Towards Interactive and Automatic Refinement of Translation Rules

Ariadna Font Llitjós

PhD Thesis Proposal

Jaime Carbonell (advisor)
Alon Lavie (co-advisor)
Lori Levin
Bonnie Dorr (Univ. Maryland)

5 November 2004
Outline

• Introduction
• Thesis statement and scope
• Preliminary Research
  – Interactive elicitation of error information
  – A framework for automatic rule adaptation
• Proposed Research
• Contributions and Thesis Timeline
Machine Translation (MT)

- **Source Language (SL) sentence:**
  
  *Gaudi was a great artist*

  **Spanish translation:**
  
  *Gaudi era un gran artista*

- MT System outputs:

  → *Gaudi estaba un artista grande*

  → *Gaudi era un artista grande*
Spanish Adjectives

General order: grande → big in size

Exception: gran → exceptional
Commercial and Online Systems

Correct Translation: *Gaudi era un gran artista*

- Systran, Babelfish (Altavista), WorldLingo, Translated.net:
  \[ \rightarrow *Gaudi \text{ era } \emptyset \text{ gran artista} \]

- ImTranslation:
  \[ \rightarrow *\text{El Gaudi era un gran artista} \]

- 1-800-Translate
  \[ \rightarrow *\text{Gaudi era un fenomenal artista} \]
Post-editing

• Current solutions:

  → Post-editing [Allen, 2003]
     by human linguists or editors (experts)

  → Automated Post-Editing prototype module (APE)
     [Allen & Hogan, 2000]
     to alleviate the tedious task of correcting most frequent errors over and over

• No solution to fully automate post-editing process
Drawbacks of Current Methods

• Manual post-editing → Corrections do not generalize
  → Gaudi era un artista grande
  → Juan es un amigo grande (Juan is a great friend)
  → Era una oportunidad grande (It is a great opportunity)

• APE → Humans need to predict all the errors ahead of time and code for the post-editing rules; given new error → ☹️
My Solution

• Automate post-editing efforts by feeding them back into the MT system.

• Possible alternatives:
  → Automatic learning of post-editing rules
    + system independent
    - several thousands of sentences might need to be corrected for the same error
  → Automatic refinement of translation rules
    + attacks the core of the problem
    - for transfer-based MT systems (need rules to fix!)
Interactive and Automatic Rule Refinement

Related Work

[Corston-Oliver & Gammon, 2003]
[Imamura et al. 2003]
[Menezes & Richardson, 2001]

[Brill, 1993]
[Gavaldà, 2000]

[Callison-Burch, 2004]

[Su et al. 1995]

[Allen & Hogan, 2000]

My Thesis
No pre-existing training data required
No human reference translations required
*Non-expert* user feedback
Resource-poor Scenarios (AVENUE)

- Lack of electronic parallel data
- Lack of computational linguists
  → Lack of manual grammar

Why bother?
- Indigenous communities have difficult access to crucial information that directly affects their life (such as land laws, plagues, health warnings, etc.)
- Preservation of their language and culture

Resource-poor Languages:
Mapudungun
Quechua
Aymara
How is MT possible for resource-poor languages?

Bilingual speakers
AVENUE Project Overview

Interactive and Automatic Rule Refinement
Interactive and Automatic Rule Refinement
Related Work

- Post-editing
- Rule Adaptation
- Fixing
- Machine Translation

My Thesis

Resource-poor languages

Interactive and Automatic Rule Refinement
Given a rule-based Transfer MT system, we can:

- Extract useful information from non-expert bilingual speakers to correct MT output.
- Automatically refine and expand translation rules, given corrected and aligned translation pairs and some error information.

So that the set of refined rules has better coverage and higher overall MT quality.
Assumptions

• No pre-existing parallel training data

• No pre-existing human reference translations

• The SL sentence needs to be fully parsed by the translation grammar.

• Bilingual speakers can give enough information about the MT errors.
Scope

Evaluate automatic refinement for the following conditions:

1. User correction information only.
2. Correction and error information.
3. Extra information is required → user interaction.

Both in manually written and automatically learned grammars [AMTA 2004].
Technical Challenges

Automatic Evaluation of Refinement process

Automatically Refine and Expand Translation Rules *minimally*

Manually written  Automatically Learned

Elicit *minimal* MT information from *non-expert* users
Preliminary Work

- Interactive elicitation of error information
- A framework for automatic rule adaptation
Error Typology for Automatic Rule Refinement (simplified)

- Missing word
- Extra word
- Wrong word order
- Incorrect word
- Wrong agreement

Local vs Long distance
- Word vs. phrase
  - + Word change
  - Sense
  - Form
  - Selectional restrictions
  - Idiom

Missing constraint
- Extra constraint
Interactive and Automatic Rule Refinement

TCTool (Demo)

Actions:  
- Add a word  
- Delete a word  
- Modify a word  
- Change word order
Interactive and Automatic Rule Refinement

1st Eng2Spa User Study

[LREC 2004]

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>error detection</td>
<td>90%</td>
<td>89%</td>
</tr>
<tr>
<td>error classification</td>
<td>72%</td>
<td>71%</td>
</tr>
</tbody>
</table>

- MT error classification $\rightarrow 9$ linguistically-motivated classes [Flanagan, 1994], [White et al. 1994]:
  - word order, sense, agreement error (number, person, gender, tense), form, incorrect word and no translation
Translation Rules

{NP,8}
NP::NP : [DET ADJ N] -> [DET N ADJ]
( (X1::Y1) (X2::Y3) (X3::Y2)
;; English parsing:
  ((x0 def) = (x1 def)) → NP definiteness = DET definiteness
  (x0 = x3) → NP = N (N is the head of the NP)
;; Spanish generation:
  ((y1 agr) = (y2 agr)) → DET agreement = N agreement
  ((y3 agr) = (y2 agr)) → ADJ agreement = N agreement
  (y2 = x3) → Pass the features of English N to Spanish N

ADJ::ADJ |: [nice] -> [bonito]
((X1::Y1)
((x0 pos) = adj)
((x0 form) = nice)
((y0 agr num) = sg) → Spanish ADJ is singular in number
((y0 agr gen) = masc)) → Spanish ADJ is masculine in number
Automatic Rule Refinement Framework

• Find **best RR operations** given a:
  • Grammar (G),
  • Lexicon (L),
  • (Set of) Source Language sentence(s) (SL),
  • (Set of) Target Language sentence(s) (TL),
  • Its Parse tree (P), and
  • Minimal correction of TL (TL’)

  such that **TQ2 > TQ1**

• Which can also be expressed as:
  \[ \max TQ(TL|TL’,P,SL,RR(G,L)) \]
Types of Refinement Operations

1. Refine a translation rule:

   \[ R_0 \rightarrow R_1 \] (change \( R_0 \) to make it more specific or more general)

R0:

- NP
  - DET a
  - ADJ nice
  - N house

R1:

- NP
  - DET a
  - ADJ nice
  - N house

- NP
  - DET una
  - N casa
  - ADJ bonito

\[ N \text{ gender} = \text{ADJ gender} \]
2. Bifurcate a translation rule:

\[ R0 \rightarrow R0 \text{ (same, general rule)} \]
\[ \rightarrow R1 \text{ (add a new more specific rule)} \]

**R0:**

- **NP:**
  - **DET:** a
  - **ADJ:** nice
  - **N:** house

**R1:**

- **NP:**
  - **DET:** a
  - **ADJ:** great
  - **N:** artist

**ADJ type:** pre-nominal
Formalizing Error Information

\[ W_i = \text{error} \]
\[ W_i' = \text{correction} \]
\[ W_c = \text{clue word} \]

\[ TL_m = (W_1, \ldots, W_i, \ldots, W_n) \]

\[ TL'_m = (W_1, \ldots, W_i', \ldots, W_c, \ldots, W'_n) \]
Triggering Feature Detection

Comparison at the feature level to detect triggering feature(s)

→ Delta function: \( \delta(W_i, W_{i'}) \)

Examples:

\[
\begin{align*}
\delta(\text{bonito}, \text{bonita}) &= \{\text{gender}\} \\
\delta(\text{comiamos}, \text{comia}) &= \{\text{person}, \text{number}\} \\
\delta(\text{mujer}, \text{guitarra}) &= \emptyset
\end{align*}
\]

If \( \delta \) set is empty, need to postulate a new binary feature
Deciding on the Refinement Op

Given:

- **Action performed by the user** (add, delete, modify, change word order)

- **Error information available** (error type, clue word, word alignments, etc.)

→ Refinement Action
Interactive and Automatic Rule Refinement

Rule Refinement Operations

Modify
- $+W_c$, $-W_c$
- $\delta \neq \emptyset$, $\delta = \emptyset$

Add
- $+W_c$, $-W_c$
- +rule, -rule
- +al, -al

Delete
- +al, -al
- $W_i$, $W_c$
- $Wi(\ldots)$, $W_c$
- $-W_c$

Change $W$ Order
- $Wi \rightarrow W_i'$
- RuleLearner
- $\delta \neq \emptyset$, $\delta = \emptyset$

POS$_i$ = POS$_i'$
POS$_i$ $\neq$ POS$_i'$
Proposed Work

- Rule Refinement Example
- Batch mode implementation
- Interactive mode implementation
- User Studies
- Evaluation
Rule Refinement Example

Change word order

SL: "Gaudí was a great artist"

MT system output:

TL: "Gaudí era un artista grande"

Goal (given by user correction):

*"Gaudí era un artista grande"

Gaudí era un gran artista
1. Error Information Elicitation
2. Variable Instantiation from Log File

Correcting Actions:

1. Word order change \((\text{artista grande} \rightarrow \text{grande artista})\):
   \[ W_i = \text{grande} \]

2. Edited \text{grande} into \text{gran}:
   \[ W_i' = \text{gran} \]
   identified \text{artist} as clue word \(\rightarrow W_c = \text{artista} \)

In this case, even if user had not identified \(W_c\), refinement process would have been the same
3. Retrieve Relevant Lexical Entries

• No lexical entry for [great → gran]

• Duplicate lexical entry [great → grande] and change TL side:

\[ ADJ::ADJ |: [great] \rightarrow [grande] \]
\((X1::Y1)\)
\((\ldots)\)
\(((y0 \text{ agr num}) = \text{sg})\)
\(((y0 \text{ agr gen}) = \text{masc})\)  

\[ ADJ::ADJ |: [great] \rightarrow [gran] \]
\((X1::Y1)\)
\((\ldots)\)
\(((y0 \text{ agr num}) = \text{sg})\)
\(((y0 \text{ agr gen}) = \text{masc})\)  

(Morphological analyzer: \textit{grande} = \textit{gran})
4. Finding triggering feature(s)

Feature $\delta$ function: $\delta(W_i, W_i') = \emptyset$

$\rightarrow$ need to postulate a new binary feature: feat1

5. Blame assignment (from MT system output)

tree: $<((S,1 (NP,2 (N,5:1 "GAUDI") )$

Grammar

S,1
...
NP,1
...
NP,8
...

(VP,3 (VB,2 (AUX,17:2 "ERA") )

(NP,8 (DET,0:3 "UN")

(N,4:5 "ARTISTA")

(ADJ,5:4 "GRANDE") ) ) ) >
6. Variable Instantiation in the Rules

\[ W_i = \text{grande} \rightarrow \text{POS}_i = \text{ADJ} = Y3, y3 \]
\[ W_c = \text{artista} \rightarrow \text{POS}_c = \text{N} = Y2, y2 \]

\{NP,8\}  ;;  Y1  Y2  Y3
NP::NP : [DET ADJ N] -> [DET N ADJ]
( (X1::Y1) (X2::Y3) (X3::Y2)
 (\(x0 \text{ def}) = (x1 \text{ def})\)
 (x0 = x3)
 (\(y1 \text{ agr}) = (y2 \text{ agr})\) ; det-noun agreement
 (\(y3 \text{ agr}) = (y2 \text{ agr})\) ; adj-noun agreement
 (y2 = x3) )
7. Refining Rules

- Bifurcate $\text{NP}_8 \rightarrow \text{NP}_8 (R0) + \text{NP}_8' (R1)$
  
  (flip order of ADJ-N)

$\{\text{NP}_8'\}$

$\text{NP::NP} : [\text{DET ADJ N}] \rightarrow [\text{DET ADJ N}]$

$\times (X1::Y1) (X2::Y2) (X3::Y3)$

$((x0 \text{ def}) = (x1 \text{ def}))$

$x0 = x3$

$((y1 \text{ agr}) = (y3 \text{ agr})) ; \text{det-noun agreement}$

$((y2 \text{ agr}) = (y3 \text{ agr})) ; \text{adj-noun agreement}$

$y2 = x3$

$((y2 \text{ feat1}) = c + )$
8. Refining Lexical Entries

ADJ::ADJ |: [great] -> [grande]
((X1::Y1)
((x0 form) = great)
((y0 agr num) = sg)
((y0 agr gen) = masc)
((y0 feat1) = -))

ADJ::ADJ |: [great] -> [gran]
((X1::Y1)
((x0 form) = great)
((y0 agr num) = sg)
((y0 agr gen) = masc)
((y0 feat1) = +))
Done? Not yet

Need to restrict application of general rule (R0) to just post-nominal ADJ

un artista grande
un artista gran
un gran artista
*un grande artista
Add Blocking Constraint

NP,8 (R0) ← ADJ(grande)
[feat1 = -] [feat1 = -]

NP,8’ (R1) ← ADJ(gran)
[feat1 =c +] [feat1 = +]

un artista grande ←
*un artista gran
un gran artista
*un grande artista

Can we also eliminate incorrect translations automatically?
Making the grammar tighter

- If $W_c = \text{artista}$
  - Add $[\text{feat1}=+]$ to $\text{N(artista)}$
  - Add agreement constraint to $\text{NP,8 (R0)}$ between $\text{N}$ and $\text{ADJ}$ ($(\text{N feat1}) = (\text{ADJ feat1})$)

*un artista grande
*un artista gran
un gran artista
*un grande artista
Batch Mode Implementation

• Given a set of user corrections, apply refinement module.

• For Refinement Operations of errors that can be refined fully automatically using:

  1. Correction information only
  2. Correction and error information

  error type, clue word
Interactive and Automatic Rule Refinement

Rule Refinement Operations

Modify
- \( +W_c \)
- \( -W_c \)
\( \delta \neq \emptyset \)
\( \delta = \emptyset \)

Add
- \( +W_c \)
- \( -W_c \)
+al
- al

Delete
- \( +W_c \)
- \( -W_c \)
+al
- al

Change
- \( W_i W_c \)
- \( W_i(\ldots)W_c \)
- \( -W_c \)

Order
- \( =W_i \)
- \( W_i \rightarrow W_i' \)

RuleLearner

POS_i = POS_i'
POS_i \neq POS_i'
1. Correction info only

Rule Refinement Operations

Modify

- $\delta \neq \emptyset$
- $\delta = \emptyset$

Add

- $+W_c$
- $-W_c$

Delete

- $+\text{al}$
- $-\text{al}$

Change W Order

- $W_i W_c$
- $W_i(\ldots) W_c$
- $-W_c$

Modify

- $+W_c$
- $-W_c$

Add

- $+W_c$
- $-W_c$

Delete

- $+\text{al}$
- $-\text{al}$

It is a nice house – *Es una casa bonito

$\Rightarrow$ Es una casa bonita

Gaudi was a great artist – *Gaudi era un artista grande

$\Rightarrow$ Gaudi era un gran artista
2. Correction and Error info

Rule Refinement Operations

Modify
- \(+W_c\) - \(-W_c\)
- \(\delta \neq \emptyset\)
- \(\delta = \emptyset\)
- \(\text{pos} = \text{pos}'\)

Add
- \(+W_c\)
- \(-W_c\)
- \(+al\)
- \(-al\)
- \(\text{PP} \rightarrow \text{PREP NP}\)

Delete
- \(-al\)
- \(W_i\)
- \(-W_c\)

Change W Order
- \(W_i(\ldots)\)
- \(-W_c\)
- \(\text{POS}_i = \text{POS}'_i\)
- \(\text{POS}_i \neq \text{POS}'_i\)

I am proud of you – *Estoy orgullosa de tu
\rightarrow Estoy orgullosa de ti
Interactive Mode Implementation

• Extra error information is required to determine triggering context automatically

→ Need to give other relevant sentences to the user at run-time (minimal pairs)

• For Refinement Operations of errors that can be refined fully automatically but:

3. require a further user interaction
3. Further user interaction

Rule Refinement Operations

Modify

Add

Delete

Change W Order

\(\delta \neq \emptyset\) \(\delta = \emptyset\) +rule –rule +al –al

\(+W_c -W_c +W_c -W_c +al -al\)

\(W_i W_c W_i(...) W_c -W_c\)

POS_i = POS_i’ \(\Rightarrow\) POS_i \(\neq\) POS_i’

\(I\ see\ them - *Veo\ los\)

\(\Rightarrow\) Los veo
Example Requiring Minimal Pair

1. Run SL sentence through the transfer engine

   \[ I \text{ see them} \rightarrow *\text{veo los} \quad \text{Correct TL: los veo} \]

2. \( W_i = \text{los} \quad \text{but no } W_i' \text{ nor } W_c \)

   \[ \rightarrow \text{Need a minimal pair to determine appropriate refinement:} \]

   \[ I \text{ see cars} \rightarrow \text{veo autos} \checkmark \]

3. Triggering feature(s):
   \[ \delta(\text{veo los, veo autos}) \]

   \[ \delta(\text{los,autos}) = \{\text{pos}\} \]

   \[ \text{PRON(los)}[\text{pos=}\text{pron}] \quad \text{N(autos)}[\text{pos=}\text{n}] \]
Refining and Adding Constraints

VP,3: VP NP $\rightarrow$ VP NP  
(veo los, veo autos)

VP,3’: VP NP $\rightarrow$ NP VP + [NP pos =c pron]  
(los veo, *autos veo)

• Percolate triggering features up to the constituent level:

NP: PRON $\rightarrow$ PRON + [NP pos = PRON pos]

• Block application of general rule (VP,3):

VP,3: VP NP $\rightarrow$ VP NP + [NP pos = (*NOT* pron)]  
*veo los, veo autos  (los veo, *autos veo)
Generalization Power

When triggering feature already exists in the feature language (pos, gender, number, etc.)

- I see them $\rightarrow$ *veo los $\rightarrow$ **los veo**
- I love him $\rightarrow$ **lo amo** (before: *amo lo)
- They called me yesterday $\rightarrow$ **me llamaron ayer**
  (before: *llamaron me ayer)
- Mary helps her with her homework
  $\rightarrow$ Maria **le ayuda con sus tareas**
  (before: *Maria ayuda le con sus tareas)
User Studies

- TCTool: new MT classification (Eng2Spa)
- Different language pair
  - Mapudungun or Quechua → Spanish
- Batch vs Interactive mode
- Amount of information elicited
  - just corrections vs + error information
Evaluation of Refined MT Output

1. Evaluate best translation $\rightarrow$ Automatic evaluation metrics (BLEU, NIST, METEOR)

2. Evaluate translation candidate list size $\rightarrow$ precision (includes parsimony)
1. Evaluate Best translation

**Hypothesis file** (translations to be evaluated automatically)

**Raw MT output:**
- **Best sentence** (picked by user to be correct or requiring the least amount of correction)

**Refined MT output:**
- Use **METEOR** score at sentence level to pick best candidate from the list

→ Run all automatic metrics on the new hypothesis file using user corrections as reference translations.
2. Evaluate Translation Candidate List

- Precision: \( \frac{tp}{tp + fp} \) binary \( \{0,1\} \) (1 = user correction)

\[
\begin{array}{cccc}
SL & SL & SL & SL \\
TL X & TL ✓ & TL ✓ & TL ✓ \\
TL X & TL X & TL X & TL X \\
TL X & TL X & TL X & TL ✓ \\
TL X & TL ✓ & TL ✓ & TL ✓ \\
\end{array}
\]

\[
0/3 < 1/5 = 1/5 < 1/3 < 1/2
\]
Expected Contributions

• An efficient online GUI to display translations and alignments and solicit pinpoint fixes from non-expert bilingual users.

• An expandable set of rule refinement operations
  – triggered by user corrections,
  – to automatically refine and expand different types of grammars.

• A mechanism to automatically evaluate rule refinements with user corrections as reference translations.
# Thesis Timeline

<table>
<thead>
<tr>
<th>Research components</th>
<th>Duration (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-end implementation</td>
<td>8</td>
</tr>
<tr>
<td>User Studies</td>
<td>3</td>
</tr>
<tr>
<td>Resource-poor language (data + manual grammar)</td>
<td>2</td>
</tr>
<tr>
<td>Adapt system to new language pair</td>
<td>1</td>
</tr>
<tr>
<td>Evaluation</td>
<td>1</td>
</tr>
<tr>
<td>Write and defend thesis</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18</strong></td>
</tr>
</tbody>
</table>

Expected graduation date: May 2006
Thanks!

Questions?
Some Questions

- What if users corrections are different (user noise)?
- More than one correction per sentence?
- $W_c$ example
- Data set
- Where is appropriate to refine vs bifurcate?
- Lexical bifurcate
- Is the refinement process deterministic?
Others

- TCTool Demo Simulation
- RR operation patterns
- Automatic Evaluation feasibility study
- AMTA paper results
- User studies map
- Precision, recall, F1
- NIST, BLEU, METEOR
Precision, Recall and F1

• Precision: \[ \frac{tp}{tp + fp} \] (selected, incorrect)

• Recall: \[ \frac{tp}{tp + fn} \] (correct, not selected)

• F1: \[ \frac{2 \cdot PR}{P + R} \]
Automatic Evaluation Metrics

- **BLEU**: averages the precision for **unigram**, **bigram** and up to **4-grams** and applies a **length penalty** [Papineni, 2001].

- **NIST**: instead of n-gram precision the **information gain from each n-gram** is taken into account [NIST 2002].

- **METEOR**: assigns most of the weight to **recall**, instead of precision and uses **stemming** [Lavie, 2004].
Proposed Work

Data Set

• Split development set (~400 sentence) into:
  – Dev set → Run User Studies
    → Develop Refinement Module
    → Validate functionality
  – Test set → Evaluate effect of Refinement operations

• + Wild test set (from naturally occurring text)

Requirement: need to be fully parsed by grammar
Refine vs Bifurcate

• Batch mode \(\rightarrow\) bifurcate (no way to tell if the original rule should never apply)

• Interactive mode \(\rightarrow\) refine (change original rule) if can get enough evidence that original rule never applies.

• Corrections involving agreement constraints seem to hold for all cases \(\rightarrow\) refine (Open research question)
More than one correction/sentence

<table>
<thead>
<tr>
<th>TL</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Tetris approach to Automatic Rule Refinement**

Assumption: different corrections to different words → different error
Exception: structural divergences

He danced her out of the room

$\rightarrow * \text{ La } \text{ bailó } \text{ fuera de la } \text{ habitación}$$

$\rightarrow \text{ her he-danced out of the room}$

$\rightarrow \text{ La } \text{ sacó } \text{ de la } \text{ habitación bailando}$$

$\rightarrow \text{ her he-take-out of the room dancing}$

Have no way of knowing that these corrections are related

$\rightarrow \text{ Do one error at a time, if TQ decreases over the test set, hypothesize that it’s a divergence}$

$\rightarrow \text{ Feed to the Rule Learner as a new (manually corrected) training example}$
Constituent order change

I gave him the tools

→ *di  a_él  las herramientas
  
  I-gave  to  him  the  tools

→ le  di  las herramientas  a_él
  
  him  I-gave  the  tools  to  him

desired refinement:

VP→VP  PP(a PRON)  NP  ⇔  VP→VP  NP  PP(a PRON)

→ Can extract constituent information from MT output (parse tree) and treat as one error/correction
More than one correction/error

A+B

TL  X

Example: edit and move same word (like: *gran*, *bailó*)

**Occam’s razor**

Assumption: both corrections are part of the same error
Wc Example

I am proud of you

→ *Estoy orgullosa de tu  \[W_c=de \text{ tu}\rightarrow ti\]

  I-am proud of you-nom

→ Estoy orgullosa de ti

  I-am proud of you-oblic

Without \(W_c\) information, would need to increase the ambiguity of the grammar significantly!

+ [you \(\rightarrow\) ti]: I love you \(\rightarrow\) *ti quiero (te quiero, …)

  you read \(\rightarrow\) *ti lees (tu lees, …)
Lexical bifurcate

• Should the system copy all the features to the new entry?
  – Good starting point
  – Might want to copy just a subset of features (possibly POS-dependent)

→ Open Research Question
Please start evaluating our translations

Our Machine Translation System translated the English sentence in the yellow box:

1. Gaudi was a great artist

as:

☐ gaudi era una artista grande
☑ gaudi era un artista grande
☐ gaudi estaba una artista grande
☐ gaudi estaba un artista grande

- If there is more than one correct translation, **click on all the boxes next to the correct ones** + NEXT SENTENCE button.

- If none are correct, please click on the **best incorrect translation** (the one that is closer to a good translation of the English sentence) + FIX TRANSLATION.

  Note that this might take a couple of seconds, so please wait.
SL + best TL picked by user
Changing word order
Changing “grande” into “gran”

The original (incorrect) word was \textit{grande}.
Please modify it in the following box:

\textbf{gran}

and tell us \textbf{what you believe are the sources of the error}.
Please \textbf{mark} all the error types that apply.
Note: To select multiple selections in the drop down box, hold CTRL and then click on the selection you wish to add, and click again to deselect.

- The word(s) “great” can be translated as “grande”, but not in this sentence.
The keyword(s) in the sentence that indicate this are: 

- The word “grande” is in the wrong form, but it is related to “gran”, and “gran” has to agree with.

- The word “grande” is simply an incorrect translation of “great”, and it has no relation to it.

- Wrong number agreement (singular/plural), this word has to agree with.

- Wrong gender agreement (fem/masc), this word has to agree with.

- The word “great” did not get translated.

Ex: they are well → \textit{ellos son bien; son = están}.
Ex: the kids are playing → \textit{ellos están jugando; el = los}.
Ex: the girls playing → \textit{el/la están jugando; el = la}.
Please make sure this is correct:

The corrected translation for **Gaudi was a great artist** is now **gaudi era un gran artista**.

The alignments show that:
* "Gaudi" translates as "gaudi"
* "was" translates as "era"
* "a" translates as "un"
* "great" translates as "gran"
* "artist" translates as "artista"
Interactive and Automatic Rule Refinement

Automatic Rule Adaptation

1. I → veo
   see → los
   them → veo

2. I
   see → los
   them → veo

3. I
   see
   them
### Input to RR module

- **User correction log file**

- **Transfer engine output (+ parse tree):**

<table>
<thead>
<tr>
<th>sl: I see them</th>
<th>tl: VEO LOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>tree: &lt;((S,0 (VP,3 (VP,1 (V,1:2 &quot;VEO&quot;) ) (NP,0 (PRON,2:3 &quot;LOS&quot;) ) ) ) ) &gt;</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sl: I see cars</th>
<th>tl: VEO AUTOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>tree: &lt;((S,0 (VP,3 (VP,1 (V,1:2 &quot;VEO&quot;) ) (NP,2 (N,1:3 “AUTOS”) ) ) ) ) &gt;</td>
<td></td>
</tr>
</tbody>
</table>
Types of RR Operations

• Grammar:
  – R0 \rightarrow R0 + R1 \[=R0' + \text{constr}\] \quad \text{Cov}[R0] \leq \text{Cov}[R0,R1]
  – R0 \rightarrow R1[=R0 + \text{constr}= -]
    \quad \rightarrow R2[=R0' + \text{constr}=c +] \quad \text{Cov}[R0] \leq \text{Cov}[R1,R2]
  – R0 \rightarrow R1[=R0 + \text{constr}] \quad \text{Cov}[R0] > \text{Cov}[R1]

• Lexicon
  – Lex0 \rightarrow Lex0 + Lex1[=Lex0 + \text{constr}]
  – Lex0 \rightarrow Lex1[=Lex0 + \text{constr}]
  – Lex0 \rightarrow Lex0 + Lex1[\approx\text{Lex0} + \not= \text{TLword}]
  – \emptyset \rightarrow Lex1 \quad \text{(adding lexical item)}
Manual vs Learned Grammars

[AMTA 2004]

• Manual inspection:

<table>
<thead>
<tr>
<th></th>
<th>number of sentences (over 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>same translation</td>
<td>17 (3 correct)</td>
</tr>
<tr>
<td>manual grammar better</td>
<td>10</td>
</tr>
<tr>
<td>learned grammar better</td>
<td>2</td>
</tr>
<tr>
<td>different error type</td>
<td>3 (1 correct)</td>
</tr>
</tbody>
</table>

• Automatic MT Evaluation:

<table>
<thead>
<tr>
<th></th>
<th>NIST</th>
<th>BLEU</th>
<th>METEOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual grammar</td>
<td>4.3</td>
<td>0.16</td>
<td>0.6</td>
</tr>
<tr>
<td>Learned grammar</td>
<td>3.7</td>
<td>0.14</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Human Oracle experiment

- As a feasibility experiment, compared raw output with manually corrected MT:

<table>
<thead>
<tr>
<th></th>
<th>NIST</th>
<th>BLEU</th>
<th>BLEU2n4</th>
<th>METEOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw MT output</td>
<td>4.2859</td>
<td>0.1560</td>
<td>0.2614</td>
<td>0.5968</td>
</tr>
<tr>
<td>manually corrected</td>
<td>5.6429</td>
<td>0.4998</td>
<td>0.5558</td>
<td>0.8014</td>
</tr>
</tbody>
</table>

statistically significant (confidence interval test)

- These is an upper-bound on how much difference we should expect any refinement approach to make.
Order deterministic?

• RR op application is not deterministic \(\rightarrow\) Order of the corrected sentences input to the system

• Example:
  
  1\(^{st}\): gran artista \(\rightarrow\) bifurcate (2 rules)
  
  2\(^{nd}\): casa bonito \(\rightarrow\) add agr constraint to only 1 rule (original, general rule)

  \(\rightarrow\) the specific rule is still incorrect (missing agr constraint)

  1\(^{st}\): casa bonito \(\rightarrow\) add agr constraint
  
  2\(^{nd}\): gran artista \(\rightarrow\) bifurcate

  \(\rightarrow\) both rules have agreement constraint (optimal order)
User noise?

Solution:
Have several users evaluate and correct the same test set
\[ \text{threshold, 90\% agreement} \]
- correction
- error information (type, clue word)

Only modify the grammar if enough evidence of incorrect rule.
Proposed Work

User Studies Map

- Only Corrections
- Corrections+error info
- Active learning
- Interactive mode
- RR module
- Batch mode
- Eng2spa
- Learned grammars
- Manual grammars
- Mapu2Spa

Interactive and Automatic Rule Refinement
Recycle corrections of Machine Translation output back into the system by refining and expanding existing translation rules
1. Correction info only

Rule Refinement Operations

- **Modify**
  - \(+W_c\)  \(-W_c\)
  - \(\delta \neq \emptyset\)  \(\delta = \emptyset\)

- **Add**
  - \(+W_c\)  \(-W_c\)
  - \(+al\)  \(-al\)

- **Delete**
  - \(-al\)

- **Change W Order**
  - \(W_i W_c\)
  - \(W_i(\ldots) W_c\)
  - \(-W_c\)

- **Rule Learner**

Examples:

- *It is a nice house – Es una casa bonito*
  - \(\rightarrow\) *Es una casa bonita*

- *John and Mary fell – Juan y María cayeron*
  - \(\rightarrow\) *Juan y María se cayeron*

- *Gaudi was a great artist – Gaudi era un artista grande*
  - \(\rightarrow\) *Gaudi era un gran ar*
1. Correction info only

Rule Refinement Operations

- **Modify**
  - $+W_c$
  - $-W_c$
  - $\delta \neq \emptyset$
  - $\delta = \emptyset$
  - $\Rightarrow$ Es una casa bonito
  - $\Rightarrow$ Es una casa bonita

- **Add**
  - $+W_c$
  - $-W_c$
  - $+al$
  - $-al$
  - $\Rightarrow$ J y M cayeron
  - $\Rightarrow$ J y M se cayeron

- **Delete**
  - $-al$
  - $\Rightarrow$ Gaudí era un artista grande

- **Change W Order**
  - $W_i W_c$
  - $W_i(\ldots) W_c$
  - $\Rightarrow$ I will help him fix the car
  - $\Rightarrow$ Ayudaré a él a arreglar el auto
  - $\Rightarrow$ Le ayudare a arreglar el auto

- **Modify**
  - $\Rightarrow$ Gaudi was a great artist
1. Correction info only

Rule Refinement Operations

Modify
- $+W_c$
- $-W_c$
- $\delta \neq \emptyset$
- $\delta = \emptyset$
- $+\text{rule}$
- $-\text{rule}$
- $+\text{al}$
- $-\text{al}$

Add
- $+W_c$
- $-W_c$
- $+\text{al}$
- $-\text{al}$

Delete
- $-\text{al}$
- $W_i$ $W_c$

Change W Order
- $W_i(\ldots) W_c$
- $-W_c$
- RuleLearner

Example:

"I would like to go – Me gustaria que ir"

$\Rightarrow$ Me

gustaria $\emptyset$ ir

"I will help him fix the car – Ayudare a él a arreglar el auto"

$\Rightarrow$ Le ayudare a arreglar el auto
2. Correction and Error info

Rule Refinement Operations

- **Modify**
  - $+W_c$ – $-W_c$
  - $\delta \neq \emptyset$ – $\delta = \emptyset$
  - $\delta \neq \emptyset$
  - $\delta = \emptyset$
  - $\delta = \emptyset$
  - $+\text{rule}$ – $-\text{rule}$
  - $+\text{al}$ – $-\text{al}$
  - $=W_i$
  - $W_i \rightarrow W_i'$
  - $\text{RuleLearner}$

- **Add**
  - $+W_c$ – $-W_c$
  - $+W_c$ – $-W_c$
  - $+\text{al}$ – $-\text{al}$
  - $=W_i$
  - $W_i \rightarrow W_i'$
  - $\text{RuleLearner}$

- **Delete**
  - $-\text{al}$
  - $W_i \rightarrow W_i'$

- **Change W Order**
  - $W_i(\ldots) W_c$
  - $-W_c$
  - $W_i(\ldots) W_c$
  - $-W_c$

- **Correction**
  - I am proud of you – Estoy orgullosa $tu$
  - $\rightarrow$ Estoy orgullosa de $ti$

- **Error info**
Focus 3
Rule Refinement Operations

Modify
+\( W_c \) -\( W_c \)
\( \delta \neq \emptyset \) \( \delta = \emptyset \)

Add
+\( W_c \) -\( W_c \)

Delete
+al -al

Change W Order
\( W_i \) \( W_c \) \( W_i(\ldots) \) \( W_c \) -\( W_c \)

Rule Learner

Wally plays the guitar – Wally juega la guitarra
\( \rightarrow \) Wally toca la guitarra

I saw the woman – Vi \( \emptyset \) la mujer
\( \rightarrow \) Vi a la mujer

I see them – Veo los
\( \rightarrow \) Los veo
Outside Scope of Thesis

Rule Refinement Operations

- Modify
  - +Wc
  - –Wc

- Add
  - +Wc
  - –Wc

- Delete
  - +al
  - –al

- Change W Order
  - W_i W_c
  - W_i(…) W_c

δ ≠ ∅  δ = ∅ +rule –rule +al –al

Æ

Wi RuleLearner

POS_i = POS_i’  POS_i ≠ POS_i’

John read the book – A Juan leyó el libro
→ ∅ Juan leyó el libro

Where are you from? – Donde eres tu de?
→ De donde eres