

# Improving Machine Translation of Null Subjects in Italian and Spanish

L. Russo, S. Loáiciga, A. Gulati



# Null subjects (pro-drop)

## ▶ Italian

- <sub>pro</sub> Penso (1.sg) che <sub>pro</sub> dobbiamo (1.pl) fare molto se <sub>pro</sub> vogliamo (1.pl) migliorare la situazione.
- *I think that **we** have to do a lot if **we** want to improve the situation.*

## ▶ Spanish

- <sub>pro</sub> Le pido (1.sg) que <sub>pro</sub> intervenga (2.sg) con el prestigio de su cargo.
- *I ask **you** to intervene with the prestige of your position.*

# Outline

- ▶ Corpus
- ▶ 1<sup>st</sup> evaluation
  - Systems (Its-2 & Moses)
  - Results
- ▶ 2<sup>nd</sup> evaluation
  - Experiment (Hybrid Its-2)
    - Preprocessing
    - Statistical post-editing
  - Results
- ▶ Work in progress

# Corpus

- ▶ Europarl release v3
- ▶ 1,000 sentences
  - aligned

	# Verbs	Personal	Impersonal	Total
Italian	3,422	18.41%	12.01%	30.42%
Spanish	3,182	23.33%	17.84%	41.17%

# Systems

## ▶ Its-2

- Rule-based

## ▶ Moses

- Statistical
- Trained on Europarl
  - 55,000 sentences
  - 3-gram language model
  - 2,000 sentences for tuning

# 1<sup>st</sup> evaluation

gender, number and person

IT → FR	Correct	Incorrect	Missing
Personal pro-drop	66.34%   71.59%	3.49%   1.11%	30.15%   27.30%
Impersonal pro-drop	16.78%   44.76%	18.97%   11.43%	64.23%   43.79%

  

ES → FR	Correct	Incorrect	Missing
Personal pro-drop	55.79%   72.64%	3.50%   2.02%	40.70%   25.34%
Impersonal pro-drop	29.29%   54.56%	11.40%   2.45%	59.29%   42.98%

Its-2 | Moses

# Experiment

<sub>pro</sub> E' la prima volta che <sub>pro</sub> intervengo in Plenaria e <sub>pro</sub> devo ammettere di essere molto emozionato [...].

E' la prima volta che **io** intervengo in Plenaria e **io** devo ammettere di essere molto emozionato [...].

Preprocessor

Its-2

Statistical  
post-editor

Est la première fois que **j'**interviens dans \*Plenaria et **j'**admettre d'est beaucoup émus [...].

**C'**est la première fois que **j'**interviens en plénière ont et **j'**admettre d'est très émus [...].

*It is the first time that I speak in the Plenary session and I admit to being [...].*

# Statistical post-editor

- ▶ Trained on Its-2 translations of Europarl
  - Italian → French
    - 976 sentences
  - Spanish → French
    - 1,005 sentences
  - 3-gram language model
    - 55,000 sentences

# Experiment

Italian	Source
Original	<sup>pro</sup> E' la prima volta che <sup>pro</sup> intervengo in Plenaria e <sup>pro</sup> devo ammettere di essere molto emozionato [...].
After preprocessing	E' la prima volta che <b>io</b> intervengo in Plenaria e <b>io</b> devo ammettere di essere molto emozionato [...].
Its-2	Translation
Baseline	Qu'est la première fois qu'intervient dans *Plenaria et admettre de m'est beaucoup émus [...].
With preprocessing ONLY	Est la première fois que <b>j'</b> interviens dans *Plenaria et <b>j'</b> admettre d'est beaucoup émus [...]. ( <i>IMPROVEMENT</i> )
With post-editing ONLY	Qu'est la première fois qu'intervient en plénière ont et admettre de s'est très émus [...]. ( <i>NO IMPROVEMENT</i> )
With both COMBINED	<b>C'</b> est la première fois que <b>j'</b> interviens en plénière ont et <b>j'</b> admettre d'est très émus [...]. ( <i>BEST TRANSLATION</i> )
	<i>It is the first time that I speak in the Plenary session and I admit to being [...].</i>

# 2<sup>nd</sup> evaluation

IT → FR	Correct	Incorrect	Missing
Personal pro-drop	66.34%⇒79.52%	3.49%⇒5.87%	30.15%⇒14.60%
Impersonal pro-drop	16.78%⇒35.76%	18.97%⇒13.13%	64.23%⇒51.09%

  

ES → FR	Correct	Incorrect	Missing
Personal pro-drop	55.79%⇒69.13%	3.50%⇒7.81%	40.70%⇒23.04%
Impersonal pro-drop	29.29%⇒41.40%	11.40%⇒8.07%	59.29%⇒50.52%

# 2<sup>nd</sup> evaluation

IT → FR	Correct	Incorrect	Missing
Personal pro-drop	79.52%   71.59%	5.87%   1.11%	14.60%   27.30%
Impersonal pro-drop	35.76%   44.76%	13.13%   11.43%	51.09%   43.79%

  

ES → FR	Correct	Incorrect	Missing
Personal pro-drop	69.13%   72.64%	7.81%   2.02%	23.04%   25.34%
Impersonal pro-drop	41.40%   54.56%	8.07%   2.45%	50.52%   42.98%

Its-2 | Moses

# Work in progress

## ▶ Null subjects

- Parallel vs. native
- Impersonal pro-drop
- ↪ LREC 2012

## ▶ Preprocessing

- Feature preservation (parsing/translation)
  - Empty category for impersonal pronouns
  - Absence/presence for personal pronouns

## ▶ Hybrid MT

**Thank You!**

