HeiMorph at SIGMORPHON 2022 Shared Task on Morphological Acquisition Trajectories

Akhilesh Kakolu Ramarao and **Yulia Zinova** and **Kevin Tang** and **Ruben van de Vijver**
Heinrich-Heine-University, Düsseldorf
{kakolura, yulia.zinova, kevin.tang, ruben.vijver}@hhu.de

Abstract

This paper presents the submission by the HeiMorph team to the SIGMORPHON 2022 task 2 of Morphological Acquisition Trajectories. Across all experimental conditions, we have found no evidence for the so-called U-shaped development trajectory. Our submitted systems achieve an average test accuracies of 55.5% on Arabic, 67% on German and 73.38% on English. We found that, bigram hallucination provides better inferences only for English and Arabic and only when the number of hallucinations remains low.

1 Introduction

Morphological inflection concerns generating the inflected word form given the lemma and a set of morphosyntactic descriptions. A morphology learner (human or machine) must be able to generalise patterns from extremely sparse data. Observations from morphology acquisition by children provides us with a glimpse of how learners generalise regular and irregular patterns differently and how the trajectories of pattern generalisations interact with a small but growing lexicon.

This paper describes our approach and results for Task 0 Part 2 of the SIGMORPHON 2022 shared task on morphological acquisition trajectories. Two main challenges of the task are that it covers two different inflectional patterns (past tense and noun plurals) over three languages and that there is only a small amount of training data ranging from as few as 100 samples to as many as 1,000 samples. This extreme data sparsity calls for the use of data hallucination techniques commonly used for low-resourced NLP development (Chen et al., 2021).

The neural baseline provided by the shared task is based on input-variant transformer (Wu et al., 2020) or the vanilla transformer with optional data augmentation (Anastasopoulos and Neubig, 2019).

We use a multi-headed self-attention Transformer with unigram-aware and bigram-aware data hallucinations.

Our models yielded an improved average test accuracy by 2.66% on Arabic, 8.69% on German, 4.5% on English, as compared with the neural baseline results.

2 Background and Data

The details of the task description can be found at `https://github.com/sigmorphon/2022InflectionST`. We use the data provided by the SIGMORPHON 2022 shared task (Part 2) (Kodner and Khalifa, 2022). The data features lemmas, inflections, and corresponding morphosyntactic description (MSD) using the uni-morph schema (Kirov et al., 2018). The data was released for English, German and Arabic. The specific inflectional patterns were the English past tense (Marcus et al., 1992), German noun plurals (Clahsen et al., 1992) and Arabic noun plurals (Dawdy-Hesterberg and Pierrehumbert, 2014).

3 System Description

In this section, we describe the neural network architecture, the data hallucination process and the submissions.

3.1 Neural Network architecture

All our models use the self-attention Transformer architecture (Vaswani et al., 2017) and implemented using the Fairseq (Ott et al., 2019) tool. Both the encoder and decoder have 4 layers with 4 attention heads, an embedding size of 256 and hidden layer size of 1,024. All models are trained with Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 0.001, batch size of 400, label smoothing as 0.1, gradient clip threshold as 1.0, and 4,000 warmup updates. All models are trained for a maximum of 3,000 optimizer updates, with checkpoints saved every 10 epochs. Beam search is used at decoding time with a beam width of 5.

The checkpoint with the smallest loss on the development data is chosen as the best model.

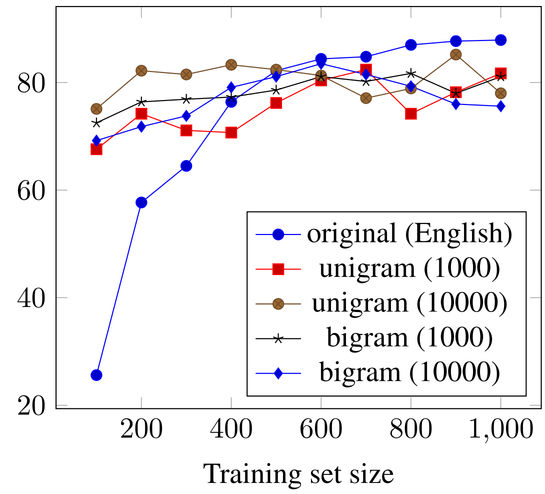
The inputs to each model are the individual characters of the lemma followed by their morphosyntactic tags separated by # symbol. For example, for the English training triple (take, took, V;PST), the input to the model is t a k e <V;PRS> # <V;PST> and the output is t o o k.

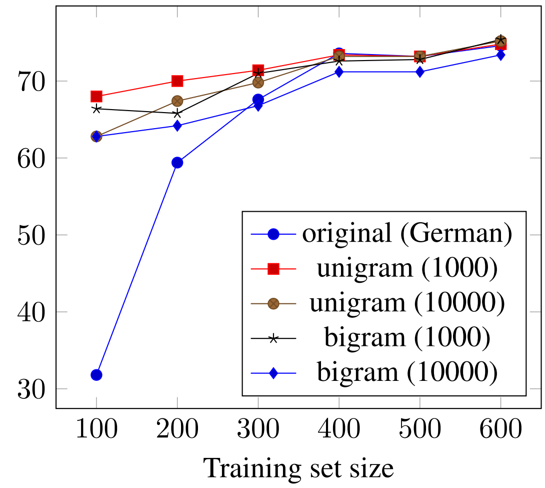
3.2 Hallucinations

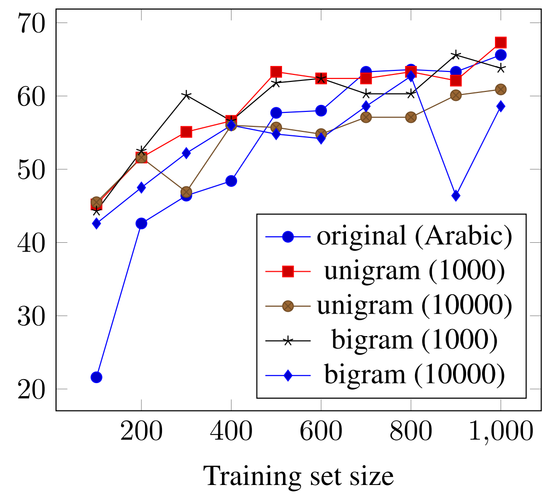
As has been shown in Anastasopoulos and Neubig (2019), adding hallucinated data boosts the learning process for small datasets. In the system described there, the hallucinated data is produced by 1) identifying a sequence of three or more consecutive characters that are aligned between the lemma and the inflected form and 2) randomly replacing the characters inside of this region by other characters from the language's alphabet. This replacement strategy produces is motivated by the label bias problem associated with the small dataset sizes. On the other hand, as a result the algorithm of Anastasopoulos and Neubig (2019) produces, among others, string pairs that lack vowels and have no resemblance to the original language data except for the inflected part. Our hypothesis was that such an approach may work well for languages with affixal morphology that is independent but may be less optimal for languages where the type of the inflection depends on the phonological properties of the stem (Haspelmath and Sims, 2013).

In order to check this hypothesis we have set up an alternative hallucination procedure. For each training set size, we compute a matrix of co-occurring characters and during the replacement step when a character from the alphabet is selected, we verify if this character occurred after the preceding character in the original train data. If yes, the replacement takes place, otherwise a new candidate character is selected.

As can be seen from the plots, the proposed bigram hallucination algorithm provides better results for English and Arabic data, if we do not produce too many hallucinations (1,000 hallucinations are better than 10,000, which was the original size in Anastasopoulos and Neubig (2019)).







3.3 Development decision

We selected our submitted system based only on a subset of the experiments, since we did not have the full picture across all experimental conditions at the time. Concretely, we based our decisions on the models performance for German. We found that the training size of less than 500 samples yielded models that performed better on 10,000 hallucinations, while training size of 500 and above yielded

	Arabic		German		English	
	Accuracy	Distance	Accuracy	Distance	Accuracy	Distance
100	42.6	2.2	62.8	0.46	69.2	2.08
200	47.5	1.99	64.2	0.45	71.8	1.9
300	52.2	1.79	66.8	0.41	73.8	1.6
400	56	1.75	71.2	0.36	79.1	1.39
500	61.8	1.5	72.8	0.34	78.6	1.83
600	62.4	1.49	75.4	0.3	81.1	1.48
700	60.3	1.61			80.2	1.62
800	60.3	1.63			81.7	1.41
900	65.6	1.35			78	1.54
1000	63.8	1.64			81.1	1.2

Table 1: Accuracy and Levenshtein distance on the development set

	Arabic		German		English	
	Accuracy	Distance	Accuracy	Distance	Accuracy	Distance
100	41.833	2.24	59	0.52	65.2	0.93
200	45.667	2.07	63.5	0.48	67.5	0.59
300	48.667	2.02	66.333	0.43	71.1	0.62
400	49.833	2.1	69	0.41	76.3	0.91
500	59.667	1.6	71	0.38	70.8	0.58
600	62.833	1.5	73.33	0.33	75.5	0.58
700	60.333	1.57			74.3	0.49
800	62.167	1.53			78.7	0.59
900	63.333	1.52			74.7	0.6
1000	59.333	1.74			80	0.48

Table 2: Accuracy and Levenshtein distance on the test set

models that performed better on only 1,000 hallucinations. Based on this finding with German, at the time of the development, we assumed this trend would hold also for English and Arabic.

3.4 Submissions

The models trained with training size less than 500 were hallucinated with 10,000 samples and the rest of the models with 1,000 samples across the three languages.

As the models trained on the proposed bigram hallucination algorithm provides better results on the development set for English and Arabic with 1000 hallucinations across all training sizes, this would have been our alternate submission.

4 Results

Table 1 shows the performance of our models on the development set. Results on the test data from SIGMORPHON 2022 Task 0 with Levenshtein distance can be found in Table 2.

5 Conclusion

How do children learn morphology? It has often been noted that children start out using correct forms, followed by a period of regularlizing irregular forms, followed by mastery of the morphology (Tessier, 2019)–often called the U-shaped development. This development can be found in Arabic (Abdalla et al., 2012; Benmamoun et al., 2014; Ravid and Farah, 1999; Saiegh-Haddad et al., 2012), German (Marcus et al., 1995) and English (Marcus et al., 1992). In our simulations we have found no evidence for such a development. This is, in fact, a good thing. Assuming a U-shaped development in morphological acquisition is too coarse, as the literature says little if anything about the question whether the (very few) forms used by very young children are used in the correct morphosynatctic environment. Moreover, this literature assumes that learning morphology involves learning how forms map onto other forms, reminiscent of the paradigm cell filling problem (Ackerman and Malouf, 2013; Guzmán, 2020; Malouf, 2017)–for example, how does a singular form map onto a plural form? The role of meaning is very limited, often not more than a contrastive label. However, the fact that children gradually reduce the number of overgeneralizations of irregular forms can be explained by the way in which children learn which word forms are used to express which particular meanings (Ramscar et al., 2013).

Our experiment with restricting the hallucination process to generate forms that are phonotactically attested (bigram) in the training data revealed that its benefit was found only in very restricted conditions depending on the amount of hallucinated samples and the specific language (and presumably the inflectional pattern). Our findings are in agreement with the detailed error analyses of data hallucination techniques by Samir and Silfverberg (2022) which concluded that hallucination is not a one-size-fits-all technique and it must be used with caution and requires closer inspection depending on the type of morphological inflections.

Acknowledgements

We gratefully acknowledge the support of the central HPC system “HILBERT” at Heinrich-Heine-University, Düsseldorf.

References

Fauzia Abdalla, Khawla Aljenaie, Abdessatar Mahfoudhi, Edith L Bavin, and Letitia R Naigles. 2012. Plural noun inflection in kuwaiti arabic-speaking children with and without specific language impairment*. *Journal of child language*, 40(1):139–168.

Farrell Ackerman and Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language*, 89(3):429–464.

Antonios Anastasopoulos and Graham Neubig. 2019. Pushing the limits of low-resource morphological inflection. *arXiv preprint arXiv:1908.05838*.

Elabbas Benmamoun, Abdulkafi Albirini, Silvina A. Montrul, and Eman Saadah. 2014. Arabic plurals and root and pattern morphology in palestinian and egyptian heritage speakers. *Linguistic Approaches to Bilingualism*, 4(1):89–123.

Jiaao Chen, Derek Tam, Colin Raffel, Mohit Bansal, and Diyi Yang. 2021. An empirical survey of data augmentation for limited data learning in nlp.

Harald Clahsen, Monika Rothweiler, Andreas Woest, and Gary F Marcus. 1992. Regular and irregular inflection in the acquisition of german noun plurals. *Cognition*, 45(3):225–255.

Lisa Garnand Dawdy-Hesterberg and Janet Breckenridge Pierrehumbert. 2014. Learnability and generalisation of arabic broken plural nouns. *Language, cognition and neuroscience*, 29(10):1268–1282.

Naranjo Matías Guzmán. 2020. Analogy, complexity and predictability in the russian nominal inflection system. *Morphology*, pages 1–44.

Martin Haspelmath and Andrea Sims. 2013. *Understanding morphology*. Routledge.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sabrina J Mielke, Arya D McCarthy, Sandra Kübler, et al. 2018. Unimorph 2.0: universal morphology. *arXiv preprint arXiv:1810.11101*.

Jordan Kodner and Salam Khalifa. 2022. SIGMORPHON-UniMorph 2022 Shared Task 0: Modeling Inflection in Language Acquisition. In *Proceedings of the SIGMORPHON 2022 Shared Task: Morphological Inflection*, Seattle. North American Chapter of the Association for Computational Linguistics.

Robert Malouf. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology*, 27(4):431–458.

Gary F Marcus, Ursula Brinkmann, Harald Clahsen, Richard Wiese, and Steven Pinker. 1995. German inflection: The exception that proves the rule. *Cognitive psychology*, 29(3):189–256.

Gary F Marcus, Steven Pinker, Michael Ullman, Michelle Hollander, T John Rosen, Fei Xu, and Harald Clahsen. 1992. Overregularization in language acquisition. In *Monographs of the society for research in child development*. JSTOR.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48–53, Minneapolis, Minnesota. Association for Computational Linguistics.

Michael Ramscar, Melody Dye, and Stewart M McCauley. 2013. Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, pages 760–793.

Dorit Ravid and Rola Farah. 1999. Learning about noun plurals in early palestinian arabic. *First Language*, 19(56):187–206.

Elinor Saiegh-Haddad, Areen Hadieh, and Dorit Ravid. 2012. Acquiring noun plurals in palestinian arabic: Morphology, familiarity, and pattern frequency. *Language Learning*, 62(4):1079–1109.

Farhan Samir and Miikka Silfverberg. 2022. One wug, two wug+ s transformer inflection models hallucinate affixes. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 31–40.

Anne-Michelle Tessier. 2019. U-shaped development in error-driven child phonology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(6):e1505.

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

Shijie Wu, Ryan Cotterell, and Mans Hulden. 2020. Applying the transformer to character-level transduction.