

## Session 4: Lexical transfer

Juan Antonio Pérez-Ortiz<sup>1,2</sup>

<sup>1</sup>Prompsit Language Engineering, S.L.,  
Av. Universitat, s/n. Edifici Quorum III. 03202 Elx, Spain

<sup>2</sup>Departament de Llenguatges i Sistemes Informàtics,  
Universitat d'Alacant, E-03071 Alacant, Spain

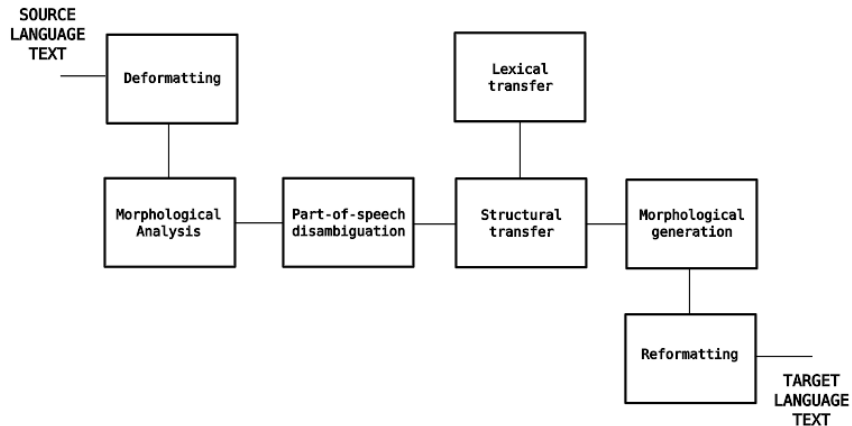
3rd May 2011



# Lexical transfer: table of contents

- 1 Transfer stage
- 2 Lexical transfer
- 3 Translation equivalents
- 4 Lexical ambiguity
- 5 Examples of bilingual entries
- 6 Lexical selection
- 7 Marking lexical features for the structural transfer
- 8 Practice

# Transfer stage /1



The **transfer module** is where the *magic* happens: the intermediate representation in source language (SL) is converted into an intermediate representation in target language (TL).

Transfer in Apertium consists of two submodules:

- **Lexical transfer:**
  - selects the most suitable equivalent in TL for a SL word;
  - marks some lexical features which will be used by the structural transfer.
- **Structural transfer:** performs syntactic operations involving groups of words (see next sessions).

# Lexical transfer

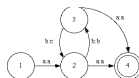
The **lexical transfer** module reads each SL lexical form and delivers the corresponding TL lexical form by looking it up in a bilingual dictionary.

## Bilingual dictionary

- No surface forms in this stage: input and output are **lexical forms** consisting of lemma, part-of-speech and inflection information.
- The dictionary contains a list of equivalent lexical forms.
- A single bilingual dictionary is used for both directions of translation.
- XML syntax similar (but simpler) to monolingual dictionaries.
- Paradigms are usually not necessary.

A simple task... apparently:

transducteur  $\longleftrightarrow$  transductor



[fr]

transducteur<n><m><s>

$\longleftrightarrow$

[es]

transductor<n><m><s>

transducteur<n><m><pl>

$\longleftrightarrow$

transductor<n><m><pl>

## A shorter representation

Only lemma and part-of-speech are mandatory if the rest of tags do not change:

transducteur<n>  $\longleftrightarrow$  transductor<n>

## XML encoding in the bilingual dictionary

```
<e><p>  
  <l>transducteur<s="n"></l>  
  <r>transductor<s="n"></r>  
</p></e>
```

These can be used for  $fr \rightarrow es$  (left to right), and  $es \rightarrow fr$  (right to left).

# Change of gender

Only the tags until the last change need to be indicated:

vallée<n><f> ↔ valle<n><m>

## XML encoding in the bilingual dictionary

```
<e><p>  
  <l>vallée<s="n"><s="f"></l>  
  <r>valle<s="n"><s="m"></r>  
</p></e>
```



# Lexical ambiguity

Real life is a little bit more complex, as SL words may have more than one equivalent in TL.

## Homography

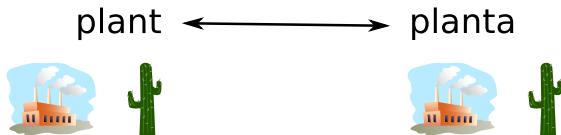
- A surface form with more than one possible morphological analysis: English *book* (noun or verb) translates into French *livre* (noun) or *réserver* (verb).
- The part-of-speech disambiguator has already selected one single alternative (recall previous session).

## Polysemy

A lemma and part-of-speech that have several meanings: English *bank* (noun) translates into Spanish *banco* or *ribera*.

# Polysemy

- Polysemy needs to be addressed in the lexical transfer when different meanings are translated in different ways; for instance, *bank* in the English-Spanish bilingual dictionary.
- This is called **lexical selection**.
- **Free-rides** do not pose any problem: English *plant* translates into Spanish *planta* when referring to both the living organism or a kind of factory/installation.



## Lexical selection in Apertium

- Currently, Apertium does not deal explicitly with polysemy (no lexical selection is performed).
- A SL lexical form can only have one TL equivalent in the dictionary (more on this later).
- We will include in the bilingual dictionary the translation of the most frequent meaning.
- This makes bilingual dictionaries domain-dependent.

# Examples of bilingual entries

## How to add bilingual entries

Let's add some entries to the French-Spanish bilingual dictionary!

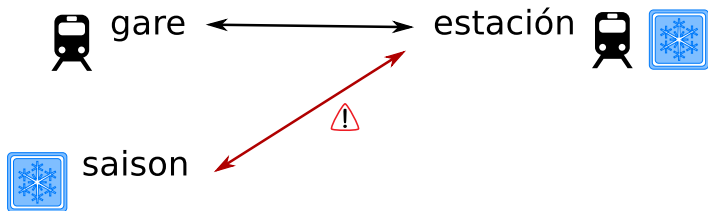
## Adding entries to the dictionary /1



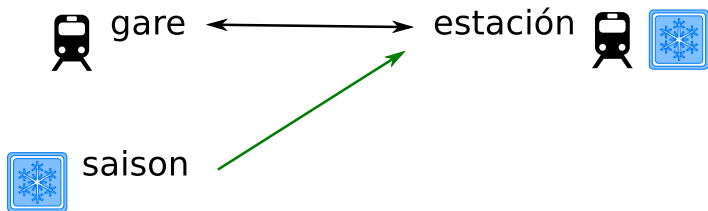
## Adding entries to the dictionary /2

```
<e><p>  
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>
```

## Adding entries to the dictionary /3



## Adding entries to the dictionary /4

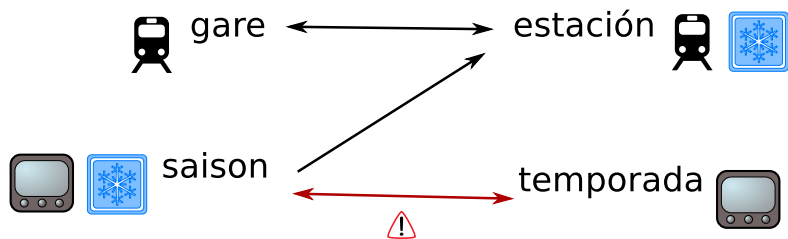




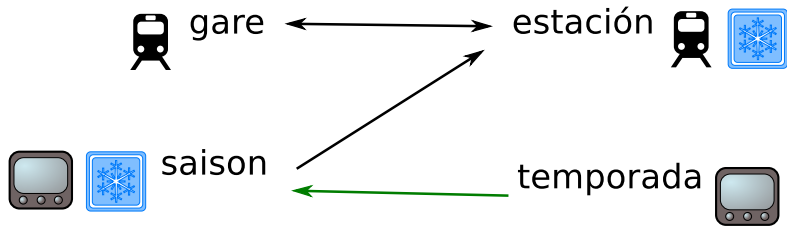
## Adding entries to the dictionary /5

```
<e><p>  
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>  
<e r="LR"><p>  
  <l>saïson<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>
```

## Adding entries to the dictionary /6






## Adding entries to the dictionary /7



## Adding entries to the dictionary /8

```
<e><p>  
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>  
<e r="LR"><p>  
  <l>saïson<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>  
<e r="RL"><p>  
  <l>saïson<s n="n"/></l> <r>temporada<s n="n"/></r>  
</p></e>
```

## Adding entries to the dictionary /9

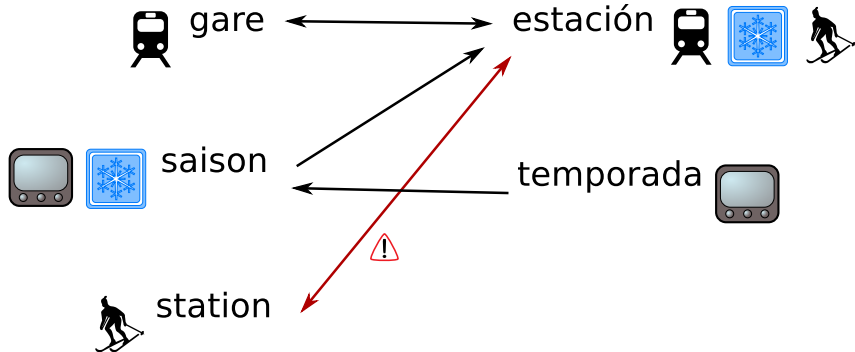
 gare  $\longleftrightarrow$  estación  

  saison  $\longleftrightarrow$  temporada 

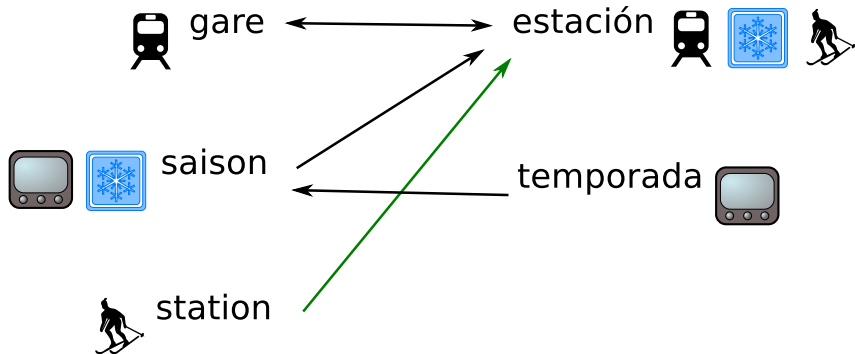
## Adding entries to the dictionary /10

```
<e><p>  
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>  
</p></e>  
<e><p>  
  <l>saïson<s n="n"/></l> <r>temporada<s n="n"/></r>  
</p></e>
```

# Adding entries to the dictionary /11



# Adding entries to the dictionary /12





## Adding entries to the dictionary /13

```
<e><p>
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>saïson<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="RL"><p>
  <l>saïson<s n="n"/></l> <r>temporada<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>station<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
```

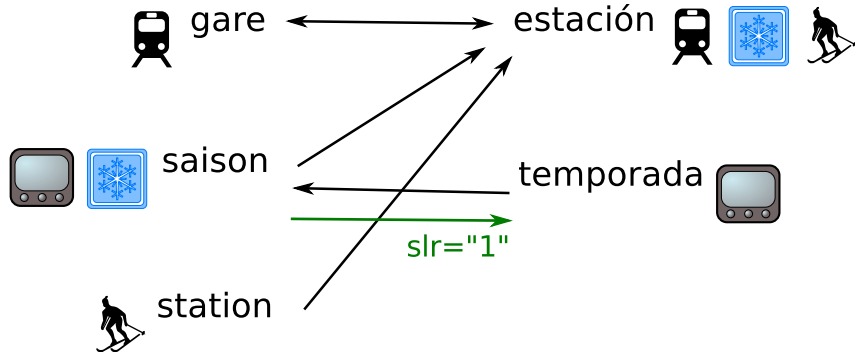
# Disambiguation of polysemy

We may cope with lexical selection of polysemous terms by using multiwords:

gare<n>	↔	estación<n>
station <g>de ski</g><n>	↔	estación <g>de esquí</g><n>

- Apertium will include a solution for lexical selection in the near future.
- In the meantime, alternative translations may be already provided in the bilingual dictionaries.

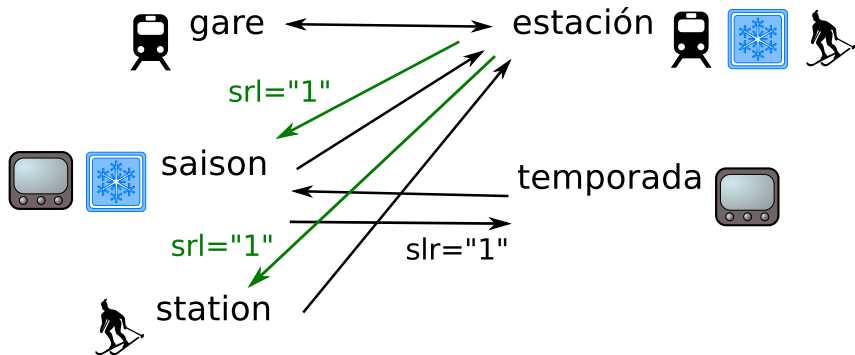
# Adding alternative translations for an entry /1



## Adding alternative translations for an entry /2

```
<e><p>
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>saison<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="RL"><p>
  <l>saison<s n="n"/></l> <r>temporada<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>station<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e slr="1" r="LR"><p>
  <l>saison<s n="n"/></l> <r>temporada<s n="n"/></r>
</p></e>
```

# Adding alternative translations for an entry /3



## Adding alternative translations for an entry /4

```
<e><p>
  <l>gare<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>saïson<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e r="RL"><p>
  <l>saïson<s n="n"/></l> <r>temporada<s n="n"/></r>
</p></e>
<e r="LR"><p>
  <l>station<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e slr="1" r="LR"><p>
  <l>saïson<s n="n"/></l> <r>temporada<s n="n"/></r>
</p></e>
<e srl="1" r="RL"><p>
  <l>saïson<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
<e srl="1" r="RL"><p>
  <l>station<s n="n"/></l> <r>estación<s n="n"/></r>
</p></e>
```

- The lexical transfer also marks some lexical features which will be used by the structural transfer.
- For instance, a noun with the same surface form for its two genders.
- Spanish monolingual dictionary:
  - estudiante* → estudiante<n><mf><sg>
  - estudiantes* → estudiante<n><mf><pl>
- The structural transfer will choose the gender by looking at the surrounding context.
- The lexical transfer simply marks this issue with the tag GD.
- Similar things hold for number (ND).

## Marking lexical features for the structural transfer /2

```
<e r="LR"><p>  
  <l>étudiant<s n="n"/><s n="m"/></l>  
  <r>estudiante<s n="n"/><s n="mf"/></r>  
</p></e>
```

```
<e r="LR"><p>  
  <l>étudiant<s n="n"/><s n="f"/></l>  
  <r>estudiante<s n="n"/><s n="mf"/></r>  
</p></e>
```

```
<e r="RL"><p>  
  <l>étudiant<s n="n"/><s n="GD"/></l>  
  <r>estudiante<s n="n"/><s n="mf"/></r>  
</p></e>
```



# Some commands

## Compiling the dictionary

```
$ xsltproc translate-to-default-equivalent.xsl ...  
$ xsltproc -stringparam ...  
$ lt-comp rl fr-es.dix es-fr.autobil.bin
```

or (when available):

```
$ make
```

or:

```
$ make es-fr.autobil.bin
```

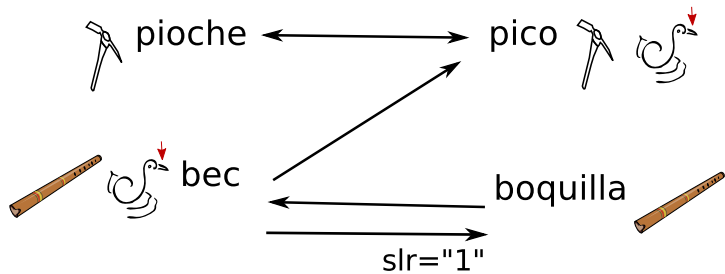
## Testing the lexical transfer

```
$ echo "Los transductores" |  
  lt-proc es-fr.automorf.bin |  
  apertium-tagger -g es-fr.prob |  
  lt-proc -b es-fr.autobil.bin
```

## Time for practice!

- Follow the workshop manual (session 4).
- Add the entries *transducteur*, *pic*, *pioche*, etc. as discussed in the manual.
- Then, add and test the entries *station*, *gare*, etc. as discussed in this presentation.
- Finally, add and test entries for a different language pair of your choice.

# Practice /2



# Practice /3

