

# From Masked Language Modeling to Translation: Non-English Auxiliary Tasks Improve Zero-shot Spoken Language Understanding

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

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?
- ▶ Do we need a machine translation model?

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?
- ▶ Do we need a machine translation model?
- ▶ How about Kazakh? or Japanese? or Indonesian? or Arabic?

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?
- ▶ Do we need a machine translation model?
- ▶ How about Kazakh? or Japanese? or Indonesian? or Arabic?
- ▶ Do we need a language model?

## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?
- ▶ Do we need a machine translation model?
- ▶ How about Kazakh? or Japanese? or Indonesian? or Arabic?
- ▶ Do we need a language model?
- ▶ How about Turkish?



## Slot and Intent Detection

- ▶ “In-domain”, in-language scores on English are generally very high
- ▶ How about performance on German? Italian? Serbian? Danish? Dutch?
- ▶ Do we need a machine translation model?
- ▶ How about Kazakh? or Japanese? or Indonesian? or Arabic?
- ▶ Do we need a language model?
- ▶ How about Turkish?
- ▶ How about Chinese? South-Tyrolean?

## Cross-lingual Slot and Intent Detection (xSID):

- ▶ 250 test and 150 dev from Facebook (Schuster et al., 2019) data
- ▶ 250 test and 150 dev from Snips (Coucke et al., 2018) data
- ▶ Concatenate them and converge annotation style
- ▶ Colleague-sourced to get higher quality/consistency as crowd-sourced

## Cross-lingual Slot and Intent Detection (xSID):

- ▶ 250 test and 150 dev from Facebook (Schuster et al., 2019) data
- ▶ 250 test and 150 dev from Snips (Coucke et al., 2018) data
- ▶ Concatenate them and converge annotation style
- ▶ Colleague-sourced to get higher quality/consistency as crowd-sourced
- ▶ Kappa agreement on Dutch slots (3 annotators): 0.92

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**'in **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 🇮🇹

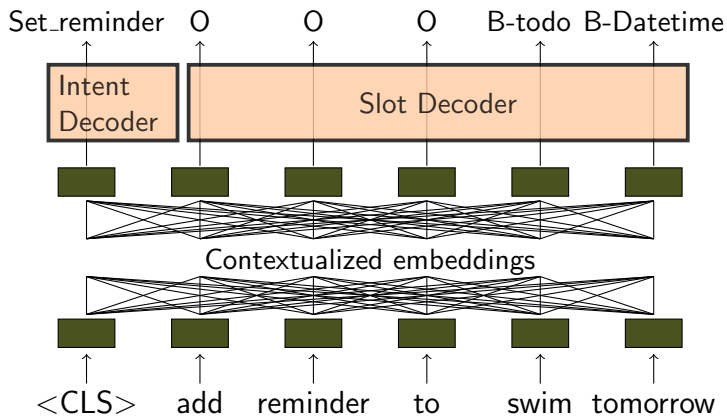
Fondazione the Microsoft Research - University of Trento COSBI 🇮🇹 Factmata 🇳🇱

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

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

# Experiments

BASELINE:



# Experiments

Evaluate 2 embeddings

- ▶ mBERT: trained on 104 languages (12/13)
- ▶ XLM15: trained on 15 languages (5/13)

# Experiments

NMT-TRANSFER

MT en-tgt

SLU en-train

Fairseq

SLU tgt-train

MaChAmp (baseline)

SLU tgt-dev/test

MT training data: Ted Talks + Opensubtitles



# Experiments

Proposed models:

- ▶ Jointly train on auxiliary task in target language:
  - ▶ Masked language modeling (AUX-MLM)

# Experiments

Proposed models:

- ▶ Jointly train on auxiliary task in target language:
  - ▶ Masked language modeling (AUX-MLM)
  - ▶ UD-parsing (AUX-UD)

# Experiments

Proposed models:

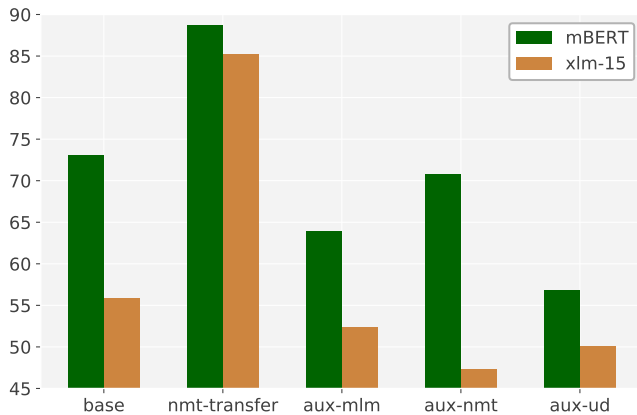
- ▶ Jointly train on auxiliary task in target language:
  - ▶ Masked language modeling (AUX-MLM)
  - ▶ UD-parsing (AUX-UD)
  - ▶ Neural machine translation (AUX-NMT)

## Results

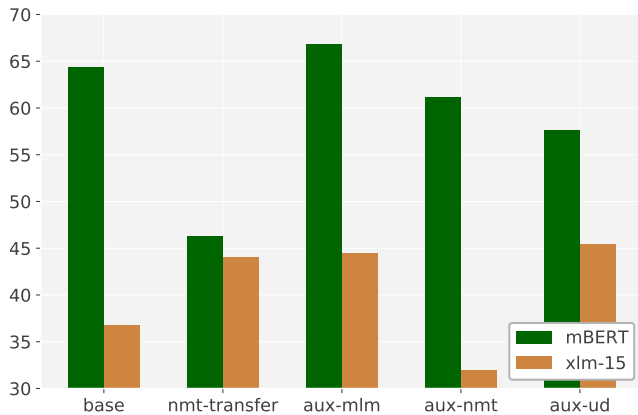
model	Time (minutes)
base	46
nmt-transfer	5,213
aux-mlm	193
aux-nmt	373
aux-ud	79

**Table:** Average minutes to train a model, averaged over all languages and both embeddings. For nmt-transfer we include the training of the NMT model.

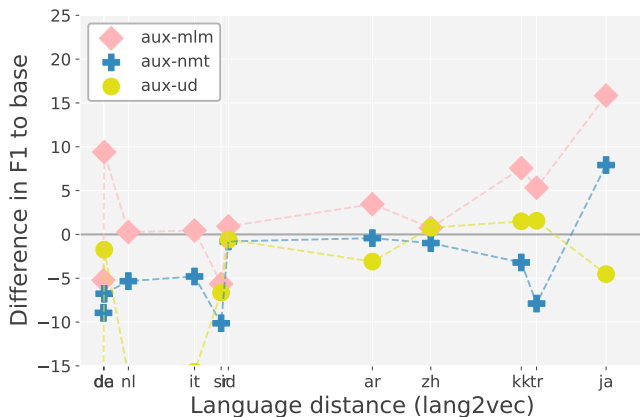
## Results (intents)



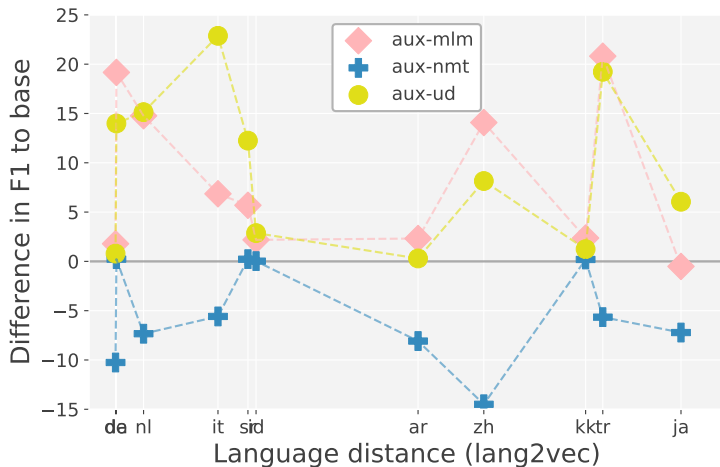
## Results (slots)



# Analysis (mBERT)



## Analysis (XLM15)



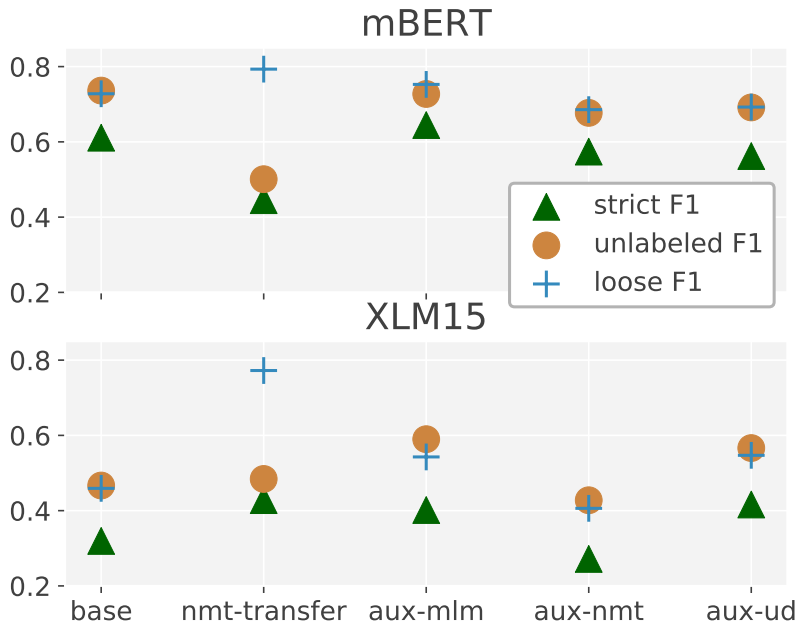


# Analysis

Analysis metrics:

- ▶ Strict F1: normal span F1
- ▶ Unlabeled F1: throw away labels, only keep BIO
- ▶ Loose F1: consider partial matches (with same label) correct

## Analysis



## A closer look at South-Tyrolean

### “Südtirolerisch”, an Austro-Bavarian Dialect in Northern Italy



[https://upload.wikimedia.org/wikipedia/commons/thumb/a/ac/Austro\\_Bavarian\\_Languages-01.png/1024px-Austro\\_Bavarian\\_Languages-01.png](https://upload.wikimedia.org/wikipedia/commons/thumb/a/ac/Austro_Bavarian_Languages-01.png/1024px-Austro_Bavarian_Languages-01.png)

## A closer look at South-Tyrolean

South Tyrolean German dialect (“Südtirolerisch”):

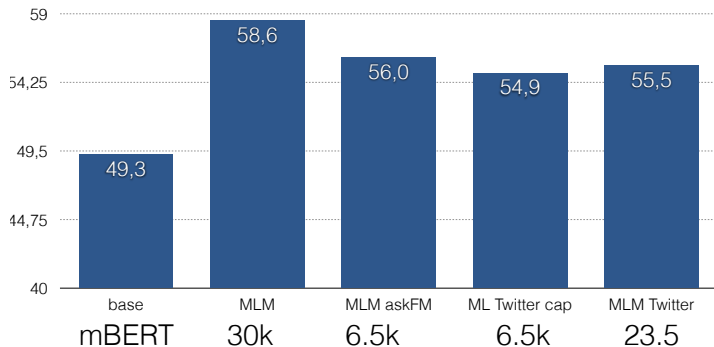
- ▶ South Bavarian (Austro-Bavarian) German vernacular
- ▶ Spoken in the northernmost Italian province (Bozen-Bolzano)
- ▶ No common orthographic standard
- ▶ Influence of Italian and Ladin:  
“Hosch is Patent schun gemocht?”  
patent=driver’s license (IT=Patente DE=Führerschein)

## Samples from the two sources

- **Twitter:**  
ich bin zwar total müde, aber ich muss noch lernen...wtf? :(  
<USER> Stimmt überhaupt et..  
<USER> <USER> Isch a imagine, schaugdo die aktion un ..  
Und Milan het a uan wegn hond krign gimisst  
<USER> Gia sveglio?
  
- **AskFM:**  
wer isch dein bestr freind/in?:)  
Laurii :P  
Wer kennt dich am besten?  
Niamend  
Welche ist die schönste Sprache?  
spanisch  
Geasch geit stodt?:)♥  
Bo keine ahnung..

## A closer look at South-Tyrolean

### De-ST: #sentences for MLM



# Conclusions

- ▶ xSID: More language (varieties), label (domain/topic) variety and high quality annotation
- ▶ For intents, nmt-transfer is hard to beat, and auxiliary tasks are not beneficial
- ▶ MLM performs robust as auxiliary task for slots
- ▶ For unseen languages, UD performs even better (XLM15)