

# **Manuální a automatické ohodnocení strojového překladu mezi evropskými jazyky**

Anna Plocková

- španělština, francouzština, němčina – angličtina
- 14 týmů, 11 institucí

<b>ID</b>	<b>Participant</b>
cmu	Carnegie Mellon University, USA (Zollmann and Venugopal, 2006)
lcc	Language Computer Corporation, USA (Olteanu et al., 2006b)
ms	Microsoft, USA (Menezes et al., 2006)
nrc	National Research Council, Canada (Johnson et al., 2006)
ntt	Nippon Telegraph and Telephone, Japan (Watanabe et al., 2006)
rali	RALI, University of Montreal, Canada (Patry et al., 2006)
systran	Systran, France
uedin-birch	University of Edinburgh, UK — Alexandra Birch (Birch et al., 2006)
uedin-phi	University of Edinburgh, UK — Philipp Koehn (Birch et al., 2006)
upc-jg	University of Catalonia, Spain — Jesús Giménez (Giménez and Màrquez, 2006)
upc-jmc	University of Catalonia, Spain — Josep Maria Crego (Crego et al., 2006)
upc-mr	University of Catalonia, Spain — Marta Ruiz Costa-jussà (Costa-jussà et al., 2006)
upv	University of Valencia, Spain (Sánchez and Benedí, 2006)
utd	University of Texas at Dallas, USA (Olteanu et al., 2006a)

# K dispozici

- Základní systém strojového překladu s datovými zdroji
- Jazykové modely pro každý jazyk
- Výcvikové a zkušební sady
- dekodér strojového překladu
  - Tréninkový skript
- Korpus Europarl
- Dvě sady paralelního textu o 2000 větách
- 29 úvodníků Project Syndicate website

# Automatické vyhodnocení

- BLEU = **bi**lingual **e**valuation **u**nderstudy

# Manuální vyhodnocení

	<b>Adequacy</b>	<b>Fluency</b>
5	All Meaning	Flawless English
4	Most Meaning	Good English
3	Much Meaning	Non-native English
2	Little Meaning	Disfluent English
1	None	Incomprehensible

## Judge Sentence

You have already judged 14 of 3064 sentences, taking 86.4 seconds per sentence.

**Source:** les deux pays constituent plutôt un laboratoire nécessaire au fonctionnement interne de l'ue .

**Reference:** rather , the two countries form a laboratory needed for the internal working of the eu .

Translation	Adequacy	Fluency
both countries are rather a necessary laboratory the internal operation of the eu .	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> 1 2 3 4 5	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> 1 2 3 4 5
both countries are a necessary laboratory at internal functioning of the eu .	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5
the two countries are rather a laboratory necessary for the internal workings of the eu .	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5
the two countries are rather a laboratory for the internal workings of the eu .	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5
the two countries are rather a necessary laboratory internal workings of the eu .	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> 1 2 3 4 5
<b>Annotator:</b> Philipp Koehn <b>Task:</b> WMT06 French-English	<input type="button" value="Annotate"/>	
Instructions	5= All Meaning 4= Most Meaning 3= Much Meaning 2= Little Meaning 1= None	5= Flawless English 4= Good English 3= Non-native English 2= Disfluent English 1= Incomprehensible

# Kdo zvítězil?

- Mnoho systémů mělo velmi podobné hodnocení
  - Těžké dělat silné závěry
  - Statistické systémy zaostávají za konkurencí
- BLEU asi citlivý na žargon a volbu správných slov
- Anotátoři měli občas tendenci hodnotit systémy

# Poznámky anotátorů

- Potíže se zachováním standardu při hodnocení
- Adekvátnost by měla být rozdělena
- Téměř nemožné ohodnotit dlouhé věty
- Nedostatek referenčních překladů (nebo nepřesné)
- Možnost vynechat věty, které nemohou posoudit

Děkuji za pozornost.



# Zdroje

- <http://www.aclweb.org/anthology/W06-3114>
- <https://en.wikipedia.org/wiki/BLEU>