *vs*. finger, toe
**f)** kidole (Swahili) *vs*. finger, toe

**2)** Classify the following contrasts according to Lado's classification presented in the Section «Lado's classification»:

**a)** asylum → asilo/asil
**b)** eventual → eventual
**c)** particular → particular
**d)** vase → vaso/got
**e)** abandon → abandonar

**3)** Following the explanations provided in the Section «Lexical mismatches», decide whether the following concepts are universal, culture-specific or decompositional:

**a)** uncle
**b)** tea
**c)** tree
**d)** privacy
**e)** yesterday

**4)** Classify the following contrasts according to Lado's classification:
**a)** preservative → preservativo/preservatiu
**b)** wedding → boda
**c)** admit → admitir/admetre
**d)** fly (N) → mosca
**e)** blanco/blanc → blank (Adj)
**f)** silla/cadira → chair (N)

**5)** Write a contrastive statement comparing the colligation of the following items in English with the equivalent in Spanish or Catalan:
**a)** continue to + INF
**b)** POSSESSIVE + part of the body
**c)** want + DO + *to*-infinitive
**d)** in QUANTIFIER cases
**e)** have got LIQUID all over ITEM OF CLOTHING

**6)** Find proverbs in Spanish/Catalan equivalent to the following English ones. For each proverb provide a suitable subject or topic to which it could be indexed: death, deception, false friendship, etc.:
**a)** Actions speak louder than words
**b)** To add insult to injury
**c)** Half a loaf is better than no bread
**d)** One man's meat is another man's poison
**e)** A leopard never changes its spots
**f)** Who will bell the cat?
**g)** Birds of feather flock together
**h)** There's no smoke without fire
**i)** Money goes where money is
**j)** It's no crime to steal from a thief
**k)** A stitch in time saves nine

**7)** Find a suitable translation for the following idioms:
**a)** Against all odds
**b)** In a jiffy
**c)** Hit the ceiling
**d)** Armed to the teeth
**e)** Down the drain

**f)** To drive a hard bargain
**g)** Make a fool of
**h)** To be a drag
**i)** To drop the ball
**j)** A dime a dozen

**8)** Find the equivalent suitable collocation in your own language:
**a)** highly skilled
**b)** extremely unhappy
**c)** bar of chocolate
**d)** draw a conclusion
**e)** to waste time
**f)** standing ovation
**g)** waste management
**h)** low cut
**i)** high season
**j)** totally awesome

**9)** Explain the phenomena that can be observed in the following pairs:
**a)** Access the building → Acceder al edificio / Accedir a l'edifici
**b)** Quark → Quark
**c)** Misappropiation → Apropiación indebida / Apropiació indeguda
**d)** Wash something away → Lavar algo / Rentar alguna cosa
**e)** Finger → Finger

**10)** Describe the processes of creation of the following neologisms and how they have been translated into Spanish and Catalan:
**a)** Kleenex → Kleenex
**b)** Reset → resetear
**c)** Sushi → sushi
**d)** Troll → trol, trolear / trol, trolejar
**e)** Roast-beef → rosbif

# Answer key

## Self-evaluation

**1.** English distinguishes between the parts of the body that are separated by the ankle whereas Finnish and Russian do not. That is, it has a different name for each: above the ankle–*leg*– and below it–*foot*–. Something similar happens with the difference between the nouns *arm* and *hand*. They are differentiated because they are located above or below the wrist.

Also, English distinguishes between the end joint members of the hand versus the end joint members of the foot whereas Spanish and Swahili do not.

**2. a)** Partial cognates because they share the meaning of «nursing home for the elderly» but they do not share all the senses. For example, in Spanish/Catalan, *asilo/asil* is not used for psychiatric institutions as it is in English.
**b)** False friends; the main senses in Spanish/Catalan are «occasionally and possible». In English, the main sense is «in the end», therefore it should be translated as *al final*.
**c)** Partial false friends. They share the meaning of «odd» or «unusual», «special» or «specific». Nevertheless, in Spanish it is commonly used with the sense of «private» or «personal», for example, the phrase *clases particulares / classes particulars* would be in English *private lessons*.
**d)** False friends. They do not hold an equivalence relation. Both words describe a kind of container but there is a difference regarding the function and the size of the object. *Vase* is translated as *jarrón* (Spanish) or *gerro* (Catalan) and *vaso* (in Spanish) and *got* (in Catalan) is translated into English as *glass*.
**e)** Cognates; they have the same form and are very similar in meaning.

**3. a)** *uncle*. This concept is culture specific. In general, the concept of family and the definition who is a family member varies across cultures.
**b)** *tea*. The word *tea* used to talk about the dried leaves or the beverage is shared by all three languages, but «the tea» as a ceremony is a decompositional concept since the act of having a cup of tea is very different in countries such as Japan or England than it is in Spain.
**c)** *tree*. A tree, as a plant, is universal. Nevertheless, depending on the location, we can find different types of trees.
**d)** *privacy*. Some concepts, as the concept of privacy, vary across cultures because the importance attached to it is not the same. For example, English people put their right to privacy before the rights of other, such as family members, to interfere on other people's lives. In our culture this is not usually seen this way.
**e)** *yesterday*. The idea that time exists is universal; the concepts of now (present), before now (past) and after now (future) are believed to be present in all the cultures.

**4. a)** False friends. They share the same form but have a completely different meaning. *Preservative* in English refers to a product use to preserve food whereas *preservativo* (*preservatiu*) is a condom.
**b)** The words are similar in meaning but different in form. It is a decompositional concept, because the parts of the ceremony or celebration can be very different depending on the culture or society.
**c)** Cognates. They share the same form and meaning: to allow something or somebody to enter, for example.
**d)** They have the same meaning in this sense, but *fly* in English has more senses; for example it is also translated as *cremallera* (in both Spanish and Catalan). For this reason if we consider the lemma they are partial cognates.
**e)** False friends. Since, as an adjective, *blank* means «empty which would be translated as *vacío* (in Spanish) o *buit* (in Catalan).
**f)** They have one sense (meaning) in common but have a completely different form. In English, *chair* is also used to describe a seat with arms, butaca in Spanish and Catalan. In addition, in English *chair* is also used to describe a position of authority, for example at the university (*cátedra* in Spanish) or the person occupying that position.

**5. a)** continue to + INF → continuar + gerund. The English verb subcategorizes an infinitive, whereas both Spanish and Catalan require a gerund after *continuar*.
**b)** Usually, body parts in English are modified by the possessive adjective whereas in Spanish and Catalan determiners are more common: *I have broken my arm vs. Me he roto el brazo / M'he trencat el braç*.
**c)** Again, the verb *want* colligates with a different grammatical category in Spanish and Catalan than it does in English. In these two languages it requires a *-that clause* in which the direct object (DO) of the English structure becomes the subject of the clause: *I want you to go → quiero que (tu) vayas / vull que (tu) hi vagis*.

**d)** This colligation is common to all three languages: *in many/several cases*; *en muchos/ bastantes casos*; *en molts/bastants casos*.

**e)** This is a colligation because *all over* collocates with the semantic class of «items of clothing». Neither in Spanish nor in Catalan does this particular grammatical collocation make much sense.

**6. a)** Actions speak louder than words → Obras son amores y no buenas razones / Fets i no paraules. (A piece of advice on behavior.)

**b)** To add insult to injury → Echar sal a la herida / Tirar sal a la ferida. (Emphasis.)

**c)** Half a loaf is better than no bread → Más vale pájaro en mano / Més val ocell al puny que una grua lluny. (Wisdom: be practical.)

**d)** One man's meat is another man's poison → Sobre gustos no hay nada escrito / Per a gustos, colors. (Wisdom: there are different opinions. Respect them.)

**e)** A leopard never changes its spots → Genio y figura hasta la sepultura / Geni i figura fins a la sepultura. (Human behavior: people don't change.)

**f)** Who will bell the cat? → ¿Quién pondrá el cascabel al gato? / Qui li posarà el cascavell al gat? (Behavior: who will take the risk?)

**g)** Birds of feather flock together → Dios los cría y ellos se juntan / Els que s'assemblen se cerquen. (Human behavior: similar people behave in a similar way.)

**h)** There's no smoke without fire → Cuando el río suena agua lleva / Quan el riu sona, alguna cosa porta. (A piece of advice, warning.)

**i)** Money goes where money is → Dinero llama dinero / Diners fan diners. (General wisdom.)

**j)** It's no crime to steal from a thief → Quien roba a un ladrón tiene cien años de perdón / Qui roba a un lladre té deu anys de perdó. (Sense of justice.)

**k)** A stitch in time saves nine → Más vale prevenir que curar / Val més prevenir que curar. (Prevention, being careful.)

**7. a)** Against all odds → Contra todo pronóstico / Contra tot pronòstic

**b)** In a jiffy → En un periquete / En un tres i no res

**c)** Hit the ceiling → Encolerizar(se)/Enrabiar-se

**d)** Armed to the teeth → Armado hasta los dientes / Armat fins a les dents

**e)** Down the drain → Por el desagüe / Anar pel desguàs

**f)** To drive a hard bargain → Saber regatear / Saber regatejar

**g)** Make a fool of → Ridiculizar a / Ridiculitzar

**h)** To be a drag → Ser una pesadez / Ser un plom

**i)** To drop the ball → Meter la pata / Dicar la pota

**j)** A dime a dozen → (los hay) a patadas / (n'hi ha) un fotimer

**8. a)** highly skilled → altamente cualificado / altament qualificat

**b)** extremely unhappy → muy desgraciado / molt desgraciat

**c)** bar of chocolate → tableta de chocolate / rajola de xocolata

**d)** to draw a conclusion → sacar una conclusión / treure una conclusió

**e)** to waste time → perder el tiempo / perdre el temps

**f)** standing ovation → ovación cerrada / ovació càlida

**g)** waste management → gestión de residuos / gestió de residus

**h)** low cut → escotado/escotat

**i)** high season → temporada alta

**j)** totally awesome → increíble/increïble

**9. a)** In English we have a transitive verb whereas in Spanish and Catalan the verb *acceder/ accedir* subcategorizes a prepositional phrase (PP).

**b)** This is a neologism that has been borrowed in Spanish and Catalan just with the same form it has in English.

**c)** English presents a derivation process (adding the prefix *mis-* a new word is created). In Spanish and Catalan, on the other hand, we observe a noun modified by an adjective, which is providing the information (meaning) contributed by the English prefix.

**d)** There is a semantic loss since some meaning is lost in Spanish and Catalan. *Wash away*, if analyzed following Talmy's proposal, incorporates manner (*wash*) and the particle *away* incorporates some meaning (path) to the event. It is not possible to translate literally this expression into Spanish or Catalan. It would be something like: *quitar una mancha (lavándola) / treure una taca (rentant-la)*, which is not a regular expression in these languages.

**e)** The lexical item *finger* has a regular translation into Spanish and Catalan, *dedo/dit*. Nevertheless, in English it is also used to define things which are similar to a finger in shape or use, for example, the walkway to a plane. In this sense, it is considered a neologism and has been transferred into Spanish and Catalan with the English form.

**10. a)** It is a brand name. It is an eponym and the proper noun has been kept in both Spanish and Catalan, In fact, currently it is used as a common noun to designate paper tissues in general.

**b)** The English neologism has been literally translated into Spanish and Catalan and the proper morphological process in these languages are applied if necessary. In Catalan the terms *buriner* (noun) and *burinar* (verb) are proposed.

**c)** It is a borrowing from Japanese into all three languages under consideration in this unit, that is, English, Spanish and Catalan.

**d)** There has been a morphological adaptation of the English word to form the noun and the verb. We can see a simplification of the double consonant in Spanish/Catalan (*trol*) when it is a noun. The verb is formed by adding the suffix *-ear/-ejar*.

**e)** Again, there has been an orthographic adaptation of the English sounds to the Spanish/ Catalan writing conventions.

# Bibliography

**Basic works**

**Lado, R.** (1957). *Linguistics across Cultures: Applied Linguistics for Language Teachers*. University of Michigan Press: Ann Arbor.

This book constitutes one of the seminal readings in the area of Applied Linguistics. In this work, Lado presents an introduction to the methodology he proposes to compare languages at phonological, lexical and syntactic level, taking into account also the culture and writing systems of the languages compared. The Contrastive Analysis Hypothesis emerges from the structuralist approach to language analysis, much en vogue at that time, and the audio-lingual methodology use at the time to teach second languages.

**Sinclair, J.** (1998). «The lexical item». In: E. Weigand (ed.) *Contrastive Lexical Semantics* (num. 1-24). Amsterdam: John Benjamins.

J. Sinclair presents his model for lexical analysis in several very well known publications. In his works (1996, 1998), Sinclair puts forward his concept of what he denominates *Extended units of meaning*. His extended units of meaning go beyond the traditional concept of word (lexical item) and also takes into account the immediate context in which words more frequently appear. He includes 5 possible combinations: core word (or words), grammatical preferences (colligations), lexical preferences (collocations), semantic preferences and semantic prosody.

**Talmy, L.** (1985). «Lexicalization patterns: Semantic structure in lexical forms». In: T. Shopen (ed.). *Language typology and syntactic description, Vol.3: Grammatical categories and the lexicon* (p. 57-149). Cambridge: Cambridge University Press.

This seminal paper by L. Talmy addresses the relation between meaning and its formal expression in surface elements (words) such as verb, subordinate clause, etc. This author is interested in studying how meaning (motion, path, figure or ground) is expressed, i.e. lexicalized, in these elements. He studies how motion events are expressed in several languages and presents a classification according to the relation established between meaning components and surface expressions.

**Further reading**

**Croft, W.** (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. London: Oxford University Press.

**Dorr, B. J.** (1993). *Machine Translation: A View from the Lexicon*. Cambridge, MA: MIT Press.

**Fernández Montraveta, A.** (2000). *The Analysis Of Verbal Lexical Items And Translation Mismatches In English And Spanish: A Theoretical Framework*. PhD, Universitat Autònoma de Barcelona.

**Firth, J. R.** (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.

**Goddard, C.; Wierzbicka, A.** (2007). «Semantic primes and cultural scripts in language learning and intercultural communication». In: G. Palmer & F. Sharifian (eds.). *Applied Cultural Linguistics: Implications for second language learning and intercultural communication* (p. 105-124). Amsterdam: John Benjamins.

**Halliday, M.** (1961). «Categories of the Theory of Grammar». *Word* (vol 17, num 3, p. 241-292).

**Hoey, M.** (2005). *Lexical Priming: a New Theory of Words and Language*. London; New York: Routledge/AHRB.

**Janssen, M.** (2004). «Multilingual lexical databases, lexical gaps, and SIMuLLDA». *International Journal of Lexicography* (vol 17, num 2).

**Kecskes, I.; Papp, T.** (2000). *Foreign Language and Modern Tongue*. London: Lawrence Erlbaum Associates, Publishers.

**Korning Zethsen, K.** (2008). «Corpus-based cognitive semantics: Extended units of meaning and their implications for Translation Studies». *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, [S.l.] (num. 7, Jun. 2013). ISSN 2295-5739. Available at: <https://lans.ua.ac.be/index.php/LANS-TTS/article/view/218>. Date accessed: 26 Jun. 2017.

**Kučera, H.; Francis, N.** (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

**Levin, B.** (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.

**Newmark, P. A.** (1988). *Textbook of Translation*. Prentice Hall International (UK).

**Pustejovsky, J.** (1991). «The Generative Lexicon». *Computational Linguistics* (vol. 17, num. 4).

**Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.** (1985). *Comprehensive English Grammar*. London: Longman.

**Sinclair, J.** (1987). «Grammar in the dictionary». In: J. Sinclair (ed.). *Looking up: An account of the COBUILD Project in Lexical Computing* (p. 104-115). London and Glasgow: Collins.

**Sinclair, J.** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

**Sinclair, J.** (1996). «The Search for Units of Meaning». *Textus* (num 9, p. 75-106).

**Wierzbicka, A.** (1996). *Semantics: Primes and Universals*. Oxford University Press.

**Sinclair, J.** (1987). «Grammar in the dictionary». In: J. Sinclair (ed.). *Looking up: An account of the COBUILD Project in Lexical Computing* (p. 104-115). London and Glasgow: Collins.

**Sinclair, J.** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

**Sinclair, J.** (1996). «The Search for Units of Meaning». *Textus* (num 9, p. 75-106).

# Annex. Useful Online Resources

- Cognates
    - <http://spanishcognates.org/>
    - <http://www.cognates.org/>
    - <http://www.esdict.com/English-Spanish-Cognates.html#.WUK6IMn-t7o>
    - <https://www.thoughtco.com/almost-but-not-quite-partial-cognates-3078345>

- False friends
    - <http://www.xtec.cat/~vfeliu/writing/false.htm>
    - <http://spanishskulduggery.tumblr.com/post/56118711025/partial-false-cognates>
    - <http://www.elearnspanishlanguage.com/vocabulary/falsosamigos.html>

- Calques
    - ésAdir: <http://esadir.cat/traduccio/interferencies/angles/literalismes>

- Collocations
    - Diccionario de colocaciones del español: <http://dicesp.com/paginas>
    - Dictionary of English Collocations: <http://www.freecollocation.com/>
    - Dictionary of English Collocations: <https://www.ozdic.com/collocation-dictionary>
    - Collocation in English: <http://corpus.byu.edu/bnc/>

- Idioms and expressions in English
    - <http://itools.com/tool/cambridge-international-dictionary-of-idioms>
    - <http://idioms.thefreedictionary.com/>
    - <http://www.smart-words.org/quotes-sayings/idioms-meaning.html>
    - <http://www.phrases.org.uk/index.html>

- Idioms in English and their equivalent in Spanish:
    - <http://www.mansioningles.com/idioms/idioms01.htm>

- Exercicis per practicar modismes en català i l'equivalent en anglès:
    - <http://www.xtec.cat/~mfarran/projecte/exercicis/idiomscat.htm>

- Proverbs in English with the equivalent in Spanish
    - <http://cogweb.ucla.edu/Discourse/Proverbs/Spanish-English.html>

-  –  &lt;http://www.lingolex.com/refranes.htm&gt;
-  –  &lt;https://bristolenos.com/2015/06/07/refranes-y-proverbios-en-ingles/&gt;

- Proverbs in English with an explanation: origin and meaning:
  - &lt;http://www.phrases.org.uk/meanings/proverbs.html&gt;

- Multilingual database of proverbs
  - &lt;http://cvc.cervantes.es/lengua/refranero/Default.aspx&gt;

- Expresiones y refranes en español:
  - &lt;https://expresionesyrefranes.com/lista-de-expresiones-espanolas/&gt;

- Dites en català
  - &lt;http://300.dites.cat/search/label/igualtat&gt;
  - &lt;http://usuaris.tinet.cat/jpv/&gt;

- Refraner català-español:
  - &lt;http://refranyer.dites.cat/&gt;

- Bilingual and Monolingual Dictionaries
  - &lt;http://dictionary.cambridge.org&gt;
  - &lt;http://www.oxfordlearnersdictionaries.com/&gt;
  - &lt;https://www.merriam-webster.com/&gt;
  - &lt;http://learnersdictionary.com/&gt;
  - &lt;http://www.diccionari.cat/&gt;
  - &lt;http://www.multilingue.cat/&gt;
  - &lt;http://www.rae.es/&gt;

- Diccionario de dudas del español:
  - &lt;http://www.fundeu.es/&gt;

**Other resources useful for language analysis**

- Sketchengine: &lt;https://www.sketchengine.co.uk/&gt;

- Corpora in Catalan, English and Spanish
  - British national corpus: &lt;http://corpus.byu.edu/bnc/&gt;
  - American Brown corpus (edited version): &lt;http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/&gt;
  - Corpus del español (Corpus del español del siglo XXI): &lt;http://www.rae.es/recursos/banco-de-datos/corpes-xxi&gt;
  - Corpus del español (Corpus escrito y oral del español actual)&lt;http://www.rae.es/recursos/banco-de-datos/crea&gt;
  - Corpus del español (El corpus del español by M. Davies; more than 2 billion words) &lt;http://www.corpusdelespanol.org/&gt;

> **Supplementary content**
>
> «Sketch Engine is the ultimate tool to explore how language works. Its algorithms analyze authentic texts of billions of words (text corpora) to identify instantly what is typical in language and what is rare, unusual or emerging usage.»

- Corpus del català (Corpus textual informatitzat de la llengua catalana): <http://ctilc.iec.cat/>
- Corpus del català (Corpus informatitzat del català antic): <http://cica.cat/>
- Neologisms in English: <http://neologisms.rice.edu/index.php?a=index&d=1>
- Neologisms in Spanish: <http://cvc.cervantes.es/lengua/banco_neologismos/busqueda.asp>
- Neologisms in Catalan: http://www.termcat.cat/neoloteca/http://www.termcat.cat/neoloteca/