Filling the Gap: Decoding of Word Embeddings for Generation of Coherent New Words

Safa AlSaidi, Amandine Decker, Stephanie Monteiro

M2 – Software Project

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

Reminder of our aim

- Apply decoder to the regression task (solving analogies)

$$A : B :: C : X \xrightarrow{X=?} A : B :: C : D$$
$$e.g.\ \ star : stars :: cat : X \rightarrow cats$$

- Current output: vectors ($\neq$ word)
- Aim: transform these vectors into words

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

Results

A	:	B	::	C	:	expected D	output D
schieben	:	schiebend	::	vergeben	:	vergebend	antisemitischer
schieben	:	vergeben	::	schiebend	:	vergebend	beschäftigter
schiebend	:	vergebend	::	schieben	:	vergeben	antisemitischer
schiebend	:	schieben	::	vergebend	:	vergeben	antisemitischer
vergebend	:	vergeben	::	schiebend	:	schieben	erschiesen
vergebend	:	schiebend	::	vergeben	:	schieben	angeregterem
vergeben	:	schieben	::	vergebend	:	schiebend	antisemitischer

Output:

- 30 different words for German
- 22 for Arabic

Discussion

1. Redundant outputs: decoder stuck on a few words
2. Possible solution to implement: VAE
3. Lack of time

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

New objective

- Regression task based on transfer

$$A : B :: C : X \xrightarrow{X=?} A : B :: C : D$$
$$e.g.\ \ dog : dogs :: chat : X \rightarrow chats$$

- Input: A and B in language 1, C in language 2
- Output: D in language 2
- Same transformation for A, B and C, D

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

Regression method

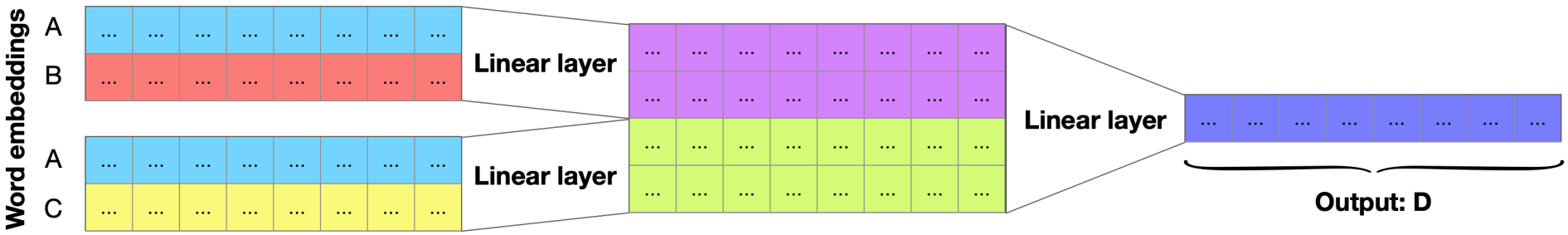
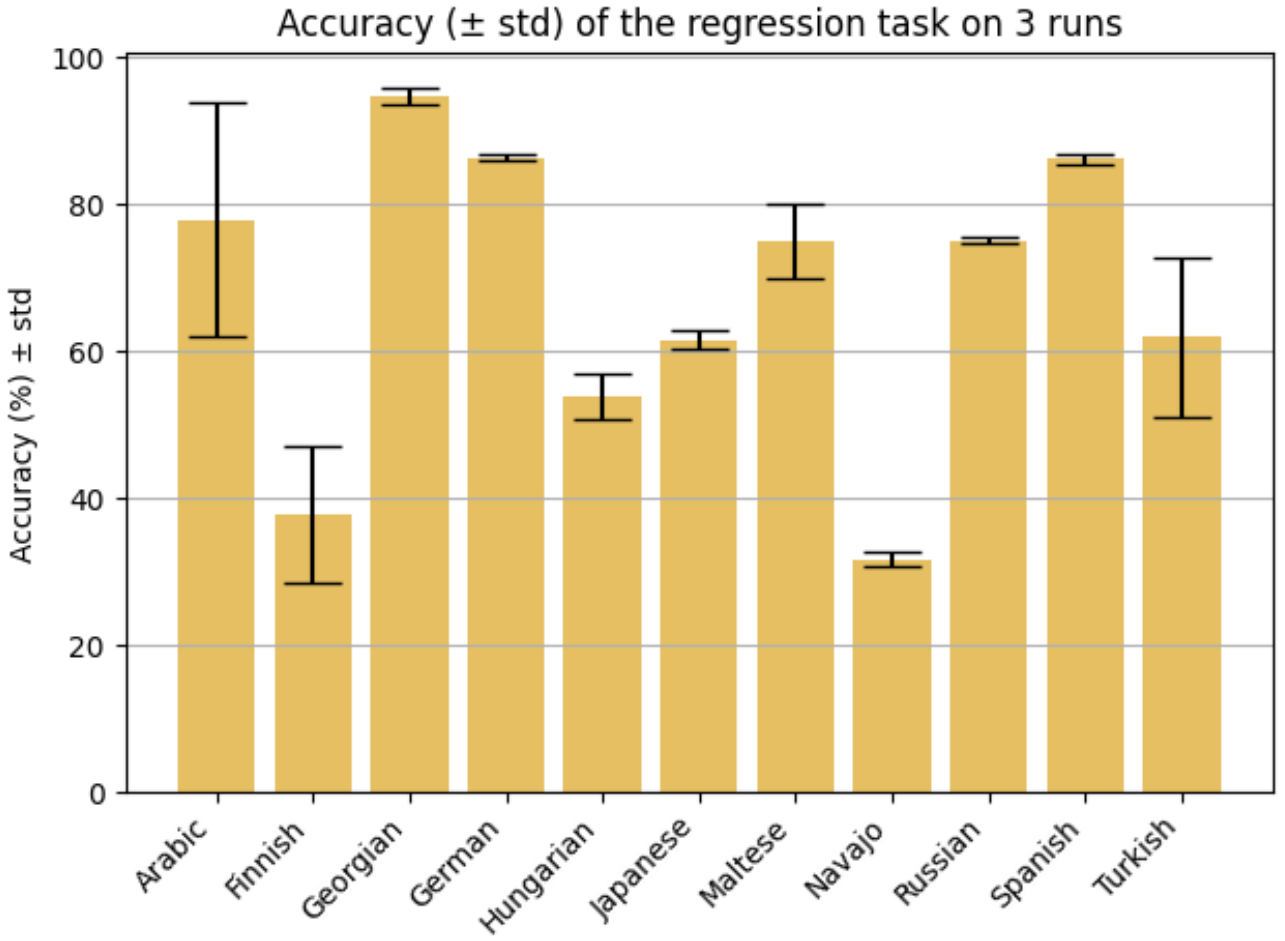


Figure: Structure of the analogy solver model

$\text{word}_D \leftarrow$ word with closest embedding to D

Model inspired by [Lim et al., 2019]

Previous results



Results from [Alsaidi et al., 2021]

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

Our approach

1. Find good pairs of languages
   - languages that have same transformations (i.e. same features)
   - languages that transfer well from one to another (based on previous experiments)
2. Build analogies only when A,B in language 1 and C,D in language 2 have the same features
3. The unknown word is either C or D

Language pairs

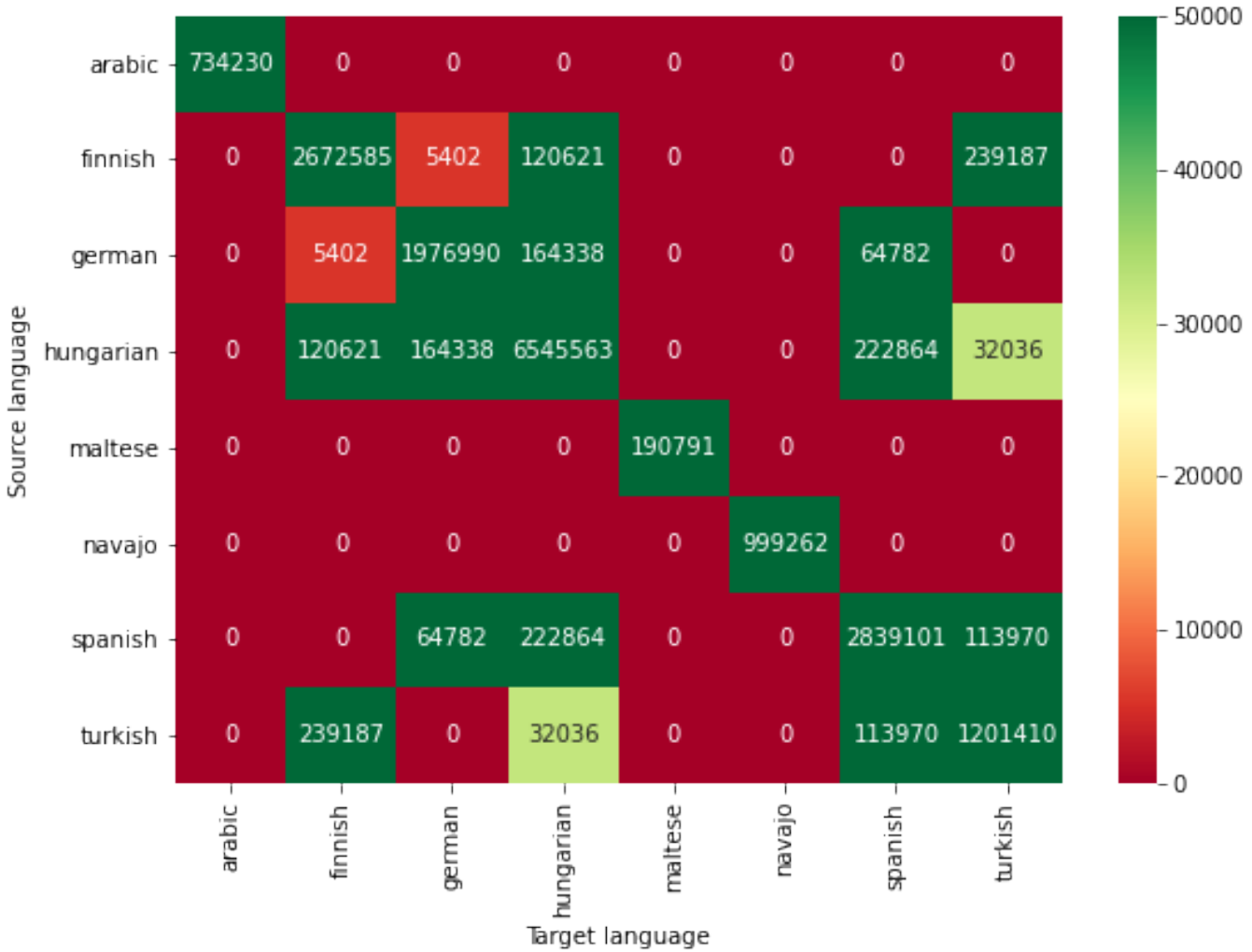


Figure: Number of possible analogies per language pair

Language pairs

Source language	Target language	Nb different features	Transfer accuracy (%)	(source, target) nb features
Hungarian	German	3	**90**	(86, 98)
Hungarian	Finnish	19	**84**	(86, 95)
Turkish	Finnish	6	**81**	(187, 95)
German	Hungarian	3	74	(98, 86)
Hungarian	Spanish	4	72	(86, 84)
Finnish	Turkish	6	70	(95, 187)

Table: Information on the language pairs (ranked by transfer accuracy).

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

First trial

	hungarian, german	turkish, finnish	hungarian, finnish
Cosine similarity	**58.9**	**39.5**	**18.9**
Euclidean distance	57.7	39.1	16.8

Table: Accuracy for the regression task on the three (source, target) language pairs.

Discussion

- Few common features
- Performance seems to decrease when number of features increases (to confirm)
- Fine-tuning the embeddings

1 Reminder

2 Issues

3 New objective

4 Regression task

5 Our approach

6 Preliminary results

7 Future work

Future work

- ~~22 Nov - Regression model + decoder / Variational auto-encoder~~
- ~~10 Dec - Qualitative analysis / Regression model + decoder~~
- 14 Jan - Application docker & webpage
- 3 Feb - Report

Future work

- ~~22 Nov - Regression model + decoder / Variational auto-encoder~~
- ~~10 Dec - Qualitative analysis / Regression model + decoder~~
- 14 Jan - Regression + Application docker & webpage
- 3 Feb - Report

شكرا جزيلا

Thank you

Merci

អរគុណ

Obrigado

References I

Alsaidi, S., Decker, A., Marquer, E., Murena, P.-A., and Couceiro, M. (2021). Tackling morphological analogies using deep learning – extended version.

Lim, S., Prade, H., and Richard, G. (2019).
Solving word analogies: A machine learning perspective.
In Kern-Isberner, G. and Ognjanovic, Z., editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 15th European Conference, ECSQARU 2019, Belgrade, Serbia, September 18-20, 2019, Proceedings*, volume 11726 of *Lecture Notes in Computer Science*, pages 238–250. Springer.