MultiCoNER II: **Multi**lingual **Co**mplex **N**amed **E**ntity **R**ecognition

Asal Shavandi, Chahyon Ku, London Lowmanstone, Josh Spitzer-Resnick

Nov 17, 2022

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Overview

- Problem Definition & Dataset
- Proposed Methods
- Limitations and Ethical Issues
- Results
- Future Direction

MultiCoNER II

Task: for each token, predict whether it is part of a named entity and the label

- 12 languages
  - our focus: **English + French**
  - v1 had 11 (7 in common)
    + multilingual + code mixed

Label space

- **36** categories defined, 33 appear
  - v1 had 6 broad categories
  - v2 tags different than traditional tags
  - medical category added,
    corporation and group merged

2022 (v1)

label	description
CORP /	Corporation
CW	Creative Work
GRP	Group
LOC	Location
PER	Person
PROD	Product

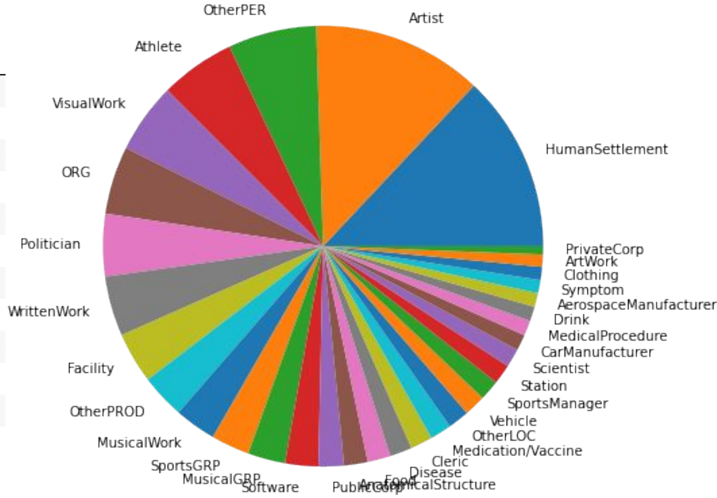
2023 (v2)

category	labels
Creative Work (CW)	ArtWork, MusicalWork, ~~OtherCW~~, Software, VisualWork, WrittenWork
Group (GRP)	AerospaceManufacturer, CarManufacturer, MusicalGRP, ORG, ~~OtherCORP~~, PrivateCORP, PublicCORP, SportsGRP, ~~TechCORP~~
Location (LOC)	Facility, HumanSettlement, OtherLOC, Station
Medical (MED) //	AnatomicalStructure, Disease, MedicalProcedure, Medication/Vaccine, Symptom
Person (PER)	Artist, Athlete, Cleric, OtherPER, Politician, Scientist, SportsManager
Product (PROD)	Clothing, Drink, Food, OtherPROD, Vehicle

Dataset

- Tags are fine grained unlike normal NER
    - could use v2 to augment v1, not the other way around
- All lowercase text - can't use as clue
- Task said would focus on misspellings / typos, not present in dataset
- Created w/ weak supervision - annotations not 100% accurate

language	samples train	dev	entities train	dev
Bangla (bn)	9,708	507	13,223	676
Chinese (zh)	9,759	506	15,228	773
English (en)	16,778	871	25,449	1,296
Farsi (fa)	16,321	855	23,662	1,236
French (fr)	16,548	857	26,377	1,352
German (de)	9,785	512	15,951	841
Hindi (hi)	9,632	514	12,870	684
Italian (it)	16,579	858	26,443	1,398
Portuguese (pt)	16,469	854	24,439	1,292
Spanish (es)	16,453	854	23,907	1,231
Swedish (sv)	16,363	856	25,414	1,391
Ukrainian (uk)	16,429	851	21,956	1,135



Distribution

- ~1.25-1.5 entities per sentence on avg
- imbalanced label classes
- varying length of entities

English: [Artist wes anderson]'s film [VisualWork the grand budapest hotel] opened the festival .
French: l [Politician amiral de coligny] réussit à s y glisser .

Proposed Methods

- 3 Novel Translation Techniques based on MulDA
  - **Full:** Linearized Labeled Sequence Translation
  - **Partial:** Linearized Partial Labeled Sequence Translation
  - **Stabilized:** Non-linearized Unlabeled Sequence Translation
    - **Full**
    - **Partial**
- Substitution using Masked Entity Language Model

MulDA

- Uses Google Cloud API as translation tool
- 3-step translation method
- No linearization during translation

Labeled sentence in the source language:
[PER Jamie Valentine] was born in [LOC London].

1. Translate sentence with placeholders:
src: PER0 was born in LOC1.
tgt: PER0 nació en LOC1.

2. Translate entities with context:
PER0
src: [Jamie Valentine] was born in London.
tgt: [Jamie Valentine] nació en Londres.

LOC1
src: Jamie Valentine was born in [London].
tgt: Jamie Valentine nació en [Londres].

3. Replace placeholders with translated entities:
[PER Jamie Valentine] nació en [LOC Londres].

Full: Linearized Labeled Sequence Translation

What MulDA Did:

Labeled sentence in the source language:
[PER Jamie Valentine] was born in [LOC London].

1. Translate sentence with placeholders:
src: PER0 was born in LOC1.
tgt: PER0 nació en LOC1.

2. Translate entities with context:
PER0
src: [Jamie Valentine] was born in London.
tgt: [Jamie Valentine] nació en Londres.

LOC1
src: Jamie Valentine was born in [London].
tgt: Jamie Valentine nació en [Londres].

3. Replace placeholders with translated entities:
[PER Jamie Valentine] nació en [LOC Londres].

What Full Does (Only Step 3 of MulDA):

src: heron was born in [B-HumanSettlement welwyn] [I-HumanSettlement garden] [I-HumanSettlement city] in 1949.
Sent to Google Translate:
heron was born in [UNK welwyn] [UNK garden] [UNK city] in 1949.
tgt: heron est né à B-HumanSettlement welwyn I-HumanSettlement garden I-HumanSettlement city en 1949.
src: barongo has undertaken children 's projects in [B-HumanSettlement sweden] [B-HumanSettlement south] [I-HumanSettlement africa] and the [B-HumanSettlement united] [I-HumanSettlement states]
Sent to Google Translate:
barongo has undertaken children 's projects in [UNK sweden] [UNK south] [UNK africa] and the [UNK united] [UNK states]
tgt: barongo a entrepris des projets pour les enfants en B-HumanSettlement suède B-HumanSettlement sud I-HumanSettlement afrique et les B-HumanSettlement unis I-HumanSettlement états

Partial: Linearized Partial Labeled Sequence Translation

What MulDA Did:

Labeled sentence in the source language:
[PER Jamie Valentine] was born in [LOC London].

1. Translate sentence with placeholders:
src: PER0 was born in LOC1.
tgt: PER0 nació en LOC1.

2. Translate entities with context:
PER0
src: [Jamie Valentine] was born in London.
tgt: [Jamie Valentine] nació en Londres.

LOC1
src: Jamie Valentine was born in [London].
tgt: Jamie Valentine nació en [Londres].

3. Replace placeholders with translated entities:
[PER Jamie Valentine] nació en [LOC Londres].

What Partial Does (Only Step 3 of MulDA):

src: heron was born in [HumanSettlement welwyn garden city] in 1949.

sent to Google Translate:
heron was born in “B-HumanSettlement welwyn” “I-HumanSettlement garden” “I-HumanSettlement city” in 1949.

tgt: heron est né à “B-HumanSettlement welwyn” “I-HumanSettlement garden” “I-HumanSettlement city” en 1949.

output: heron est né à [HumanSettlement welwyn garden city] en 1949.

Stabilized: Non-linearized Partial Unlabeled Sequence Translation

What MulDA Did:

Labeled sentence in the source language:
[PER Jamie Valentine] was born in [LOC London].

1. Translate sentence with placeholders:
src: PER0 was born in LOC1.
tgt: PER0 nació en LOC1.

2. Translate entities with context:
PER0
src: [Jamie Valentine] was born in London.
tgt: [Jamie Valentine] nació en Londres.

LOC1
src: Jamie Valentine was born in [London].
tgt: Jamie Valentine nació en [Londres].

3. Replace placeholders with translated entities:
[PER Jamie Valentine] nació en [LOC Londres].

What **Stabilized** Does (Only Step 2 of MulDA):

ORG0
src: they teamed from 1989 to 1991 in the [national wrestling alliance] (nwa) and world championship wrestling (wcw) .
tgt: ils ont fait équipe de 1989 à 1991 dans la [national wrestling alliance] (nwa) et le championnat du monde de lutte (www).

PrivateCorp1
src: they teamed from 1989 to 1991 in the national wrestling alliance (nwa) and [world championship wrestling] (wcw) .
tgt: ils ont fait équipe de 1989 à 1991 dans l'alliance nationale de lutte (nwa) et [la lutte du championnat du monde] (wcw).

***The tgts don't match** (translation isn't stable) - skip this example.*

Stabilized: Non-linearized Partial Unlabeled Sequence Translation

What MulDA Did:

Labeled sentence in the source language:
[PER Jamie Valentine] was born in [LOC London].

1. Translate sentence with placeholders:
src: PER0 was born in LOC1.
tgt: PER0 nació en LOC1.

2. Translate entities with context:
PER0
src: [Jamie Valentine] was born in London.
tgt: [Jamie Valentine] nació en Londres.

LOC1
src: Jamie Valentine was born in [London].
tgt: Jamie Valentine nació en [Londres].

3. Replace placeholders with translated entities:
[PER Jamie Valentine] nació en [LOC Londres].

If the translations do match, then combine them to create **orig** (partial) and **trans** (full) examples:

MedicalProcedure0
src: the [polymerase chain reaction] was developed in the 1980s by kary mullis .
tgt: la [réaction en chaîne par polymérase] a été développée dans les années 1980 par kary mullis .

OtherPER1
src: the polymerase chain reaction was developed in the 1980s by [kary mullis] .
tgt: la réaction en chaîne par polymérase a été développée dans les années 1980 par [kary mullis] .

Full Output: la [MedicalProcedure réaction en chaîne par polymérase] a été développée dans les années 1980 par [OtherPER kary mullis] .
Partial Output: la [MedicalProcedure polymerase chain reaction] a été développée dans les années 1980 par [OtherPER kary mullis] .

Limitations

Sentence-to-sentence Translation

- Tags were translated
    - "PROD-Vehicle" -> “PROD-véhicule” and "MED-Symptom" -> “MED-symptôme”
- Sometimes words inside the brackets were translated sometime not
    - city -> city and sweden -> suède
- Words were capitalized after translation
    - james -> James
- Plural s was dropped
- Words inside brackets were swapped
    - [tag1 united] [tag2 states] -> [tag1 unis] [etats tag2]

Cost

- Without batching, takes a long time to augment the training set
- Only the first 500,000 characters are free for translation in Google Cloud

Limitations

Bad Dataset Annotations

id 156a76ec-4fe3-42ed-9bcc-ec550894cf08

he _ _ O
holds _ _ O
wins _ _ O
over _ _ O
tito _ _ B-Athlete
ortiz _ _ I-Athlete
masakatsu _ _ B-Athlete
funaki _ _ I-Athlete
yuki _ _ B-OtherPER
kondo _ _ I-OtherPER
semmy _ _ B-Athlete
schilt _ _ I-Athlete
and _ _ O
minoru _ _ B-Athlete
suzuki _ _ I-Athlete
. _ _ O

id 1de39462-f3b7-41b2-8605-5a8cf6e4e220

ray _ _ B-Athlete
ferraro _ _ I-Athlete
(_ _ O
select _ _ O
games _ _ O
) _ _ O
jamie _ _ B-Athlete
mclennan _ _ I-Athlete
(_ _ O
select _ _ O
games _ _ O
) _ _ O
mike _ _ B-OtherPER
johnson _ _ I-OtherPER
(_ _ O
select _ _ O
games _ _ O
) _ _ O

Ethical Issues

Environmental Risks

- Model training and development costs enormous amounts of **compute time and energy**
  - We use a "base" model with relatively short # of epochs (20)
- We take for granted **the resource** that went into building necessary **infrastructure (Agate!)**
  - **Rare metal mining** for high performance chips
  - **Carbon consumption** for maintaining Internet, servers, etc.

Injecting Bias

- The task organizer, or more generally **NLP researchers, inject their own set of values to the task**
  - Who are we to decide if a person is a politician, an artist, or or just simply "other person"
- Broadly speaking, the elites with technological expertise will **decide and enforce ideas to all users**
  - Loosely tied to censorship in social media

Who Benefits? Who Doesn't?

- Many believe that the research efforts will benefit all humanity equally, but that's not true
  - **Who are we really helping?** Corporates like Amazon? Researchers like DK?
  - **Who are we sacrificing?** 3rd-world mining bases? Maybe multilingual systems help…

Experiments and Results

Experimental Setup

Finetune pretrained XLM-RoBERTa-base + Conditional Random Field (CRF) Classifier

- Adapted from MultiCoNER 1 baseline code
- 20 Epochs
- AdamW Optimizer w/ 1e-5 Learning Rate

Takes about 2.5 hours (per model) to train on MSI's Agate Cluster

Train, Test, and Compare 12 models trained using different datasets.

Experimental Setup Cont'd

Evaluate on Macro Averaged F1 Score (From Competition)

- An NE tag prediction is "correct" if equal to ground truth tag
- Average of F1 score for each tag (without considering number of each tag)

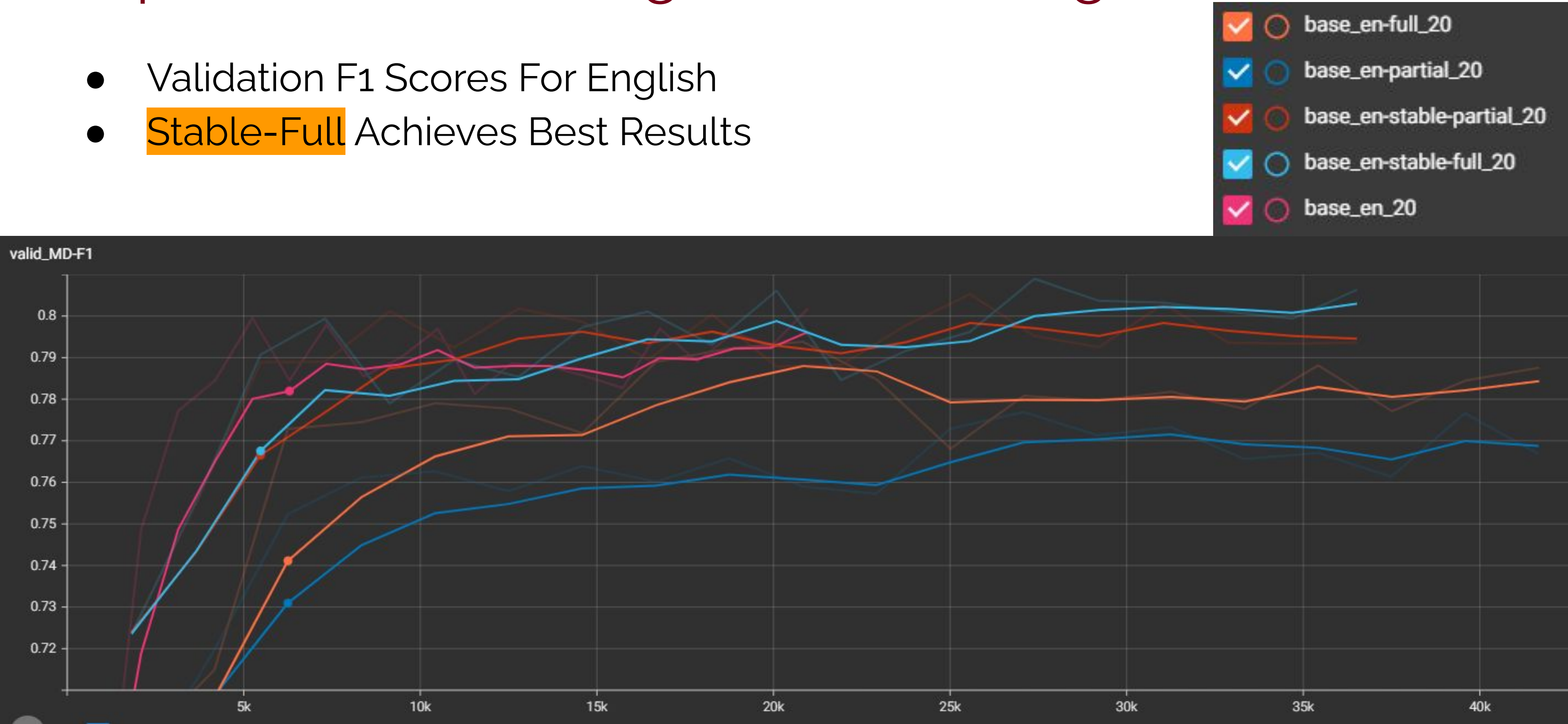
```
# id 8a8e516d-e4ba-42e3-bf62-f2994db69d55   domain=en
it              O              O
stars           O              O
tomokazu        B-Artist       B-Artist
sugita          I-Artist       I-Artist
daisuke         B-OtherPER     B-Artist
sakaguchi       I-OtherPER     I-Artist
rie             B-Artist       B-Artist
kugimiya        I-Artist       I-Artist
among           O              O
others          O              O
.               O              O
```

Experiment 1: Translation

- 1 Language Pair: English - French
    - Train with (English-Train) + Translated (French-Train -> English)
    - Train with (French-Train) + Translated (English-Train -> French)
- Evaluate on English-Dev / French-Dev

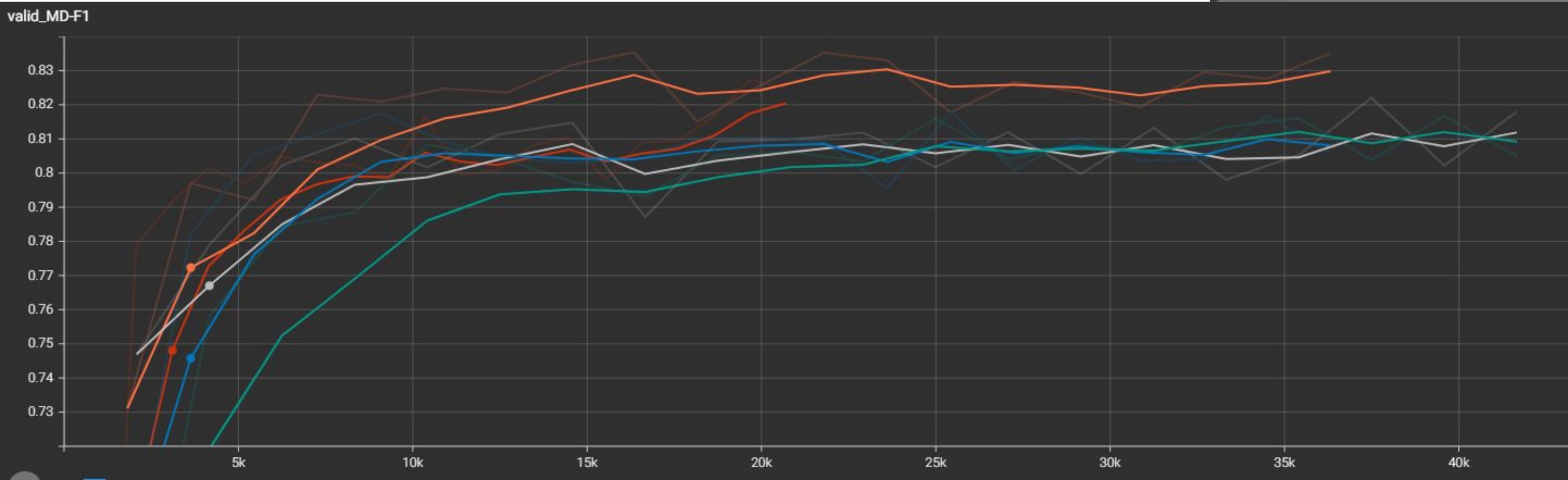
Experiment 1: Training (French → English)

- Validation F1 Scores For English
- Stable-Full Achieves Best Results



Experiment 1: Training (English → French)

- Validation F1 Scores For French
- Stable-Partial Achieves Best Results



Experiment 1: Quantitative Results

	Valid F1	Valid P at Max Valid F1	Valid R at Max Valid F1
EN	0.802	0.800	0.804
EN-Full	0.794	0.796	0.792
EN-Partial	0.777	0.766	0.789
EN-Stable-Orig	0.805	**0.810**	0.801
EN-Stable-Trans	**0.809**	0.795	**0.823**
FR	0.827	0.825	0.829
FR-Full	0.817	0.811	0.822
FR-Partial	0.822	0.828	0.817
FR-Stable-Orig	**0.835**	**0.836**	**0.835**
FR-Stable-Trans	0.818	0.820	0.816

Experiment 1: Error Analysis (Label vs. Pred)

```
# id 5239d808-f300-46ea-aa3b-5093040213a3   domain=en
eli             B-OtherPER      B-OtherPER
lilly           I-OtherPER      I-OtherPER
founder         O               O
president       O               O
of              O               O
pharmaceutical  O               O
company         O               O
eli             B-PublicCorp    B-PublicCorp
lilly           I-PublicCorp    I-PublicCorp
and             I-PublicCorp    I-PublicCorp
company         I-PublicCorp    O
```

Experiment 1: Error Analysis (Label vs. Pred)

```
# id d7d47dfc-7e5d-48e8-9390-019a3e9476c1	domain=en
christoph	B-OtherPER	B-OtherPER
haberland	I-OtherPER	I-OtherPER
designed	O	O
a	O	O
new	O	O
marble	O	O
pulpit	B-OtherPROD	O
for	O	O
the	O	O
church	O	O
which	O	O
was	O	O
built	O	O
in	O	O
italy	B-HumanSettlement	B-HumanSettlement
in	O	O
1793	O	O
.	O	O
```

Experiment 1: Error Analysis (Label vs. Pred)

```
# id 7051b30d-a8e5-4bc3-a83a-eacc863f94d0   domain=en
he                            O                             O
was                           O                             O
succeeded                     O                             O
as                            O                             O
chancellor                    O                             O
by                            O                             O
sir                           B-OtherPER                    B-Politician
frank                         I-OtherPER                    I-Politician
kitto                         I-OtherPER                    I-Politician
.                             O                             O


# id 6c63b565-b3d4-4c2d-b4a7-6a00460f0d32   domain=en
it                            O                             O
was                           O                             O
described                     O                             O
by                            O                             O
edward                        B-OtherPER                    B-OtherPER
meyrick                       I-OtherPER                    I-OtherPER
in                            O                             O
1915                          O                             O
.                             O                             O
```

Experiment 2: Masked Entity Language Model

- Inspired by Masked Language Modelling from BERT
  - Finetune XLM-RoBERTa-base + Masked Language Modelling Head
  - Mask named entities and learn to recover them from context
- Masks are greedily recovered one-by-one

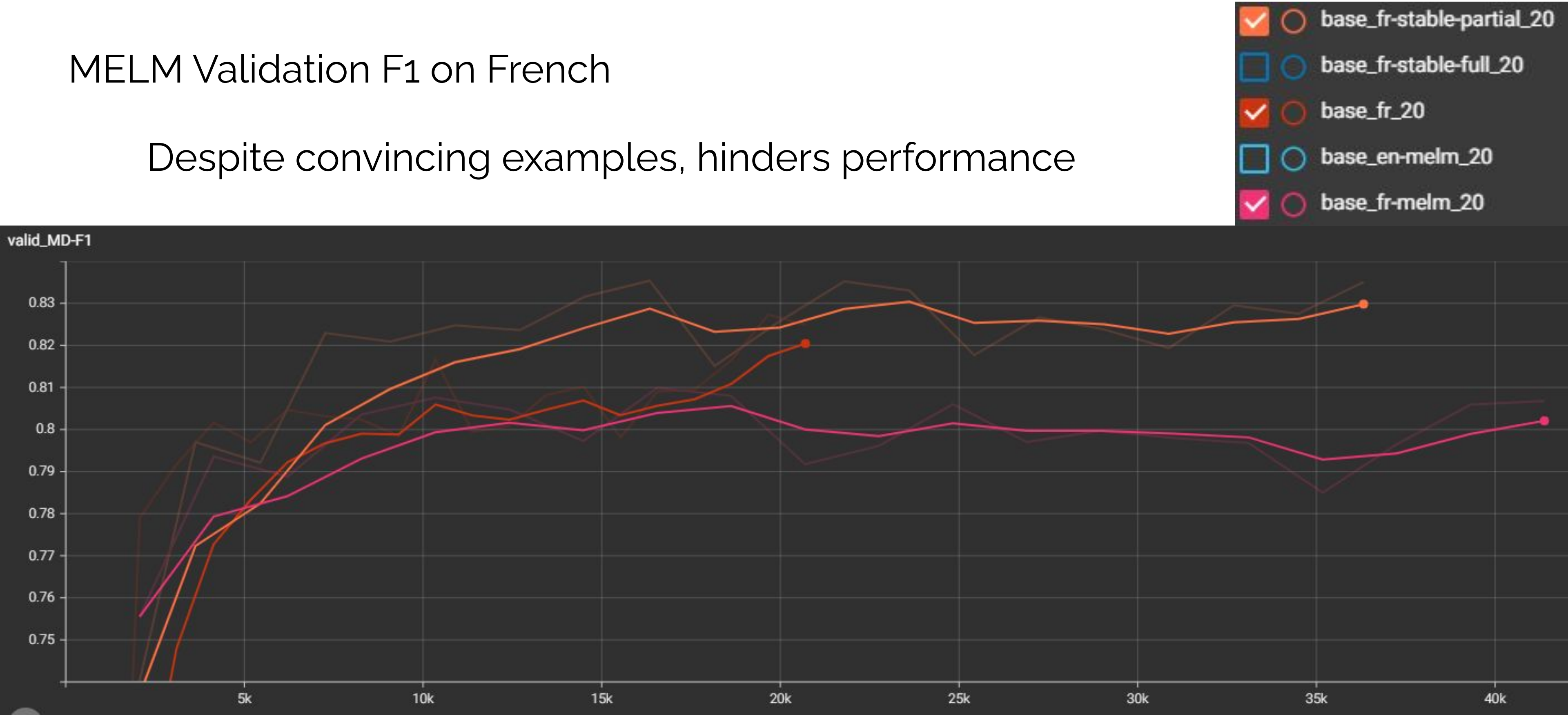

```
"source": "the ideas were introduced by B-OtherPER william I-OtherPER burnside at the end of the nineteenth century.",
"masked": "the ideas were introduced by B-OtherPER <mask><mask> I-OtherPER <mask><mask> at the end of the nineteenth century.",
"sample": "the ideas were introduced by B-OtherPER robert I-OtherPER walker at the end of the nineteenth century."

"source": "the alphabet was reworked by B-OtherPER sarsen I-OtherPER amanzholov and accepted in its current form in 1940.",
"masked": "the alphabet was reworked by B-OtherPER <mask><mask> I-OtherPER <mask><mask><mask><mask> and accepted in its current form in 1940.",
"sample": "the alphabet was reworked by B-OtherPER george I-OtherPER ljubvinsky and accepted in its current form in 1940."
"source": "he is voiced by B-Artist mitsuko I-Artist horie in the first anime and by B-OtherPER motoko I-OtherPER kumai in the second.",
"masked": "he is voiced by B-Artist <mask><mask><mask> I-Artist <mask><mask> in the first anime and by B-OtherPER <mask><mask> I-OtherPER <mask><mask> in the second.",
"sample": "he is voiced by B-Artist s\u014dsuke I-Artist shinichi in the first anime and by B-OtherPER mitsumi I-OtherPER yoshi in the second."
"source": "jointly with B-OtherPER robert I-OtherPER gompf he discovered four dimensional models of space time topology.",
"masked": "jointly with B-OtherPER <mask><mask> I-OtherPER <mask><mask> he discovered four dimensional models of space time topology.",
"sample": "jointly with B-OtherPER george I-OtherPER wilson he discovered four dimensional models of space time topology."
```

Experiment 2: Quantitative Results

MELM Validation F1 on French

Despite convincing examples, hinders performance



Future Direction

- Improving Translation
  - Make use of more
  - Exact copy of MulDA
- Improving MELM
  - Not-greedy sampling
  - Longer training
- Comparative analysis
  - T-SNE plot of samples obtained from different methods
  - Similarity vs. Difficulty plot

References

- https://multiconer.github.io/
- SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition (MultiCoNER) (Malmasi et al., SemEval 2022)
- A Survey of Data Augmentation Approaches for NLP (Feng et al., ACL Findings 2021)
- MulDA: A Multilingual Data Augmentation Framework for Low-Resource Cross-Lingual NER (Liu et al., ACL 2021)
- MELM: Data Augmentation with Masked Entity Language Modeling for Low-Resource NER (Zhou et al., ACL 2022)
- DAGA: Data Augmentation with a Generation Approach for Low-resource Tagging Tasks (Ding et al., EMNLP 2020)
- On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (Bender et al., FAccT 2021)