

Dealing with lexical semantic mismatches between Spanish and English

G. Vázquez (*), A. Fernández (**), M.A. Martí (***)

(*) University of Lleida gvazquez@udl.es; (**) University Autònoma de Barcelona Ana.Fernandez@uab.es (***) M.A. Martí Univeristy of Barcelona marti@lingua.fil.ub.es

1 Abstract

We are going to present a classification of verb centered mismatches between Spanish and English. This classification is based on a proposal for lexical representation of linguistic knowledge and therefore falls within the scope of lexical semantics. Our proposal is based on a model for lexical description that takes into account meaning components, event structure and diathesis alternations. In this paper we describe how these elements allows us to create a framework for the analysis of mismatches. We also suggest a way to use conceptual transfer in order to deal with some of these mismatches in a Machine Translation system.

1. Introduction

In this paper¹ we present a classification of verb mismatches that arise from the analysis of sentences in English and Spanish that pose a translation problem, that is to say, the two languages use different syntactic constructions to express the same meaning.

We believe that in order to carry out a detailed analysis of the mismatches that exist between English and Spanish, it is not sufficient to treat the topic at a strictly syntactic level. Our intention is to provide something more than just a relation of verbs that present a different syntactic realization of the arguments in each language. Since our proposal is carried out within a lexical-semantic approach, we do not intend to account for pragmatic or cultural mismatches.

In the present work we establish translation equivalences between languages at a semantic level. We can thus link inter-linguistically expressions of the type:

- (1) He got bored
Se aburrió

We analyze this pair of sentences as semantically equivalent. Both focus on the resulting state of an event, regardless of the way they express it syntactically. They are both anticausative constructions that relate to the corresponding causative structure (2):

- (2) X bored Y

¹ This research has been supported by the X-tract project, DGICYT PB98-1226.

X aburrió a Y

In [Vázquez et al, 2000] we propose a model for the description of verb entries that takes into account event structure, meaning components and diathesis alternations. This framework allows us to deal with lexical mismatches in a systematic fashion. In this paper we present lexical mismatches between Spanish and English following this proposal.

2. Classification Proposal

In this section we present first those mismatches that are due to different lexicalization patterns (section 2.1). Secondly, we examine those resulting from a specific compositional process, only possible in English, from which a complex lexical item is derived (section 2.2). Finally, we discuss mismatches related to argument structure (section 2.3). This classification is based on the three main components of our model for lexical description.

2.1 Mismatches based on meaning components

In this group we have included those mismatches that occur when in the source language a verb expresses a meaning component that is not lexicalized in the equivalent item in the target language (lexical semantic mismatch).

2.1.1 Lexicalization of a component

We consider that, as a general rule, English allows lexicalization processes more frequently than Spanish [Talmy, 1986]. The type of elements that can be conflated (as used by [Talmy, 1985] to refer to the lexicalization of a meaning component) are semantically varied in nature. An example of this type of phenomena is the lexicalization of the component manner:

- (3)
- | | |
|---------------|-----------------------------|
| a. to stride: | andar a zancadas |
| b. to fax: | enviar un fax |
| c. to e-mail: | enviar un mensaje de e-mail |

In addition to this component, it is also possible to lexicalize many others. In (4) we can see an example of the conflation of the affected entity in Spanish whereas in English the verb has a more general sense:

- (4)
- clap*: 1: to strike (as two flat hard surfaces) together so as to produce a sharp percussive noise 2 a: to strike (the hands) together repeatedly usu. in applause [Webster's, 1986].
- aplaudir*: 1: dar palmadas en señal de aprobación en un espectáculo o acto público o en prueba de admiración [Moliner, 1991].

Since *aplaudir* conflates the component hands (palmas) in Spanish, this component is not specified at the syntactic level whereas in English it can be syntactically realized as the nominal phrase (Direct Object of the VP):

- (5)
- | |
|-------------------------------------|
| a. Martha clapped (her hands) |
| b. Marta aplaudió (*sus/las palmas) |

Other meaning components that can be lexicalized are the cause (6) and the instrument (7):

- (6) a. They starved to death
 b. Se murieron de hambre
- (7) a. They fenced the property with wire
 b. Alambraron la propiedad

2.1.2. Incorporation of a component by pronominalization

In Spanish it is usual to incorporate a meaning component in the verbal lexical item by means of a pronominalization process. Since this process is nonexistent in English, this language has other resources to express this incorporation, usually making use of a different lexical item. This is the case of certain verbs that in Spanish incorporate either the source (8a,b) or the goal (8c) when used pronominally:

- (8) a. ir / irse vs. go / leave
 b. marchar / marcharse vs. go / leave
 c. acercar / acercarse vs. bring closer / come closer

2.1.3. Incorporation of a component by derivation

In this section we deal with those processes of word formation that in one of the languages are carried out by means of a derivative process and in the other require a syntactic construction. Semantically, the new lexical item is constituted by the addition of the value of the suffix or the prefix plus the lexical base upon which the process of derivation has taken place. This mechanism of incorporation is a regular and systematic process and consequently it is highly productive.

As can be observed in the examples, in the case of derivational processes of verbal items the original meaning of the verb that is taken as the semantic basis is partially modified by the information provided by the prefix:

prefix + verb	verb + adverb
mishear	oír mal
overcharge	cobrar demasiado/en exceso
prefix + verb	auxiliary verb + verb
rewrite	volver a escribir, reescribir
prefix + verb	auxiliary verb + adjective
aligerar	make lighter
inutilizar	make useless
prefix + verb	auxiliary verb + noun
abaratar	cut the price

In addition, it is sometimes also possible to find two equivalent structures in one of the languages, even though the use of one or the other is usually determined by the context of use:

misinform	informar mal
	malinformar
mistreat	maltratar

tratar mal

2.1.4. Conflation of a component through the context

In this case, the verb acquires a different meaning according to the linguistic context and in this way moves on to specify the manner in which the event is carried out. For instance, in dialogued sequences of a narrative work, English allows verbs of the type smile, grin, etc. to act as verbs reporting the interventions of the participants in a dialogue. In this particular context, the verb contributes information about how the main action, in this case conveying a message, is carried out. In Spanish, to express this type of semantic composition we require a generic verb (taken from the context of use) and another verb that adds information about the manner:

- (9) a. He laughed
a'. Dijo sonriendo

2.2 Lexical semantic compositional mismatches

We include in this group those mismatches that occur when in the source language there is a process of lexical semantic composition that has no equivalent construction in the target language (lexical semantic compositional mismatch). This process is typical of English. The verbal lexeme is always one of the elements to participate in this construction, together with an adverb, an adjective, a preposition, etc. As we can see in the following examples:

- (10) a. He ran out of the room
a'. Salió corriendo de la habitación
b. He ran
b'. Corrió.

In English, the verb run implies movement plus the manner in which this movement is carried out. On the other hand, the preposition out of is expressing the path. When together, a more complex item is created and it expresses the movement, the manner and the path. In Spanish this is not possible, and we require a movement verb (movement and path) plus an adjunct indicating the manner.

Other verbs that present this type of mismatch are those belonging to the group of sound emission, as shown in (11) and (12):

- (11) a. Peter tiptoed
a'. Peter se puso de puntillas
b. Peter tiptoed out of the room
b'. Peter salió de la habitación de puntillas
- (12) a. The cat purred
a'. El gato ronroneaba
b. The cat purred out of the room
b'. El gato salió de la habitación ronroneando

We could also include in this section those mismatches corresponding to the resultant and x-way English constructions. It is not always easy to find an equivalent Spanish translation (13,14 and 15):

- (13) a. John hammered the sculpture flat
a'. John aplanó la escultura con el martillo
- (14) a. The baby cried himself to sleep
a'. El niño acabó durmiéndose de tanto llorar

- (15) a. She talked her way out of the class
a'. Ella les convenció para lograr salir de la clase

Some verbs that express the meaning components: movement and contact, as for example to kick, to hit, etc., can participate in what is known as the conative construction². In this construction, these verbs appear with the preposition *at* and they focus only on the movement, leaving aside the contact information. In Spanish, this sort of structure is not possible and to convey a similar meaning we must use a periphrasis like the following (16):

- (16) a. John hit his brother
a'. John pegó a su hermano
b. John hit at his brother
b'.?John intentó pegar a su hermano

Thus, to express the same meaning we must resort to syntactically more complex structures or else leave the meaning incompletely expressed. As a result, the Spanish conceptualization of this meaning is not natural as it must be translated by periphrases of the kind: *intentó golpear*, *lanzó golpes en dirección a*, etc.

2.3 Mismatches based on argument structure

In this section we are going to deal with mismatches that arise from the interaction between the verb and its arguments. In other words, we observe the same semantic representation that was seen for the number of constituents, but the meaning components are differently distributed in the source language and in the target language.

On the one hand, we present those mismatches that result from two like meaning components being realized in a different manner (2.3.1 simple mismatches), therefore implying only one structure in each language. On the other, we examine those that imply distinct structures in both languages (2.3.2 mismatches based on diathesis alternations).

2.3.1. Simple mismatches

As we have indicated, in this section we are going to present mismatches that arise from the different syntactic realization of the same meaning components. We have further subdivided this section according to the way in which this realization differs.

2.3.1.1. Different distribution of meaning components

The resulting sentences in both languages are semantically equivalent but syntactically they realize the participants in switched positions:

- (17) a. He (subject) likes reading mystery books (object)
a'. Le (object) gusta leer novelas de misterio (subject)

² Some authors consider that this is not an alternation but two different lexical items since they express different meanings and they also have a distinctive form (for a discussion see [Dixon, 1991], [Gawron, 1983]).

Apart from the commonly recognized case of the verbs like/gustar, it is possible to find this type of mismatch in other verbs, although sometimes there is more than one possible translation in the target language:

- (18) a. Me duele mucho la cabeza
a'. I have a terrible headache
c. My head aches

As can be seen in (18), in Spanish we have only one possibility to express this event whereas in English we have two possibilities depending on the argument we are focalizing on.

Other interesting examples are the verbs pagar and cobrar, which describe the same action but from different perspectives, as established by authors such as [Fillmore, 1968, 1977]. We can see in example (24) that Spanish allows us to describe the cognitive scene from the two different perspectives by using one verb or the other. In English, on the other hand, we must resort to two different syntactic constructions (active versus passive) in order to adopt one perspective or the other:

- (19) a. ¿Cuánto cobras?
a'. How much do you get paid?
b. ¿Cuánto te pagan?
b'. How much do they pay you?

2.3.1.2. Different realization of argument-2

The mismatches depicted in this section correspond mainly to structures that in the target language are transitive while in the source language they are intransitive, or vice versa.

- (20) a. He entered the room
a' Entró en la habitación
(21) a. He easily accessed the information
a' Accedió fácilmente a la información

In our description these predicates present the same argument structure, since the intransitivity of the verb is not dependent on the absence or presence of one of the obligatory arguments in one of the languages but rather on its realization as a prepositional phrase.

2.3.1.3. Different expression of possession

This type of mismatch can be found in those constructions that imply the expression of the concept of possession. As basic meaning components we have an object or person that can be viewed as a whole made up of parts (properties, qualities or characteristics). As we have pointed out, the number of components is identical in both languages, a possessor and a possessed object, but in Spanish they are usually expressed in two syntactic constituents, one to express the possession relation and the other to express the object. In English, on the other hand, the first component appears to determine the possessed object within the same constituent.

- (22) a. Me lo he dejado en el coche
a' I have left it in my car
(23) a. Le he peinado (el pelo)
a' I have combed his hair

2.3.2. Mismatches corresponding to diathesis alternations

In this section we describe mismatches that are due to the different mechanisms a language disposes of to express the diathesis alternations we have defined in our framework [Fernández et al, 1999].

We account for those lexical items that in one of the languages can alternate in different syntactic structures but appear in only one structure in the target language. The other syntactic alternations can only be expressed by means of a morphological or periphrastic resource or we must use a different lexical item.

2.3.2.1. Mismatches regarding causativity

Firstly we examine mismatches that correspond to verbs characterized by the presentation of two alternative argument structures that denote the opposition cause-anticause. The manner in which these verbs express the different structures can vary in each language. For example (24), the verb *aburrir* uses a pronoun to express the anticausativity, whereas the equivalent construction in English requires a periphrastic construction:

- (24) a. *Mirar la televisión me aburre*
 a'. Watching TV bores me
 b. *Se aburrió*
 b'. He got bored

In this alternation, there are two possible readings of an event. The first one is a causative interpretation of the event (24a), and the second is the anticausative (24b). In the former, we focalize on the process, in the latter, on the subevent known as the resulting state.

Next we present a series of tables in which we show the different ways in which the expression of the cause and expression of the anticause can be conveyed. The first table expresses the different possibilities we have in English to express the opposition between two structures: one that expresses the cause and another in which it is not expressed (table 1). Secondly, we present the same information for Spanish (table 2) and in the third table (table 3) we present all the possibilities of equivalence we have found. Finally we present in table 3 examples for all the equivalencies between English and Spanish constructions:

ENGLISH (E)		
<i>Code</i>	<i>Causative</i>	<i>Anticausative</i>
1E	Item	Item
2E	Syntactic	Syntactic
3E	Item	Syntactic
4E	Item 1	Item 2

<i>Code</i>	<i>Causative</i>	<i>Anticausative</i>
1S	Item	Item
2S	Syntactic	Item
3S	Item	Morphological
4S	Item	passive + se

Table 1: Alternation causative/anticausative in English

SPANISH (S)

Table 2: Alternation causative/anticausative in Spanish

In Spanish, this alternation is possible using the four different linguistic mechanisms shown in table 2.

We would like to conclude this section with a table presenting some of the equivalencies established between the mechanisms allowed by each one of the languages, followed by examples. It should be noted that the indexes (e.g. 1E) make reference to the tables above (code). In this case 1E refers to table 1 in which the alternation is expressed using the same lexical item:

English	Spanish	Examples
1E	1S	1
1E	2S	2
1E	3S	3
3E	3S	4
4E	3S	5

Table 3: Equivalencies causative/anticausative/middle in Spanish and English

- | | |
|---|--|
| 1.1 John <i>boiled</i> the water for too long. | 3.1' <i>Cerró</i> la puerta de golpe. |
| 1.2 The water <i>boiled</i> for too long. | 3.2' La puerta <i>se cerró</i> de golpe. |
| 1.1' Juan <i>hirvió</i> el agua demasiado tiempo. | |
| 1.2' El agua <i>hirvió</i> demasiado tiempo. | 4.1 The dog <i>scared</i> the little children. |
| | 4.2 The little children <i>got scared</i> . |
| 2.1 He <i>rang</i> the bell at dusk. | 4.1' El perro <i>asustó</i> a los niños. |
| 2.2 The bell <i>rang</i> at dusk. | 4.2' Los niños <i>se asustaron</i> . |
| 2.1' <i>Hizo sonar</i> la campana al anochecer. | |
| 2.2' La campana <i>sonó</i> al anochecer. | 5.1 Su opinión no me <i>preocupa</i> nada. |
| | 5.2 <i>Se preocupa</i> mucho (de su opinión). |
| 3.1 He <i>closed</i> the door abruptly. | 5.1' His opinion <i>doesn't worry</i> me. |
| 3.2 The door <i>closed</i> abruptly. | 5.2' She <i>cares</i> a lot (about his opinion). |

2.3.2.2. Mismatches based on different order

Another possibility is that one of the languages allows focalization on one participant or the other by simply changing the syntactic order of these elements whereas the other language requires a more complex syntactic construction.

In Spanish, it is possible to change the topic of the sentence (focalize) altering the order of the participants in the predication (25a,b). This possibility does not exist in English (25a',b'), which is more strict a language with regard to the order of the constituents. Translating sentences that take advantage of this resource in Spanish is not always easy.

- (25) a. Sus compañeros han regalado un reloj a Juan por su cumpleaños
a'. His colleagues have given John a watch for his birthday
b. A Juan le han regalado un reloj por su cumpleaños

b'. He has been given a watch for his birthday

It should be noted that English admits the passive construction in which the IO of the transitive is the passive subject. This type of resource allows us to translate structures in which the indirect object is the topic of the sentence. We have included this mismatch here since it requires a syntactic construction of the type seen before to solve the problem.

Even though the sentences we are about to discuss would not properly be considered diathesis alternations in our framework, we mention them here because they are related to the previous ones (25) in that in Spanish we use the structural order as the focalization mechanism. In this sense, Spanish sentences (26a,b) are perfectly normal uses whereas a construction such as the one seen in (26b) is not very habitual in English and it can only be used in certain contexts, in this case in an oral context:

- (26)
- a. Le gusta María
 - a'. He likes Mary
 - b. María le gusta
 - b'. It's Mary he likes (oral context)
 - c. Le gusta mucho más María que su hermana
 - d. María le gusta mucho más que su hermana
 - c'/d'. He likes Mary much better than her sister

2.4 Mismatches due to a different event structure

In this section we address those mismatches that occur because of the different information regarding event structure that is represented by equivalent items in both languages. This kind of mismatch can be inherent to the lexical item or can arise as a result of a different composition process in one of the languages with respect to the other.

2.4.1. Different lexeme according to aspectual information

In this mismatch we describe a divergence in the temporal information reflected by the verb. This is the case of some verbs belonging to the semantic class verbs of contact by impact: hit verbs. In English, these verbs present the peculiarity of including aspectual information or, more precisely, information about the iteration of the action. In this way, for example, beat expresses reiteration whereas strike and hit require a progressive tense to express this information. The latter used in the simple form has a punctual reading.

- (27)
- a. He hit John
 - b. He struck John
 - c. He beat John

The possibilities of translation that these sentences offer are diverse, but it is not possible to make reference to the iteration of the event if we do not want to make use of a periphrastic structure:

- (28)
- a. Dio un golpe a John
 - b. Golpeó a John
 - c. Estuvo golpeando a John

In Spanish, as can be observed in the examples above, if we choose the verb golpear then we are not specifying if it is an iterative action or if it is not (28b). Conversely, if our

intention is to make sure that we are expressing a simple, punctual event then we resort to the use of a verbal periphrasis *dar + NP*, in this case *un golpe* (28a).

3. Resolution of a mismatch

The treatment we propose when dealing with mismatches between Spanish and English provides several advantages for their resolution because it goes further from the purely syntactic divergences. All the mismatches presented so far can be resolved in a transfer module from the framework we propose.

In order to demonstrate our approach to the problem, we are going to present in this section how one of the mismatches is dealt with. We have chosen mismatches based on diathesis alternations because we believe they constitute a clear illustration of both the validity of linguistic generalizations carried out from a lexical perspective and the application of the syntactic-semantic interface in a machine translation system.

As we have seen, diathesis alternations are considered as semantic oppositions that can be expressed by means of morphological, lexical or syntactic mechanisms. Thus, alternations contribute to define verbal semantic classes, so each verb belongs to one or more semantic classes and participates in one or more diathesis alternations.

When analyzing a sentence in the source language, the syntactic structure plus the semantic class it has been assigned in the lexicon provides relevant information so as to allow the system to hypothesize the diathesis alternation it is expressing. In this way, it also allows us to know if the syntactic structure expresses causativity, underspecification, or other communicative intentions.

The approximation we propose to the problem goes as follows. First, we start from a sentence which constitutes the input. This sentence is morpho-syntactically tagged, and this syntactic structure is given a semantic representation which forms the basis of our proposal.

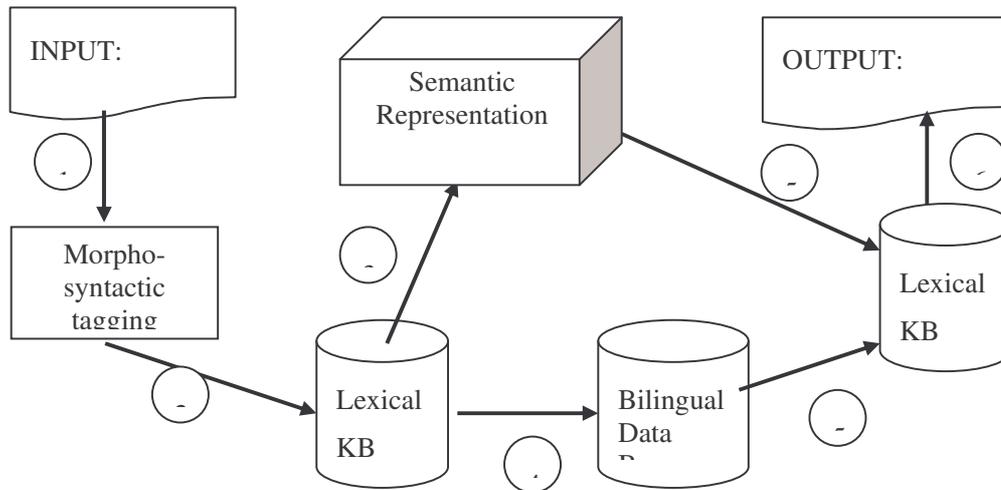
Continuing with this hypothesis, we resort to the lexical translation of the lexical item, then from a bilingual dictionary a translation is provided. The next step is to see how the semantic representation hypothesized for the structure in the source language is realized in the target language and in particular what mechanism each lexical item takes to convey precisely this meaning.

Next, we are going to exemplify and explain the process step by step with an example. This example constitutes an anticausative reading interpretation which requires a lexical mechanism in English whereas in Spanish it uses the pronominal process.

Let's examine a sentence similar to the one considered in the introduction:

He gets drunk every night
Se emborracha cada noche

In the process that we propose for generating translation equivalencies, we have established six steps. First of all we analyze the input structures, both morphologically and syntactically, and thus we establish the lexical items involved. In this case, we establish that there exists an event of a drink type, that it has an argument he and that the event of drinking is modified by a frequency complement every night.



From step one, we move to step two in which each item is looked up in the LKB. In this LKB, we have provided the possible subcategorization frames linked to the meaning components and their realization. From this information we obtain all the possible semantic representations linked to syntactic structures. Following this example (intransitive structures with just one argument) we could have two possible interpretations: the middle construction and the anticausative construction.

Next (step 4), we establish the translation link at a lexical level for each item participating in the sentence. The verb will constitute the engine for the target language equivalent sentence and therefore we have to get the information declared for the target language in that particular lexical item (step 5). Once we have the information for the item, we need to know how the hypothesized syntactic structure is expressed in the language for that particular item (step 5).

The sentence that constitutes the output is built from both, the assignation of a lexical value and the information it possesses about the mechanism required to convey a semantic interpretation (step 6) for the hypothesis established.

4. Conclusions

In this paper we have presented a classification of mismatches found in English and Spanish that are lexically-motivated and verb-centered. The resolution to the problem we propose has a lexical dimension since we pretend to resolve mismatches from the information declared in both the entry (structure) that constitutes the input and its equivalent, the output.

This work assumes the verb to be the head of the sentence and we have tried to declare all the morphological, syntactic and semantic information relevant to the mismatches in the lexicon in order to provide this element with all the information needed for mismatch resolution.

Sentences are interpreted in terms of the semantic representation they convey and, then, they are assigned a syntactic frame depending on each particular item for each of the languages involved. We have further exemplified our approach with the analysis of a sentence and we have outlined how this approach can be extended to other linguistic phenomena that could be explained and dealt with more efficiently within the proposed framework.

References

[Dixon, 1991]

Dixon, R.M. *A new Approach to English Grammar on Semantic Principles*, Oxford University Press, 1991.

[Fillmore, 1968]

Fillmore, J. C. "Lexical entries for verbs", *Foundations of Language*, 4, 373-393, 1968.

[Gawron, 1983]

Gawron, M.J. *Lexical Representatiopn and the Semantics of Complementation*, Doctoral dissertation, University of California, 1983.

[Fernández et al, 1999]

Fernández, A., M.A. Martí, G. Vázquez, I. Castellón, "Establishing Semantic Oppositions for the Typification of Predicates", in *Language Design*, 2, 1999.

[Talmy, 1985]

Talmy, L. "Lexicalization patterns: semantic structure in lexical forms", en T. Shopen (ed.), *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press. p. 57-149. 1985

[Vázquez et al., 2000]

Vázquez, G., A. Fernández y M.A. Martí, *Clasificación verbal: Alternancias de diátesis*. Universidad de Lleida, 2000.