Using unsupervised corpus-based methods to build rule-based machine translation systems

Felipe Sánchez Martínez

fsanchez@dlsi.ua.es

Ph.D. thesis
supervised by
Mikel L. Forcada
Juan Antonio Pérez Ortiz

Universitat d’Alacant
Universidad de Alicante
Departament de Llenguatges i Sistemes Informàtics
Departamento de Lenguajes y Sistemas Informáticos

30th June 2008
1 Motivation & goal

2 Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training

3 Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method

4 Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering

5 Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules

6 Concluding remarks
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Motivation

- Experience in the development of shallow-transfer MT systems
  - interNOSTRUM  Spanish↔Catalan
  - Traductor Universia  Spanish↔Portuguese
  - Apertium  Several language pairs available

- Huge human effort to code all the linguistic resources

- Resources usually needed by shallow-transfer MT systems
  - Monolingual dictionaries
  - Part-of speech (PoS) taggers
  - Bilingual dictionaries
  - Structural transfer rules
Goal

Goal:
- To reduce the human effort
- Using corpus-based methods
- In an unsupervised way

Focus on:
- the PoS taggers used in the analysis phase
- the set of shallow structural transfer rules used in translation

⇒ Benefiting from the rest of resources ⇐

http://apertium.org
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Part-of-speech tagging /1

**Problem:** Selecting the correct PoS tag for those words with more than one (ambiguous words)

⇒ *Hidden Markov models* (HMM) are one of the standard statistical solutions

- Each HMM state corresponds to a different PoS tag
- Each input word is replaced by its corresponding ambiguity class
In MT PoS tagging becomes crucial:

- Translation may differ from one PoS tag to another

<table>
<thead>
<tr>
<th>English</th>
<th>PoS</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>book</td>
<td>noun</td>
<td>libro</td>
</tr>
<tr>
<td>verb</td>
<td></td>
<td>reservar</td>
</tr>
</tbody>
</table>

- Structural transformations may be applied (or not) for some PoS tag

<table>
<thead>
<tr>
<th>English</th>
<th>PoS</th>
<th>Spanish</th>
<th>reordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>the green house</td>
<td>green--adj</td>
<td>la casa verde</td>
<td>←rule</td>
</tr>
<tr>
<td></td>
<td>green--noun</td>
<td>* el césped casa</td>
<td>applied</td>
</tr>
</tbody>
</table>
General-purpose HMM training methods:

- **Supervised** (hand-tagged corpora available):
  - Maximum-likelihood estimate (MLE)
- **Unsupervised** (only untagged corpora available):
  - Baum-Welch (expectation-maximization, EM)

Main features:

- Only use information from the language being tagged
- Independent of the natural language processing application
- To get high tagging accuracy **supervised** resources (hand-tagged corpora) must be built
MT-oriented HMM training method

- PoS tagging is just an intermediate task for the whole translation procedure
- Good translation performance, rather than PoS tagging accuracy, becomes the real objective

**Idea:** As the goal is to get good translations into TL, let a TL model decide whether a given “construction” in the TL is good or not
Unsupervised training

Resources required:
- an SL untagged text automatically obtained from an SL raw corpus
- the other modules of the MT system following the PoS tagger
- a TL model trained from a raw TL corpus
MT-oriented HMM training method: overview /2

Procedure:
1. SL corpus is segmented
2. All possible disambiguations of each segment are translated into TL
3. A TL model is used to score each translation
4. HMM parameters are computed according to the likelihood of the corresponding translations into TL

⇒ The resulting tagger is tuned to the translation fluency ⇐
Example: English $\rightarrow$ Spanish

- SL segment (English):
  - He-prn rocks-noun | verb the-art table-noun | verb
Example: English → Spanish

- **SL segment (English):**
  - He-prn rocks-noun | verb the-art table-noun | verb

- **Possible translations (Spanish) according to each disambiguation and their normalized likelihoods according to a TL model:**
  - Él-prn mece-verb la-art mesa-noun 0.75
  - Él-prn mece-verb la-art presenta-verb 0.15
  - Él-prn rocas-noun la-art mesa-noun 0.06
  - Él-prn rocas-noun la-art presenta-verb + 0.04

The HMM parameters involved in these 4 disambiguations are updated according to their likelihoods in the TL.
Example: English → Spanish

- **SL segment (English):**
  - He-prn rocks-noun|verb the-art table-noun|verb

- **Possible translations (Spanish) according to each disambiguation and their normalized likelihoods according to a TL model:**
  - Él-prn mece-verb la-art mesa-noun 0.75
  - Él-prn mece-verb la-art presenta-verb 0.15
  - Él-prn rocas-noun la-art mesa-noun 0.06
  - Él-prn rocas-noun la-art presenta-verb + 0.04 1.00

- The HMM parameters involved in these 4 disambiguations are updated according to their likelihoods in the TL
Experiments /1

- **Task**: training PoS tagger for Spanish, French and Occitan to be used in MT into Catalan

- **TL model**: trigram language model trained from a Catalan corpus with $\approx 2 \cdot 10^6$ words

- **Experiments** conducted with
  - 5 disjoint corpora with $0.5 \cdot 10^6$ words for Spanish
  - 5 disjoint corpora with $0.5 \cdot 10^6$ words for French
  - Only one corpus with $0.3 \cdot 10^6$ words for Occitan
Experiments /2

- **Reference results:**
  - **Baum-Welch** expectation maximization on $10 \cdot 10^6$ words corpora
  - **Supervised**: MLE from a hand-tagged corpus $\approx 21.5 \cdot 10^3$ words (only for Spanish)
  - **TLM-best**: when a TL model is used at translation time to select always the most likely translation
    - approximate indication of the best results the MT-oriented method could achieve
Some results: Spanish→Catalan

Mean and std. dev. of the translation performance, WER (% of words)

---

Felipe Sánchez Martínez (Univ. d'Alacant)
Some results: Spanish→Catalan /2

Mean and std. dev. of the PoS tagging error rate (% of words)
Some results: Spanish → Catalan / 3

Why are the translation performances for the supervised and the MT-oriented method comparable, but no the PoS tagging error rates?

- TL information does not discriminate among the SL analyses of a segment leading to the same translation

<table>
<thead>
<tr>
<th>French</th>
<th>PoS</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>la ville</td>
<td>la-art</td>
<td>la ciudad</td>
</tr>
<tr>
<td></td>
<td>la-prn</td>
<td></td>
</tr>
</tbody>
</table>

- **Free-ride**: phenomenon by which choosing the incorrect interpretation for an ambiguous word does not result in a translation error
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Disadvantages of the MT-oriented method

- The number of possible disambiguations to translate grows exponentially with segment length
- Translation is the most time-consuming task
- **Goal**: To overcome this problem
- **How**: Pruning unlikely disambiguation paths by using *a priori* knowledge
Pruning method /1

- Based on an initial model of SL tags

Assumption: Any reasonable model of SL tags may be useful to choose a subset of possible disambiguation paths so that the correct one is in that subset

- The model used for pruning can be updated dynamically during training
1. The *a priori* likelihood of each possible disambiguation path of SL segment $s$ is calculated using the pruning model.

2. The set of disambiguation paths to take into account is determined by using a mass probability threshold $\rho$ - Only the minimum number of paths to reach the mass probability threshold $\rho$ are taken into account.
Example (English→Spanish)

- SL segment (English):
  - He-prn rocks-noun|verb the-art table-noun|verb
Example (English→Spanish)

- **SL segment (English):**
  - He-prn rocks-noun|verb the-art table-noun|verb

- **Normalized *a priori* likelihoods:**
  
  \[
  \begin{align*}
  g_1 &= (\text{prn, verb, art, noun}) & 0.69 \\
  g_2 &= (\text{prn, verb, art, verb}) & 0.14 \\
  g_3 &= (\text{prn, noun, art, noun}) & 0.10 \\
  g_4 &= (\text{prn, noun, art, verb}) & + 0.07 \\
  \end{align*}
  \]

  \[
  = 1.00
  \]
Example (English → Spanish)

- **SL segment (English):**
  - He-prn rocks-noun|verb the-art table-noun|verb

- **Normalized *a priori* likelihoods:**
  - \( g_1 = (\text{prn, verb, art, noun}) = 0.69 \)
  - \( g_2 = (\text{prn, verb, art, verb}) = 0.14 \)
  - \( g_3 = (\text{prn, noun, art, noun}) = 0.10 \)
  - \( g_4 = (\text{prn, noun, art, verb}) + 0.07 \)

With \( \rho = 0.8 \), \( g_3 \) and \( g_4 \) are discarded because \( 0.69 + 0.14 \geq 0.8 \)
Some results: Spanish→Catalan

Mean and std. dev. of the translation performance, WER (% of words)
Some results

Ratio of translated words

Ratio of translated words (%)
Probability mass (\( \rho \))
Spanish–Catalan
French–Catalan
Occitan–Catalan

Felipe Sánchez Martínez (Univ. d’Alacant)
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Best HMM topology for taggers used in MT

- Large tagsets (set of PoS tags) for richly-inflected languages
  - fine PoS tags convey lot of information
    e.g. verb.pret.3rd.pl, noun.m.sg

- A reduced tagset **manually defined** following linguistic guidelines is usually used
  - Maps fine tags into coarse ones
  - Should allow for better parameter estimation

- **Goal**: To automatically determine the set of states to be used
  - Avoid the human intervention in defining the tagset

  ⇒ *Model merging* approach (Stolcke and Omohundro, 1994) cannot be applied using untagged corpora
**Bottom-up agglomerative clustering**

1. Place each object in its own cluster (singleton)

2. Iteratively **compare** all pairs of clusters and choose the two **closest** clusters according to a distance measure
   - If the distance between the selected clusters is below a certain threshold, **merge** both clusters
   - Otherwise, **stop**
Clustering of PoS tag

- First model trained using the large tagset via the MT-oriented method
- Distance between cluster based on the state-to-state transition probabilities
- An additional constraint ensures that it is possible to restore the information about the fine tag from the coarse one
Some results: Spanish → Catalan

Mean and std. dev. of the translation performance, WER

![Graph showing word error rate (WER) vs. number of coarse tags and threshold.]

Felipe Sánchez Martínez (Univ. d'Alacant)
Some results: French→Catalan

Mean and std. dev. of the translation performance, WER

Word error rate (WER, % of words)

Threshold

Number of coarse tags

Wer

# tags

Felipe Sánchez Martínez (Univ. d’Alacant)
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Automatic inference of transfer rules

Goal:
- To automatically learn those transformations that produce correct translations in the TL

How:
- Adapting the alignment templates (ATs) already used in statistical MT to the shallow-transfer approach
  - AT \( z = (S_n, T_m, G) \)
    - \( S_n \): sequence of \( n \) SL word classes
    - \( T_m \): sequence of \( m \) TL word classes
    - \( G \): alignment information
Resources required:
- A SL–TL parallel corpus
- The morphological analyzers and PoS taggers of the MT system
- The bilingual dictionary of the MT system

Procedure:
1. Analyze both sides of the training corpus
2. Compute word alignments
3. Extract bilingual phrase pairs and derive ATs from them
4. Generate shallow-transfer rules
Word class: part-of-speech (including all the inflection information)

Exception: lexicalized words are placed in single-word classes

Lexicalized categories: categories that are known to be involved in lexical changes, such as prepositions

the method can learn not only syntactic changes
ATs are extended with a set $R$ of restrictions over the TL inflection information of non-lexicalized words

$$AT\ z = (S_n, T_m, G, R)$$
ATs are extended with a set $R$ of restrictions over the TL inflection information of non-lexicalized words

$$AT\ z = (S_n, T_m, G, R)$$

Restrictions are derived from the bilingual dictionary

- Bilingual entry that does not change inflection information
  $$<e><p>
  \langle l\rangle castigo\langle s\ n="noun"/\rangle</l>
  \langle r\rangle c\`astig\langle s\ n="noun"/\rangle</r>
  </p></e>$$

  $R: \ w=\text{noun.}^*$

- Bilingual entry that does change inflection information
  $$<e><p>
  \langle l\rangle calle\langle s\ n="noun"/\rangle<s\ n="f"/\rangle</l>
  \langle r\rangle carrer\langle s\ n="noun"/\rangle<s\ n="m"/\rangle</r>
  </p></e>$$

  $R: \ w=\text{noun.m.}^*$

- The bilingual dictionary is also used to discard phrase pairs that cannot be reproduced by the MT system
Alignment template example /1

Bilingual phrase:

\[
\begin{align*}
\text{Alacant} & \quad \text{viure} & \quad \text{van} & \quad \text{Alicante} \\
& \quad \text{anar} & \quad \text{en} & \quad \text{vivieron}
\end{align*}
\]

Alignment template:

\[
\begin{align*}
\text{(noun.loc)} & \quad \text{(verb.pret.3rd.pl)} & \quad \text{(verb.inf)} & \quad \text{(vbaux.pres.3rd.pl)} & \quad \text{(noun.loc)} \\
\text{a-(pr)} & \quad \text{viure-(verb.inf)} & \quad \text{anar-(vbaux.pres.3rd.pl)} & \quad \text{en-(pr)} & \quad \text{Alacant-(noun.loc)}
\end{align*}
\]

Spanish analysis:
\[
vivieron en Alicante \rightarrow
vivir-(verb.pret.3rd.pl) \quad en-(pr) \\
Alicante-(noun.loc)
\]

Catalan analysis:
\[
van viure a Alacant \rightarrow
anar-(vbaux.pres.3rd.pl) \quad viure-(verb.inf) \\
a-(pr) \quad Alacant-(noun.loc)
\]

Restrictions: \( w_2 = \text{verb.\ast}, \ w_4 = \text{noun.\ast} \)

\(^1\)Translated into English as *They lived in Alicante*
Alignment template example /2

Bilingual phrase:

Spanish analysis:  

\[ \text{la calle estrecha} \rightarrow \text{el}-(\text{art.f.sg}) \]
\[ \text{calle}-(\text{noun.f.sg}) \text{ estrecho}-(\text{adj.f.sg}) \]

Catalan analysis:  

\[ \text{el carrer estret} \rightarrow \text{el}-(\text{art.m.sg}) \]
\[ \text{carrer}-(\text{noun.m.sg}) \text{ estret}-(\text{adj.m.sg}) \]

Restrictions:  

\[ w_2 = \text{noun.m.} *, \ w_3 = \text{adj.} * \]

\[ \text{Translated into English as } \text{The narrow street} \]
Generation of Apertium transfer rules

Procedure:

1. Discard useless AT
2. Select the AT to use according to their frequency
3. For all ATs with the same SL part a rule is generated

Rule generation:

- The rule matches the SL part all ATs have in common
- In decreasing order of AT frequency counts code is generated to
  - test the restrictions $R$ over the TL inflection information
  - if they hold, apply the AT and stop rule execution
- code that translates word-for-word is added
  - it is executed only if none of the AT were applicable
AT applicability test /1

Restrictions $R$ are tested by looking at the bilingual dictionary.

Example:

- **$R$:** $w_2 = \text{noun.m.} \ast$, $w_3 = \text{adj.} \ast$
- **Input string (Spanish):** *la señal roja* $\rightarrow$
  - el-(art.f.sg) *señal*-(noun.f.sg)
  - *rojo*-(adj.f.sg)
- **Translation of non-lexicalized words:**
  - *señal*-(noun.f.sg) $\rightarrow$ *senyal*-(noun.m.sg)
  - *rojo*-(adj.f.sg) $\rightarrow$ *vermell*-(adj.f.sg)
- **Restriction holds, AT can be applied**
AT applicability test /2

Restrictions $R$ are tested by looking at the bilingual dictionary

Example:

- $R$: $w_2 = \text{noun.m.} \ast$, $w_3 = \text{adj.} \ast$
- Input string (Spanish): *la silla blanca* $\rightarrow$
  - *el* (art.f.sg) *silla* (noun.f.sg)
  - *blanco* (adj.f.sg)
- Translation of non-lexicalized words:
  - *silla* (noun.f.sg) $\rightarrow$ *cadira* (noun.f.sg)
  - *blanco* (adj.f.sg) $\rightarrow$ *blanc* (adj.f.sg)
- Restriction does not hold, AT cannot be applied
Alignment templates application, an example

Spanish (input): \text{permanecieron en Alemania}^{3} \rightarrow 
\text{permanecer}-(\text{verb.pret.3rd.pl}) \text{ en}-(\text{pr}) 
\text{Alemania}-(\text{noun.loc})

Catalan (output): \text{anar}-(\text{vbaux.pres.3rd.pl}) 
\text{romandre}-(\text{verb.inf}) \text{ a}-(\text{pr}) 
\text{Alemanya}-(\text{noun.loc}) \rightarrow 
\text{van romandre a Alemanya}

Word-for-word translation:
\text{romangueren *en Alemanya}

\text{R: } w_1 = \text{verb.*}, w_3 = \text{noun.*}

\text{3 Translated into English as They remained in Germany}
Experiments

- **Task**: Inference of shallow-transfer rules for Spanish↔Catalan, Spanish↔Galician and Spanish→Portuguese

- $\approx 8$ lexicalized categories

- Two different training corpora:
  - One with $2 \cdot 10^6$ words
  - Another with only $0.5 \cdot 10^6$ words

- Two different evaluation corpora:
  - *post-edit* reference translation is a *post-edited* version of the MT performed using *hand-coded* transfer rules
  - *parallel* text to translate and reference translation comes from a parallel corpus *analogous* to the one used for training
## Some results

### Spanish → Catalan, WER ± 95% confidence interval

<table>
<thead>
<tr>
<th>Training</th>
<th>Test</th>
<th>Word-for-word</th>
<th>AT transfer</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \cdot 10^6$</td>
<td>post-edit</td>
<td>12.6 ± 0.9</td>
<td>8.7 ± 0.7</td>
<td>6.7 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>parallel</td>
<td>26.4 ± 1.2</td>
<td>20.3 ± 1.1</td>
<td>20.7 ± 1.0</td>
</tr>
<tr>
<td>$0.5 \cdot 10^6$</td>
<td>post-edit</td>
<td>12.6 ± 0.9</td>
<td>9.9 ± 0.7</td>
<td>6.7 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>parallel</td>
<td>26.4 ± 1.2</td>
<td>21.4 ± 1.1</td>
<td>20.7 ± 1.0</td>
</tr>
</tbody>
</table>

### Spanish → Portuguese, WER ± 95% confidence interval

<table>
<thead>
<tr>
<th>Training</th>
<th>Test</th>
<th>Word-for-word</th>
<th>AT transfer</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \cdot 10^6$</td>
<td>post-edit</td>
<td>11.9 ± 0.8</td>
<td>12.1 ± 0.9</td>
<td>7.0 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>parallel</td>
<td>47.9 ± 1.7</td>
<td>46.5 ± 1.7</td>
<td>47.6 ± 1.8</td>
</tr>
<tr>
<td>$0.5 \cdot 10^6$</td>
<td>post-edit</td>
<td>11.9 ± 0.8</td>
<td>12.1 ± 0.9</td>
<td>7.0 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>parallel</td>
<td>47.9 ± 1.7</td>
<td>47.4 ± 1.7</td>
<td>47.6 ± 1.8</td>
</tr>
</tbody>
</table>
Some results

Why such a large difference between Spanish→Catalan and Spanish→Portuguese?

- Because of how training corpora have been built
  - Spanish→Catalan, by translating one language into another (newspaper *El Periódico de Catalunya*)
    - 22% of discarded ATs
  - Spanish→Portuguese, by translating from a third language (*JRC-ACQUIS* parallel corpus)
    - 53% of discarded ATs
Outline

1. Motivation & goal
2. Part-of-speech taggers for machine translation
   - Part-of-speech tagging
   - MT-oriented hidden Markov model training
3. Pruning of disambiguation paths
   - Disadvantages of the MT-oriented method
   - Pruning method
4. Part-of-speech tag clustering
   - Best HMM topology for taggers used in MT
   - Bottom-up agglomerative clustering
5. Automatic inference of transfer rules
   - Alignment templates for shallow-transfer machine translation
   - Generation of Apertium transfer rules
6. Concluding remarks
Concluding remarks

Steps towards more efficient development of RBMT systems

- A new method to train PoS tagger to be used in MT
  - focuses on the task in which it will be used
  - uses TL information without using parallel corpora
  - benefits from information in the rest of modules
  - using a priori knowledge saves around 80% of the translations to perform while training
  - better translation quality than tagging accuracy

- PoS tags clustering
  - has not provided the expected results, but
  - may be useful if the number of states is crucial
Concluding remarks /2

- A method to infer shallow-transfer rules from parallel corpora
  - extends the definition of alignment template
  - small amount of information provided by human is used
  - the process followed to build the parallel corpus deserves special attention
  - inferred rules are human-readable
  - they can coexist with hand-coded rules
Open-source software

- Can be downloaded from sf.net/projects/apertium
  - Packages apertium-tagger-training-tools and apertium-transfer-tools
- Ensures reproducibility
- Allows other researchers to improve them
- Eases the development of new language pairs for Apertium

apertium-tagger-training-tools is being used by Prompsit Language Engineering S.L.
Future research lines

This thesis opens several research lines:

- the use of TL information to train other statistical models that run on the SL
- the use of more than one TL (triangulation)
- the use of a TL model of different nature
- linguistically-driven extraction of bilingual phrases
- a more flexible way to use lexicalized categories
- a bootstrapping method to learn both the PoS tagger and the set of transfer rules cooperatively
- ...
Acknowledgments

- Spanish Ministry of Education and Science, and European Social Fund; research grant BES-2004-4711
- Autonomous Government of Catalonia; project Traducció automàtica de codi obert per al català
- Spanish Ministry of Education and Science; project TIN2006-15071-C03-01

⇒ Thank you very much for your attention⇐