Transformers on Clause-Level Morphology

KUIS AI Submission for the 1st Shared Task on Multilingual Clause-level Morphology

Emre Can Acikgoz
December 8th, 2022

KUIS | KOÇ UNIVERSITY | MRL

The 2nd Workshop on Multilingual Representation Learning

Personal

2022-2024

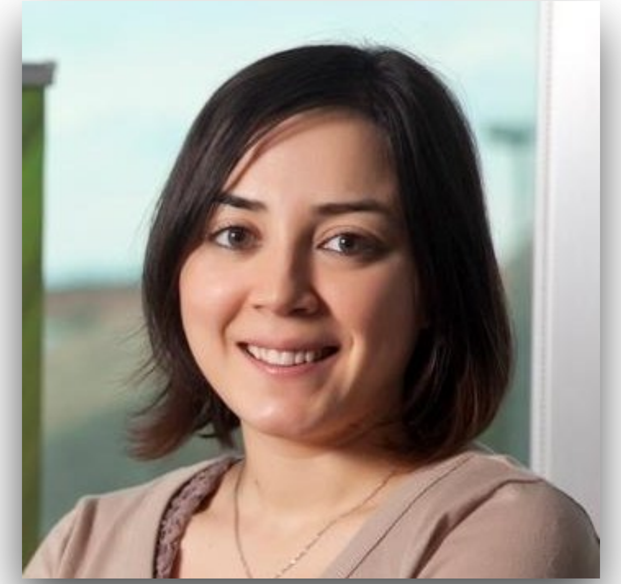
MSc at Koc University in CS and Fellow at KUIS AI Center
Advisor: Deniz Yuret
Topic: LLMs, Multimodal Learning, Grounded Language Learning

2018-2022

BSc at Koc University major in EEE.
Undergraduate Advisor: Deniz Yuret
Topic: Supervised/Unsupervised Morphological Analysis

Tilek Chubakov,
VS

Muge Kural,
PhD



Gozde Gul Sahin,
Asst. Prof

Deniz Yuret,
Prof

Motivation

The Reason Why?

- Language generalization problems
- Methods in morphological tasks sometimes be “old-fashioned”
- The era of Language Models
- Multilingual/Monolingual models

Shared-Task

2022 (MRL): Multilingual Clause-level Morphology

- Task1: Inflection

  give + IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) → I will give him to her

- Task2: Reinflection

  I will give him to her
  + IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) → We don't give you to them
  + IND;PRS;NOM(1,PL);ACC(2);DAT(3,PL);NEG

- Task3: Analysis

  I will give him to her → give + IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)

- Languages: Ger, Eng, Fra, Heb/Heb-unvoc, Rus, Swa, Spa, Tur

[1] Omer Goldman and Reut Tsarfaty. 2022. Morphology without borders: Clause-level morphological annotation.

Shared-Task

Task1: Inflection

- **Task:** Inflection

  give + IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) → I will give him to her

- **Metric:** Edit Distance (ED)

- **Method:** Vanilla Transformer **[2]** + Data Hallucination **[1]**

- **Tricks:** Batch Size 400 **[3]**, layer normalization before self-attention and feed-forward layers **[3]**

System	Inflection
Transformer Baseline	3.278
mT5 Baseline	2.577
KUIS AI	**0.292**

Figure: Task1 Averaged Results

[1] Antonios Anastasopoulos and Graham Neubig. 2019. Pushing the limits of low-resource morphological inflection.
[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.
[3] Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction.

Shared-Task

Data Hallucination

- Good choice for low-resource languages.
- Add noise to the stem parts.
- Increase the training set with the hallucinated samples.

celebrate + IND;PRS;NOM(1,PL);NEG;ACC(3,SG,MASC) → we don't celebrate him
cjuexua te + IND;PRS;NOM(1,PL)NEG;ACC(3,SG,MASC) → we don't cjuexuate him.

Example of Hallucinated Data (English)

cevap vermek + NEC;PST;NOM(3,SG);NEG;Q;DAT(3,PL) → onlara cevap vermemeli miydi?
cevCp vDOme k + NEC;PST;NOM(3,SG);NEG;Q;DAT(3,PL) → onlara cevCp vDOmemeli miydi?

Example of Hallucinated Data (Turkish)

[1] Antonios Anastasopoulos and Graham Neubig. 2019. Pushing the limits of low-resource morphological inflection.

Shared-Task

Task2: Reinflection

- **Task:** Reinflection

  I will give him to her
  ~~+ IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)~~ → We don't give you to them
  + IND;PRS;NOM(1,PL);ACC(2);DAT(3,PL);NEG

- **Metric:** Edit Distance (ED)

- **Method:** Vanilla Transformer **[1]**

- **Tricks:** Batch Size 400 **[2]**, layer normalization before self-attention and feed-forward layers **[2]**. We didn’t use the input clause features.

System	Reinflection
Transformer Baseline	4.642
mT5 Baseline	2.826
KUIS AI	**0.705**

Figure: Task2 Averaged Results

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.
[2] Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction.

Shared-Task

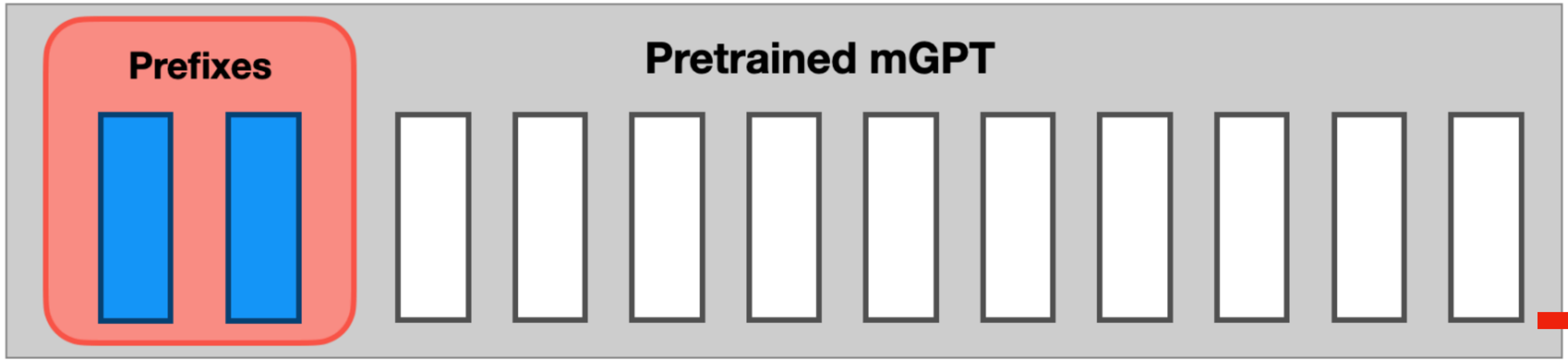
Task3: Analysis

- **Task:** Analysis

  I will give him to her → give + IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)

- **Metric:** F1 Score

- **Method:** mGPT-based prefix tuning **[1]**, **[2]**



I will give him to her. **<SEP> give IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)**
Input (clauses) **Output (lemma and features)**

Figure: Prefix Tuning example for Task3

System	Analysis
Transformer Baseline	80.00
mT5 Baseline	84.50
KUIS AI	**94.17**

Figure: Task3 Averaged Results

[1] Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation.
[2] Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. mgpt: Few-shot learners go multilingual.

Results

Submitted Detailed Results Table

Number of Tokens:

- Eng: ~52B
- Deu: ~50B
- Spa: ~30B
- Heb: ~0.69B

	Task1: Inflection			Task2: Reinflection			Task3: Analysis		
Model	**Transformer + D.A.**			**Transformer**			**Prefix Tuning**		
Metrics	F1↑	EM↑	ED↓	F1↑	EM↑	ED↓	F1↑	EM↑	ED↓
Deu	97.71	91.80	0.241	92.40	66.50	0.788	95.89	83.40	0.991
Eng	98.02	88.90	0.221	95.42	72.30	0.477	99.61	98.50	0.064
Fra	98.59	93.20	0.124	92.64	68.30	0.758	95.63	81.90	0.933
Heb	97.73	89.80	0.550	94.00	83.30	0.796	92.84	73.50	1.322
Heb-Unvoc	97.96	94.20	0.113	86.70	57.70	1.002	82.09	36.20	2.044
Rus	97.57	87.70	0.828	97.29	84.90	0.854	97.51	88.60	3.252
Swa	99.72	99.61	0.019	92.05	84.47	0.182	90.51	62.63	3.114
Spa	98.79	92.00	0.199	96.42	77.60	0.480	98.11	89.40	0.560
Tur	97.50	89.80	0.333	95.36	84.70	0.593	95.36	84.70	0.593
Average	98.18	91.89	**0.292**	93.14	74.72	**0.705**	**94.17**	77.65	1.430

Figure: D.A. indicates Data Augmentation

Conclusion

Summary and Future Work

- **Summary**
  - No single method achieves best results in all tasks.
  - Recent NLG methods provide promising results on morphological tasks.
  - Data hallucination, multilingual models, and lightweight tuning methods are the game changers.
- **Future work**
  - Prefix-Tuning in all types of architectures (autoencoding, autoregressive, seq2seq).
  - Hallucination for reinflection and analysis tasks.
- **Github Code:** https://github.com/emrecanacikgoz/mrl2022

Q/A for 5 min. ?

Appendix

Transformer Architecture

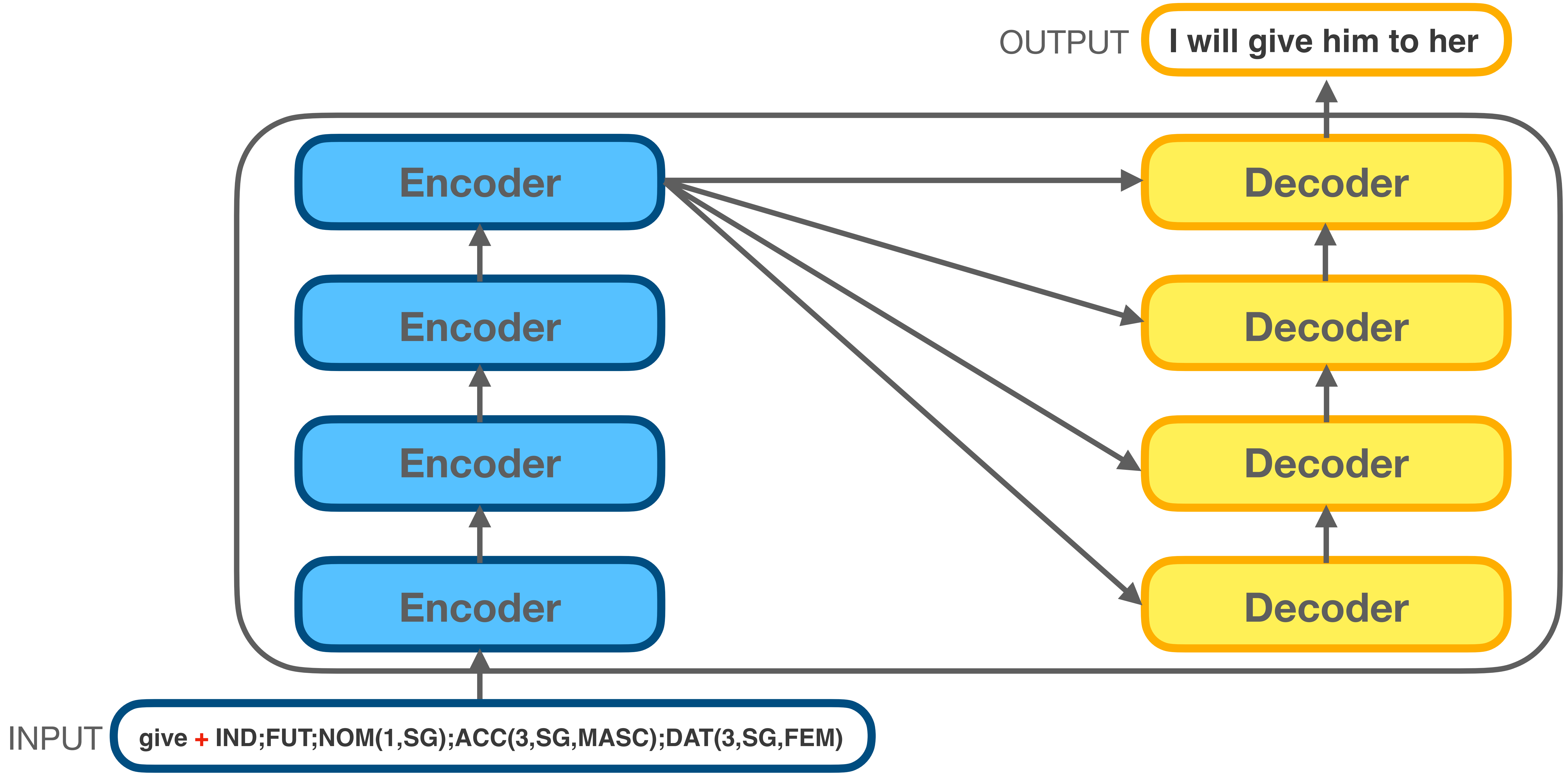


Figure1: Our Vanilla Transformer Architecture

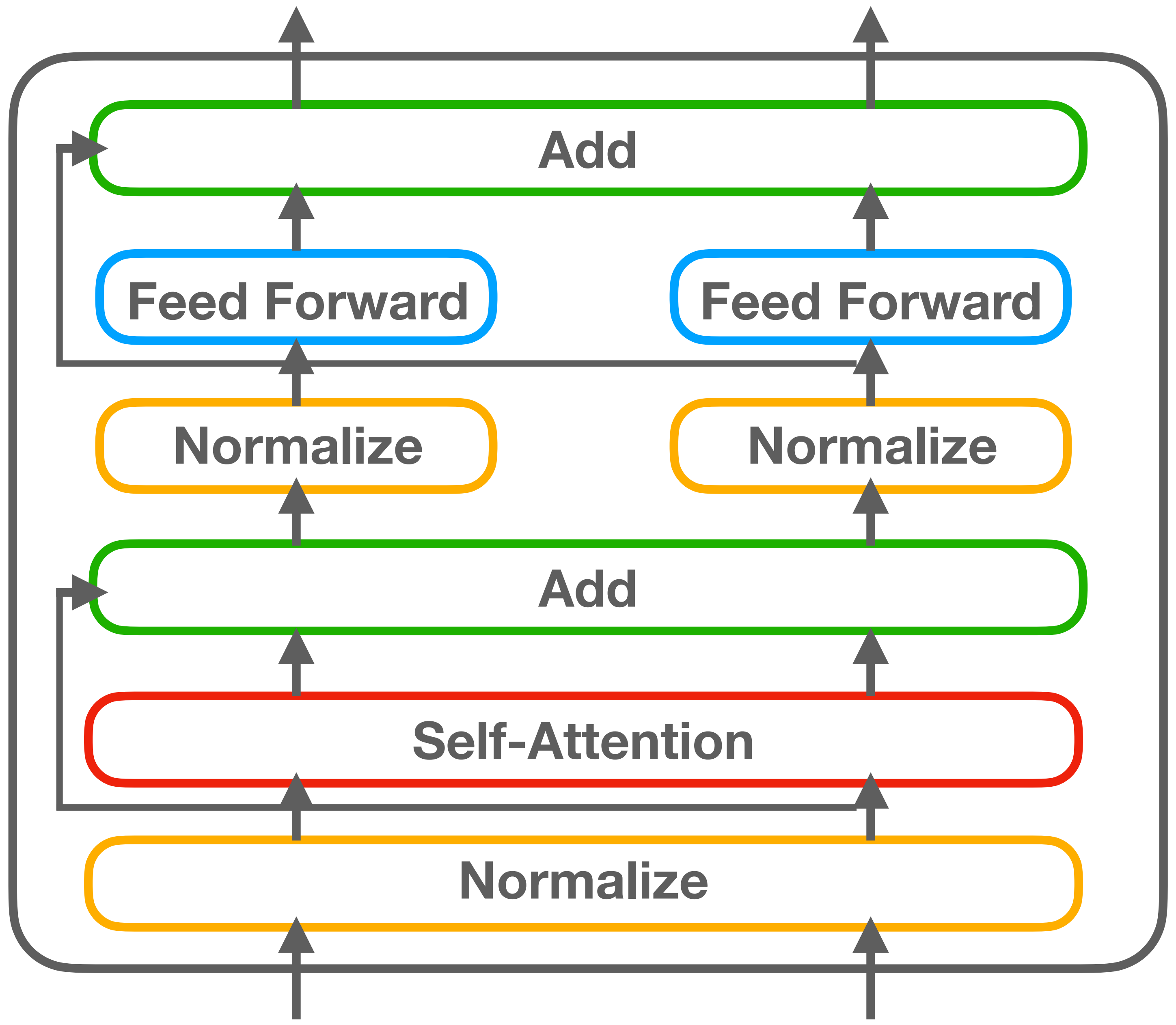


Figure2: Layer Normalization Trick **[2]**

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

[2] Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction.

Appendix

Prefix-Tuning

Prefix | Source Input | Target

P1, P2 | I will give him to her. | <SEP> give IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)

[Prefix; x; y]

Figure: Auto-regressive Prefix-Tuning set-up

[1] Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation.
[2] Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. mgpt: Few-shot learners go multilingual.

Appendix

mGPT Language Corpus

Afrikaans, Azerbaijani, Belarusian, Bengali, Chuvash, German, English, Basque, Finnish, Hebrew (modern), Hungarian, Indonesian, Japanese, Kazakh, Kirghiz, Kyrgyz, Latvian, Mongolian, Malay, Dutch, Polish, Romanian, Moldavan, Yakut, Swahili, Telugu, Thai, Turkish, Tuvinian, Urdu, Vietnamese, Yoruba, Arabic, Bashkir, Bulgarian, Buriat, Danish, Greek, Modern, Spanish; Castilian, Persian, French, Hindi, Armenian, Italian, Georgian, Korean, Lithuanian, Malayalam, Marathi, Burmese, Ossetian, Ossetic, Portuguese, Russian, Swedish, Tamil, Tajik, Turkmen, Tatar, Ukrainian, Uzbek, Kalmyk, Chinese

Figure: 60 different languages

[2] Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. mgpt: Few-shot learners go multilingual.

Appendix

mGPT Language Corpus

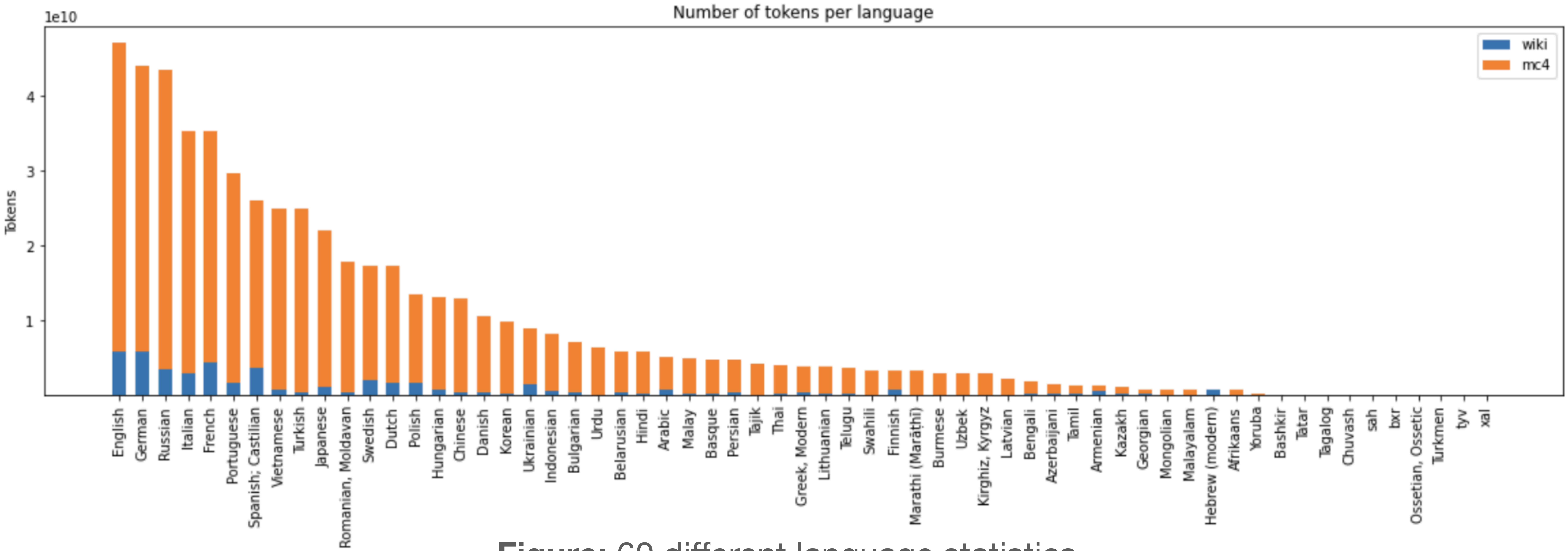


Figure: 60 different language statistics

[2] Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. mgpt: Few-shot learners go multilingual.

Appendix

Extra Results

Inflection	Deu		Eng		Fra		Heb		Rus		Tur	
Models	EM	ED	EM	ED	EM	ED	EM	ED	EM	ED	EM	ED
T	80.8%	0.645	92.1%	0.129	92.4%	0.270	92.5%	0.488	92.8%	0.763	95.2%	0.083
T + H(N=1000)	89.8%	0.467	96.6%	0.132	94.0%	0.273	93.6%	0.289	93.6%	0.709	**99.4%**	**0.010**
T + H(N=5000)	**92.0%**	**0.422**	**97.0%**	**0.113**	**95.3%**	**0.121**	**96.0%**	**0.112**	**93.8%**	0.907	99.3%	0.018
T + H(N=10000)	89.7%	0.474	96.8%	0.130	94.6%	0.159	95.2%	0.181	93.7%	**0.899**	98.7%	0.270

Figure 1: Results for varying number of hallucinated data for Task1

	Eng	
Models	**EM**	**ED**
GPT-2	83.5% ± 0.007	0.660 ± 0.026
T5	90.4% ± 0.016	0.316 ± 0.073
mGPT	93.8% ± 0.011	0.121 ± 0.070

Figure 2: Results for monolingual vs. multilingual for Task1

[1] Antonios Anastasopoulos and Graham Neubig. 2019. Pushing the limits of low-resource morphological inflection.
[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.
[3] Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction.