On the Transferability of Neural Models of Morphological Analogies

Safa Alsaidi*, Amandine Decker*, Puthineath Lay*, Esteban Marquer*, Pierre-Alexandre Murena**, **Miguel Couceiro***

AIMLAI 2021 - 13 of September

*Université de Lorraine, CNRS, LORIA
**HIIT, Aalto University

Morphological Analogies

Proportional Analogies

Notation:

- "A is to B as C is to D"
- $A : B :: C : D$

3 Core properties:

1. $A : B :: C : D \longrightarrow C : D :: A : B$ (symmetry)
2. $A : B :: C : D \longrightarrow A : C :: B : D$ (central permutation)
3. $A : B :: A : B$ (reflexively)

Morphology:

- "*read* is to *unreadable* as *count* is to *uncountable*"
- *aalto* : *aalloksi* :: *spirituaali* : *spirituaaliksi* (Finnish)

Detecting or Solving Analogies

Detecting analogies:

- take A, B, C, and D
- is $A : B :: C : D$ a valid analogy?
- output: ***yes*** or ***no*** (binary classification)

Solving analogies:

- take A, B, and C
- find x to have $A : B :: C : x$ a valid analogy
- output: $\boldsymbol{x}$ (regression)

Our Approach

Proposed Solution: General Process

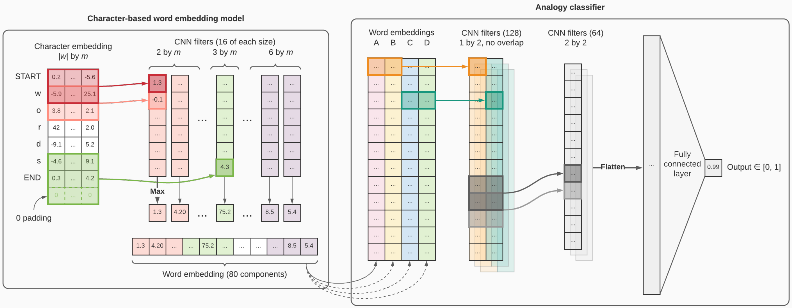
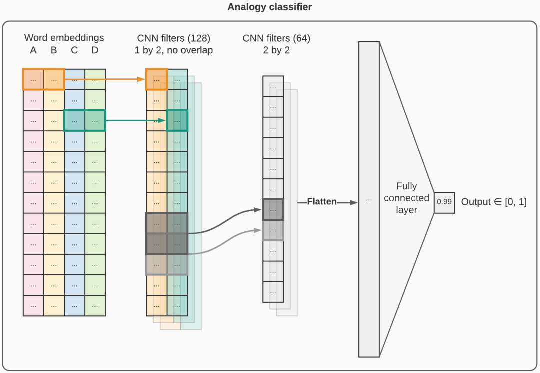


Figure 1: Morphological embedding and classifier

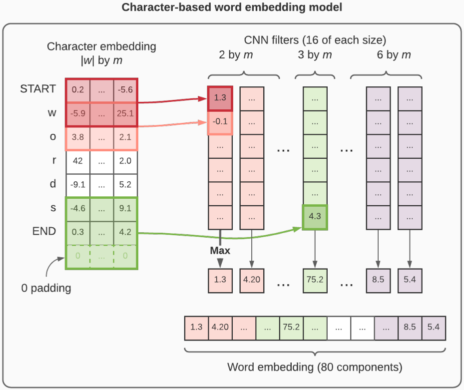
Classifier CNN

- Inspired from Lim et al. (2019)'s model of analogy[1] (semantic analogies).
- Compares each feature on A and B, C and D.
- Aggregates the result.



[1] Suryani Lim et al. (2019). "Solving Word Analogies: A Machine Learning Perspective". In: *15th ECSQARU*. ed. by Gabriele Kern-Isberner and Zoran Ognjanovic. Vol. 11726. Belgrade, Serbia: Springer, pp. 238–250

Morphological Embedding Model Using CNN



Pre-trained word embeddings (*word2vec*, Glove, *fasttext*, *etc.*) not adapted to our use-case:

- not available for our languages;
- semantics and not morphology (leads to poor performance).

Inspired from Vania (2020)'s work on embeddings for morphology[1].

Uses CNN filters to capture interesting sub-words (morphemes).

[1] Clara Vania (2020). "On understanding character-level models for representing morphology". PhD thesis. University of Edinburgh

Training Procedure

For all analogies $A : B :: C : D$ in dataset:

1. embed A, B, C, D;
2. augment embeddings using properties of analogical prop.;
3. aggregate Binary Cross-Entropy (BCE) loss over all forms.

Training: invalid analogies from base form (3 invalid)
Evaluation: equivalent forms of the invalid analogies (8×3 invalid)

$A : B :: C : D$	$C : D :: A : B$	$A : C :: B : D$	$B : A :: D : C$
$D : B :: C : A$	$D : C :: B : A$	$C : A :: D : B$	$B : D :: A : C$

Table 1: 8 equivalent forms per analogy

$B : A :: C : D$	$C : B :: A : D$	$A : A :: C : D$

Table 2: 3 invalid analogies per analogy

Experiments

Datasets: Sigmorphon 2016 (10 languages) and Japanese Bigger Analogy Test Set.

Subtasks: Inflection, reinflection and unlabeled reinflection.

Data: pairs of words $\langle A, B \rangle$ (ex: $\langle$"do","doing"$\rangle$) with B the morphological transformation of A w.r.t. a set F of morph. features (ex: present participle).

Analogy: pairs of words $\langle A, B \rangle, \langle A', B' \rangle$ with the same features ($F = F'$)

Datasets

Datasets: Sigmorphon 2016 (10 languages) and Japanese Bigger Analogy Test Set.

Subtasks: Inflection, reinflection and unlabeled reinflection.

Data: pairs of words $\langle A, B \rangle$ (ex: $\langle$"do","doing"$\rangle$) with B the morphological transformation of A w.r.t. a set F of morph. features (ex: present participle).

Analogy: pairs of words $\langle A, B \rangle, \langle A', B' \rangle$ with the same features ($F = F'$)

Baselines

Baselines:

- Kolmogorov Complexity Based Approach by Murena *et al.*[1]
- Alea algorithm by Langlais *et al.*[2]
- Lepage's Classifier from Fam and Lepage's toolset[3]

[1] P.-A. Murena, *et al.* Solving analogies on words based onminimal complexity transformation, IJCAI 2020, 1848–-1854.
[2] P. Langlais, *et al.* Improve-ments in analogical learning: Application to translating multi-terms of the medical domain, EACL 2009, 487–-495.
[3] R. Fam and Y. Lepage. Tools for the production of analogical grids and a resource of *n*-gram analogical grids in 11 languages, LREC 2018.

Classification

Language	CNN (ours) Valid	CNN (ours) Invalid	Best Baseline Valid		Best Baseline Invalid	
Arabic	99.89	**97.52**	34.21	(Alea@10)	**97.79**	(Kolmo@1)
Finnish	99.44	82.62	25.60	(Lepage)	98.78*	(Alea@1)
Georgian	99.83	91.71	93.20	(Kolmo@10)	95.21	(Alea@1)
German	99.48	89.01	86.90	(Alea@10)	97.19	(Alea@1)
Hungarian	99.99	**98.81**	36.80	(Kolmo@10)	**98.40**	(Kolmo@1)
Maltese	99.96	77.83	78.05	(Alea@10)	69.29	(Kolmo@1)
Navajo	99.53	90.82	21.45	(Kolmo@10)	94.93	(Kolmo@1)
Russian	97.95	79.85	42.37	(Alea@10)	93.88	(Lepage)
Spanish	99.94	78.33	85.90	(Alea@10)	86.62	(Lepage)
Turkish	99.48	92.63	44.76	(Alea@10)	91.40	(Kolmo@1)
Japanese	99.99	**98.65**	19.20	(Kolmo@10)	**98.13**	(Lepage)

123 no significant difference;

123 best result, significant difference between baselines and ours;

*Obtained on 4000 base analogies (too slow for 50000);

In most cases, no significant difference between the baselines.

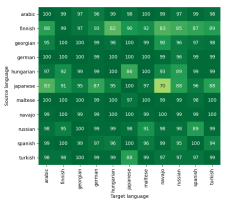
Datasets: Sigmorphon 2016 and Japanese Bigger Analogy Test Set.

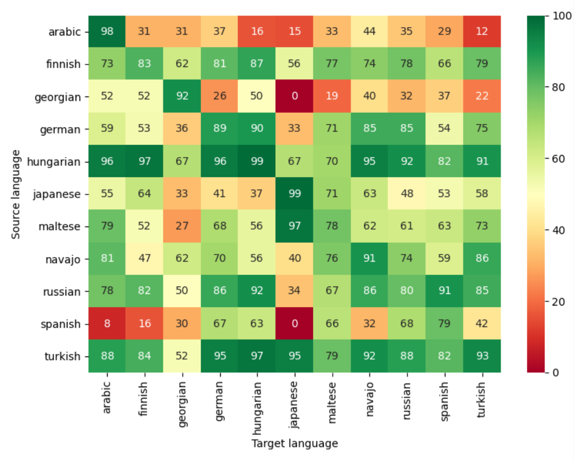
Transferability: Partial & Full

We experimented with two transferability settings:

- Partial transfer: using the classifier (but not the embeddings) of a language on another language;
- Full transfer: using both the embeddings and the classifier of a language on another language.

Transferability: Partial Transfer

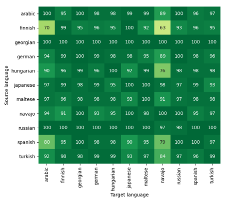


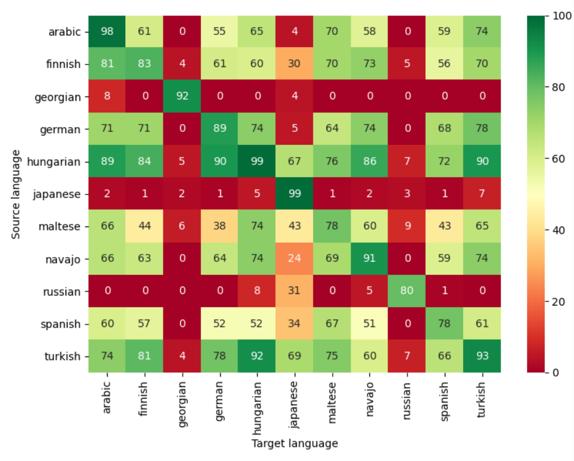


Transferability: Full Transfer

Alphabet gap:

- Georgian;
- Japanese;
- Russian.





Bilingual models: Hungarian - Finnish

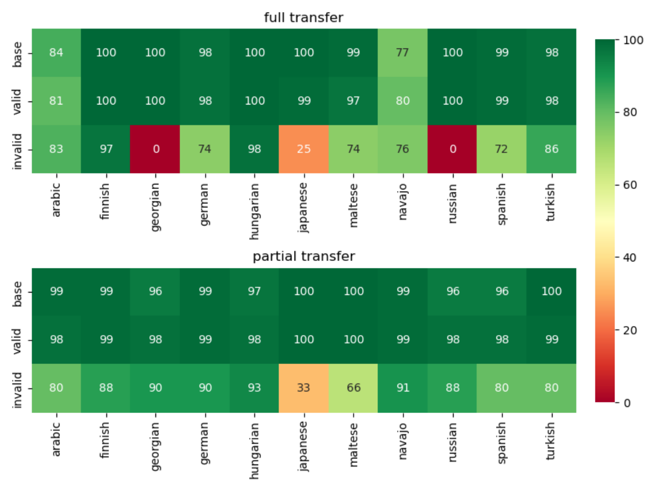


Figure 2: Accuracy (in %) of fully transferred bilingual models on Sigmorphon 2016 and Japanese Bigger Analogy Test Set.

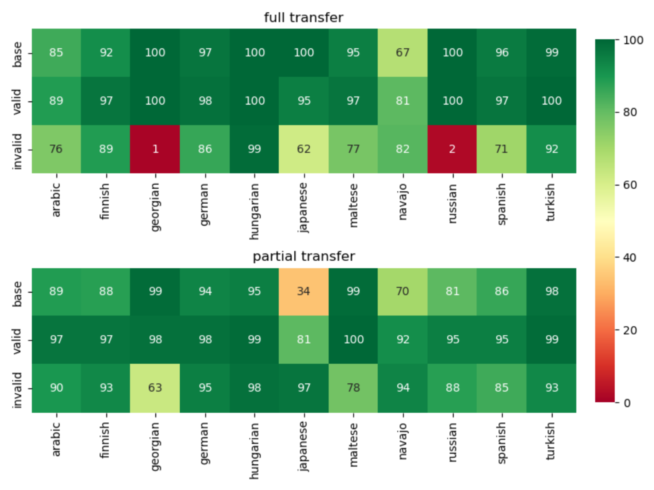


Figure 3: Accuracy (in %) of fully transferred bilingual models on Sigmorphon 2016 and Japanese Bigger Analogy Test Set.

Multilingual models: one embedding

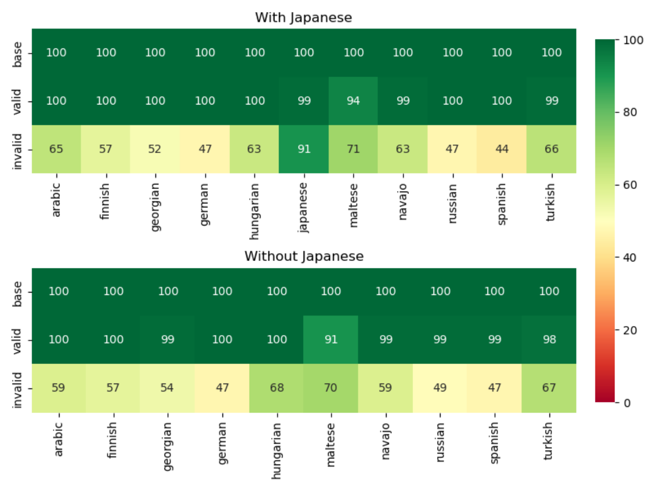


Figure 4: Accuracy (in %) of fully transferred multilingual models on SIGMORPHON 2016and JAPANESE BIGGER ANALOGY TEST SET.

Multilingual models: Multiple embeddings

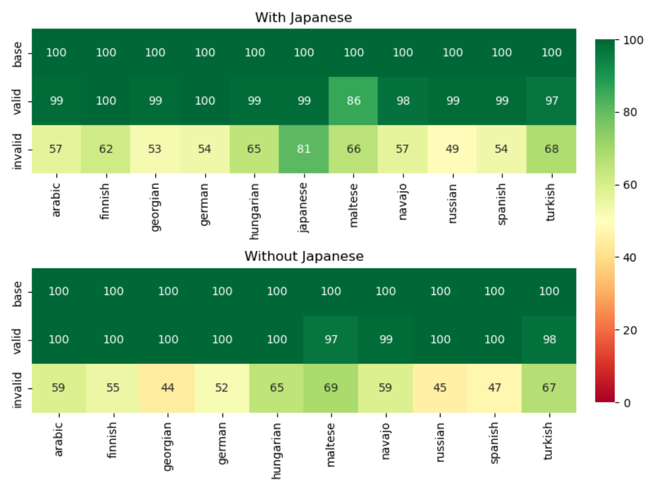


Figure 5: Accuracy (in %) of fully transferred multilingual models on SIGMORPHON 2016and JAPANESE BIGGER ANALOGY TEST SET.

Conclusion

Conclusion

Our model[4]:

- outperforms symbolic baselines on valid analogies;
- performance comparable to baselines on invalid;
- promising transferability between languages.

Trained model of analogy:

- training time;
+ inference time;
+ flexibility in terms of the properties of analogical proportions;
+ most likely adaptable to any setting, given a good embedding space.

[4] Safa Alsaidi et al. (2021a). "A Neural Approach for Detecting Morphological Analogies". In: *The 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Porto/Online, Portugal. URL: https://hal.inria.fr/hal-03313556

Further Work

- √ explore transferability[5];
- √ balancing data;
- √ multilingual setting;
- √ regression (analogy solving);
- ■ qualitative analysis of embedding model;
- ■ other domains.

[5] Safa Alsaidi et al. (2021b). "On the Transferability of Neural Models of Morphological Analogies". In: *AIMLAI, ECML PKDD 2021: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. URL: https://hal.inria.fr/hal-03313591

Thank you for your attention!

Questions now or to:

- esteban.marquer@loria.fr
- pierre-alexandre.murena@aalto.fi
- miguel.couceiro@loria.fr

References

Alsaidi, Safa et al. (2021a). "A Neural Approach for Detecting Morphological Analogies". In: *The 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Porto/Online, Portugal. URL: https://hal.inria.fr/hal-03313556.

– (2021b). "On the Transferability of Neural Models of Morphological Analogies". In: *AIMLAI, ECML PKDD 2021: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. URL: https://hal.inria.fr/hal-03313591.

Lim, Suryani, Henri Prade, and Gilles Richard (2019). "Solving Word Analogies: A Machine Learning Perspective". In: *15th ECSQARU*. Ed. by Gabriele Kern-Isberner and Zoran Ognjanovic. Vol. 11726. Belgrade, Serbia: Springer, pp. 238–250.

Vania, Clara (2020). "On understanding character-level models for representing morphology". PhD thesis. University of Edinburgh.