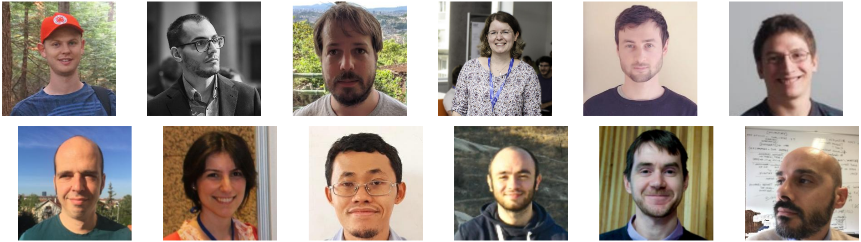
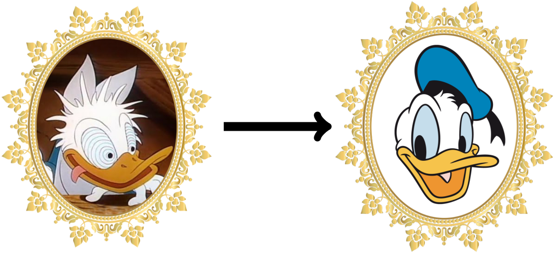
MultiLexNorm: A Shared Task on Multilingual Lexical Normalization

Rob van der Goot, Alan Ramponi, Arkaitz Zubiaga, Barbara Plank, Benjamin Muller, Iñaki San Vicente Roncal, Nikola Ljubešić, Özlem Çetinoğlu, Rahmad Mahendra, Talha Çolakoğlu, Timothy Baldwin, Tommaso Caselli and Wladimir Sidorenko



Lexical Normalization

Lexical normalization is the task of transforming an utterance into its standard form, word by word, including both one-to-many (1-n) and many-to-one (n-1) replacements.



Lexical Normalization

State before shared task:

- Most work on English
- Also work on single other languages
- Varieties in task definitions, guidelines and metrics
- No common evaluation benchmark

MultiLexNorm

- combination of existing datasets
- annotation style and file format converged
- “new” evaluation metric
- external evaluation (UD)

MultiLexNorm

Lang.	Language name	Normalization example
DA	Danish	De skarpe lamper gjorde destromindre ek bedre . De skarpe lamper gjorde destro mindre ikke bedre .
DE	German	ogäj isch hätts auch dwiddern könn Okay ich hätte es auch twittern können
EN	English	u hve to let ppl decide what dey want to do you have to let people decide what they want to do
ES	Spanish	@username cuuxamee sii peroo veen yaa eem @username escúchame sí pero ven ya eh
HR	Croatian	svi frendovi mi nešto rade , veceras san osta sam . svi frendovi mi nešto rade , večeras sam ostao sam .
ID-EN	Indonesian-English	pdhal not fully bcs those ppl jg sih . padahal not fully because those people juga sih .
IT	Italian	a Roma è cosí primavera che sembra gia giov a Roma è così primavera che sembra già giovedì
NL	Dutch	Kga me wss trg rolle vant lachn Ik ga me waarschijnlijk terug rollen van het lachen
SL	Slovenian	jst bi tud najdu kovanec vreden veliko denarja . jaz bi tudi našel kovanec vreden veliko denarja .
SR	Serbian	komunalci kace pocne kaznjavanje ? komunalci kad počne kažnjavanje ?
TR	Turkish	He o dediyin suala cvb verdim He o dediğin suale cevap verdim
TR-DE	Turkish-German	@username Yerimm senii , damkee schatzymm :-* @username Yerim seni , danke Schatzym :-*

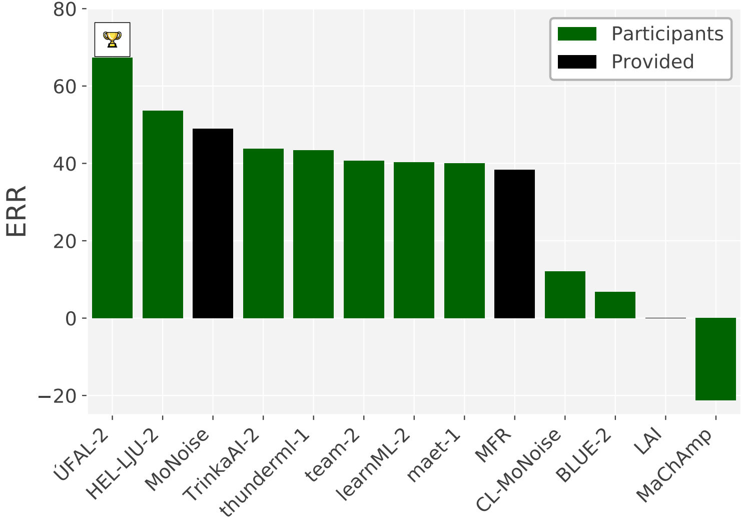
Is this sample biased?

Metric

- Previously: accuracy, accuracy over OOV words, F1 score, BLEU, word error rate, character error rate, etc.
- Now: accuracy normalized for amount of words to be normalized. Error Reduction Rate:

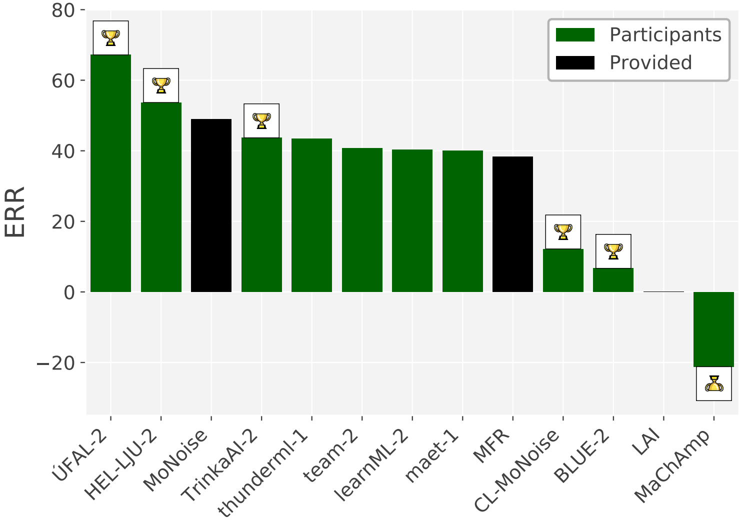
$$ERR = \frac{\%\text{accuracy} - \%\text{words\_not\_normed}}{100 - \%\text{words\_not\_normed}}$$

Results



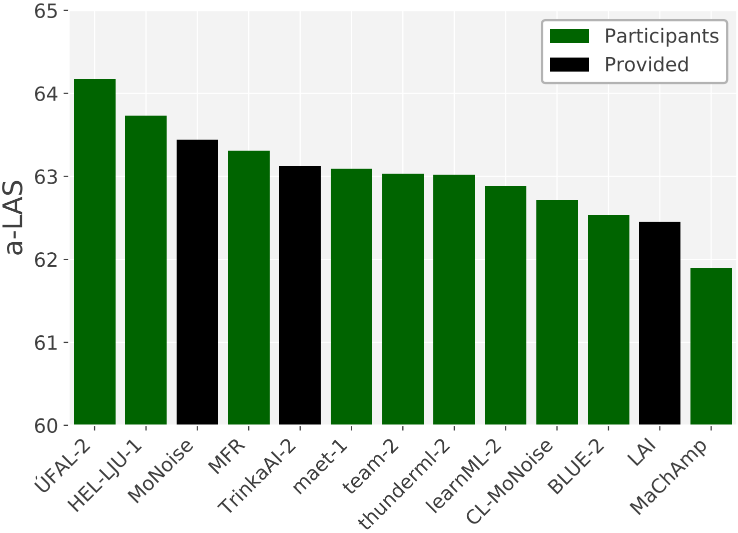
ERR
80
60
40
20
0
−20
🏆
Participants
Provided
ÚFAL-2
HEL-LJU-2
MoNoise
TrinkaAI-2
thundermI-1
team-2
learnML-2
maet-1
MFR
CL-MoNoise
BLUE-2
LAI
MaChAmp

Results

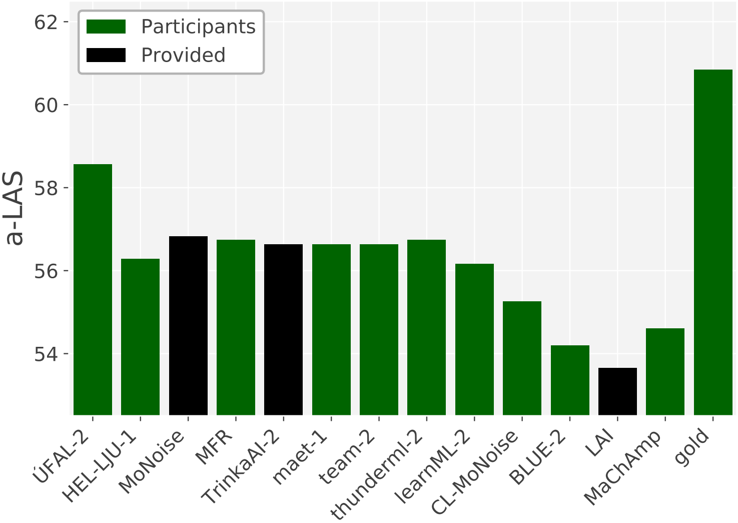


ERR
80
60
40
20
0
−20
🟩 Participants
⬛ Provided
🏆
🏆
🏆
🏆
🏆
ÚFAL-2
HEL-LJU-2
MoNoise
TrinkaAI-2
thunderml-1
team-2
learnML-2
maet-1
MFR
CL-MoNoise
BLUE-2
LAI
MaChAmp

Extrinsic Evaluation (avg.)



Extrinsic Evaluation (EN-MoNoise)



Thanks!

- `https://bitbucket.org/robvanderg/multilexnorm/`