

Ma. Antònia Martí, Carmen Soler  
*Universitat de Barcelona,*  
*Secció de Lingüística General \**

## **Dealing with Lexical Mismatches**

### **Abstract**

A first approach to lexical translation mismatches reveals that for some subsets a correspondence with morphological and semantic processes can be established which enables the treatment by means of translation links to represent multilingual information. A deeper study of the phrases that constitute the mismatch involves selectional restrictions criteria related to the types in the LKB1.<sup>1</sup>

### **1. Introduction**

Languages are known to exhibit distinct preferences in lexicalization patterns (Talmy 1985) in such a way that some word senses in one language have to be translated by a phrase in the other, giving rise to a mismatch. The study of such patterns shows that some kind of mismatches in one language correspond to morphological processes of word formation (such as derivation and compounding) and semantic processes (such as metaphoric and metonymic sense extensions) in the other language. According to this, some subsets of mismatches can be established and, therefore, treated together and generalised (Briscoe & Copestake 1991).

This document presents some of the lexical mismatches observed in the bilingual *English–Spanish Spanish–English dictionary VOX–HARRAP’S ESENCIAL*, their organization into different subsets according to the correspondences mentioned above and their import on translation equivalence by means of a link rule. The first part of this document concentrates on nominal lexical mismatches. The second part consists of a description of different subsets showing distinct preferences in lexicalization patterns. Finally we outline how we intend to study the phrases that constitute a mismatch taking into account the relationship between its components.

### **2. Spanish–English nominal mismatches**

We have concentrated our study on nominal lexical mismatches, starting with an examination of some 33% of the Spanish–English part of the dictionary. We have classified some of the mismatches according to certain correspondences in order to allow the generalization of lexical rules.

## 2.1. Differences in plurality

Some of the mismatches are due to differences in plurality as a reflection of a count–mass discrepancy (Copestake & Sanfilippo 1993). This can be represented by establishing a link between a word and its translation in the plural form. In this case the equivalence holds between a basic lexical entry and a lexical entry after plural rule application (Fig. 1):

SPANISH singular	ENGLISH plural
<i>arenal</i> (sandy area)	<i>sands</i>
<i>clientela</i> (group of customers)	<i>customers</i>
<i>comparsa</i> (group of extras)	<i>extras</i>
<i>competencia</i> (group of competitors)	<i>competitors</i>

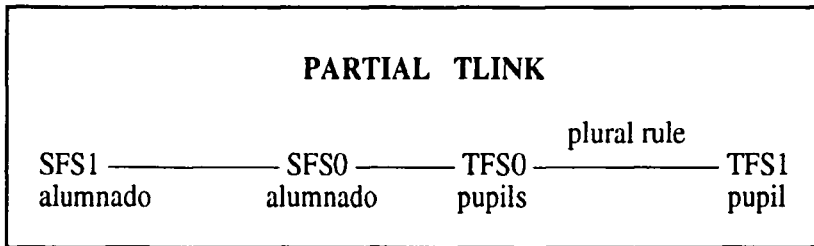


Fig. 1

A similar count–mass discrepancy gives rise to individuating phrases such as (Copestake & Sanfilippo, 1993):

SPANISH	ENGLISH
<i>consejo</i>	<i>a piece of advice</i>
<i>información</i>	<i>a piece of news</i>
<i>noticia</i>	<i>a piece of news</i>

## 2.2. Derivational morphology

Other mismatches in English correspond to a process of derivational morphology in Spanish:

**2.2.1.** In this first example the Spanish suffix *-ero, -a* attached to a fruit name creates a new word meaning a ‘fruit tree’. This derivational process in Spanish corresponds to a mismatch in English:

SPANISH	ENGLISH
fruit + <i>-ero, -a</i> fruit tree	fruit + <i>tree</i>

<i>limón</i>	<i>limon-ero</i>	<i>lemon</i>	<i>lemon tree</i>
<i>albaricoque</i>	<i>albaricoqu-ero</i>	<i>apricot</i>	<i>apricot tree</i>
<i>melocotón</i>	<i>melocoton-ero</i>	<i>peach</i>	<i>peach tree</i>

Here, as in most of the following examples, a **PHRASAL TLINK** is used (Fig.2):

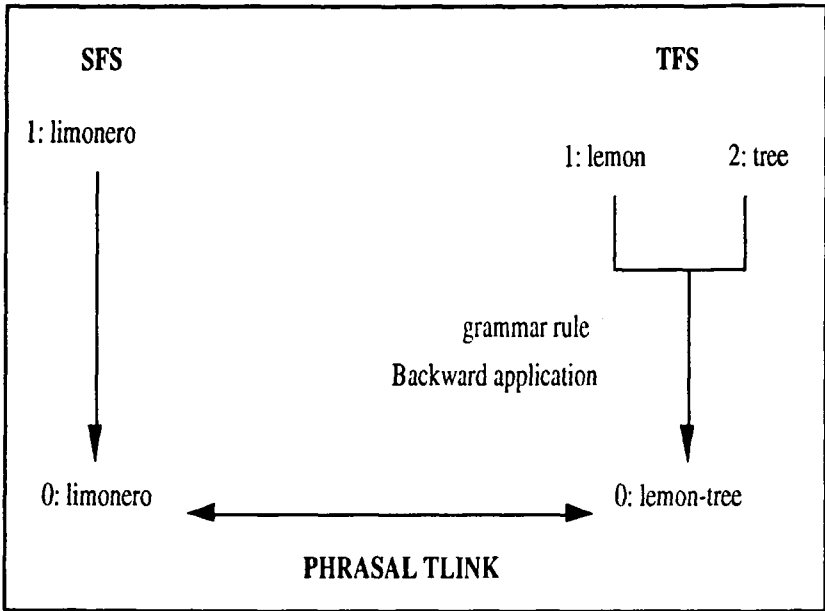


Fig. 2

2.2.2. This case is similar to the previous one, the only difference being the suffix, here *-a*:

<i>fruit + -a</i>	<i>fruit tree</i>	<i>fruit + tree</i>	
<i>almendr-a</i>	<i>almendr-o</i>	<i>almond</i>	<i>almond tree</i>
<i>aceitun-a</i>	<i>aceitun-o</i>	<i>olive</i>	<i>olive tree</i>
<i>manzan-a</i>	<i>manzan-o</i>	<i>apple</i>	<i>apple tree</i>

2.2.3. Here the suffix *-al* added to a plant creates a mass noun. Note that while Spanish only lexicalizes the idea of constituting a group of plants, the English corresponding phrases seem to go further, marking a difference on the kind of plant:

plant + <i>-al</i>	plantation	fruit + <i>plantation</i> etc.
<i>aguacate</i>	<i>aguacat-al</i>	<i>avocado plantation</i>
<i>abeto</i>	<i>abet-al</i>	<i>fir plantation</i>
<i>abrojo</i>	<i>abroj-al</i>	<i>thistle patch</i>
<i>almendro</i>	<i>almendr-al</i>	<i>almond grove</i>
<i>alcornoque</i>	<i>alcornoc-al</i>	<i>cork oak grove</i>
<i>trigo</i>	<i>trig-al</i>	<i>wheat field</i>
<i>arroz</i>	<i>arroz-al</i>	<i>rice field</i>

2.2.4. In this case the suffix *-ía* added to a noun which denotes a certain job creates the noun of the shop where it takes place. The English equivalent produces a change in the case (Fig.3):

job + <i>-ía</i>	shop	job + 's	( <i>shop</i> )
<i>joyero</i>	<i>joyer-ía</i>	<i>jeweller</i>	<i>jeweller's (shop)</i>
<i>calderero</i>	<i>calderer-ía</i>	<i>boilmaker</i>	<i>boilmaker's (shop)</i>
<i>abacero</i>	<i>abacer-ía</i>	<i>grocer</i>	<i>grocer's (shop)</i>

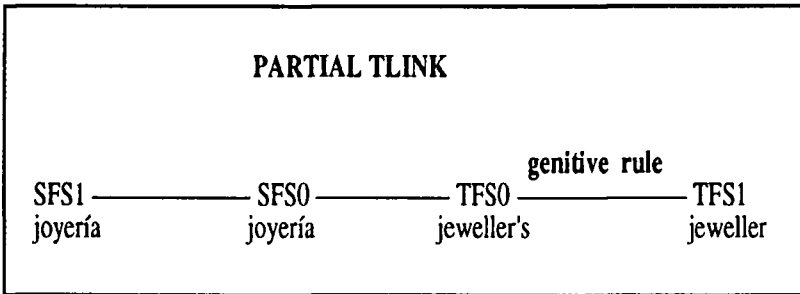


Fig.3

2.2.5. In the following examples some words with Spanish diminutive suffixes have resulted in frozen forms which have to be considered different lexical entries. As English lacks this derivative process an adjective is used to express the correspondence:

<b>noun + diminutive suffix</b>		<i>small/little + noun</i>
<i>caballo + -ito</i>	<i>caballito</i>	<i>small horse</i>
<i>bandera + -ita</i>	<i>banderita</i>	<i>little flag</i>
<i>campana + -illa</i>	<i>campanilla</i>	<i>small bell</i>
<i>tabla + -illa</i>	<i>tablilla</i>	<i>small board</i>

<i>aro + -ete</i>	<i>arete</i>	<i>small ring</i>
<i>arca + -eta</i>	<i>arqueta</i>	<i>small chest</i>
<i>botella + -ín</i>	<i>botellín</i>	<i>small bottle</i>

### 2.3. Sense extensions

The following mismatches in English correspond to metonymic and metaphoric sense extensions in Spanish.

**2.3.1. Metonymic sense extension from a fruit to a fruit tree presenting the same kind of mismatch in English as the ones in 2.2.1 and 2.2.2.**

fruit & tree	fruit + tree
<i>albaricoque</i>	<i>apricot tree</i>
<i>anacardo</i>	<i>cashew tree</i>
<i>membrillo</i>	<i>quince tree</i>

**2.3.2. Metaphoric sense extension in which a noun denoting an animal is also used to express both a quality of a person (therefore CAT=Adj) and the person itself (CAT=N). The mismatch is produced when, for some animals, the sense extension in Spanish has no correspondence in English, and also by the fact that the double category Adj-N is not as frequent in English as it is in Spanish. The adjective referring to the quality of the animal needs a support noun:**

animal & person	animal	adj.+ person
<i>lince</i>	<i>lynx</i>	<i>sharp-eyed person</i>
<i>burro,-a</i>	<i>donkey</i>	<i>ignorant person</i>
<i>jirafa</i>	<i>giraffe</i>	<i>tall person</i>
<i>ganso</i>	<i>goose</i>	<i>slow/lazy person</i>

**2.3.3. Metaphoric sense extension similar to the previous one. English takes a support noun while Spanish extends the quality to the person:**

quality (adj) & person (n)	quality (adj) + person
<i>agonizante</i>	<i>dying person</i>
<i>ahogado</i>	<i>drowned person</i>
<i>beligerante</i>	<i>belligerent person</i>

### 3. English–Spanish nominal mismatches: compounds

We have also examined 33% of the English–Spanish part of the dictionary. Apart from the mismatches produced by a difference in plurality we have observed that a great number of nominal mismatches in Spanish correspond to the compounding process, which seems to be much more productive in English than in Spanish:

<i>airlift</i>	<i>puente aéreo</i>
<i>airline</i>	<i>línea aérea</i>
<i>birthmark</i>	<i>marca de nacimiento</i>
<i>birthplace</i>	<i>lugar de nacimiento</i>
<i>bloodbath</i>	<i>baño de sangre</i>
<i>bloodshed</i>	<i>derramamiento de sangre</i>
<i>bellyache</i>	<i>dolor de barriga</i>
<i>headache</i>	<i>dolor de cabeza</i>
<i>sundial</i>	<i>reloj de sol</i>
<i>sunglasses</i>	<i>gafas de sol</i>

### 4. Non–nominal mismatches

We present here some of the subsets we found before dealing with nominal mismatches.

#### 4.1 Support verbs

This is a case of verbal diathesis. In Spanish some active transitive verbs change to pronominal verbs, so-called a "pasiva refleja", by means of the particle *se*. Of the three different equivalences found in English one gives rise to a mismatch as a support verb is necessary.

- (a) the same English verb is used in both cases:  
*quemar(se)* t–p *to burst*
- (b) the verb changes:  
*adjudicar(se)* t *to award*  
p *to appropriate*
- (c) a support verb is needed:  
*anegar* t *to flood*  
*anegarse* p *to be flooded*  
*fundar* t *to found*  
*fundarse* p *to be founded*

## 4.2 More derivational morphology: prefixes

4.2.1 The negative English prefix *-un* can produce a mismatch in the Spanish equivalence:

<i>unaccommodating</i>	<i>poco sociable</i>
<i>unadventurous</i>	<i>poco atrevido</i>
<i>unaccompanied</i>	<i>sin compañía</i>
<i>unafraid</i>	<i>sin miedo</i>

Note that the difference between ‘poco/sin’ can be regarded as a matter of selectional restriction as ‘*sociable, atrevido, ambicioso*’ are gradable adjectives.

4.2.2 The English prefixes *mis-*, *pre-* and *re-* are here equivalent to the Spanish adverbs *mal*, *de antemano* and *de nuevo*.

<i>misbehave</i>	<i>comportarse mal</i>
<i>miscalculate</i>	<i>calcular mal</i>
<i>prearrange</i>	<i>arreglar de antemano</i>
<i>pre-establish</i>	<i>establecer de antemano</i>
<i>re-cover</i>	<i>forrar de nuevo</i>
<i>re-examine</i>	<i>examinar de nuevo</i>

4.2.3. In this case the prefix *re-* corresponds to a construction with a support verb in Spanish:

<i>reassemble</i>	<i>volver a montar</i>
<i>reawaken</i>	<i>volver a despertar</i>

## 5. Mismatches and selectional restrictions

As the kind of correspondences we have treated appear to be insufficient to cover most of the mismatches, we have gone on to study the phrases which constitute the mismatch, i.e. its elements and the kind of relationship established between its members. The two most frequent constructions are:

(1) NOUN + prep + NOUN	<i>headache</i> <i>afonía</i>	<i>dolor de cabeza</i> <i>loss of voice</i>
(2)a. NOUN + MODIFIER (adj in Spanish)	<i>chip</i>	<i>patata frita</i>
b. MODIFIER + NOUN (English)	<i>alianza</i> <i>bermejo</i> <i>camerino</i>	<i>wedding ring</i> <i>bright red</i> <i>dressing room</i>

The relationship which holds between the head and the modifier implies selectional restrictions. In our view the modifiers fill one of the features of the type associated with the head-noun and That at the same time they specify the noun type they can modify. There is a selectional restriction in both directions, from the head to the modifier it can support, and from the modifier to the noun it can accept.

If we take, as an example, “*chip* → *patata frita* (fried potato)” the resulting mismatch resulting is not just the sum of *patata* + *frita*, as a phrasal TLINK would suggest, it is in fact a more complex relationship where *frita* fills one of the features associated with the type of the head *patata*. On the other hand *frita* associates with a certain type which can support being “fried”. The study of such phrases (noun + adj) and relationship will be the next step in our work.

#### Notes

\* This research has been supported by the AcquiLEX Esprit Project(7315) and by the PB-91-0854 of the DGICYT.

#### References

- Briscoe, T. and A. Copestake 1991. “Sense extensions as lexical rules” in D. Fass, E. Hinkelman and J. Martin (eds.), *Computational approaches to non-literal language: metaphor, metonymy, idiom, speech acts, implicature*, Proc. of IJCAI Workshop, Sydney, pp.12–20.
- Copestake et al. 1992. “Multilingual Lexical Representation” in ESPRIT BRA-3030 ACQUILEX WP NO.043.
- Copestake, A., Briscoe T. 1991. “Lexical Operations in a Unification Based Framework”, Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation, Berkeley, California, pp.88–101.
- Copestake, A., Sanfilippo, A 1993. “Multilingual Lexical Representation”, ACQUILEX II – Esprit Project 7315 , January 1993.
- Diccionario VOX HARRAP’S ESENCIAL INGLES-ESPAÑOL ESPAÑOL-INGLES, 2 Edición, Bibliograf S.A., Barcelona 1992.
- Talmy, L. 1985. “Lexicalization Patterns: Semantic Structure in Lexical Forms”, in T. Shapen, ed., *Language Typology and Syntactic Description: Grammatical categories and the Lexicon*. vol. III. Cambridge: C.U.P. pp. 57–149.