Asiya: An Open Toolkit for Automatic Machine Translation and (Meta-)Evaluation

Meritxell Gonzàlez
February 18th, 2013
MT Development cycle

- MT system developer
- Evaluation Methods
- Error Analysis
  - Identify Type of Error
  - Analyze Possible Causes
- System Refinement
- Evaluation
MT Development cycle

- Error Analysis
  - Identify Type of Error
  - Analyze Possible Causes
  - Metric Refinement
  - Evaluation

- MT metric developer
- Meta-Evaluation Methods
MT Development cycle

Error Analysis

- Identify Type of Error
- Analyze Possible Causes
- System Refinement
- Evaluation

Evaluation Methods

MT metric developer

- Identify Type of Error
- Analyze Possible Causes
- Metric Refinement
- Evaluation

Error Analysis

Meta-Evaluation Methods

MT system developer

Evaluation

Methods
Metrics and Meta-metrics

Languages:
- English, Spanish, Catalan
- Also: Arabic, Czech, French, German, Czech, Romanian

Similarity principles
- Precision, recall, overlap

Linguistic layers:
- Lexical, Syntactic, Semantic
- Confidence estimation
The captain rejects that the coach leaves the Barça bench due problems with some of the players.

The captain descarta that the technician abandon the banquillo of the Barça by problems with some of his players.

The captain discards that the technician leaves the bench of the Barça by problems with some of his players.

The captain dismisses the technician leaves the Barca bench due to problems with some of their players.

The master ruled that the technician leaves the Barca bench by problems with some of his players.

The captain rejects that the technician leaves the bench of the Barça for problems with some of his(her,your) players.

The captain discards that the technician leaves the bench of the Barça by problems with some of his players.
The captain rejects that the coach leaves the Barça bench due problems with some of the players.

{ 'the' = 4; 'captain' = 1; 'coach' = 1; 'rejects' = 1; 'due' = 1; 'of' = 1; 'players' = 1; 'that' = 1; 'barça' = 1; 'of' = 1; 'bench' = 1; 'some' = 1; 'with' = 1; 'problems' = 1 } (15 different tokens)

The captain descarta that the technician abandon the banquillo of the Barça by problems with some of his players.

{ 'the' = 4; 'captain' = 1; 'technician' = 1; 'descarta' = 1; 'of' = 2; 'players' = 1; 'that' = 1; '.' = 1; 'his' = 1; 'banquillo' = 1; 'problems' = 1; 'abandon' = 1; 'by' = 1; 'barça' = 1; 'some' = 1; 'with' = 1 } (16 different tokens)

Ol => hits: 13 and total: 25 = 13/25 = 0.52
Example - Overlap

✦ The captain rejects that the coach leaves the Barça bench due problems with some of the players.

✦ The captain descarta that the technician abandon the banquillo of the Barça by problems with some of his players.

✦ Ol => hits: 13 and total: \( \frac{25}{13/25} = 0.52 \)
The Test Suit

- Asiya operates over test suites (or test beds).
  - a test suite is a collection of test cases:
    - Source segment
    - Candidate translation(s)
    - Reference translation(s)
The Test Suit

- Asiya.pl Asiya.config
- Asiya.config:

```plaintext
# MANDATORY DEFINITIONS
# -----------------------------------
input=raw
src=es
src=cs
trg=en
trg=cs

#SOURCE definition ================
src=./data/src.txt

#REFERENCE definition =============
ref=./data/ref.txt

#SYSTEM OUTPUT definition =========
sys=./data/google.txt
sys=./data/bing.txt
sys=./data/reverso.txt
sys=./data/systran.txt
sys=./data/apertium.txt
sys=./data/babelfish.txt
```

General Options

✧ Input format
  ✧ Raw
  ✧ Nist

✧ Language pair
  ✧ Srclang
  ✧ Trglang

✧ Predefined sets of metrics, systems and references
Eval

- Eval <schema>
  - Single
  - ULC
  - Queen [AGPV05]

- Meta-Eval
- Learn

- Output format
  - Metric matrix
  - System matrix
  - Nist
  - Granularity
    - System, document, segment
  - Pdf, tex
Meta-Eval

- Eval
- Meta-Eval <schemas> <criteria>
  - Correlation with assessments
    - Pearson
    - Spearman
    - Kendall tau
  - Orange [LO04]
  - King [AGPV05]
  - Consistency
- Learn
Meta-Eval

- Eval
- Meta-Eval <schemas> <criteria>
  -ci <method> Asiya.config
    - Fisher [Fis24]
    - Bootstrap resampling [ET86]
    - Paired bootstrap resampling [Koe04]
    - Orange [LO04]
    - Options:
      - significance level
      - Asiya.pl -v -optimize <schemas> <criteria> number of resamplings
- Learn
Learning

- Eval
- Meta-Eval
- Learn <scheme>
  - Perceptron
  - model <s>

- Asiya.pl -eval single -model <s>
Asiya Interfaces

- **Asiya Online Interface**
  - A graphical interface to access an on-line version of Asiya.

- **Asiya tSearch**
  - Online interface that allows to search for output translations (of a given testbed) that match some specific criteria related to their quality (as assessed by the automatic scores).

- **Asiya Web Service**
  - A RESTful web service to access the Asiya evaluation.
Asiya Online Interface
Asiya Web Service

Asiya Server

Web Service

Asyncronous Job Operator

Cluster

Asiya Instance

Input and Output Data

I/O Data
References


References


References

