

# Rob van der Goot

Bla bla bla ...



Frisian  
Dutch  
Danish  
Chinese  
English



(Most) interesting projects:

- ▶ Lexical normalization (MultiLexNorm)
- ▶ Cross-lingual slot and intent detection (xSID)
- ▶ Multi-task learning (MaChAmp)

# Lexical normalization in 2015

- ▶ Most work on English, some isolated work on other languages
- ▶ Usually assume words to be normalized are known
- ▶ Different evaluation metrics

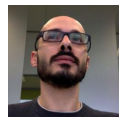
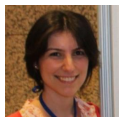
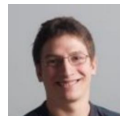
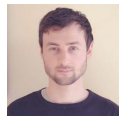
Example:

social	ppl	r	troublesome	.
IV	OOV	OOV	IV	IV
	people	are		

# Lexical normalization in 2021

## 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

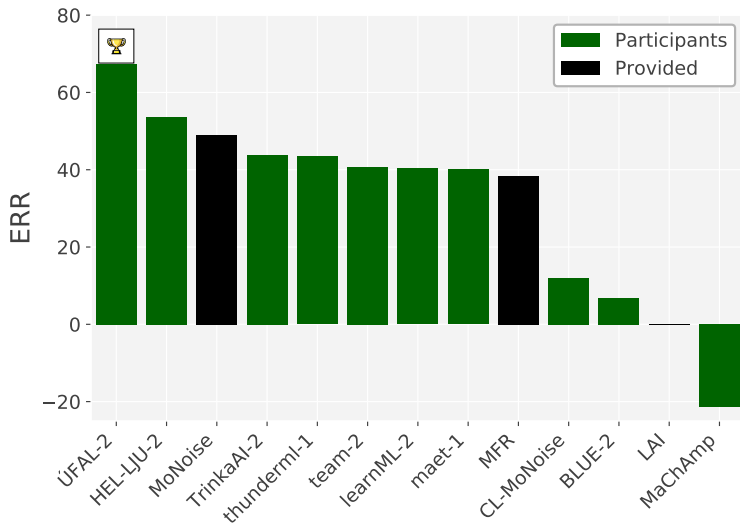


# MultiLexNorm

- ▶ Combination of existing datasets
- ▶ Annotation style and file format converged
- ▶ “new” evaluation metric

Lang.	Language name	Normalization example
DA	Danish	De skarpe lamper gjorde destromindre ek bedre . De skarpe lamper gjorde destro mindre ikke bedre .
DE	German	ogāj isch hātts 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 pero 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 così primavera che sembra già 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 Schatzymm :-*

# MultiLexNorm: Results



# xSID: Cross-lingual Slot and Intent Detection

Rob van der Goot, Ibrahim Sharaf, Aizhan Imankulova, Ahmet Üstün, Marija Stepanović, Alan Ramponi, Siti Oryza Khairunnisa, Mamoru Komachi and Barbara Plank





## Slot and Intent Detection

I'd like to see the showtimes for Silly Movie 2.0 at the movie house

Intent: SearchScreeningEvent

ar أود أن أرى مواعيد عرض فيلم **Silly Movie 2.0** في **دار السينما**  
 da Jeg vil gerne se spilletiderne for **Silly Movie 2.0** i **biografen**  
 de Ich würde gerne den Vorstellungsbeginn für **Silly Movie 2.0** im **Kino** sehen  
 de-st I mecht es Programm fir **Silly Movie 2.0** in **Film Haus** sechn  
 en I'd like to see the showtimes for **Silly Movie 2.0** at the **movie house**  
 id Saya ingin melihat jam tayang untuk **Silly Movie 2.0** di gedung **bioskop**  
 it Mi piacerebbe vedere gli orari degli spettacoli per **Silly Movie 2.0** al **cinema**  
 ja 映画館の **Silly Movie 2.0** の上映時間を見せて。  
 kk Мен **Silly Movie 2.0** бағдарламасының **кинотеатрда** көрсетілім уақытын көргім келеді  
 nl Ik wil graag de speeltijden van **Silly Movie 2.0** in het **filmhuis** zien  
 sr Želela bih da vidim raspored prikazivanja za **Silly Movie 2.0** u **bioskopu**  
 tr **Silly Movie 2.0**'ın **sinema salonundaki** seanslarını görmek istiyorum  
 zh 我想看 **Silly Movie 2.0** 在 **影院** 的放映

## Massive Choice, Ample Tasks (MACHAMP):



### A Toolkit for Multi-task Learning in NLP



**Rob van der Goot** 🌐 **Ahmet Üstün** 🌐 **Alan Ramponi** 🌐🌐 **Ibrahim Sharaf** 🌐

**Barbara Plank** 🌐

IT University of Copenhagen 🌐 University of Groningen 🌐 University of Trento 🌐

Fondazione the Microsoft Research - University of Trento COSBI 🌐 Factmata 🌐

robv@itu.dk, a.ustun@rug.nl, alan.ramponi@unitn.it

ibrahim.sharaf@factmata.com, bapl@itu.dk

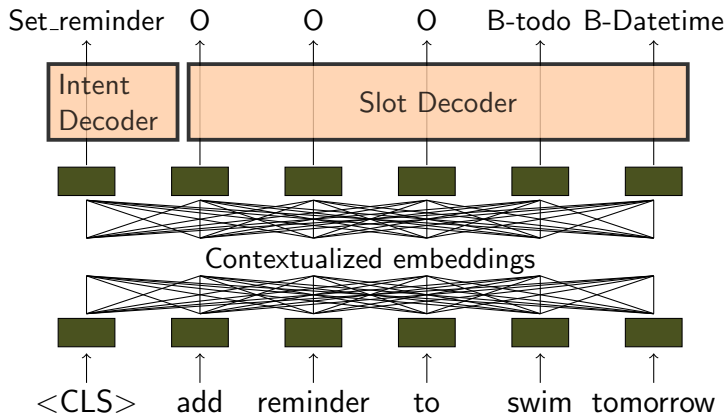


Multi-task learning within and across datasets.

Task-types:

Input:	Gary	Oak	:	Smell	ya	later
seq	PROPN	PROPN	PUNCT	VERB	PRON	ADV
string2string	Gary	Oak	:	smell	you	later
seq_bio	B-PER	I-PER	0	0	0	0
dependency	4-vocative	1-flat	4-punct	0-root	4-obj	5-obl
classification	Greeting					
mlm	4.3e-34					
seq2seq	Gary Oak : Ruik je later					
...						

# MaChAmp



# MaChAmp

