

# Translation Quality Evaluation in the MOLTO Project

Maarit Koponen (University of Helsinki),  
Cristina España-Bonet and Lluís Màrquez  
(Universitat Politècnica de Catalunya)

2<sup>nd</sup> MOLTO Project Meeting - Open Day  
Gothenburg, March 9, 2011

The logo for the MOLTO project, featuring the word "MOLTO" in a bold, black, sans-serif font. The letter "O" is significantly larger than the other letters and is filled with a dark red, textured pattern, resembling a wood grain or a similar natural material.

# Overview

1. Introduction: MOLTO and Quality

2. Manual Evaluation

(Maarit Koponen, University of Helsinki)

1. Evaluation Methods

2. Example of a Manual Evaluation Study

3. Automatic Evaluation

(Cristina España-Bonet, Universitat Politècnica de Catalunya)

## Introduction: MOLTO and Quality

- MOLTO will provide tools for automatic translation on the web, with high quality in 15 languages.
- The tools are targeted for producers of information.

### → User requirements

- Text quality
  - high **precision**: message not distorted
  - **fluency**: publication quality
- Usability
  - **speed**: instant translation, new system can be built in months or even days
  - **ease**: no deep technological knowledge required

# Evaluation Methods

- **Manual Evaluation**

- error analysis
- scalar evaluation (e.g. fluency, adequacy, clarity), ranking
- post-editing (with source text, without source text)
- comprehension
- usability evaluation
- ✓ reliability
- ✗ cost (money, time, labor), subjectivity, non-reusability

- **Automatic Evaluation**

# Example of a Manual Evaluation Study

- **Test subjects:** students of translation. **Material:** newspaper articles, machine translated from English into Finnish by two systems (one statistical and one rule-based).
- **Tasks:**
  - Edit the machine translation as necessary without source text. Options for “nothing to correct” or “not able to correct”.
  - Rate the translation for fluency and clarity, suitability for use.
- **Assessment of the corrected sentences:**
  - Were the corrections successful?
  - How well did the subjects decipher meaning?
- **Detailed error analysis on the MT:**
  - What types of errors are easy to correct?
  - What types of errors lead to incomprehensible sentences?
  - What types of errors lead to misunderstandings?

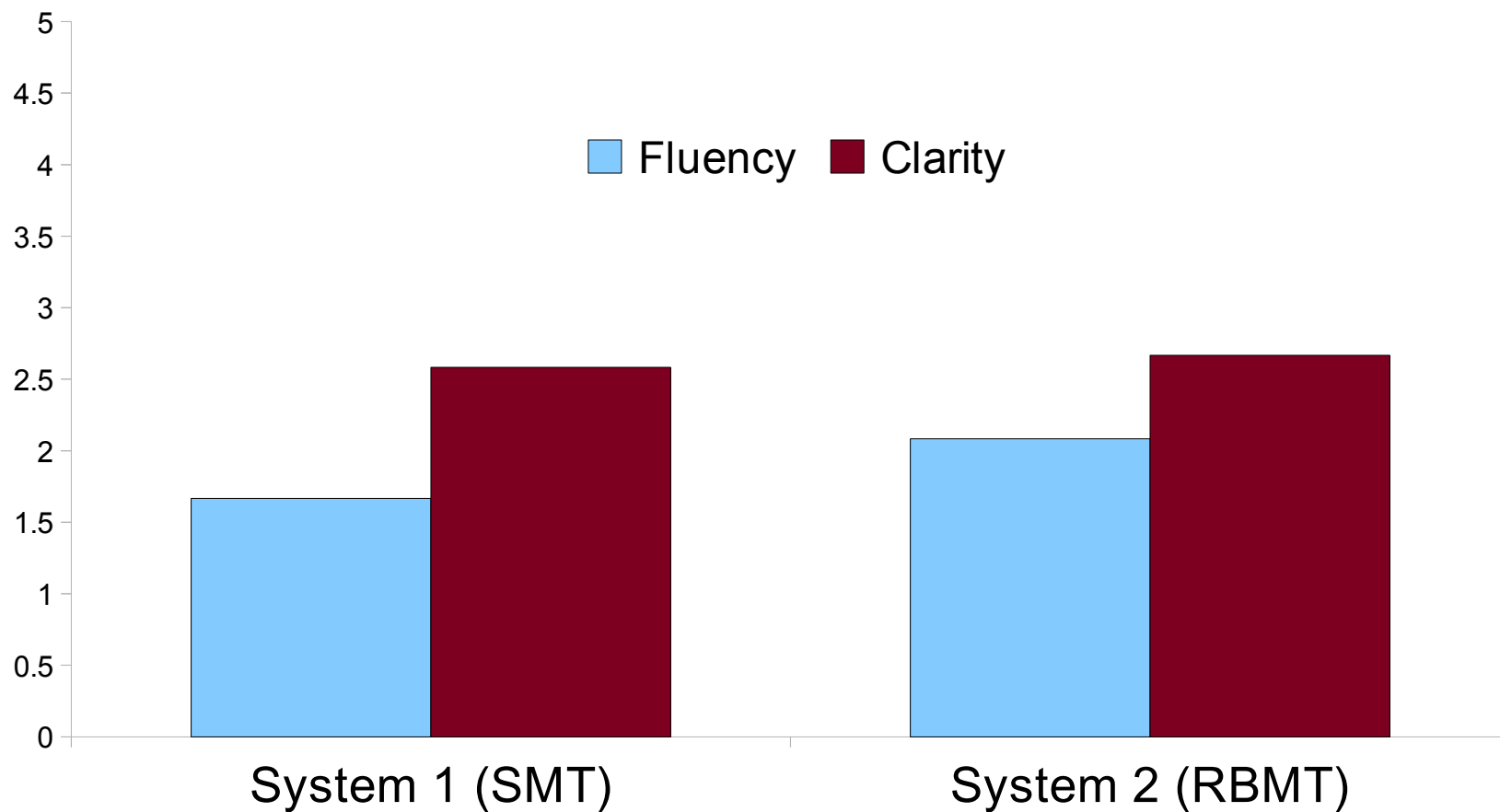
# Preliminary results 1: Fluency and clarity

“How fluent is the text?”

(5 flawless – 1 incomprehensible)

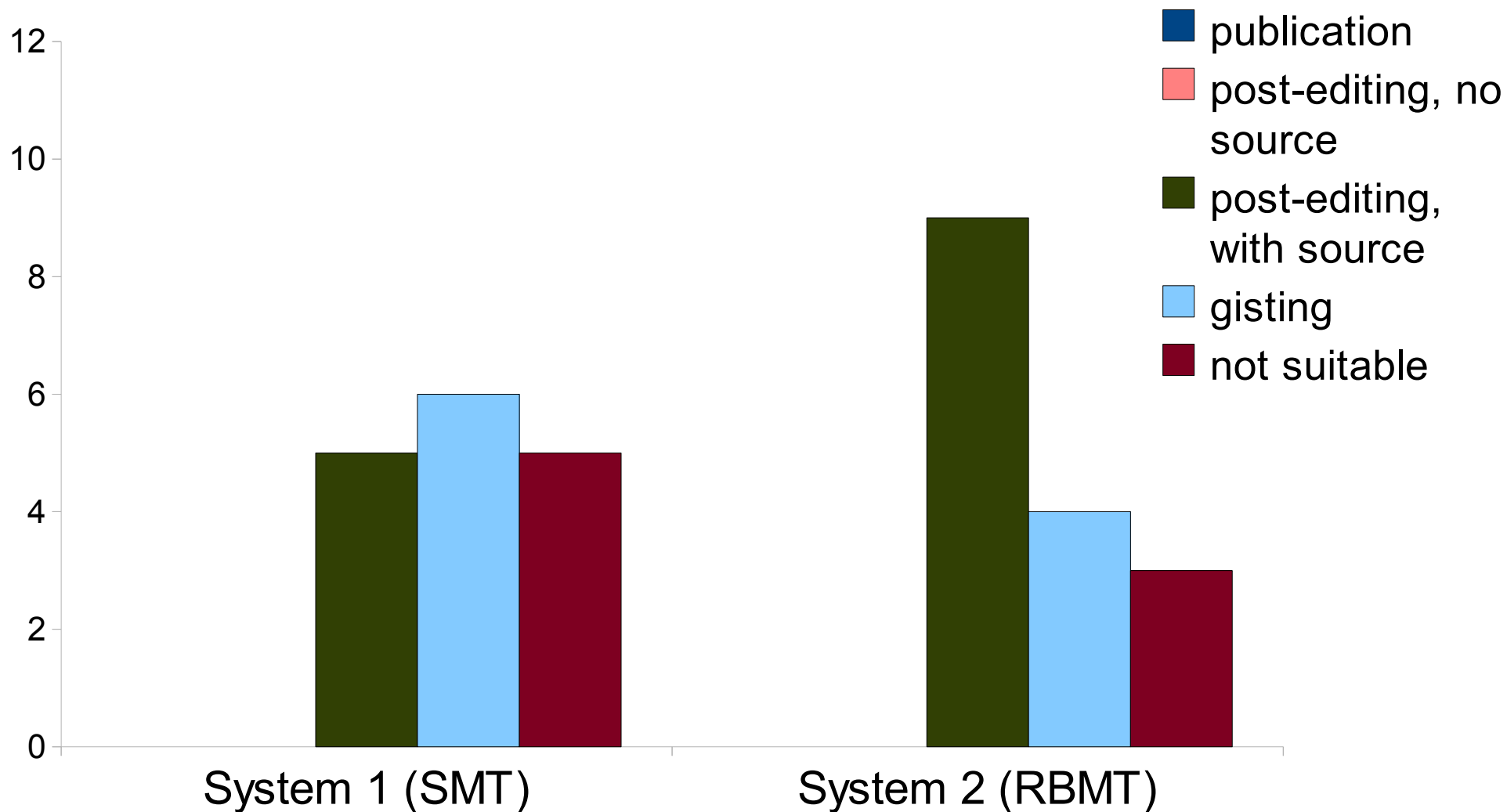
“How clear is the meaning?”

(5 entirely clear on first reading – 1 incomprehensible after several readings)

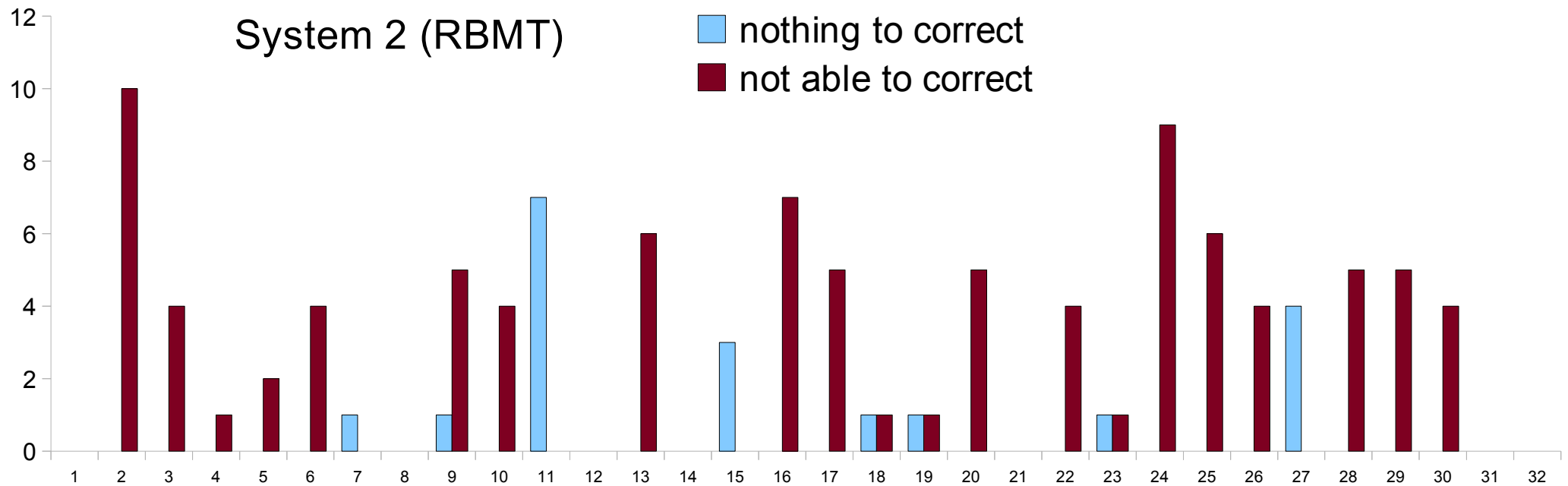
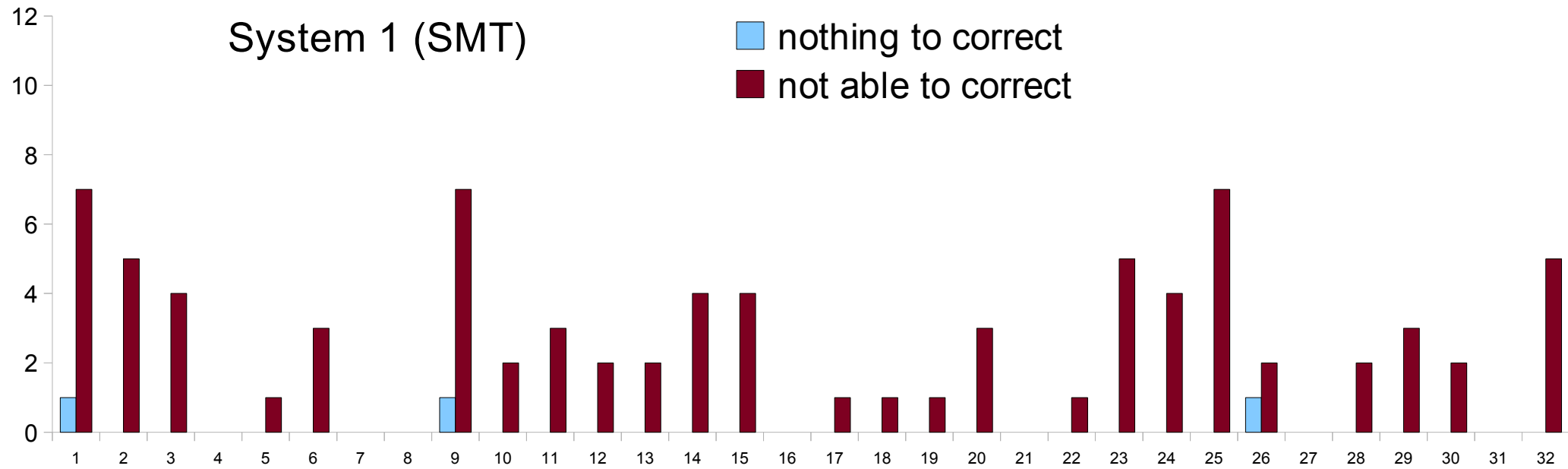


# Preliminary results 2: Usefulness of the translation

Do you consider the text suitable for...



# Preliminary results 3: Nothing to correct/not able to correct





# Manual Evaluation: Example

- Source  
*Home truths about telecom*
- System 1 (SMT)  
*Koti totuuksia tietoliikenne*  
Home SG.NOM. truth PL.PART. telecommunication SG.NOM.
  - 7 “not able to correct”, 1 “nothing to correct”
  - “Facts about telecommunication” (2), “Telecommunication for home use”, “Facts about telecommunication at home”
- System 2 (RBMT)  
*Karvas totuus tietoliikenteesta*  
Bitter SG.NOM. truth SG.NOM. telecommunication SG.\*ELAT.
  - “tietoliikenteesta → tietoliikenteestä” (11)
  - “karvas (bitter) → karu (harsh)” (2)