

Chunk Accuracy: A Simple, Flexible Metric for Translation Quality

Lars Ahrenberg

Department of Computer and Information Science
Linköping University

MTE workshop, Reykyavik
May 26, 2014

Overview

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

Introduction

Error-based approaches

The proposal

Translation Quality Assessment through metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ Equivalence metrics e.g. Tarvi, 2004

Translation Quality Assessment through metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ Equivalence metrics e.g. Tarvi, 2004
- ▶ Criteria-based
 - ▶ tabulated (yes, no, N/A)

Translation Quality Assessment through metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ Equivalence metrics e.g. Tarvi, 2004
- ▶ Criteria-based
 - ▶ tabulated (yes, no, N/A)
- ▶ Error-based
 - ▶ error count
 - ▶ error profile

Translation Quality Assessment through metrics

- ▶ Equivalence metrics e.g. Tarvi, 2004
- ▶ Criteria-based
 - ▶ tabulated (yes, no, N/A)
- ▶ Error-based
 - ▶ error count
 - ▶ error profile
- ▶ Reference-based
 - ▶ a single metric, or
 - ▶ a set of metrics

Problems with reference-based metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

They are indirect!

Problems with reference-based metrics

They are indirect!

- ▶ quality is usually not assessed

Problems with reference-based metrics

They are indirect!

- ▶ quality is usually not assessed
- ▶ purpose of reference may differ from system translations

Problems with reference-based metrics

They are indirect!

- ▶ quality is usually not assessed
- ▶ purpose of reference may differ from system translations
- ▶ references cannot cover the space of possible translations

Overview

Chunk Accuracy

Lars Ahrenberg

Introduction

**Error-based
approaches**

The proposal

Introduction

Error-based approaches

The proposal

Lessons from Stymne & Ahrenberg, 2012

Chunk Accuracy

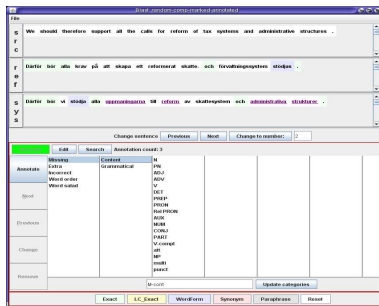
Lars Ahrenberg

Introduction

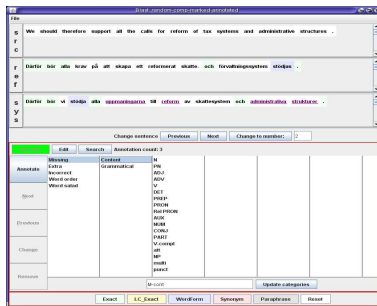
Error-based
approaches

The proposal

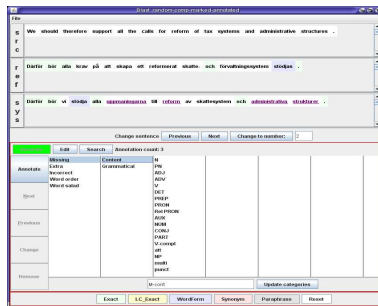
The screenshot displays the MIRA chunk-based parser interface. At the top, a text editor window shows three sentences with green annotations marking chunks and parts of speech. Below the text editor is a table with columns for 'Annotate', 'Store', 'Display', 'Change', and 'Remove'. The 'Annotate' column is expanded, showing a list of parts-of-speech tags such as N, PRE, ADJ, ADV, V, DET, PRON, and so on. The 'Annotate' column also includes a search bar and an 'Update categories' button. The 'Store' column has an 'Update categories' button. The 'Change' column has an 'Exact' button. The 'Remove' column has an 'S.C.Exact' button. The 'Display' column has a 'Spencer' button. The 'Change' column has a 'Paraphrase' button. The 'Remove' column has a 'Reset' button.



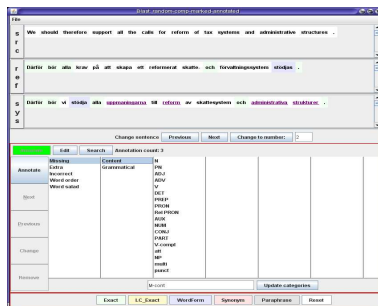
- ▶ disagreement on which token was affected



- ▶ disagreement on which token was affected
- ▶ disagreement on whether error is present



- ▶ disagreement on which token was affected
- ▶ disagreement on whether error is present
- ▶ the more detail the less agreement (cf. Bayerl & Paul, 2011)



- ▶ disagreement on which token was affected
- ▶ disagreement on whether error is present
- ▶ the more detail the less agreement (cf. Bayerl & Paul, 2011)
- ▶ short sentences didn't represent the full set well

Quality aspects

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

Translation

Quality aspects

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

SourceText ——— Translation

Quality aspects

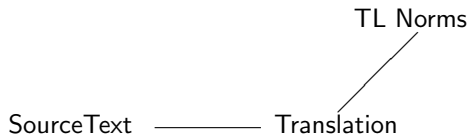
Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal



Quality aspects

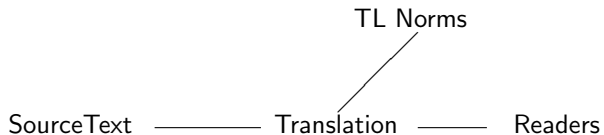
Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal



Quality aspects

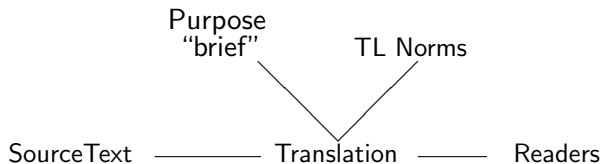
Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based approaches

The proposal



General problems with error-based metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ the error taxonomy is not selected in relation to purpose

General problems with error-based metrics

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low

General problems with error-based metrics

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low
 - ▶ errors may often be corrected (and thus classified) in many ways

General problems with error-based metrics

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low
 - ▶ errors may often be corrected (and thus classified) in many ways
 - ▶ large, detailed taxonomies are useless for really bad translations

General problems with error-based metrics

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low
 - ▶ errors may often be corrected (and thus classified) in many ways
 - ▶ large, detailed taxonomies are useless for really bad translations
 - ▶ boring task

General problems with error-based metrics

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low
 - ▶ errors may often be corrected (and thus classified) in many ways
 - ▶ large, detailed taxonomies are useless for really bad translations
 - ▶ boring task
- ▶ global (macro-level) properties are not considered

General problems with error-based metrics

- ▶ the error taxonomy is not selected in relation to purpose
- ▶ inter-annotator agreement tends to be low
 - ▶ errors may often be corrected (and thus classified) in many ways
 - ▶ large, detailed taxonomies are useless for really bad translations
 - ▶ boring task
- ▶ global (macro-level) properties are not considered
- ▶ it is not obvious how to define informative metrics (cf. Williams, 2001)

Overview

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

Introduction

Error-based approaches

The proposal

Main features

- ▶ Allocate issues to chunks – instead of words

Main features

- ▶ Allocate issues to chunks – instead of words
- ▶ Make the taxonomy small and relevant

Main features

- ▶ Allocate issues to chunks – instead of words
- ▶ Make the taxonomy small and relevant
- ▶ Each issue type should contribute to the translation profile

Example

- ▶ *These star-forming region, called by scientists unromantisch*

Example: agreement issues

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

- ▶ **These *star-forming* region, ...**

Example: word order issue

Word order uncertainties

- ▶ *..., called by scientists unromantisch*

Example: post-editing

- ▶ Good
- ▶ Editable
- ▶ Useless

Example: assimilation / reading support

- ▶ Comprehensible without effort
- ▶ Comprehensible with effort
- ▶ Incomprehensible

- ▶ A profile: the percentage of chunks falling into any given category
- ▶ A single metric: a weighted combination of these percentages

Example

Category	Weight	Result
Comprehensible	1.0	0.40
Compr. with effort	0.5	0.25
Incomprehensible	0	0.35

Single metric: $0.40 + 0.5 \cdot 0.25 = 0.525$

Chunk selection

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

Chunk selection

- ▶ a single clause, or a long phrase

Chunk selection

- ▶ a single clause, or a long phrase
- ▶ a discourse unit (?)

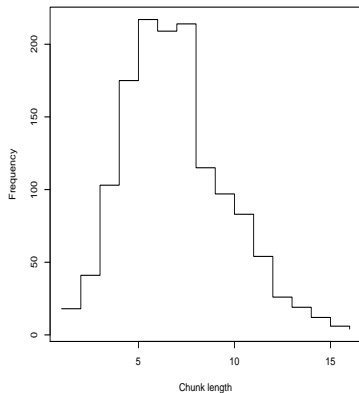
Chunk selection

- ▶ a single clause, or a long phrase
- ▶ a discourse unit (?)
- ▶ based on source chunks (?)

Chunk selection

- ▶ a single clause, or a long phrase
- ▶ a discourse unit (?)
- ▶ based on source chunks (?)
- ▶ goal: automatic

Chunk lengths - manual sample



mean: 6.5 words

The End

Chunk Accuracy

Lars Ahrenberg

Introduction

Error-based
approaches

The proposal

Thank you!