Meteor Universal: Language Specific Translation Evaluation for Any Target Language

Michael Denkowski **Alon Lavie**

{mdenkows,alavie}@cs.cmu.edu

Carnegie Mellon

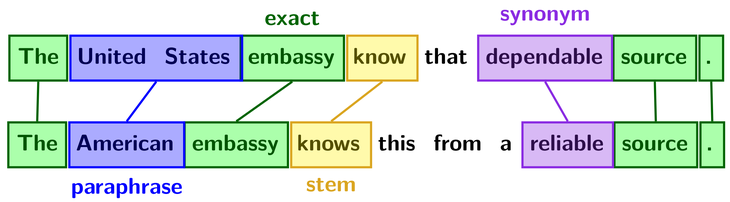
Meteor Universal

Support for **any target language** using only **bitext** used to build MT systems:

- Linguistic resources (**paraphrases** and **function words**) extracted from **bitext**.
- **Universal parameter set** learned by pooling data from **all WMT languages**.
- **Significantly outperforms** baseline metrics on **unseen languages** with **no development data**.

Meteor Scoring

Align MT hypothesis to reference translation using **flexible matching**:



Compute **similarity score** based on **precision**, **recall**, and **fragmentation** (measure of gaps and reordering)

Parameters tuned to maximize agreement with human judgments:

- α: balance between precision and recall
- β, γ: shape and severity of fragmentation penalty
- δ: balance between content and function words
- w_{stem}, $w_{synonym}$, $w_{paraphrase}$: weights of approximate matches

Universal Parameter Set

Parameter set learned using **all WMT12 data** (Callison-Burch et al., 2012):

- **100,000** binary rankings covering **8 language directions**.
- Restrict scoring for **all languages** to **exact** and **paraphrase** matching.

Parameters encode human preferences that **generalize** across languages:

- Prefer **recall** over **precision**.
- Prefer **word choice** over **word order**.
- Prefer correct translations of **content words** over **function words**.
- Prefer **exact matches** over **paraphrase matches**, while still giving significant credit to paraphrases.

Visualization

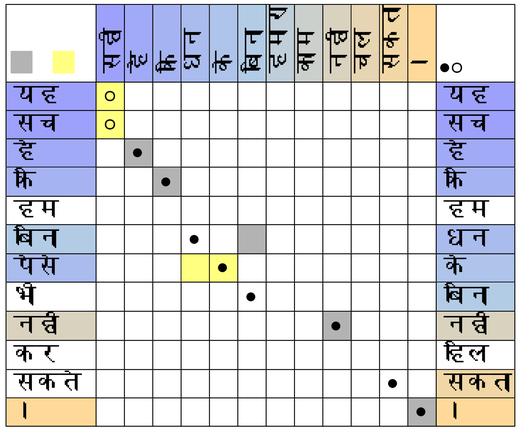
Meteor X-ray visualizes translation scoring for **any language**:

System 1 Reference System 2

	это	заставляет	меня	тосковать	по	многим	обещаниям	,	которые	улетучились	.	●○
это	●											это
делает		●										заставляет
меня			●									меня
жаждут				●								тосковать
многочисленные					●							по
обещания						●						многим
,							●					обещаниям
которые								●				,
исчезли									●			которые
.												исчезли
											●	.

Example output of Meteor-Xray comparing two translations into Russian

System 1 Reference System 2



Example output of Meteor-Xray comparing two translations into Hindi

Visually inspect individual MT outputs or **compare** multiple systems:

- Highlights **exact and approximate** matches and **word fragmentation**.
- Uses **Meteor Universal** and **Unicode rendering** to fully support any language.
- **LaTeX** and **PDF** output format for inclusion in reports and presentations.

Linguistic Resources

Paraphrases extracted using phrase pivoting (Bannard and Callison-Burch, 2005) and **filtered** for precision (Denkowski and Lavie, 2011):

After a sharp **drop** in the morning ...
Después de la rápida **caída** de la mañana ...
... Una **caída** de volumen parecido se registró por última vez ...
... having registered a similarly-ranged **fall** the last time ...

Learning paraphrase ("drop", "fall") by pivoting through "caída"

Function words learned by relative frequency in monolingual text:

The weight **of** one **of** **the** world's longest-running conflicts **is** resting on ...
All **of** **this** **is** designed to reinforce one point: **the** Gaza withdrawal belongs ...
For **the** source **of** **the** problem **is** neither **the** European Central Bank ...
So it **is** surprising that **this** choice **is** not at **the** center **of** **the** political ...

Learning function words "the", "of", "is", "this" by high frequency

Experiments

Compare to (language-specific) **Meteor 1.3** on WMT13 data (Macháček and Bojar, 2013):

τ	Meteor 1.3	Meteor Universal	BLEU
English	0.214	0.206	0.124
Czech	0.092	0.085	0.044
German	0.163	0.157	0.097
Spanish	0.106	0.101	0.068
French	0.150	0.137	0.099

Sentence-level agreement with human judges on WMT13 data

Compare to **BLEU** on two **unseen** languages, **Russian** and **Hindi**:

τ	Meteor Universal	BLEU
Russian (WMT13)	0.128	0.068
Hindi (WMT14)	0.264	0.227

Sentence-level agreement with human judges on data for unseen languages

Software

Meteor 1.5 (including **Meteor Universal**) is freely available under the **GNU Lesser General Public License**.

Meteor Universal tutorial:
http://www.cs.cmu.edu/~mdenkows/meteor-universal.html

Meteor download:
https://www.cs.cmu.edu/~alavie/METEOR/