

# 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 State of the project
- 3 Results
- 4 Software

# Reminder of our aim

- 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 State of the project**
- 3 Results
- 4 Software

# Since the last time ...

- ❶ Run the experiments again
  - ▶ More runs
  - ▶ Typo in our code: two regression models now
- ❷ Improve our software
- ❸ Write the report

# Analogy solver model

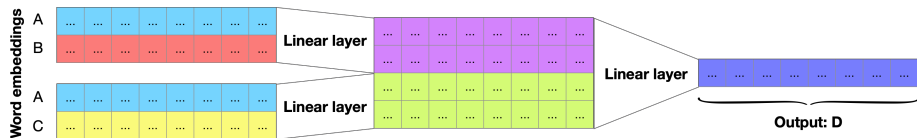


Figure: Analogy solver model

- 1 Reminder
- 2 State of the project
- 3 Results**
- 4 Software

# Regression model

<b>Language</b>	<b>ANNr (previous)</b> (mean $\pm$ std.)	<b>old model</b> (mean $\pm$ std.)	<b>new model</b> (mean $\pm$ std.)
Arabic	77.97 $\pm$ 16.03	59.14 $\pm$ 1.76	61.13 $\pm$ 0.83
Finnish	37.78 $\pm$ 9.28	76.61 $\pm$ 1.15	76.46 $\pm$ 1.58
Georgian	94.66 $\pm$ 1.13	85.51 $\pm$ 2.00	84.67 $\pm$ 2.78
German	86.38 $\pm$ 0.45	89.26 $\pm$ 0.51	88.70 $\pm$ 0.58
Hungarian	53.83 $\pm$ 3.12	78.49 $\pm$ 0.65	78.72 $\pm$ 0.53
Maltese	75.00 $\pm$ 5.08	77.37 $\pm$ 2.27	78.04 $\pm$ 1.44
Navajo	31.74 $\pm$ 0.90	46.14 $\pm$ 0.54	45.74 $\pm$ 0.99
Russian	75.15 $\pm$ 0.44	72.51 $\pm$ 0.46	72.23 $\pm$ 0.44
Spanish	86.27 $\pm$ 0.71	91.18 $\pm$ 0.51	91.72 $\pm$ 0.43
Turkish	61.95 $\pm$ 10.86	80.37 $\pm$ 0.82	80.37 $\pm$ 1.00
Japanese	61.60 $\pm$ 1.33	74.75 $\pm$ 1.33	72.58 $\pm$ 2.47

Table: Accuracy (in %) of 10 runs of the regression models (3 runs for previous results).



# Omnilingual model

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	61.85±1.79	3.22±1.73	28.58±2.69	/	55.23±3.85
German	3.15±1.47	64.38±1.27	<b>66.91±4.49</b>	62.07±3.58	/
Hungarian	36.26±4.52	55.25±1.47	<b>73.33±1.31</b>	<b>78.36±1.33</b>	<b>32.00±3.03</b>
Spanish	/	61.16±2.54	<b>74.05±1.77</b>	69.38±1.65	70.67±4.03
Turkish	54.12±1.48	/	25.38±3.94	65.72±6.09	<b>52.23±1.09</b>

Table: Accuracy (in %) of 10 runs of the new omnilingual regression model

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	60.95±1.75	3.34±1.16	30.34±2.02	/	53.82±2.74
German	3.05±1.43	63.78±0.88	60.99±4.86	62.65±3.47	/
Hungarian	35.06±4.49	56.35±1.96	71.69±1.84	70.61±7.90	27.10±3.17
Spanish	/	61.90±2.42	67.32±5.05	67.90±1.97	66.12±6.86
Turkish	54.00±2.41	/	24.19±2.21	64.94±5.61	50.39±1.64

Table: Accuracy (in %) of 10 runs of the old omnilingual regression model

- 1 Reminder
- 2 State of the project
- 3 Results
- 4 Software**

# Last time

The screenshot shows a web application interface for generating word analogies. It features a light purple background with white input fields and purple buttons. At the top, there are two input fields for 'Source language' and 'Target language', both containing the word 'Hungarian'. To the right of these fields is the text 'Bilingual analogies?'. Below these fields is a large purple button labeled 'Generate an example'. To the right of this button is the text 'What is a valid example?'. Below the button, there are four input fields arranged horizontally. The first three fields are white and contain the words 'Hungarian', 'Paris', and 'Berlin' respectively. The fourth field is greyed out. Between the first and second fields is a colon ':', and between the second and third fields is a double colon '::'. Below the input fields, there are two purple buttons: 'Get closest result' and 'Shuffle the words'. To the left of the 'Get closest result' button is the text 'How is an analogy solved?', and to the right of the 'Shuffle the words' button is the text 'Does the order matter?'.

Source language: Hungarian Target language: Hungarian Bilingual analogies?

Generate an example What is a valid example?

How is an analogy solved? Get closest result Shuffle the words Does the order matter?

Figure: Preview of our previous software

# No answer given

## DANNa

[About Us](#) [Our app](#) [More about analogies](#) [More about morphology](#)

### Didactics for Analogies and Neural Networks

This website shows you several features of analogies. You can click on ⓘ to get more information.  
 Click on *Generate an example* to get a valid example, you can then solve it manually or see what our neural network proposes with *Get closest result*.  
 Eventually, *Get the right answer* will give you the right answer if yours was wrong.

ⓘ Source language: Spanish
Target language: German
Advanced options ▼

Generate an example ⓘ

Spanish to German: Verb, present tense

olvidar

:

olvidando

::

kremieren

:

ⓘ Get closest result
Get the right answer
Shuffle the words ⓘ

Please write an answer.

Figure: Preview of our software

# Valid analogy

## DANNa

[About Us](#) [Our app](#) [More about analogies](#) [More about morphology](#)

### Didactics for Analogies and Neural Networks

This website shows you several features of analogies. You can click on ⓘ to get more information.  
 Click on *Generate an example* to get a valid example, you can then solve it manually or see what our neural network proposes with *Get closest result*.  
 Eventually, *Get the right answer* will give you the right answer if yours was wrong.

ⓘ Source language: Hungarian
 ⓘ Target language: Any language
 ⓘ Advanced options ⌵

ⓘ Transformation: Noun, accusative case, plural ⓘ

Generate an example ⓘ

Hungarian to Finnish: Noun, accusative case, plural

adomány

 ⓘ

adományokat

 ⓘ

kabardi

 ⓘ

kabardit

ⓘ Get closest result ⓘ
 ⓘ Get the right answer ⓘ
 ⓘ Shuffle the words ⓘ

Figure: Preview of our software

# Invalid analogy

## DANNa

[About Us](#) [Our app](#) [More about analogies](#) [More about morphology](#)

### Didactics for Analogies and Neural Networks

This website shows you several features of analogies. You can click on ⓘ to get more information.  
 Click on *Generate an example* to get a valid example, you can then solve it manually or see what our neural network proposes with *Get closest result*.  
 Eventually, *Get the right answer* will give you the right answer if yours was wrong.

ⓘ Source language: Hungarian Target language: Turkish Advanced options ▼

Generate an example ⓘ

**Hungarian to Turkish: Noun, nominative case, plural**

vánkös

:

vánkösök

::

silindir

:

sionisteista

ⓘ Get closest result ⓘ
 ⓘ Get the right answer ⓘ
 ⓘ Shuffle the words ⓘ

This result is not the expected one (silindirler).

Figure: Preview of our software

شكرا جزيلا

Thank you

Merci

អរគុណ

Obrigado

# Monolingual VS Bilingual (old)

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	42.06±2.82	80.13±1.37	/	81.53±1.57
German	94.51±0.60	/	71.94±2.84	71.94±2.84	/
Hungarian	42.99±4.69	84.22±3.33	/	84.22±3.33	40.75±1.89
Spanish	/	93.75±0.38	93.75±0.38	/	96.40±0.42
Turkish	67.43±1.06	/	<b>73.61±0.77</b>	95.17±1.12	/

Table: Accuracy (in %) of 10 runs of the old regression model

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	83.24±0.35	35.58±0.74	/	28.14±1.14
German	80.13±0.57	/	30.14±0.29	12.92±6.03	/
Hungarian	51.21±3.09	78.09±0.74	/	94.05±0.12	34.55±0.60
Spanish	/	36.79±11.74	79.11±0.63	/	41.83±0.85
Turkish	47.23±0.91	/	15.29±0.85	70.79±0.06	/

Table: Accuracy (in %) of 5 runs of the old bilingual regression models



# Monolingual VS Bilingual (new)

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	39.75±2.67	80.33±1.95	/	79.89±2.43
German	94.26±0.63	/	70.55±2.61	70.55±2.61	/
Hungarian	43.93±3.44	85.39±1.76	/	85.39±1.76	40.98±3.42
Spanish	/	94.26±0.53	94.26±0.53	/	95.83±0.24
Turkish	64.98±2.76	/	70.74±2.23	94.03±3.70	/

Table: Accuracy (in %) of 10 runs of the new regression model

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	66.01±33.01	36.28±0.76	/	28.34±0.48
German	64.43±32.23	/	30.53±0.57	11.92±3.30	/
Hungarian	50.61±2.26	77.98±1.24	/	94.02±0.29	33.89±0.77
Spanish	/	32.42±17.01	78.99±0.13	/	40.45±1.52
Turkish	46.43±1.24	/	16.07±0.90	70.86±0.04	/

Table: Accuracy (in %) of 5 runs of the new bilingual regression models

# Bilingual

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	66.01±33.01	36.28±0.76	/	28.34±0.48
German	64.43±32.23	/	30.53±0.57	11.92±3.30	/
Hungarian	50.61±2.26	77.98±1.24	/	94.02±0.29	33.89±0.77
Spanish	/	32.42±17.01	78.99±0.13	/	40.45±1.52
Turkish	46.43±1.24	/	16.07±0.90	70.86±0.04	/

Table: Accuracy (in %) of 5 runs of the new bilingual regression models

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	/	83.24±0.35	35.58±0.74	/	28.14±1.14
German	80.13±0.57	/	30.14±0.29	12.92±6.03	/
Hungarian	51.21±3.09	78.09±0.74	/	94.05±0.12	34.55±0.60
Spanish	/	36.79±11.74	79.11±0.63	/	41.83±0.85
Turkish	47.23±0.91	/	15.29±0.85	70.79±0.06	/

Table: Accuracy (in %) of 5 runs of the old bilingual regression models

# Omnilingual model: Full dataset vs Bilingual features only

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	61.85±1.79	3.22±1.73	28.58±2.69	/	55.23±3.85
German	3.15±1.47	64.38±1.27	<b>66.91±4.49</b>	62.07±3.58	/
Hungarian	36.26±4.52	55.25±1.47	<b>73.33±1.31</b>	<b>78.36±1.33</b>	<b>32.00±3.03</b>
Spanish	/	61.16±2.54	<b>74.05±1.77</b>	69.38±1.65	70.67±4.03
Turkish	54.12±1.48	/	25.38±3.94	65.72±6.09	<b>52.23±1.09</b>

Table: Accuracy (in %) of 10 runs of the new omnilingual regression model trained on the full dataset

	Finnish	German	Hungarian	Spanish	Turkish
Finnish	<b>52.41±1.35</b>	24.79±1.89	38.22±2.26	/	48.81±1.33
German	18.40±4.44	31.24±1.25	<b>79.08±1.85</b>	55.48±1.09	/
Hungarian	<b>64.11±2.06</b>	68.52±1.10	<b>10.56±0.37</b>	<b>79.29±1.81</b>	<b>41.26±2.96</b>
Spanish	/	51.27±1.50	<b>80.04±1.64</b>	<b>28.91±0.65</b>	74.57±5.75
Turkish	<b>59.89±1.54</b>	/	26.52±2.69	83.49±2.41	21.86±1.04

Table: Accuracy (in %) of 10 runs of the new omnilingual regression model trained only on the bilingual features

# Analogy solver model

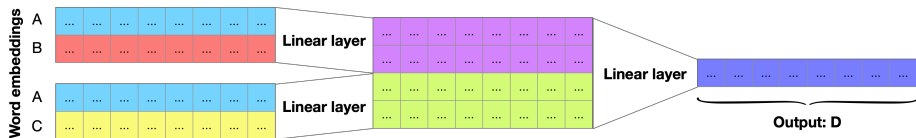


Figure: Analogy solver model