

Quality Estimation

Christian Buck,
University of Edinburgh



In this lecture you will ...

- Lose trust in MT
- Learn how to trust some MT
- Learn how to build a **complete** confidence estimation system
- Be surprised how **easy** that is
- Be also surprised how **hard** it is

MT - what is it good for?

- Making Websites available
- Skyping with foreign landlords
- Post-Editing
- Trading (including HFT)
- Information Retrieval

Easy to fail at any of these



Translate

From: German - detected ▼



To: English ▼

Translate

English

Spanish

French

German - detected

Facebook wurde Opfer eines Hacker-Angriffs. ✕

Stattdessen habe der Angriff bereits im Januar diesen Jahres.

Nach Angaben von Facebook hätten mehrere Mitarbeiter eine kompromittierte Developer-Webseite besucht und sich über eine zu diesem Zeitpunkt unbekannte Lücke in der Java-Sandbox Schad-Software eingefangen.

English

Spanish

Arabic

Facebook has been the victim of a hacker attack.

Instead I found the attack in January this year.

According to Facebook several employees had visited a compromised developer website and trapped on an unknown at this point gap in the Java sandbox malware.



Translate

From: German - detected ▼



To: English ▼

Translate

English

Spanish

French

German - detected

Facebook wurde Opfer eines Hacker-Angriffs. ✕

Stattdessen habe der Angriff bereits im Januar diesen Jahres.

Nach Angaben von Facebook hätten mehrere Mitarbeiter eine kompromittierte Developer-Webseite besucht und sich über eine zu diesem Zeitpunkt unbekannte Lücke in der Java-Sandbox Schad-Software eingefangen. |

English

Spanish

Arabic

Facebook has been the victim of a hacker attack.

Instead I found the attack in January this year.

According to Facebook several employees had visited a compromised developer website and trapped on an unknown at this point gap in the Java sandbox malware.



Translate

From: German - detected ▼



To: English ▼

Translate

English

Spanish

French

German - detected

English

Spanish

Arabic

Facebook wurde Opfer eines Hacker-Angriffs. ✕

Stattdessen habe der Angriff bereits im Januar diesen Jahres.

Nach Angaben von Facebook hätten mehrere Mitarbeiter eine kompromittierte Developer-Webseite besucht und sich über eine zu diesem Zeitpunkt unbekannte Lücke in der Java-Sandbox Schad-Software eingefangen.

Facebook has been the victim of a hacker attack.

Instead I found the attack in January this year.

According to Facebook several employees had visited a compromised developer website and trapped on an unknown at this point gap in the Java sandbox malware.



568185

To configure remote authentication with LDAP, complete these steps:



Per configurare l'autenticazione remota con LDAP, completare i seguenti passaggi:

00m: 01s

568186

Select Configure Remote Authentication.



Selezionare Configura Autenticazione Remota.

00m: 00s

T→

DRAFT

TRANSLATED

Translation matches

Select Configure Remote Authentication.

Selezionare Configura Autenticazione Remota.

Source: Anonymous 2012-10-05 100%

Select Configure Remote Authentication.

Seleziona Configura Remote Authentication.

Source: MT-FBK-IT (EN->IT) 2012-11-12 MT

Remote authentication

Autenticazione remota

Source: Anonymous 2012-10-25 67%

568187

Select the type of LDAP server that is used for authentication.



Selezionare il tipo di server LDAP che viene utilizzato per l'autenticazione.

00m: 01s

568188

The system supports IBM Tivoli Directory Server, Microsoft Active Directory, and other LDAP servers.



Il sistema supporta IBM Tivoli Directory Server, Microsoft Active Directory e altri server LDAP.

00m: 00s

Progress: 67%

Total Words: 1,021

To-do: 330

Speed: 674 Words/h

Completed in: 0h 29m

[Editing Log](#)

(Sentence Level) Quality Estimation

Produce quality score

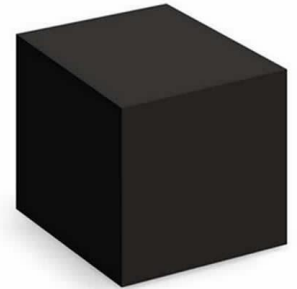
- Given source and (machine) translation
- *Without reference* translation



Applications:

- Good enough for publishing (print signs)?
- Inform readers
- Hide terrible translation from post-editors
- Decide between different systems

$$Q = f(\text{source}, \text{target})$$



$$Q = f(\text{source}, \text{target}, \text{MT})$$





2003 Summer Workshop @ JHU

What is good quality?

Early work: Predict automatic scores

- BLEU (~TrustRank)
- WER
- [many other scores not yet invented]

Problem: noisy on sentence level

Good quality for gisting

Content should be comprehensible

Accuracy over Fluency?

Gold standard:

- Collect feedback from users
 - Likert scores 1-4, 1-5, ...
- Answer questions

Good quality for post-editing

Time is money

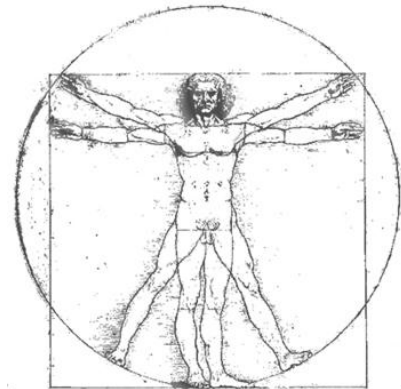
Avoid making translators hate their job

Fit with workflow

Only show MT if **speedup** expected

Measure time, collect interface actions

Humans are complicated



Summary

1. Specify objective
2. Get training data
3. Extract features
4. Train classifier / regression model
5. Profit!



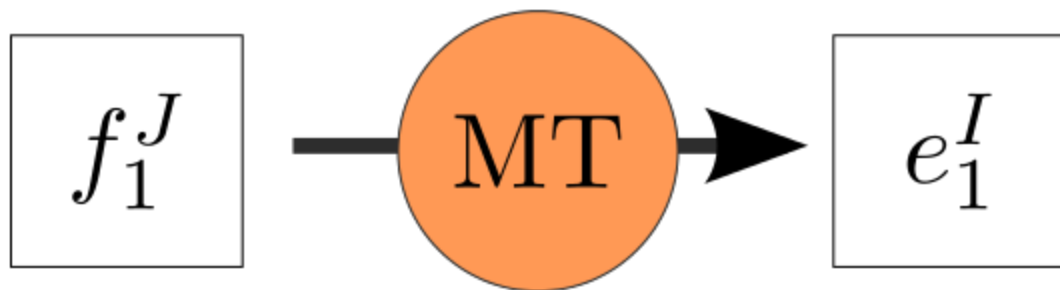
Necessary tool for human trials

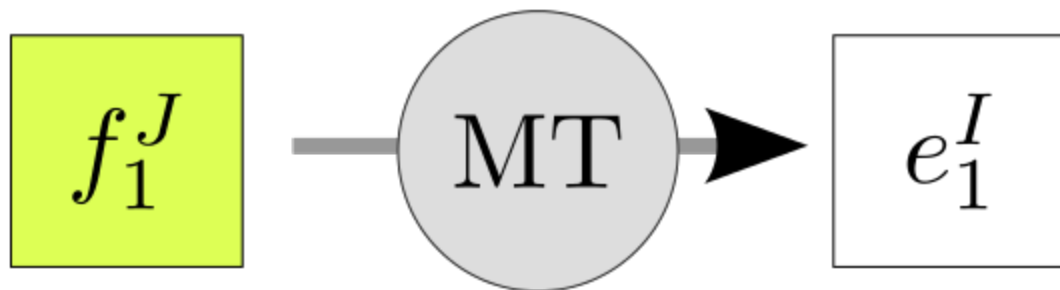
Features

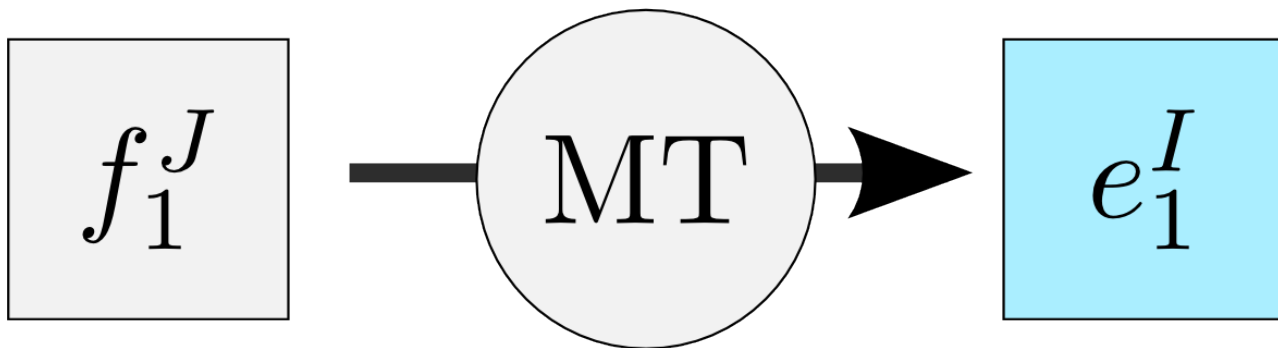
Think of some features!

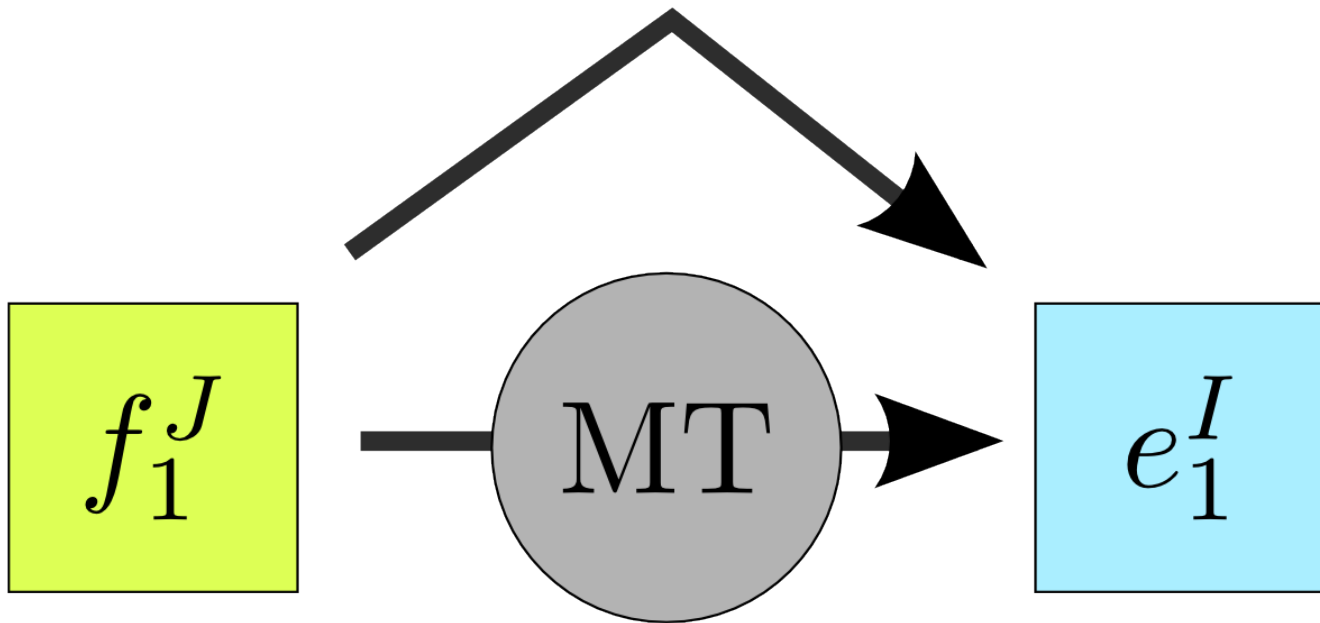
Common good features

- Source sentence perplexity
- Number of out-of-vocabulary words
- Number of words with many translations
- Number of words in source
- Mismatched question marks









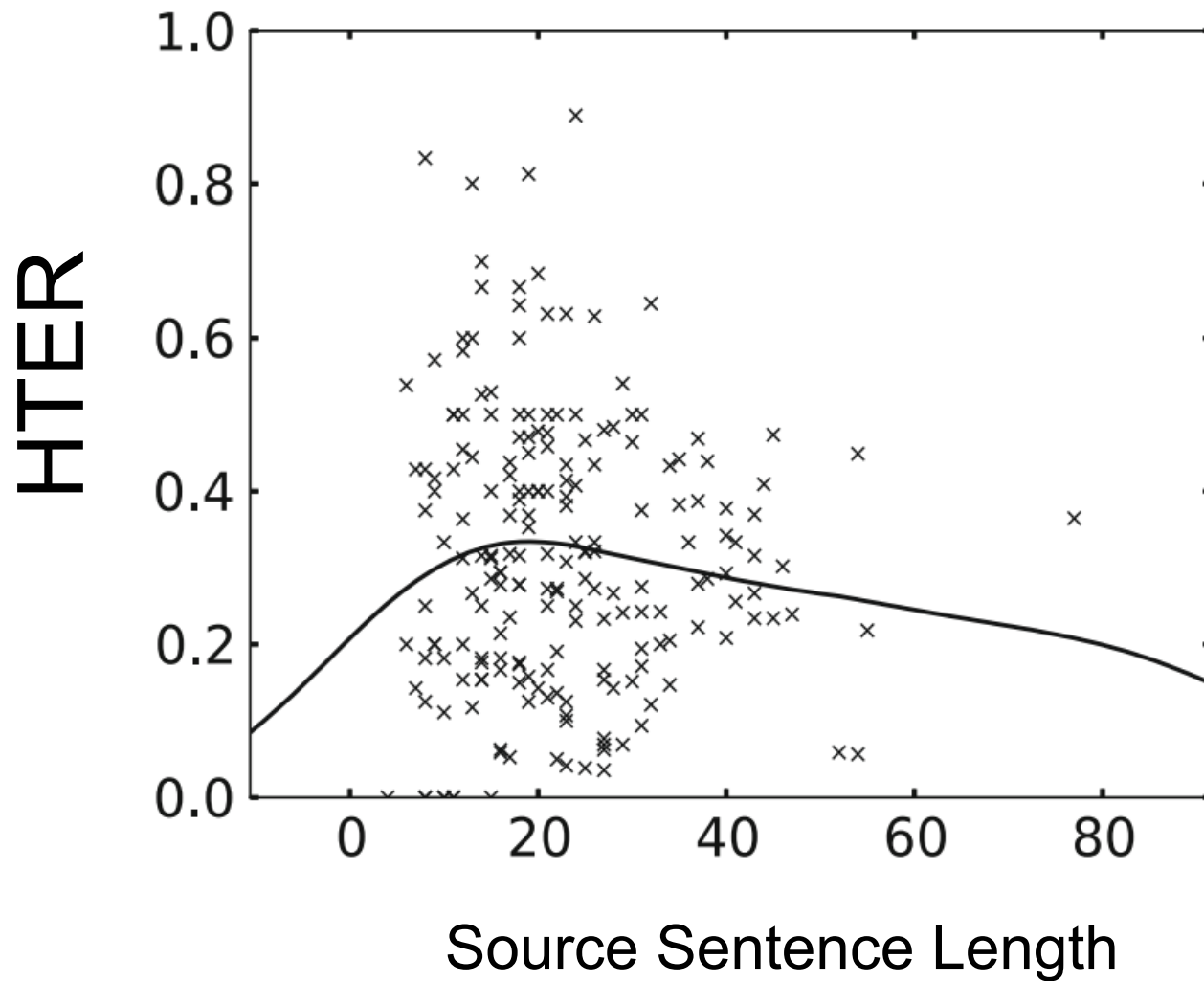
Simple source side features

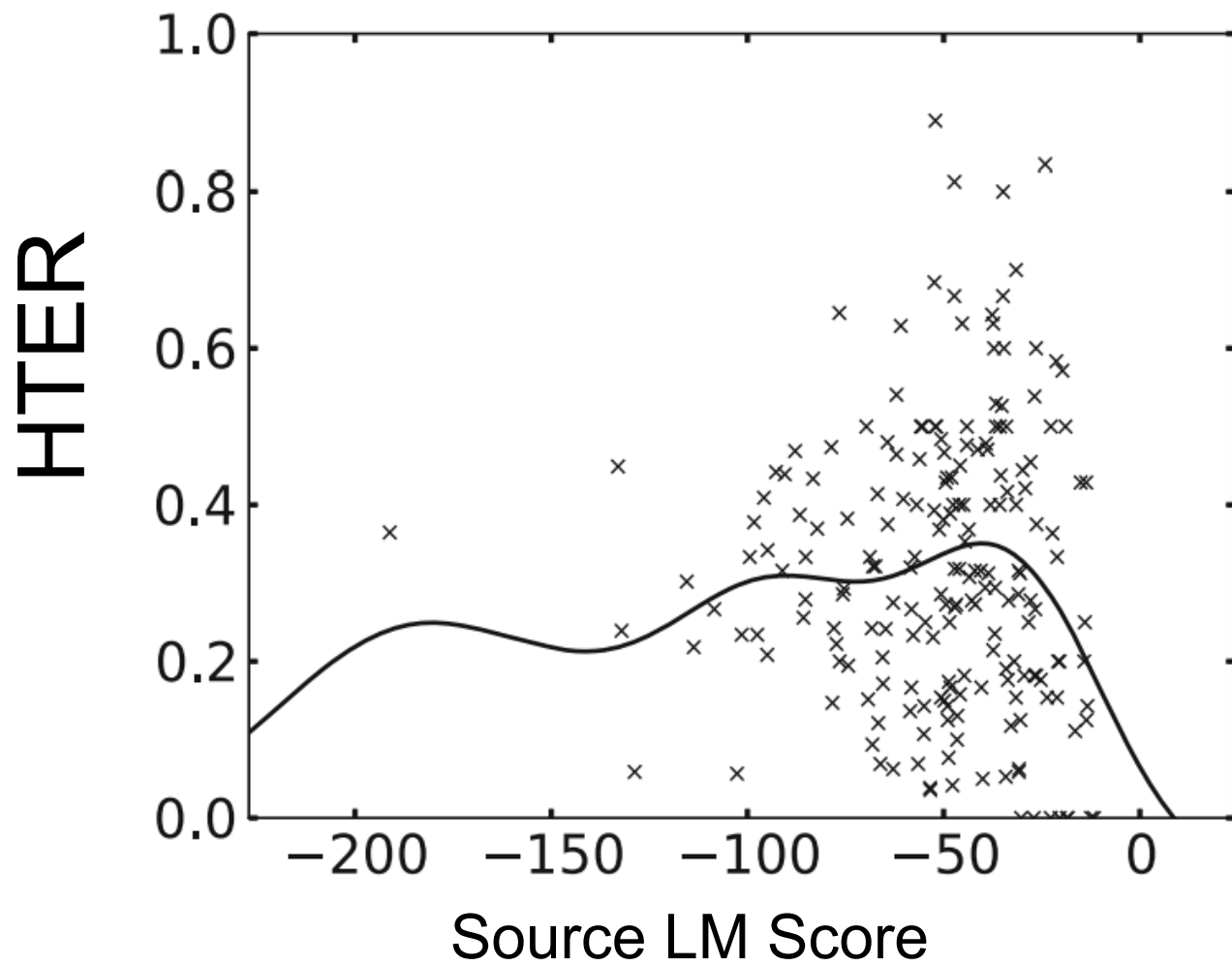
- Language model score
- Number of
 - Words
 - Characters
- Percentage of
 - Proper names
 - Numbers
 - Punctuation characters
 - Very rare/common words/ngrams

Simple source side features

- Language model score
- Number of
 - Words
 - Characters
- Percentage of
 - Proper names
 - Numbers
 - Punctuation characters
 - Very rare/common words/ngrams

Things that make
MT difficult





Hard to translate?

"Zora told it like it was," said Ella Dinkins, 90, one of the Johnson girls Hurston immortalized by quoting men singing off-color songs about their beauty.

Hard to translate?

"**Zora** told it like it was," said **Ella Dinkins**, 90, one of the **Johnson girls Hurston** immortalized by quoting men singing off-color songs about their beauty.

Hard to translate?

"**Zora** told it like it was," said **Ella Dinkins**, 90, one of the **Johnson girls Hurston immortalized** by quoting men singing off-color songs about their beauty.

Hard to translate?

"Zora told it like it was," said Ella Dinkins, 90, one of the Johnson girls Hurston immortalized by quoting men singing off-color songs about their beauty.

Spanish English French English - detected ▼



English Spanish German ▼

Translate

"Zora told it like it was," said Ella Dinkins, 90, ✕
one of the Johnson girls Hurston
immortalized by quoting men singing off-
color songs about their beauty.



"Zora gesagt, dass es, wie es war", sagte Ella
Dinkins, 90, einer der Johnson Mädchen
Hurstons zitieren Männer singen off-color
Lieder über ihre Schönheit verewigt.



 Wrong?

More source side features

Words with many possible translations

English	German	$P(\text{German} \text{English})$
work	Arbeit (job, physics, object)	0.4
	arbeiten (to work)	0.2
	Aufgabe (task)	0.2
	Werk (work of art)	0.1
	Arbeitsplatz (workplace)	0.1

Rare and common n-grams

Zora told it like it was,

Zora told it

told it like

it like it

like it was

it was ,

Rare and common n-grams

[Zora told it] [told it like] [it like it] [like it was] [it was ,]

infrequent

frequent

n-grams from large corpus, sorted by count

Rare and common n-grams

[Zora told it] [told it like] [it like it] [like it was] [it was ,]

infrequent

frequent



Rare and common n-grams

[Zora told it] [told it like] [it like it] [like it was] [it was ,]

infrequent

frequent



Linguistic features: POS

- Part of speech (POS) LM
 - on source or target side
- LEPOR (~BLEU on POS Tags)

LEPOR

its	ratification	would	require	226	votes

seine	Ratifizierung	erfordern	wuerde	226


LEPOR

its	ratification	would	require	226	votes
PRON	NOUN	VERB	VERB	NUM	NOUN


seine	Ratifizierung	erfordern	wuerde	226
PRON	NOUN	NOUN	VERB	NUM

LEPOR

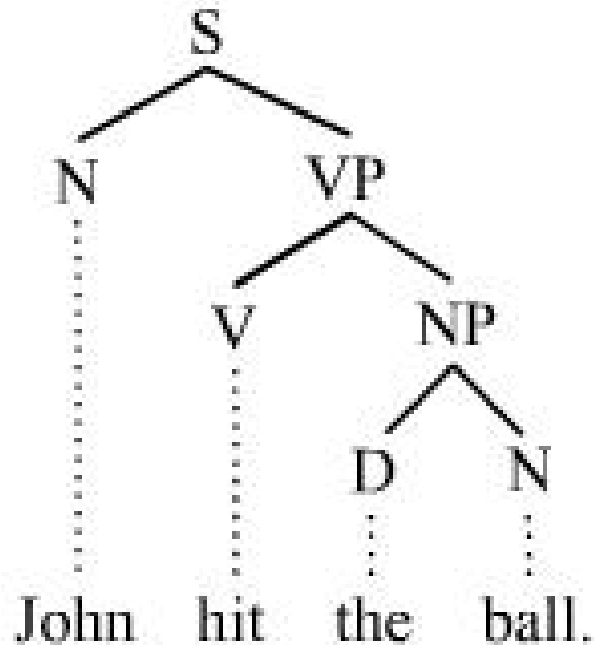
its	ratification	would	require	226	votes
PRON	NOUN	VERB	VERB	NUM	NOUN



seine	Ratifizierung	erfordern	wuerde	226
PRON	NOUN	NOUN	VERB	NUM

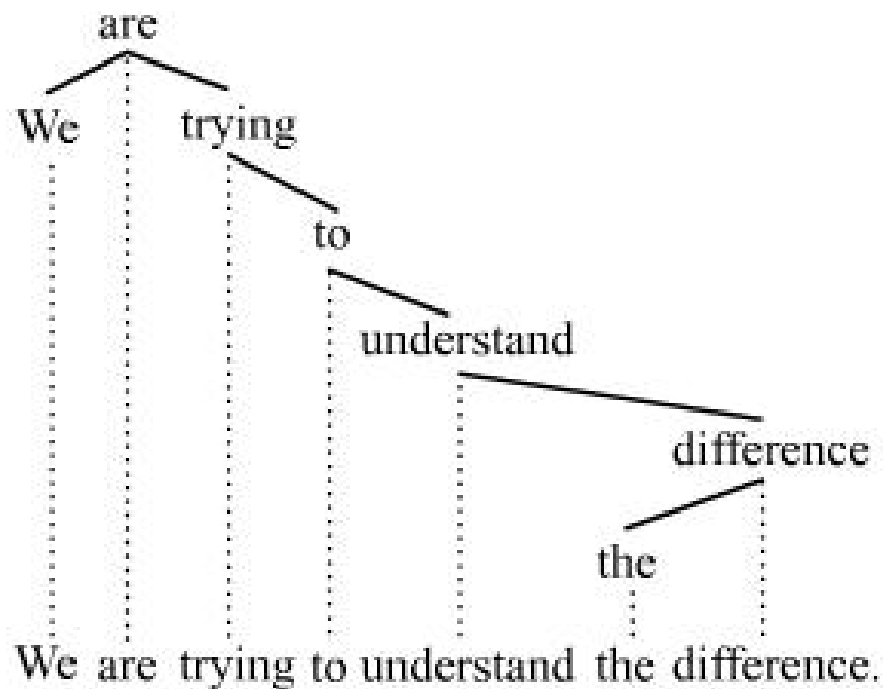


Linguistic features II

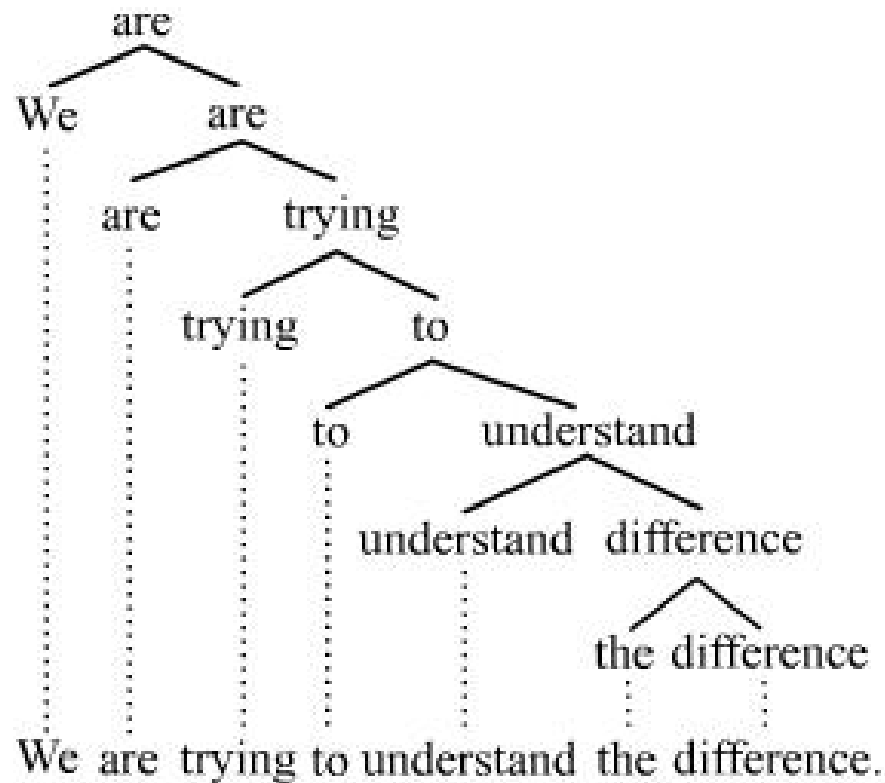


Constituency-based parse tree

Linguistic features II



Dependency



Constituency (BPS)

Pseudo-References

The “How much does it look like the Google translation?”-feature

Applicability questionable

Back-Translation

Idea:

1. Translate target back to source language
2. Compare with original (using BLEU, TER)

Back-Translation

German Spanish English Detect language ▼



English German Spanish ▼

Translate

In Deutschland wird scheinbar kontrovers
über Europas Rettungspolitik diskutiert.



In Germany is discussed seemingly
controversial about Europe's bailout policy.



 Wrong?

Back-Translation

German Spanish **English** Detect language ▼



English **German** Spanish ▼

Translate

In Germany is discussed seemingly
controversial about Europe's bailout policy.



In Deutschland ist scheinbar kontroverse über
Europas Rettungspolitik diskutiert.



 **Wrong?**

Back-Translation

German Spanish English Detect language ▼



English German Spanish ▼

Translate

In Germany is discussed seemingly controversial about Europe's bailout policy.



In Deutschland ist scheinbar kontroverse über Europas Rettungspolitik diskutiert.



 Wrong?

Back-Translation

German (Auto-Detected) ▼

In Deutschland wird scheinbar kontrovers über Europas Rettungspolitik diskutiert.

81/5000



Translate

English ▼

In Germany, rescue policy is apparently controversy about Europe.



Back-Translation

English (Auto-Detected) ▼

In Germany, rescue policy is apparently controversy about Europe.

65/5000



Translate

German ▼

In Deutschland ist die Rettung-Politik offenbar Kontroverse über Europa.



Original: In Deutschland wird scheinbar kontrovers über Europas Rettungspolitik diskutiert.

Cross-Translation



Word level errors

German Spanish English Detect language ▼

↔

English German Spanish ▼

Translate

In Deutschland wird scheinbar kontrovers über Europas Rettungspolitik diskutiert. ✕

✎

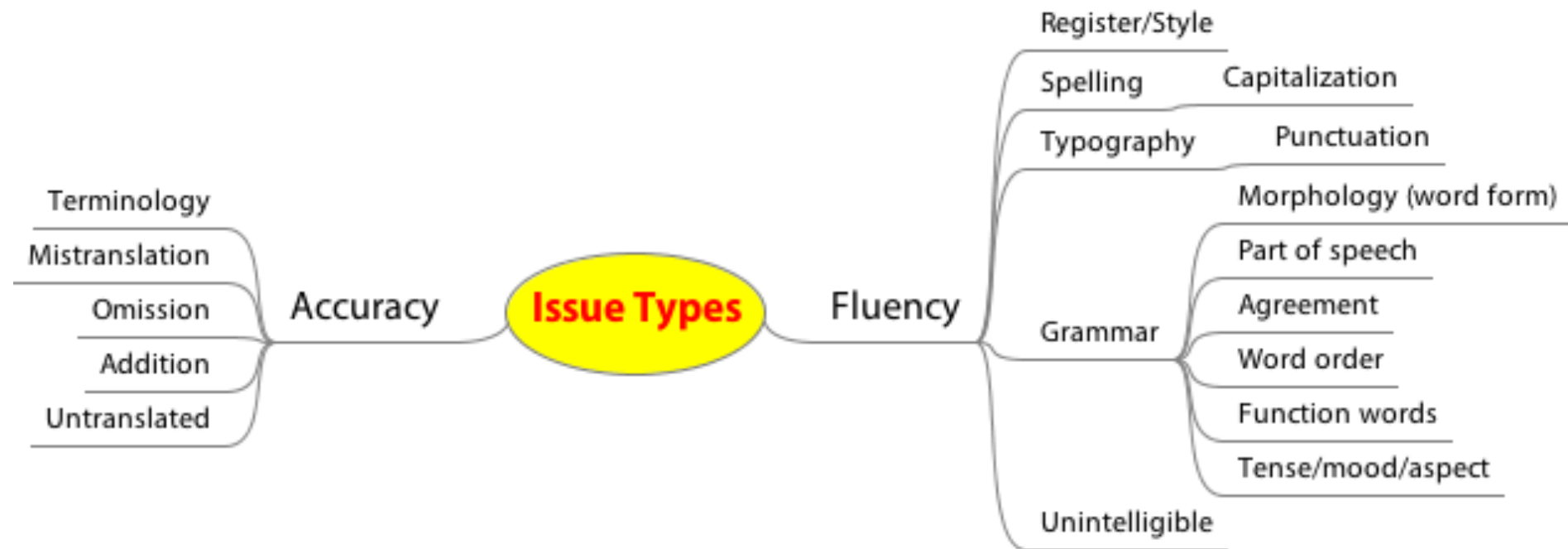
In Germany is discussed seemingly controversial about Europe's bailout policy.

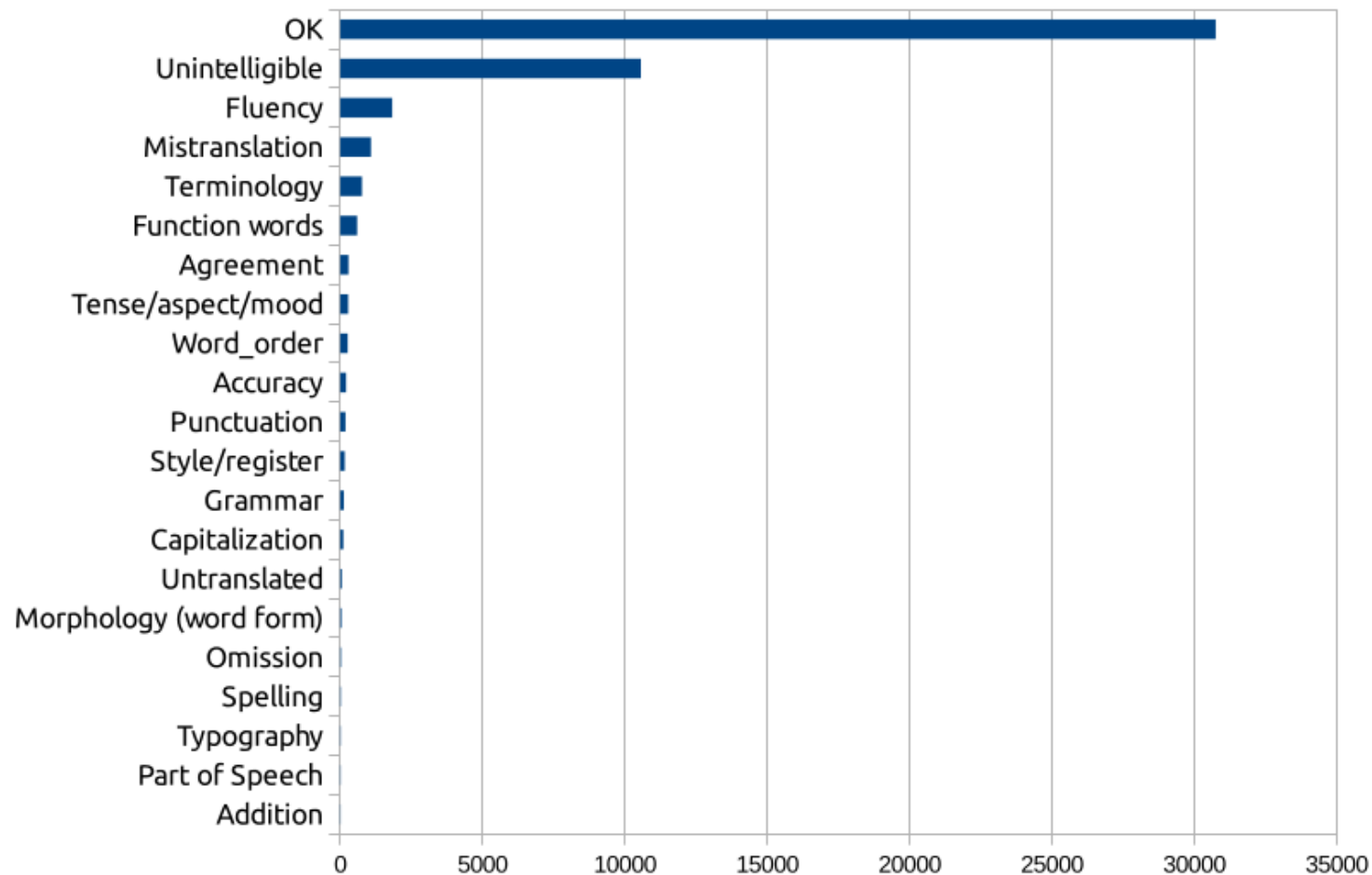
☆ ☰ 🔊

✎ Wrong?

Roughly: Germany is seemingly controversially discussing Europe's bailout policy

Word level error annotation





Word Posterior Probabilities (WPP)

	p
Mary slapped the green witch.	0.7
Mary did slap the green witch.	0.2
It was Mary who slapped the green witch.	0.1

Feature Selection

Find best subset of 24 features

- How many subsets?

Feature Selection

Find best subset of 24 features

- 2^{24} subsets
- Testing 1 subset takes 1m. How long?

Feature Selection

Find best subset of 24 features

- 2^{24} subsets
- Testing 1 subset takes 1m.
- Wait **32** years

Feasible!

Greedy feature selection

Forward selection

- Add feature that gives best improvement on dev set

Backward selection

- Remove feature that gives best improvement on dev set (when it's gone)

Alternatives

Gaussian Processes

Sparsity inducing regularization (L_1)

Hand picking

Random search

Get your hands dirty

<http://statmt.org/wmt15/quality-estimation-task.html>

- Sentence level (predict HTER)
- Word level (predict Good/Bad)
- Paragraph level (predict METEOR)

Submission: May 25, 2015

