Tools and Guidelines for Principled Machine Translation Development

MT-ComparEval and MQM

Nora Aranberri, Eleftherios Avramidis, Aljoscha Burchardt,
Ondřej Klejch, Martin Popel, Maja Popović

nora.aranberri@ehu.eus, {eleftherios.avramidis,aljoscha.burchardt}@dfki.de,
{klejch,popel}@ufal.mff.cuni.cz, maja.popovic@hu-berlin.de

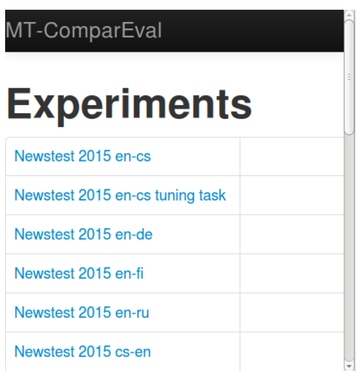
University of the Basque Country (UPV/EHU); German Research Center for Artificial Intelligence (DFKI, Language Technology Lab);
Charles University in Prague (Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, ÚFAL); Humboldt University of Berlin

MT-ComparEval

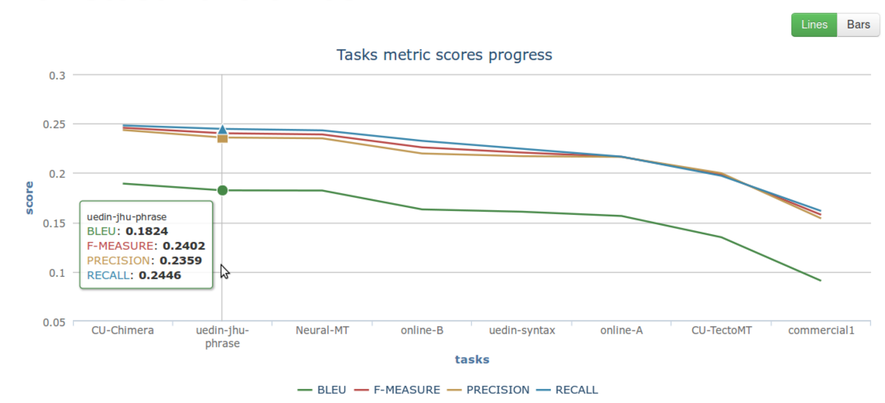
- **Try it now (WMT 2014–2016 results)**
  http://wmt.ufal.cz
- **Upload and analyze your translations**
  http://mt-compareval.ufal.cz
- **Install it** (and report issues or contribute):
  https://github.com/choko/MT-ComparEval

- **web-based tool for MT developers**
- **check progress of a system over time or compare several MT systems**
- **focus on analyzing system differences**
- **can be integrated into your workflow (import translations: disk/git/REST-API)**

① Select an "Experiment", e.g. English-Czech WMT15



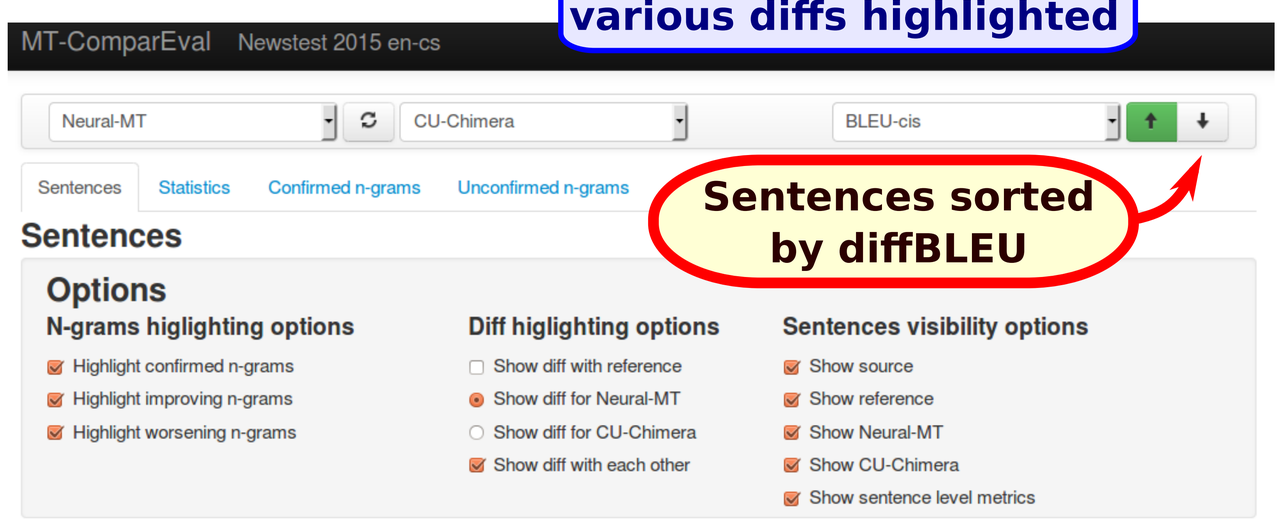
② See BLEU, F-measure,... Select two "Tasks", i.e. system translations



compare

name	description	BLEU ↓	F-MEASURE	PRECISION	RECALL	
CU-Chimera		0.1893	0.2458	0.2435	0.2481	hide
uedin-jhu-phrase		0.1824	0.2402	0.2359	0.2446	hide
Neural-MT		0.1822	0.239	0.2351	0.2431	hide

③ Sentences pane with various diffs highlighted

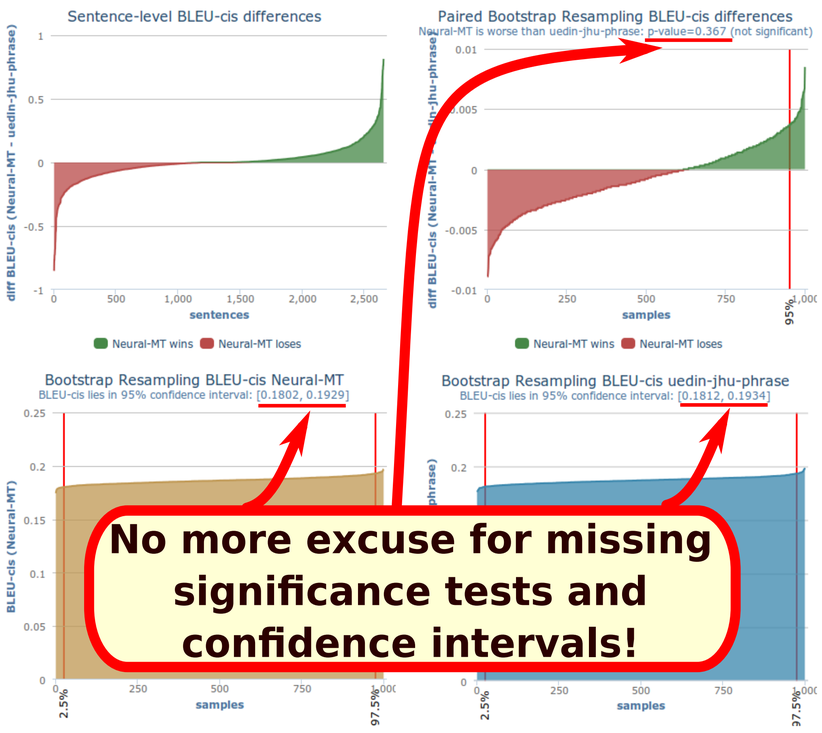


Source	The soundtrack is a mix of reggae and rap and the tunes are upbeat .
Reference	Soundtrack je směsicí reggae a rapu a melodie je optimistická .
Neural-MT	The je směs , a rap a melodie jsou optimistická .
CU-Chimera	Soundtrack je směsicí reggae a rapu a melodie jsou optimistické .

No more excuse for missing examples!

	BLEU	BLEU-cis	F-MEASURE	F-MEASURE-cis	PRECISION	PRECISION-cis	RECALL	RECALL-cis
Neural-MT	0.1173	0.1173	0.1864	0.1864	0.1864	0.1864	0.1864	0.1864
CU-Chimera	0.6989	0.6989	0.7025	0.7025	0.7025	0.7025	0.7025	0.7025
Diff	-0.5816	-0.5816	-0.5161	-0.5161	-0.5161	-0.5161	-0.5161	-0.5161

④ Statistics pane (paired bootstrap resampling, ...)



No more excuse for missing significance tests and confidence intervals!

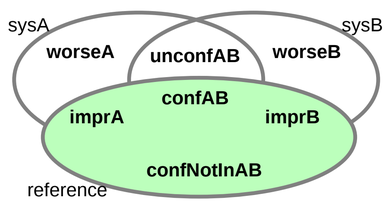
- See the sentences with biggest improvement/worsening
- See the 1-grams...4-grams with biggest improvement/worsening
- Hints for improving the systems

⑥ Unconfirmed n-grams

1-gram

uedin-nmt loses		metamind loses	
,	634 - 499 = 135	-	312 - 263 = 49
die	502 - 405 = 97	eine	191 - 161 = 30
der	492 - 404 = 88	dass	136 - 109 = 27
von	285 - 225 = 60	einen	127 - 105 = 22
den	238 - 178 = 60	einem	98 - 77 = 21
'	55 - 7 = 48	Kirche	18 - 0 = 18
sich	128 - 81 = 47	sei	46 - 29 = 17
an	118 - 72 = 46	Es	54 - 38 = 16
zu	289 - 244 = 45	Als	23 - 7 = 16
ein	123 - 80 = 43	seine	67 - 52 = 15

- **confirmed n-gram** = occurres in the reference (light yellow and blue highlight)
- **improving n-gram** = confirmed n-gram occurring in only one of the systems (dark yellow and blue highlight)
- **worsening n-gram** = unconfirmed, occuring in only one of the systems (red highlighting)
- **diff** (LCS underlined in green)



⑤ Confirmed n-grams pane

3-gram

Pilot 0.00 wins		Pilot 1.00 wins	
I do to	9	How do I	28
how can I	6	I change the	14

Click on any n-gram to see all its occurrences

Source	Wie öffne ich ein Dokument in Libreoffice ?
Reference	How do I open a document in LibreOffice ?
Pilot 0.00	As I open a document in LibreOffice ?
Pilot 1.00	How do I open a document in Libreoffice ?

Source	Wie verschicke ich eine Datei über Skype ?
Reference	How do I send a file using Skype ?
Pilot 0.00	As I always send a file on Skype ?
Pilot 1.00	How do I send a file above Skype ?

MQM (Multidimensional Quality Metrics)

explicit error annotation in translate5 software

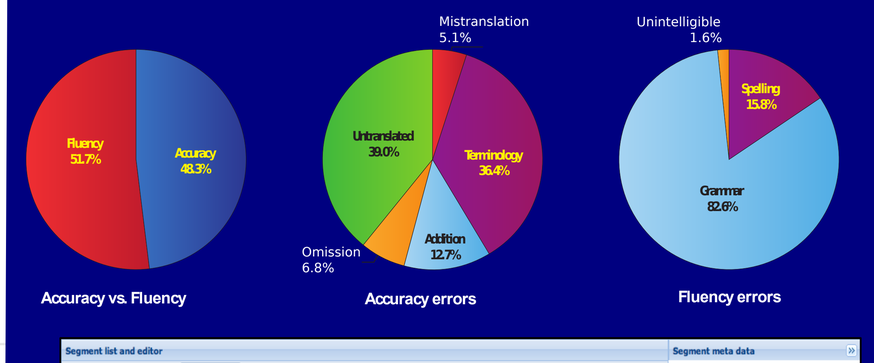
3	(DE)	**Gehen Sie zu Tools und wählen Sie dann Browsingchronik Löschen.., können Sie dann vorziehen, Ihre Internet-Cookies zu löschen.**	
	deA	[[1] Gehen Sie zu] [[2] Tools] und wählen Sie dann [[3] Browsingchronik] [[4] Löschen].., [[5] können] Sie dann [[6] vorziehen], Ihre Internet-Cookies zu löschen.	6

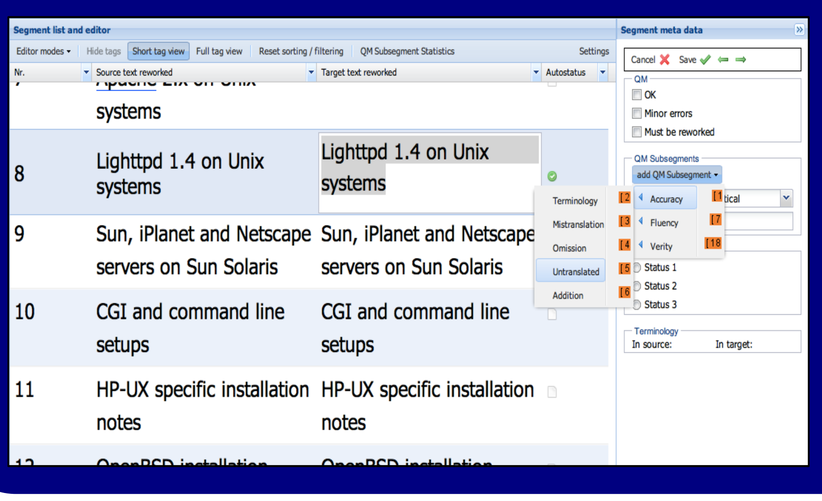
1. **Mistranslation** [Gehen Sie zu]
2. **Untranslated** [Tools]
3. **Mistranslation** [Browsingchronik]
4. **Part of speech** [Löschen]
5. **Word order** [können]
6. **Mistranslation** [vorziehen]

3	(DE)	**Sprung zu Extras und wählen Sie dann Browserverlauf löschen,.., Sie können dann Ihre Internet-Cookies löschen.**	
	deA	[[1] Sprung zu] Extras und wählen Sie dann Browserverlauf löschen,..[[2] ,] Sie können dann[[3]]Ihre Internet-Cookies löschen.	3

1. **Mistranslation** [Sprung zu]
2. **Typography** [,]
3. **Omission** [ ]

- **Detailed error analysis when needed**
- **Bridge to quality assurance in industry**
- **Enables "profiling" of systems/languages, etc.**





Presented at LREC 2016, Portorož, Slovenia. Supported by the FP7 EU grant QTLeap (No. 610516) and GAČR grant GA15-10472S. Using language resources hosted by the LINDAT/CLARIN Research Infrastructure, Project No. LM2010013.