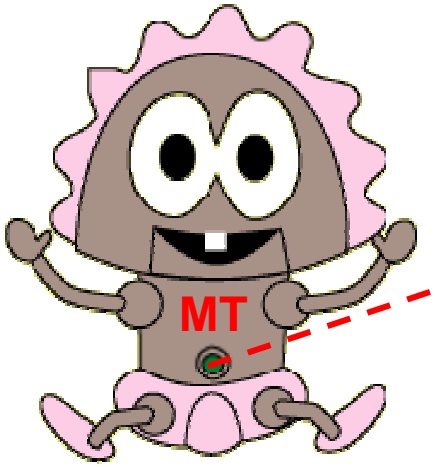
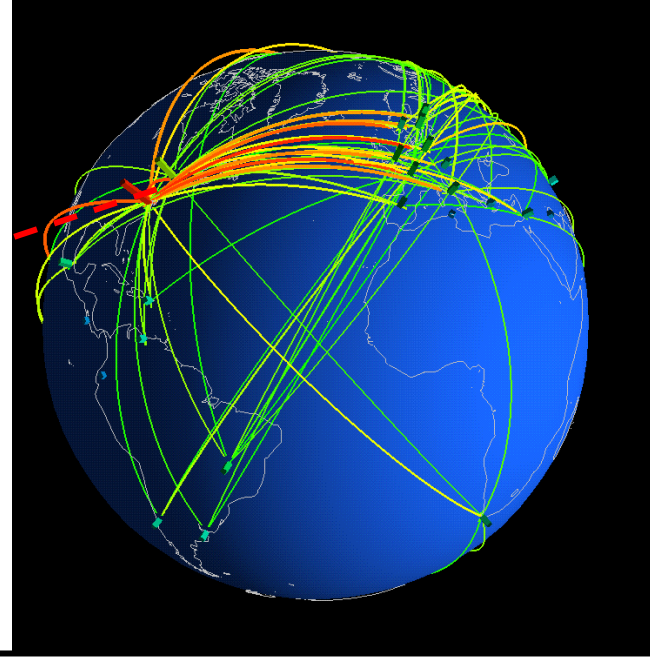




Can the Internet help Improve Machine Translation?

Ariadna Font Llitjós

June 4, 2006

Problem and Motivation

MT output: `Assassinated a diplomat Russian and kidnapped other four in Bagdad`

Problem and Motivation

MT output: `Assassinated a diplomat Russian and kidnapped other four in Bagdad`

- Could hire experts to correct machine translations

Problem and Motivation

MT output: Assassinated a diplomat Russian and kidnapped other four in Bagdad

- Could hire experts to correct machine translations

- Expensive + time consuming

Problem and Motivation

MT output: Assassinated a diplomat Russian and kidnapped other four in Bagdad

- Not feasible for large amounts of data (google, yahoo, etc.)
- Does not generalize to new sentences

My Solution

SL: Asesinado un diplomático ruso y secuestrados otros cuatro en Bagdad

TL: Assassinated a diplomat Russian and kidnapped other four in Bagdad

- Get *non-expert* bilingual speakers to provide correction feedback online
  - Make correcting translations easy and fun
  - 5-10 minutes a day → **Large amounts of correction data**

My Solution

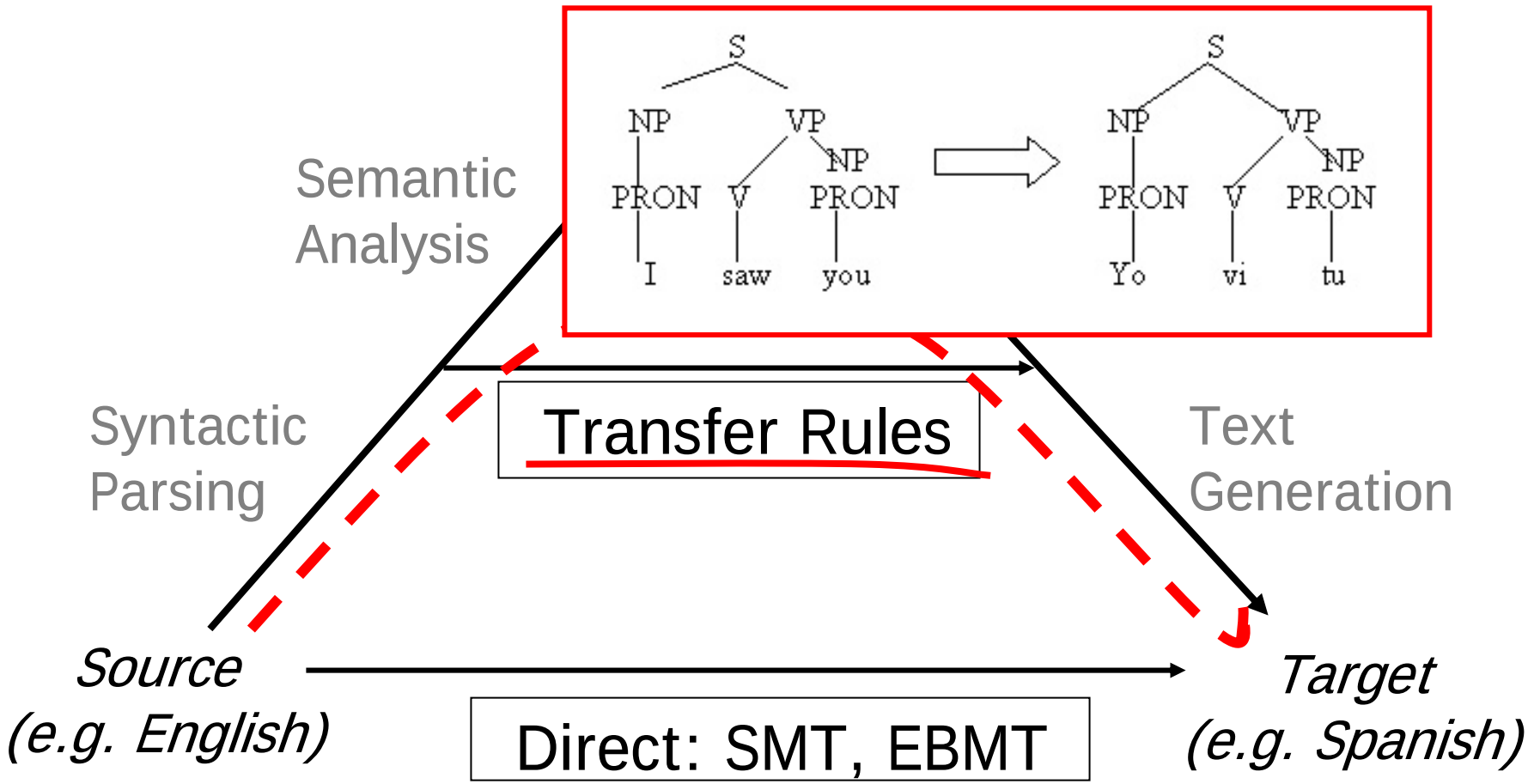
SL: Asesinado un diplomático ruso y secuestrados otros cuatro en Bagdad

TL: Assassinated a diplomat Russian and kidnapped other four in Bagdad

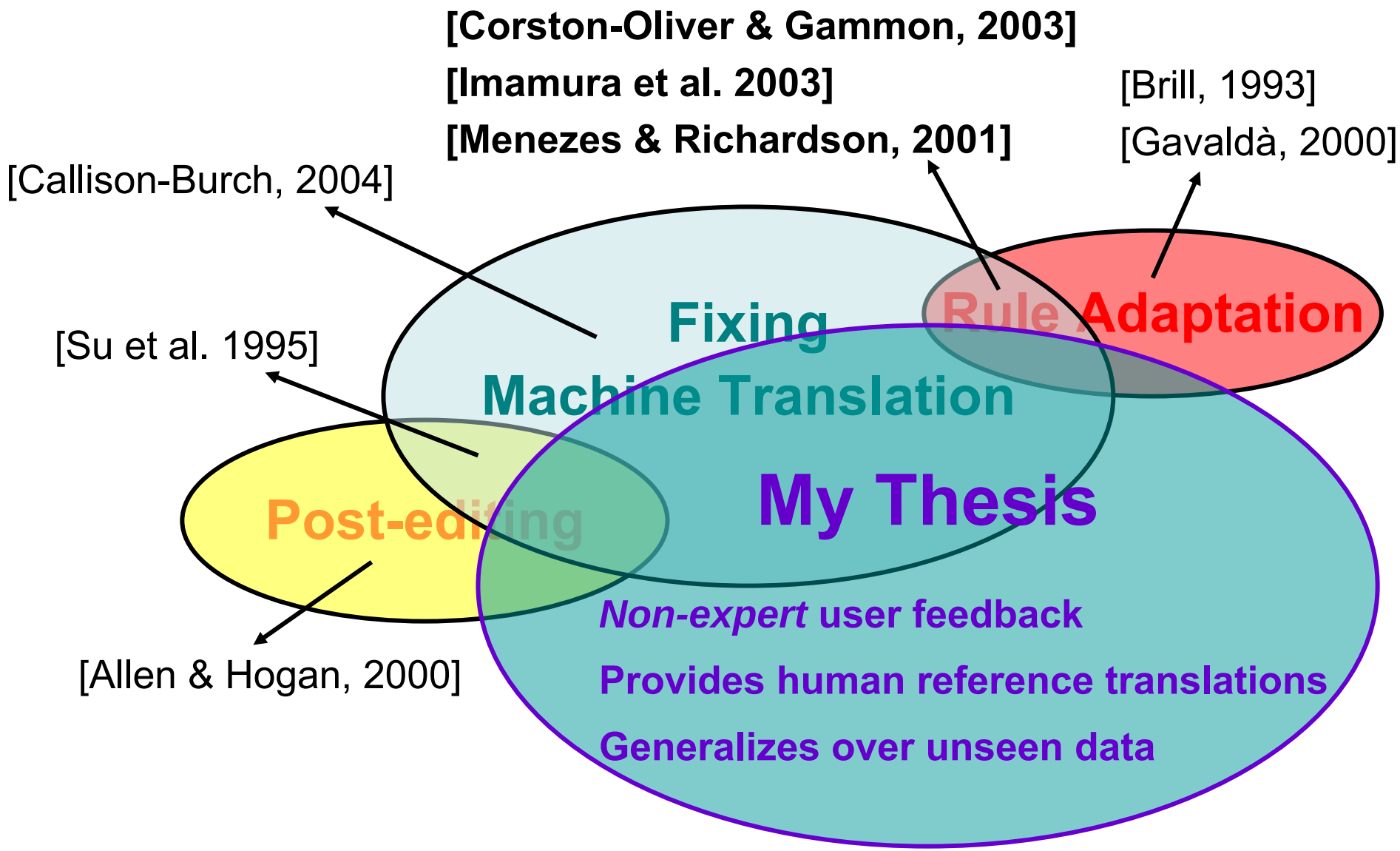
- Get *non-expert* bilingual speakers to provide correction feedback online
  - Make correcting translations easy and fun
  - 5-10 minutes a day → **Large amounts of correction data**
- Feed corrections back into the MT system, so that they can be generalized

→ **System will translate new sentences better**

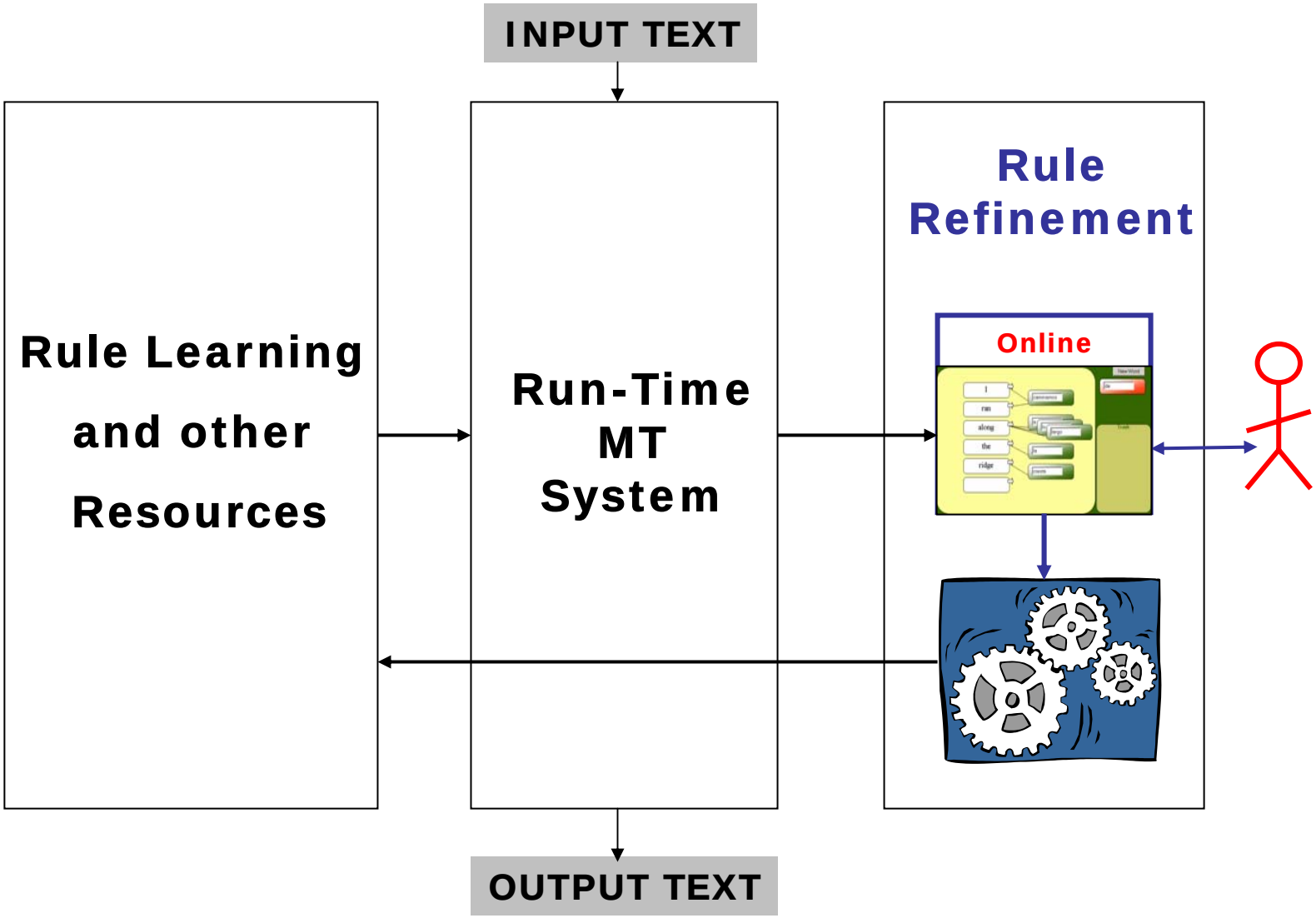
MT Approaches



Related Work



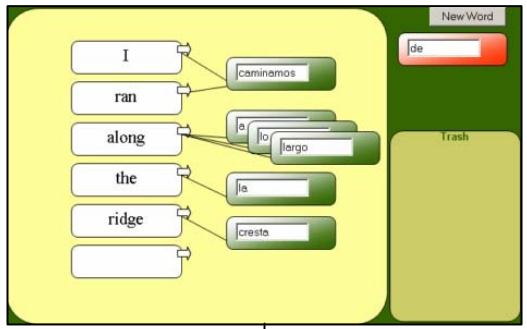
System Architecture

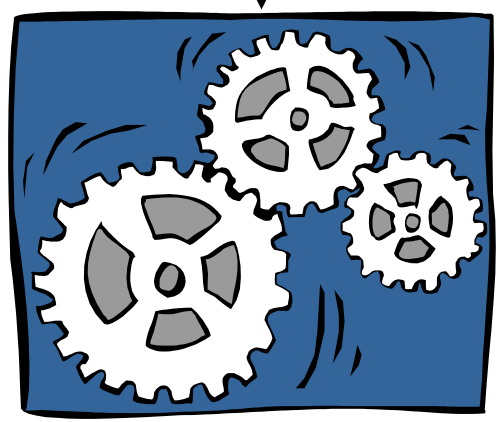


Main Technical Challenge

Simple user edits to MT output

Mapping between





Blame assignment

Rule Modifications

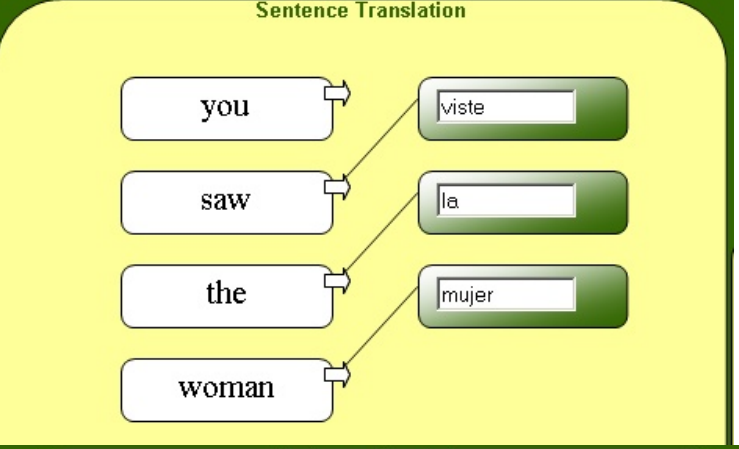
Lexical Expansions

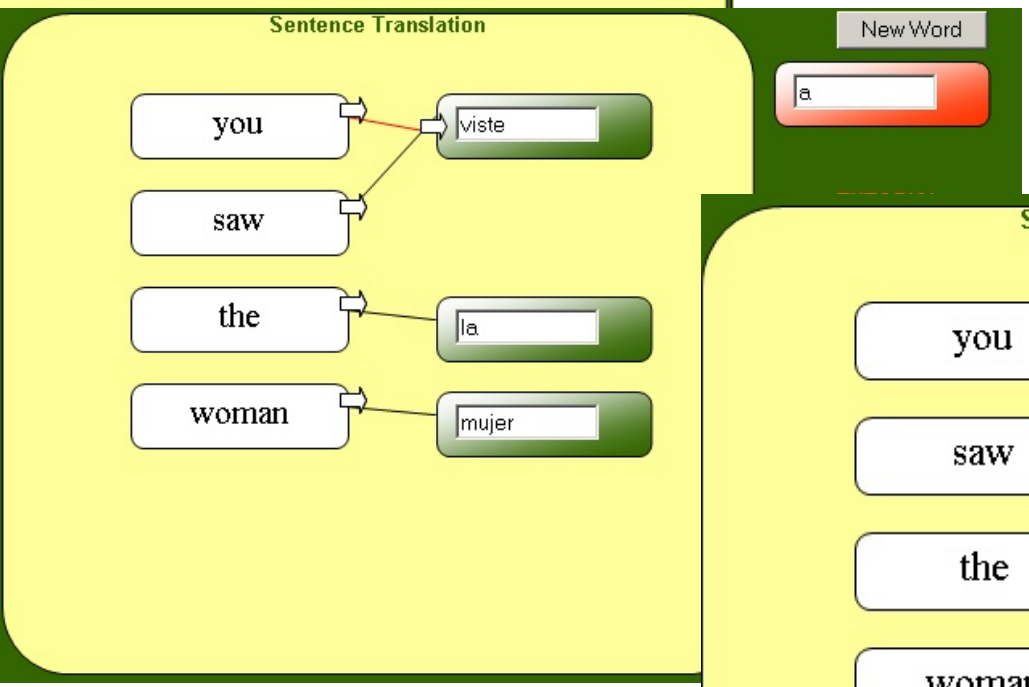
Improved Translation Rules

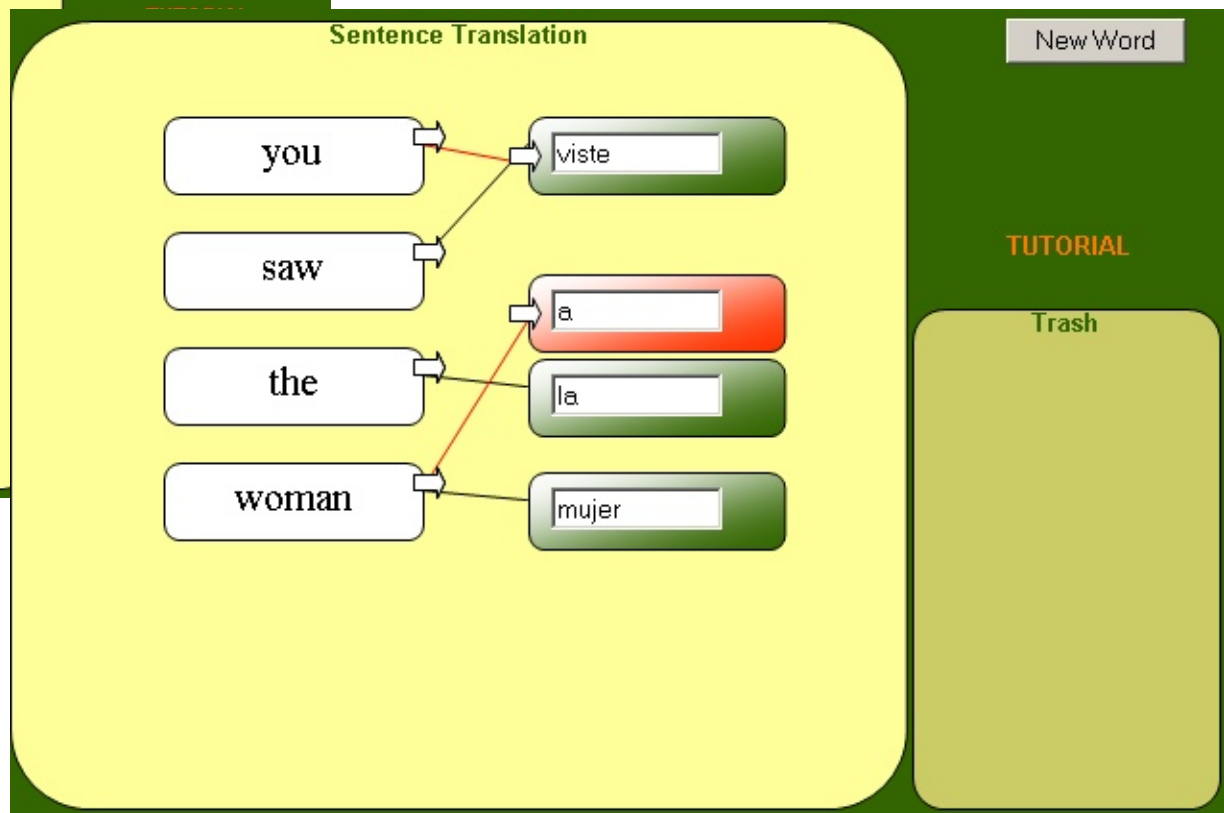
TCTool (Demo)

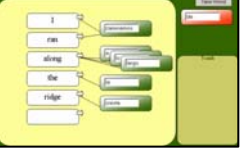
Actions:

- Add a word
- Delete a word
- Modify a word
- Change word order









Eng2Spa User Study

[LREC 2004]

	precision	recall
error detection	90%	89%
error classification	72%	71%

- MT error classification → 9 linguistically-motivated classes [Flanagan, 1994], [White et al. 1994]:

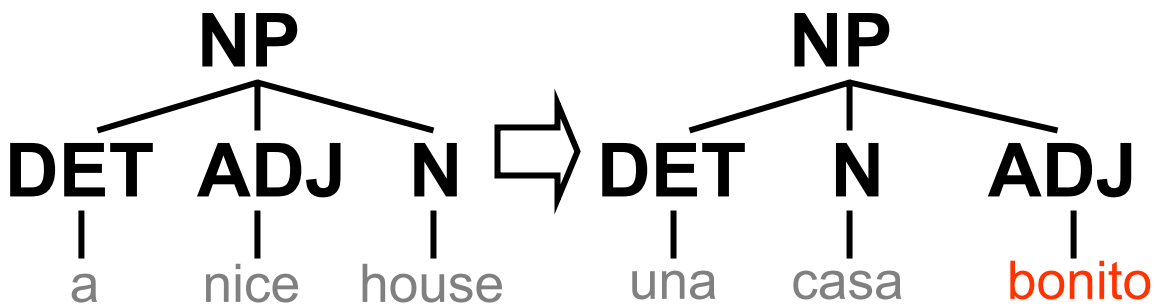
  word order, sense, agreement error (number, person, gender, tense), form, incorrect word and **no translation**

Types of Refinement Operations

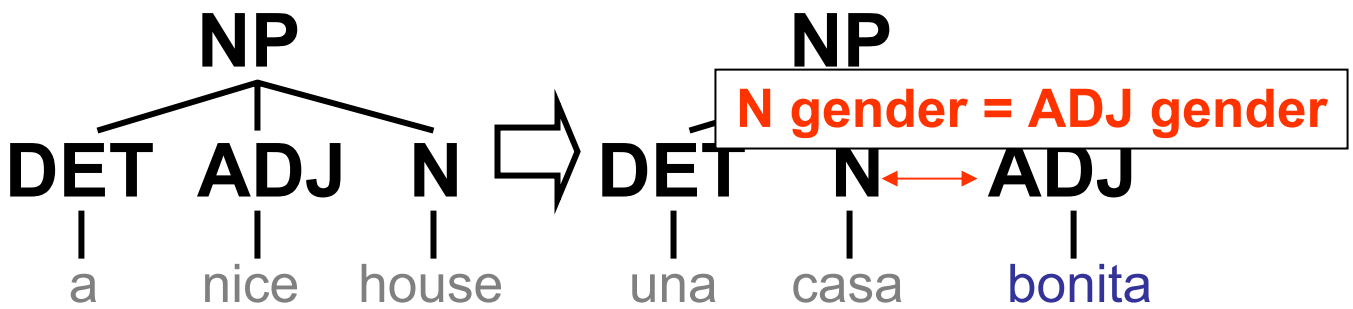
1. Refine a translation rule:

R0 → R1 (change R0 to make it more specific or more general)

R0:



R1:



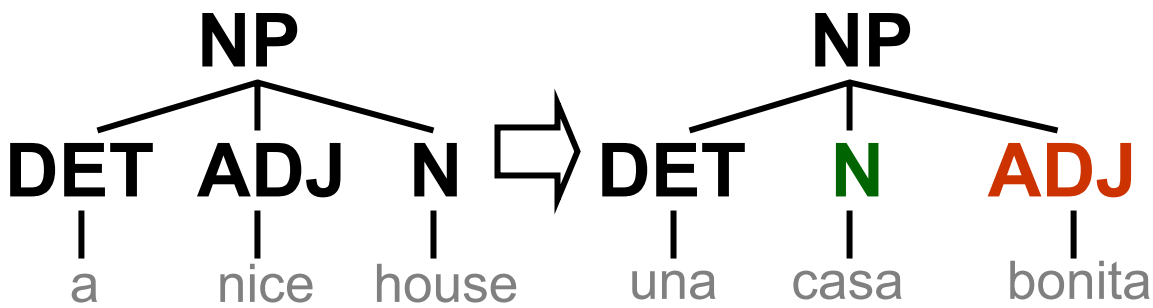
Types of RefinementOperations (2)

2. Bifurcate a translation rule:

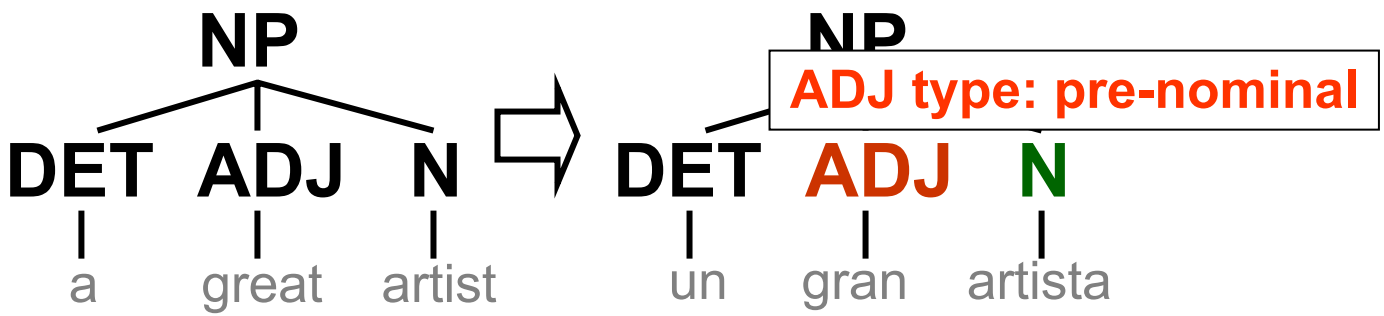
R0 → R0 (same, general rule)

→ R1 (add a new more specific rule)

R0:



R1:



Formalizing Error Information

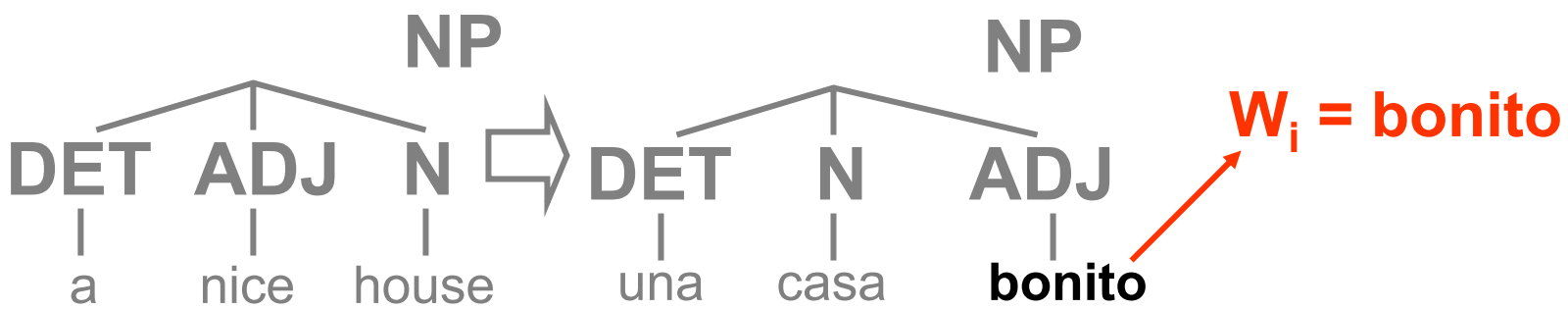
W_i = error

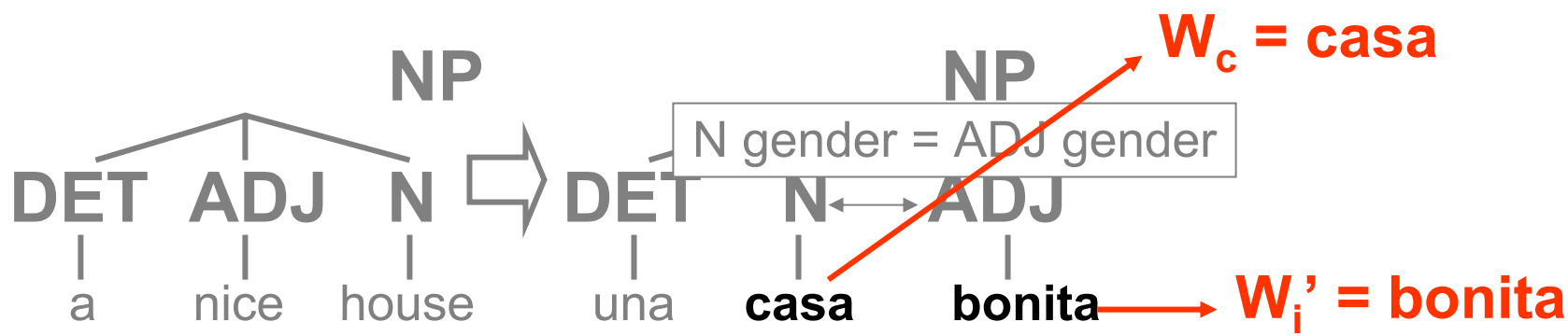
W_i' = correction

W_c = clue word

$$\overrightarrow{TL_m} = (W_1, \ldots W_i \ldots, W_n)$$

$$\overrightarrow{TL'_m} = (W_1, \ldots, W_i', \ldots, W_c, \ldots, W'_n)$$





Triggering Feature Detection

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

→ Delta function: $\delta(W_i, W_i')$

Examples:

gen = masc | gen = fem

δ(bonito,bonita) = {gender}

δ(comiamos,comia) = {person,number}

δ(mujer,guitarra) = {∅}

If δ set is empty, need to

postulate a new binary feature (*feat_i* = {+,−})

Refinement Steps

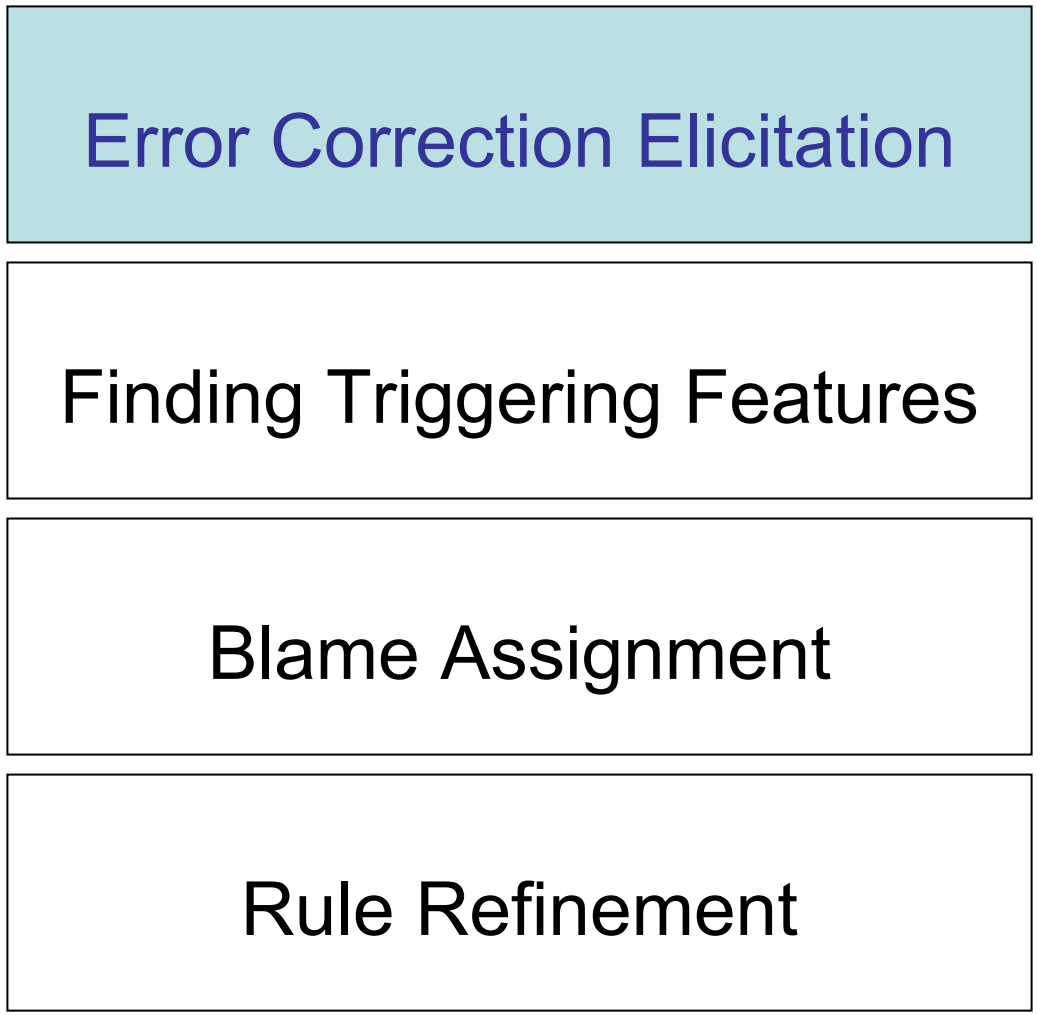
Error Correction Elicitation

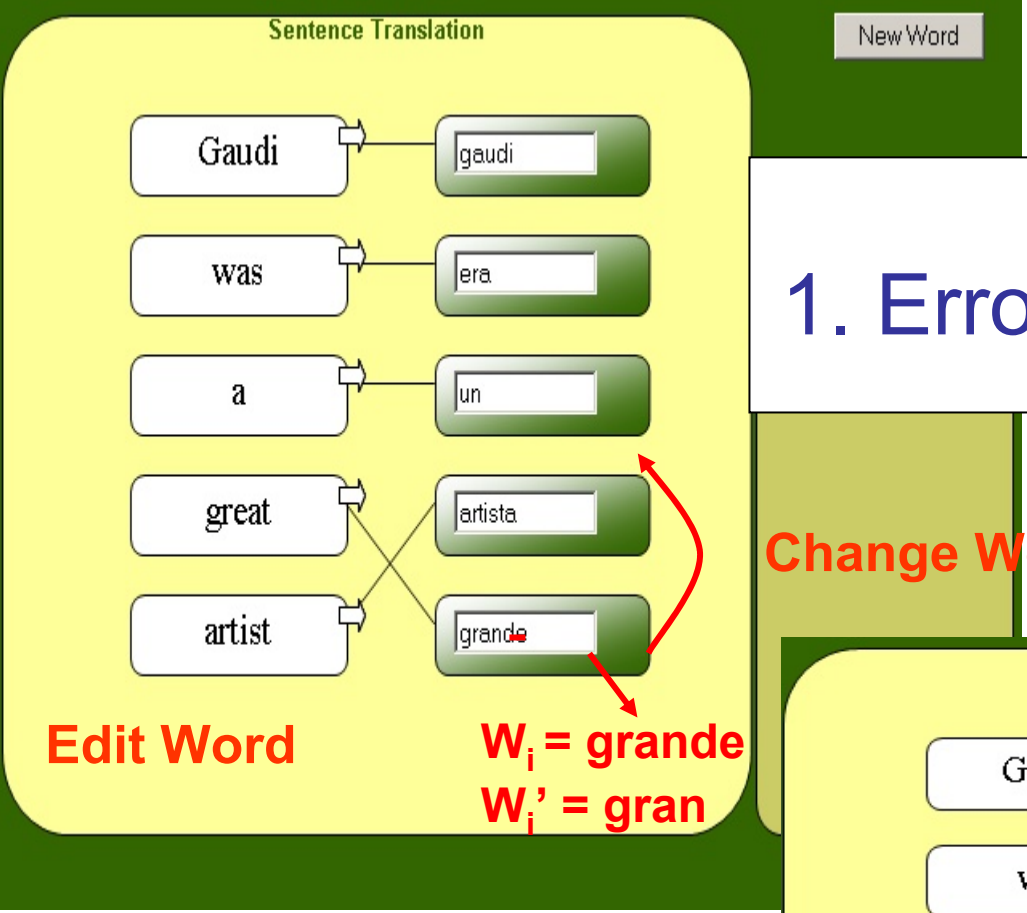
Finding Triggering Features

Blame Assignment

Rule Refinement

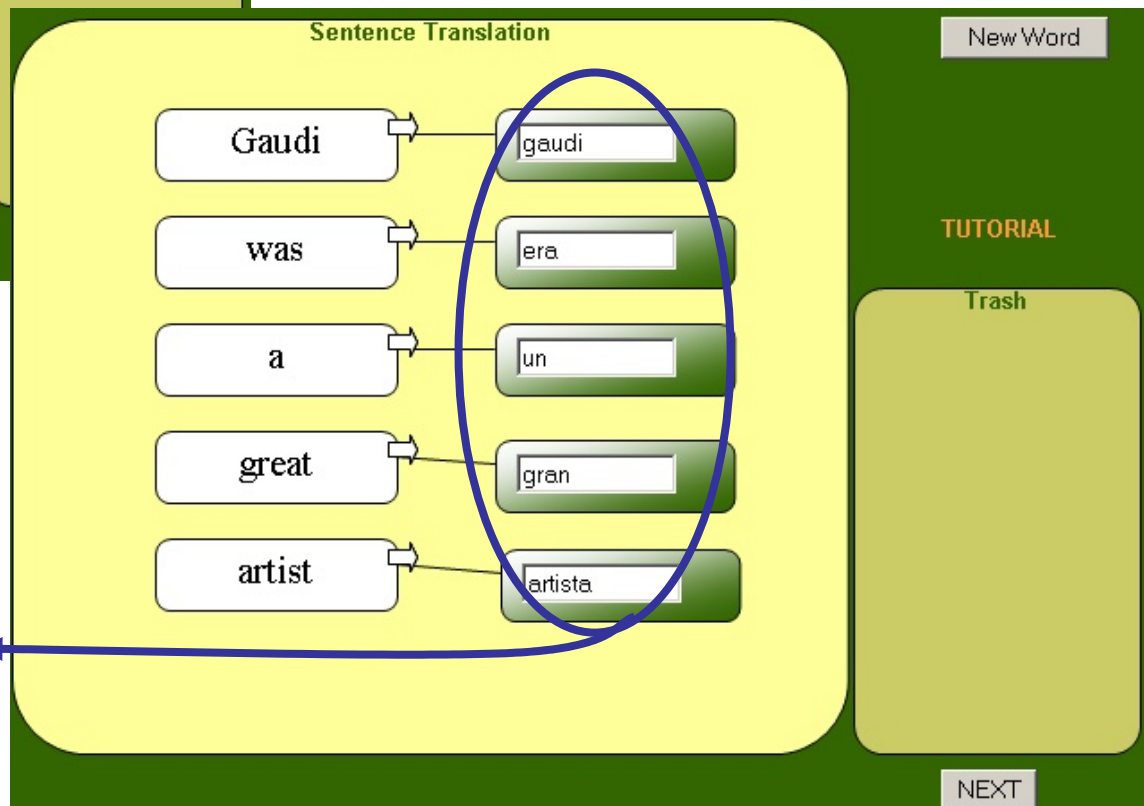
Refinement Steps





1. Error Correction Elicitation

Change Word Order



SL: ***Gaudí was a great artist***

TL: ***Gaudí era un artista grande***

CTL: ***Gaudí era un gran artista***

Refinement Steps

Error Correction Elicitation

Finding Triggering Feature

Blame Assignment

Rule Refinement

1. **Edit:**
   W_i = grande W_i' = gran
2. **Change Word Order:**
   artista gran → gran artista

Refinement Steps

Error Correction Elicitation

Finding Triggering Features

Blame Assignment

Rule Refinement

1. **Edit:**
   W_i = grande W_i' = gran
2. **Change Word Order:**
   artista gran → gran artista

2. Finding Triggering Features

Delta function — difference at the feature level?

ADJ	ADJ
[great] → [grande]	[great] → [gran]
agr num = sg	agr num = sg
agr gen = masc	agr gen = masc

δ(grande, gran) = $\varnothing$

2. Finding Triggering Features

Delta function difference at the feature level?

ADJ	ADJ
[great] → [grande]	[great] → [gran]
agr num = sg	agr num = sg
agr gen = masc	agr gen = masc

δ(grande, gran) = $\varnothing$

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

2. Finding Triggering Features

Delta function difference at the feature level?

ADJ	ADJ
[great] → [grande]	[great] → [gran]
agr num = sg	agr num = sg
agr gen = masc	agr gen = masc

δ(grande, gran) = $\varnothing$

→ need to postulate a new binary feature: feat1

[feat1 = +] = [type = pre-nominal]

[feat1 = −] = [type = post-nominal]

2. Finding Triggering Features

Delta function — difference at the feature level?

ADJ	ADJ
[great] → [grande]	[great] → [gran]
agr num = sg	agr num = sg
agr gen = masc	agr gen = masc

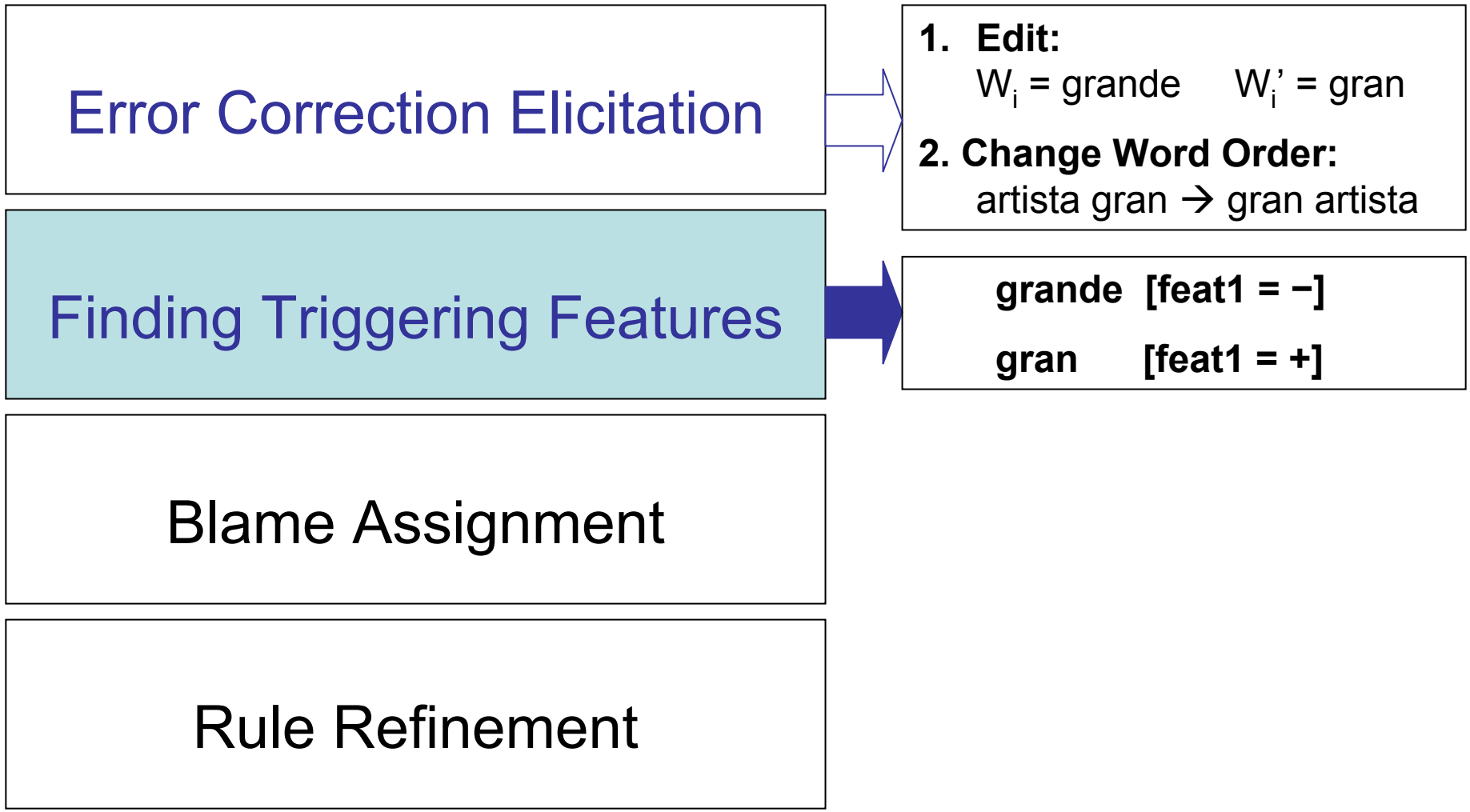
δ(grande, gran) = $\varnothing$

new binary feature: feat1

REFINE

ADJ	ADJ
[great] → [grande]	[great] → [gran]
agr num = sg	agr num = sg
agr gen = masc	agr gen = masc
feat1 = −	**feat1 = +**

Refinement Steps



Refinement Steps

Error Correction Elicitation

1. **Edit:**
   W_i = grande W_i' = gran
2. **Change Word Order:**
   artista gran → gran artista

Finding Triggering Features

grande [feat1 = −]

gran [feat1 = +]

Blame Assignment

Rule Refinement

3. Blame Assignment

(from transfer and generation tree)

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

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

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

(N,4:5 "ARTISTA")

(ADJ,5:4 "GRANDE"))))>

Grammar

S,1

…

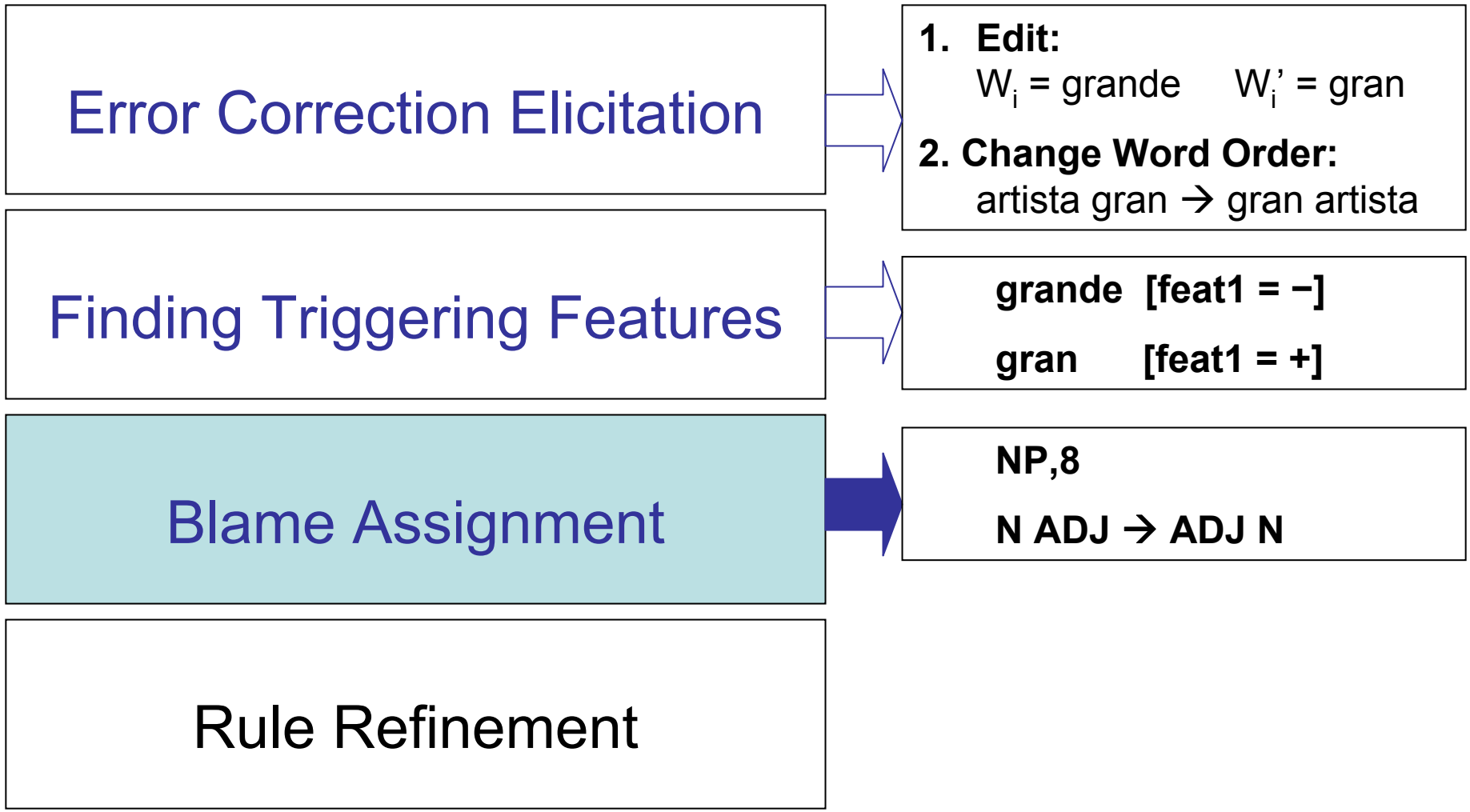
NP,1

…

NP,8

…

Refinement Steps



Refinement Steps

Error Correction Elicitation

1. **Edit:**
   W_i = grande W_i' = gran
2. **Change Word Order:**
   artista gran → gran artista

Finding Triggering Features

grande [feat1 = −]

gran [feat1 = +]

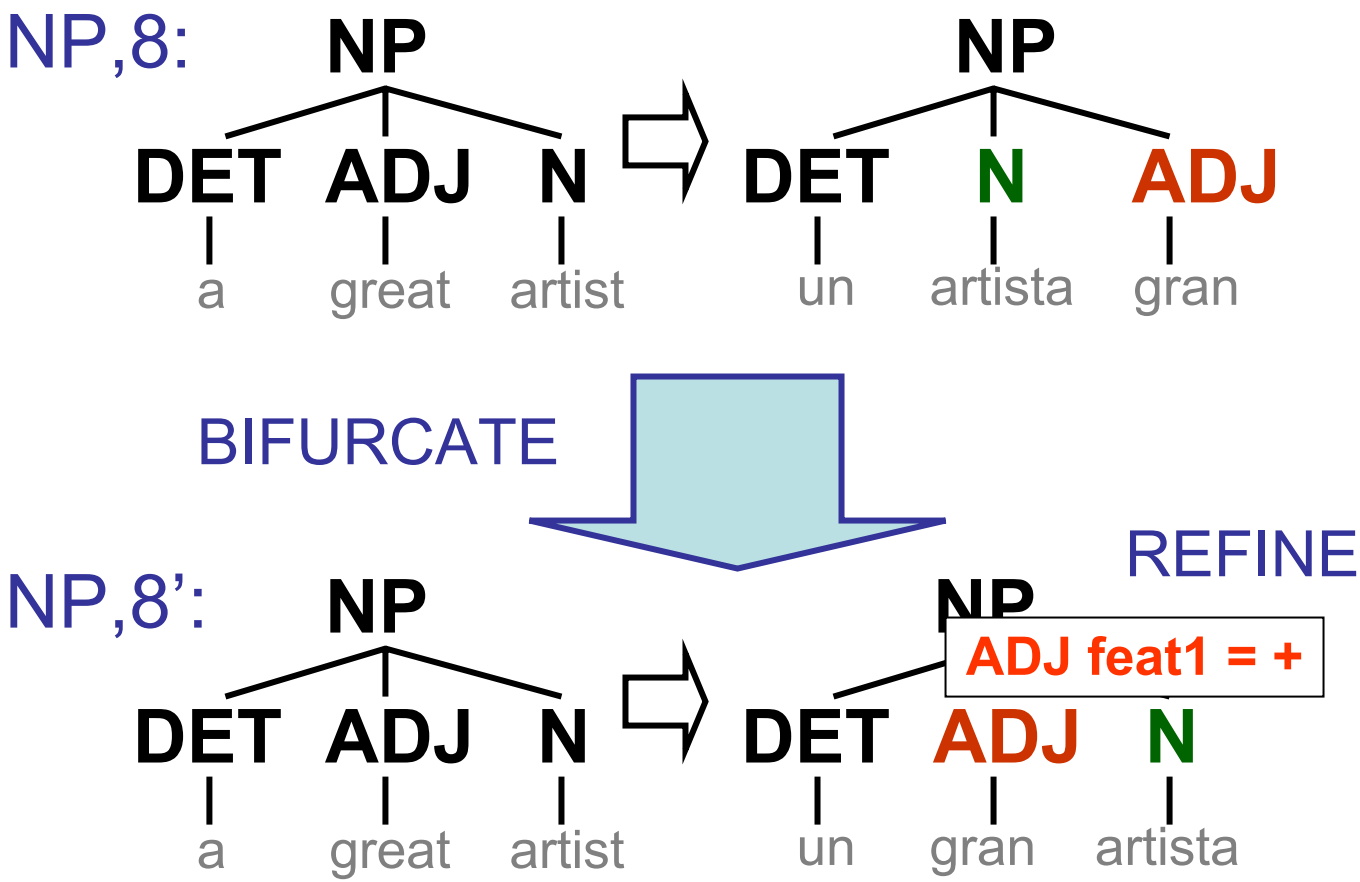
Blame Assignment

NP,8

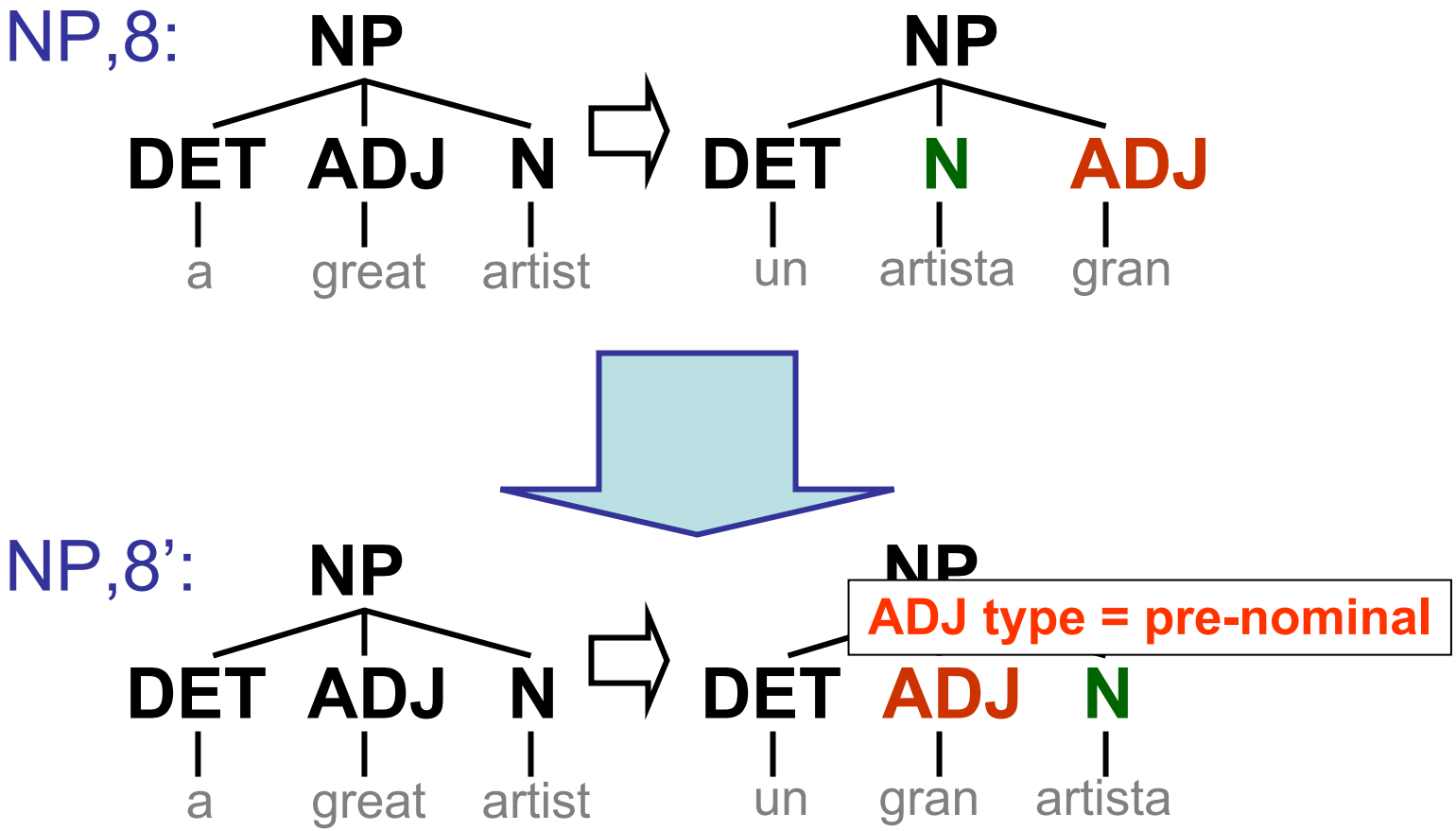
N ADJ → ADJ N

Rule Refinement

4. Rule Refinement



4. Rule Refinement



Refinement Steps

Error Correction Elicitation

1. **Edit:**
   W_i = grande W_i' = gran
2. **Change Word Order:**
   artista gran → gran artista

Finding Triggering Features

grande [feat1 = −]

gran [feat1 = +]

Blame Assignment

NP,8

(N ADJ → ADJ N)

Rule Refinement

NP,8 ADJ N → **N ADJ**

NP,8' ADJ N → **ADJ N**
[ADJ feat 1 = +]

Correct Translation Output

NP,8 ⟷ ADJ(great-grande)

[feat1 = -]

NP,8' ⟷ ADJ(great-gran)

[ADJ feat1 =c +] **[feat1 = +]**

Gaudi era un artista grande

Gaudi era un gran artista

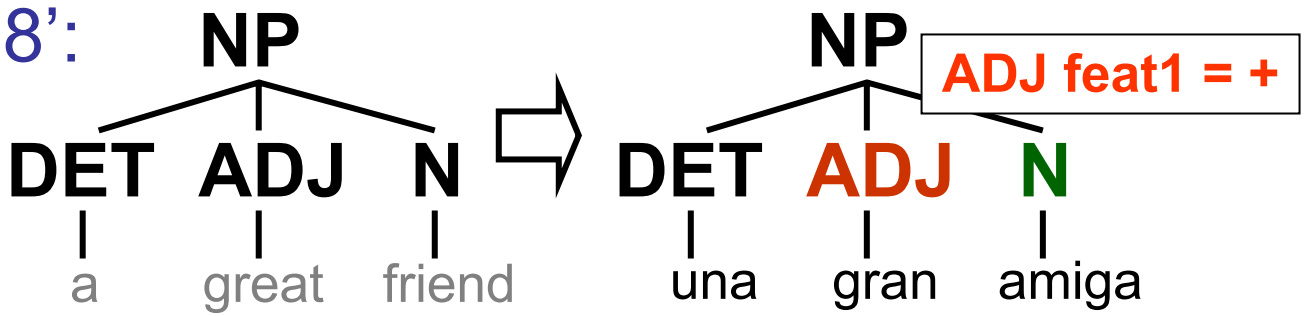
**Gaudi era un grande artista*

Generalization Power: abstract feature (*feat_i*)

Irina is a great friend → *Irina es una gran amiga*

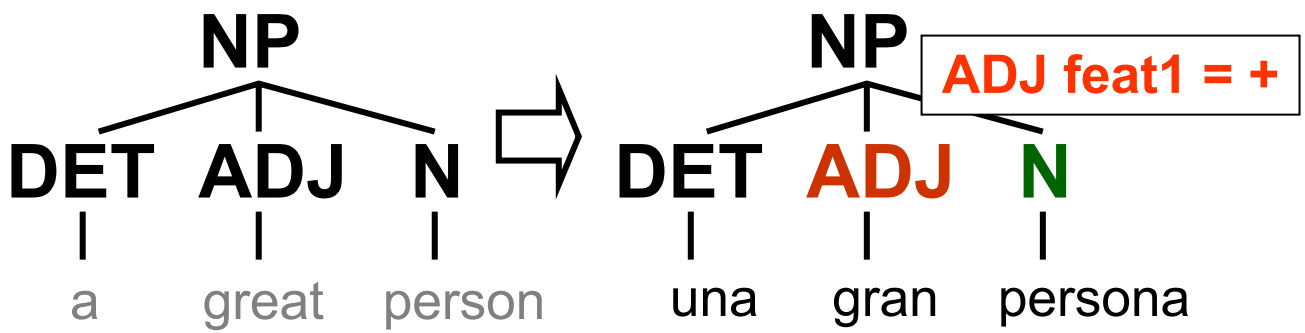
(instead of **Irina es una amiga grande*)

NP,8':



Juan is a great person → *Juan es una gran persona*

(instead of **Juan es una persona grande*)



Generalization Power++

When triggering feature already exists in the grammar/lexicon (POS, gender, number, etc.)

ADJ gender = N gender

I see the red car → **veo un auto roj**a*** / **o** → *veo un* ***auto rojo***

gender = masc **gender = fem** **gender = masc**

Refinements generalize to all lexical entries that have that feature (*gender*):

gender = fem

The yellow houses are his → *las* ***casas amarillas*** *son suyas*

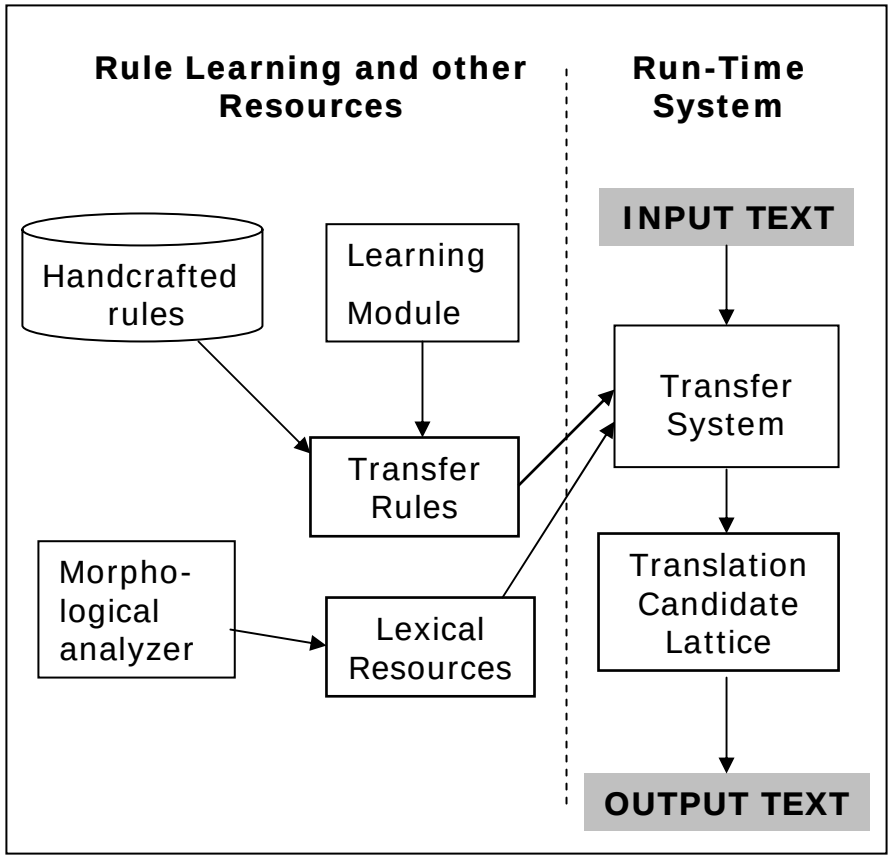
(before: **las casas amarillos son suyas*)

gender = fem

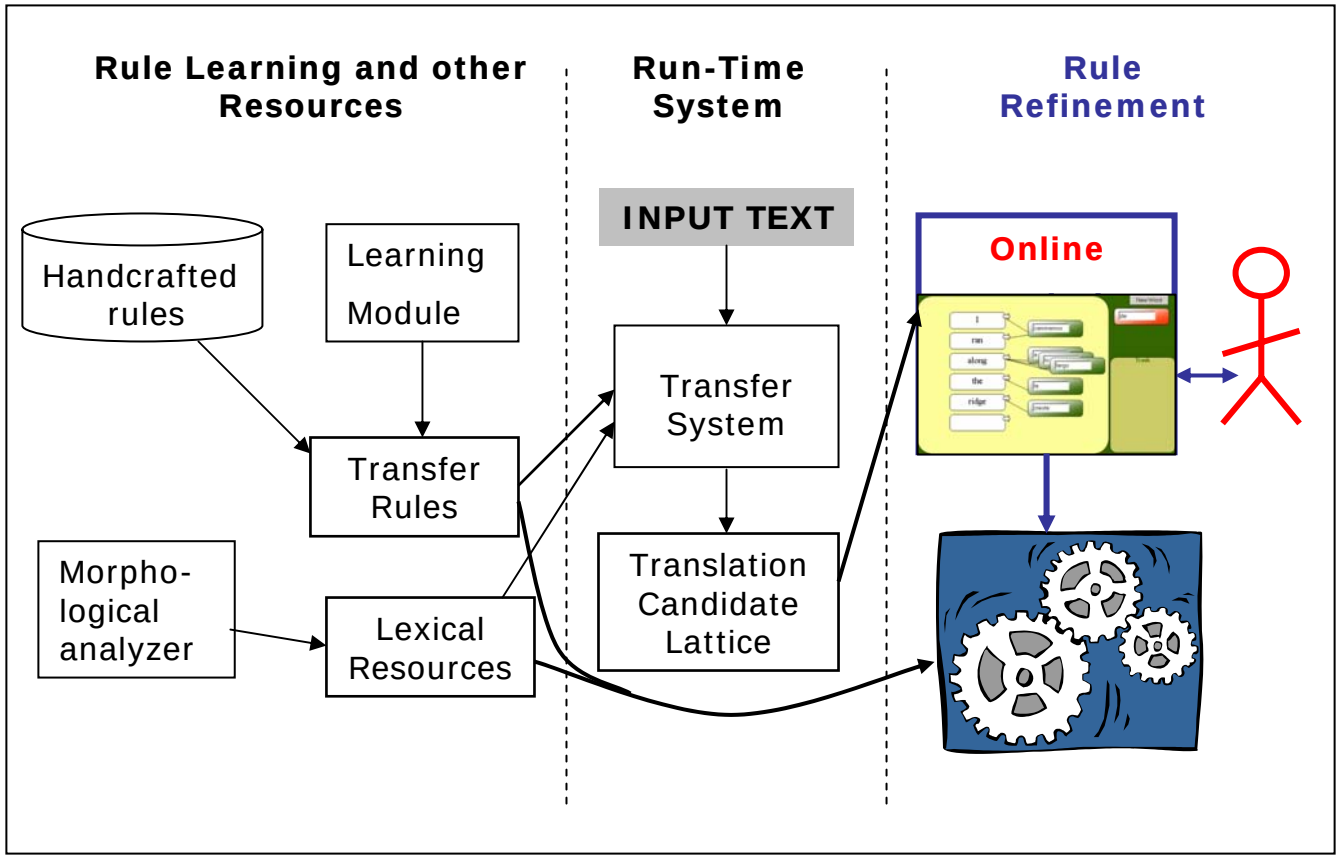
We need to go to a dark cave → *tenemos que ir a una* ***cueva oscura***

(before: **cueva oscuro*)

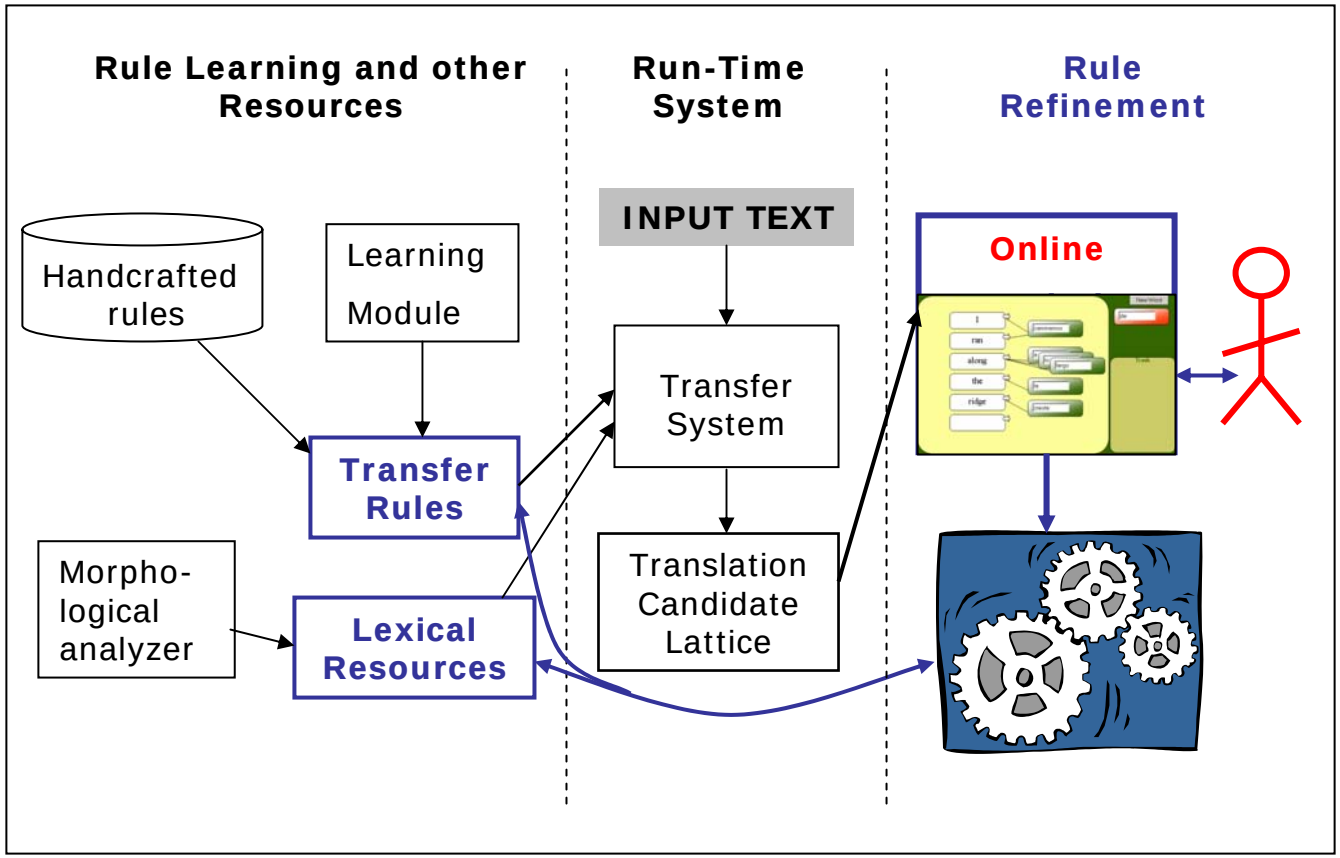
Impact on Transfer-Based MT



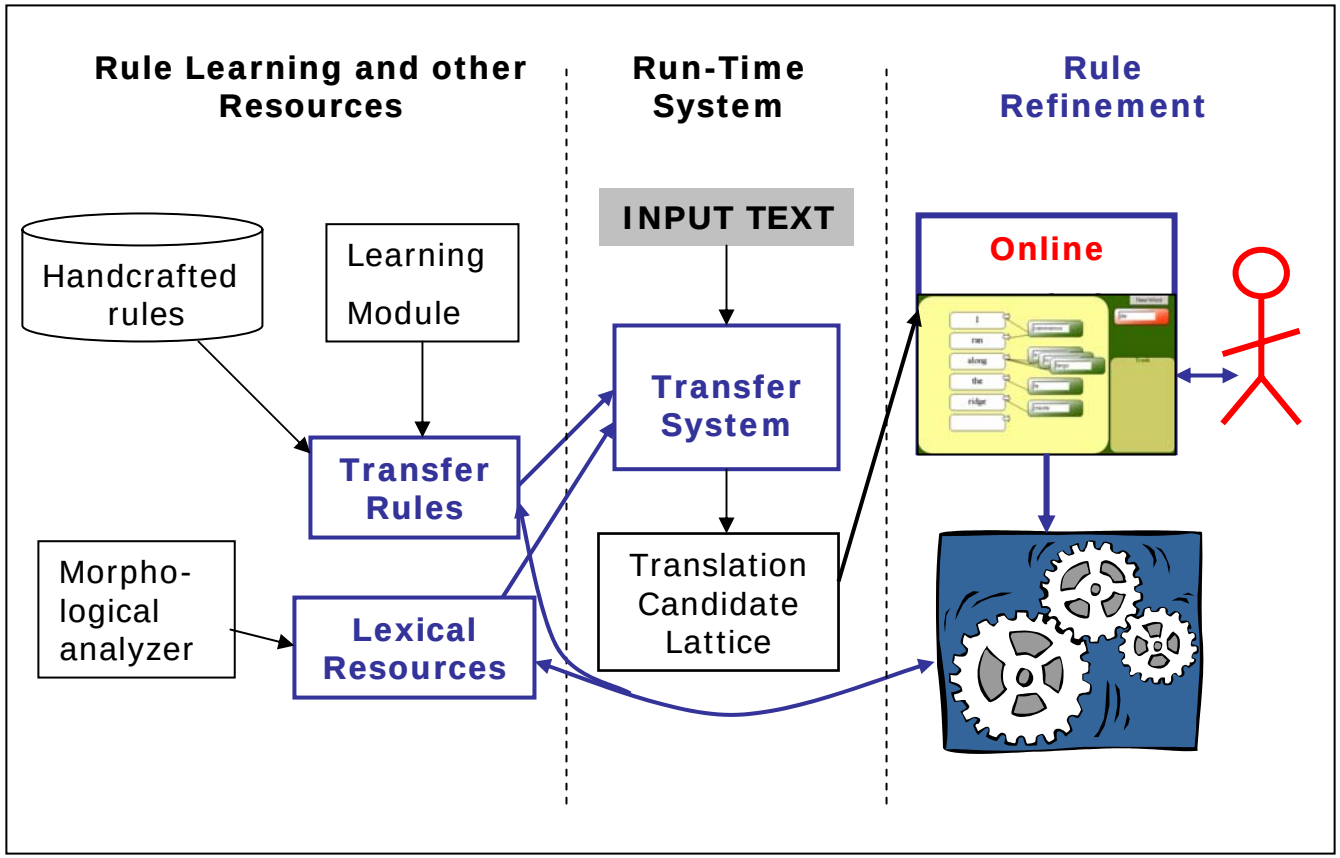
Impact on Transfer-Based MT



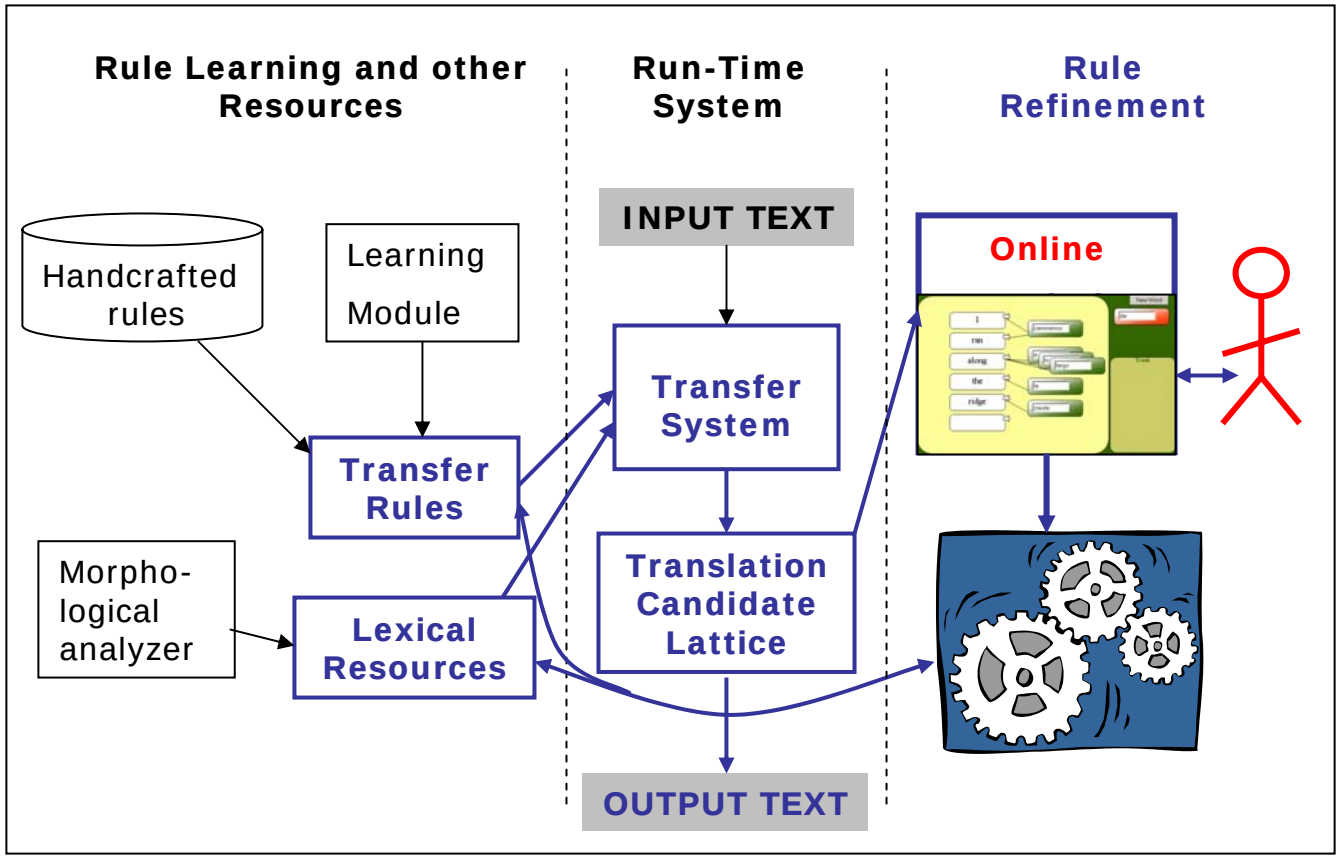
Impact on Transfer-Based MT



Impact on Transfer-Based MT



Impact on Transfer-Based MT



TCTool can help improve

- Rule-based MT (grammar, lexicon, LM)
- EBMT (examples, lexicon, alignments)
- Statistical MT (lexicon, and alignments)

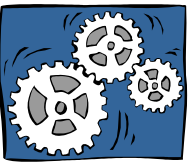
Work in Progress

- We have identified 10 different types of refinement operations, integrating last 2 into the system.
- Handling *incorrect* Correction Instances
  - Have multiple users correct the same set of sentences

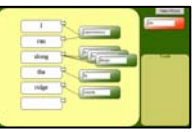
    → filter out noise (threshold: 90% users agree)
- User study with multiple users

  → evaluate improvement after refinements.

Contributions so far



New Framework to improve MT quality: an expandable set of rule refinement operations



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

Future work

- Explore other ways to make the interaction with users more fun

  - Games with a purpose

    [Von Ahn and Blum, 2004 & 2006]

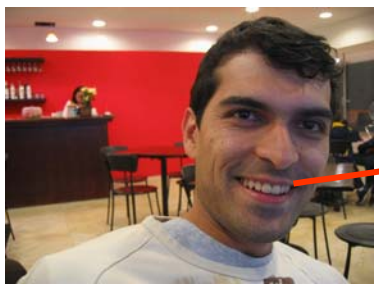
Future work

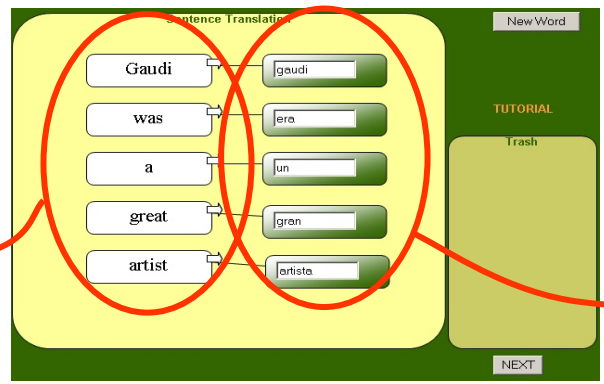
- Explore other ways to make the interaction with users more fun

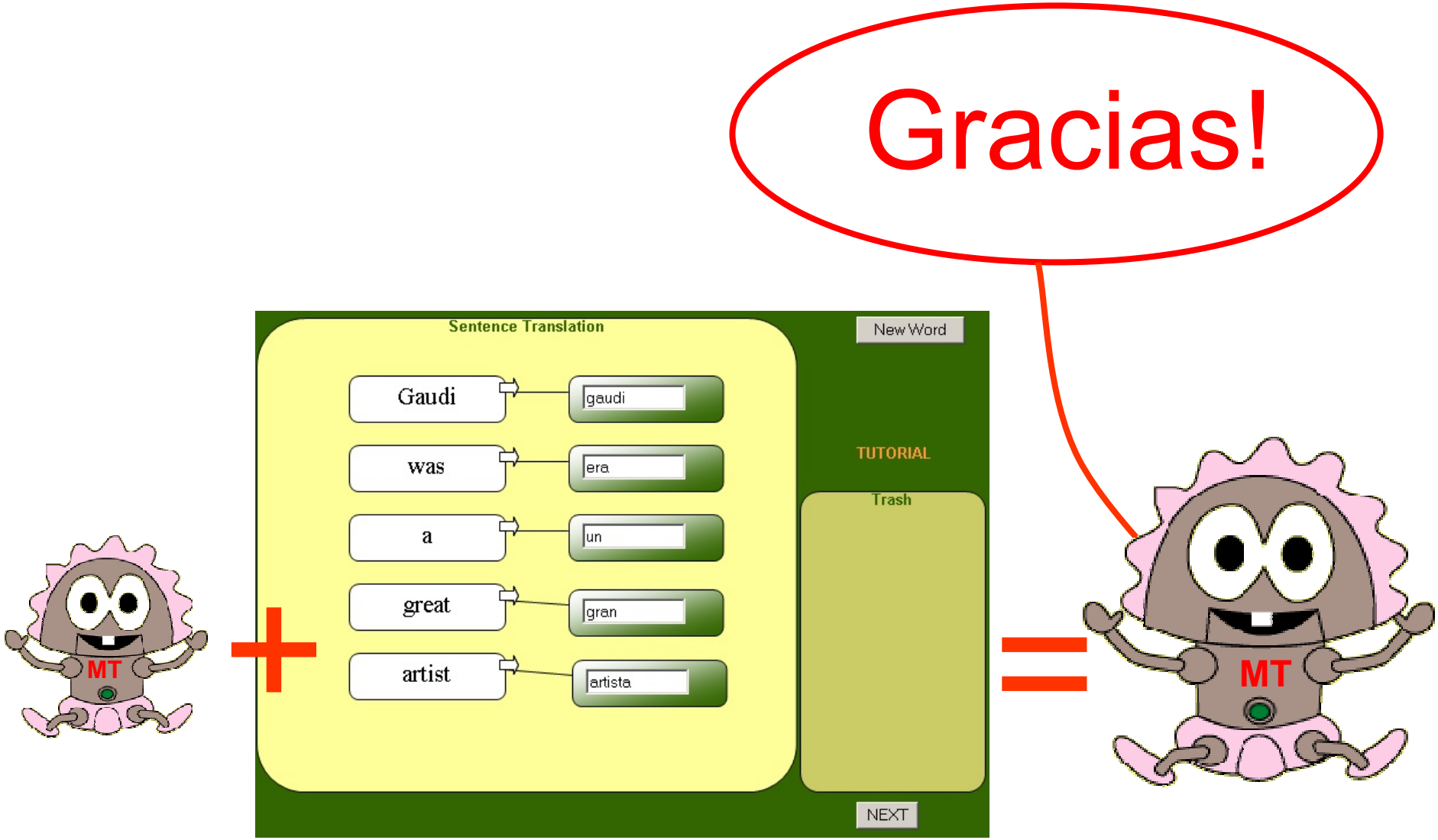
- Games with a purpose

[Von Ahn and Blum, 2004 & 2006]

- Language Learning






Gracias!
Sentence Translation
New Word
Gaudi
gaudi
was
era
a
un
great
gran
artist
artista
TUTORIAL
Trash
NEXT
MT
MT