# Modeling irregular morphological patterns with Transformers
# The case of L-shaped morphomes

Kevin Tang, Akhilesh Kakolu Ramarao, Dinah Baer-Henney

Heinrich-Heine-University Düsseldorf, Germany

{kevin.tang,akhilesh.kakolu.ramarao,dinah.baer-henney}@hhu.de

## The problem

- State-of-the-art machines cannot cope with irregular inflectional morphology (as opposed to regular)
- Why? The architectures ignore cognitive factors, while humans don't
- The factor of interest in the present study is type frequency

## The idea

- Computational approach
- Probing a deep learning model with frequency?
- Is the model sensitive to frequency?
- Can the model pick up the irregular pattern and make correct predictions?

## Test case: Spanish

We take Spanish as a test case, focussing on the L-shape morphome pattern within the verbal paradigm [1, 2].

90 % of all verbs are regular
Non L-shaped verbs (NL)

| 'to eat' | IND | SUBJ |
|---|---|---|
| 1sg | komo | koma |
| 2sg | komes | komas |
| 3sg | kome | koma |

10 % are irregular verbs
L-shaped verbs (L)

| 'to say' | IND | SUBJ |
|---|---|---|
| 1sg | digo | diga |
| 2sg | dises | digas |
| 3sg | dise | diga |

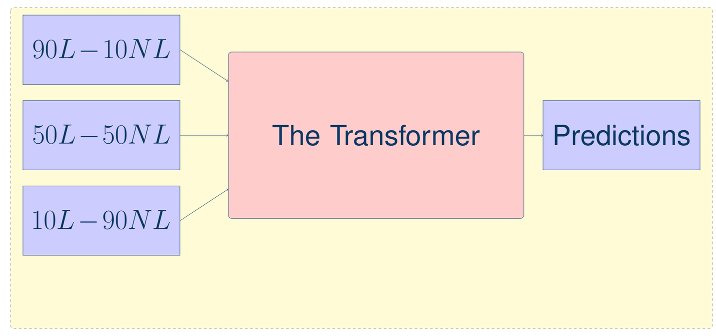
## The data

- The entries of the raw dataset include lemma/form pair (represented in IPA) and a Morpho-Syntactic tag Description (MSD). The paradigm is constructed (as shown above).
- Two-slot combinations followed by MSD for the slot to be filled is treated as input and the inflected form for the target slot as the output.
- In our setting, each lemma produces around 600 combinations. We considered 333 lemmas and downsampled the combinations by 25 %. This gave us a training set of 40000, development set of 4500 and test set of 44000 samples.
- These combinations are generated for 10L-90NL, 50L-50NL, 90L-10NL conditions such that they are split at the lemma level, combination level and between combinations level.

## Method

Research rationale:

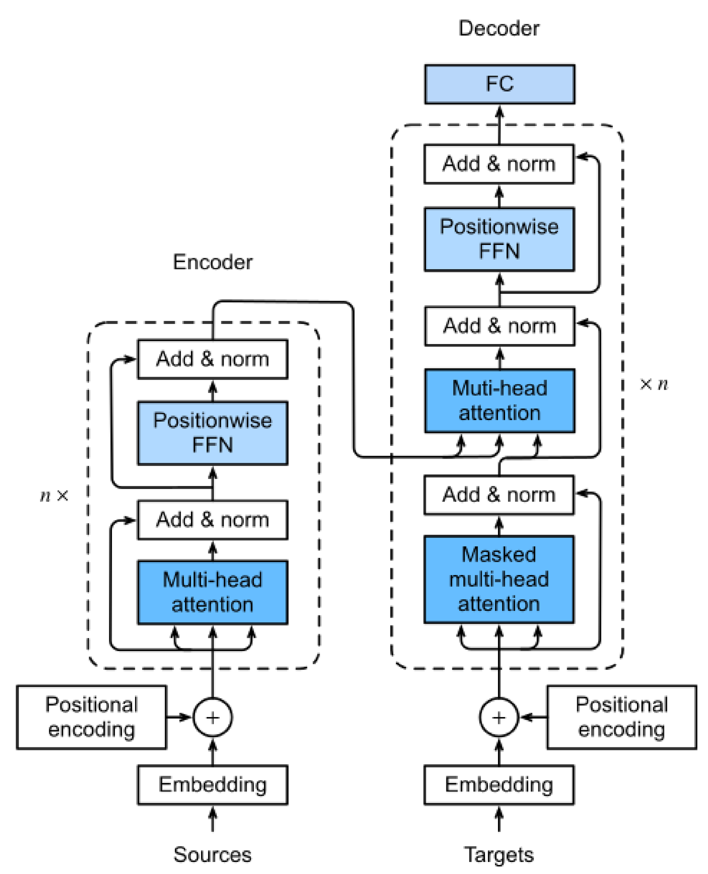
- Across conditions, the ratio of NL- vs. L-shaped forms used for training varies
- The model is trained with, and asked to perform on all cell combinations in the test set
- Every model receives input of the below kind. The table illustrates general patterns (and is not an exhaustive list)




| examples of training and test data | Input: 1st source | Input: 2nd source | Input: target tag | Output |
|---|---|---|---|---|
| NL-shape | k o m e # ⟨V;IND;PRS;3;SG⟩ # | k o m a # ⟨V;SBJV;PRS;3;SG⟩ # | ⟨V;SBJV;PRS;2;SG⟩ | k o m a s |
| L-shape training no mix | d i g a # ⟨V;SBJV;PRS;1;SG⟩ # | d i g a s # ⟨V;SBJV;PRS;2;SG⟩ # | ⟨V;IND;PRS;1;SG⟩ | d i g o |
| L-shape training mix | d i g a # ⟨V;SBJV;PRS;1;SG⟩ # | d i s e # ⟨V;IND;PRS;3;SG⟩ # | ⟨V;SBJV;PRS;2;SG⟩ | d i g a s |

## Machine Learning architecture

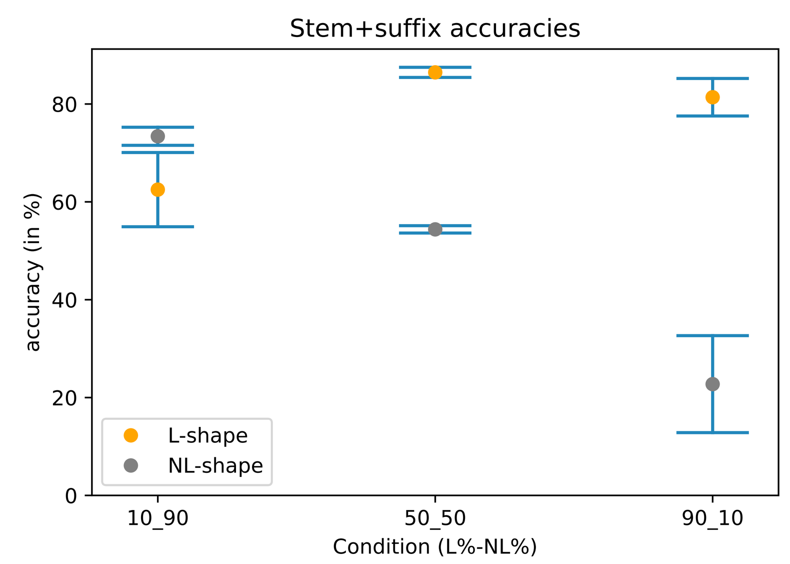
- The transformer used is a self-attention-based encoder-decoder model. Both encoder and decoder have 4 layers, embedding dimension size of 256, hidden layer size of 1024 and 4 attention heads with ReLU activation.
- We use Adam [3] with a learning rate of 0.001, label smoothing 0.1, batch size 400, dropout 0.3, clip-norm 1.0, adam-betas (0.9, 0.98). The label smoothed cross entropy as loss function.

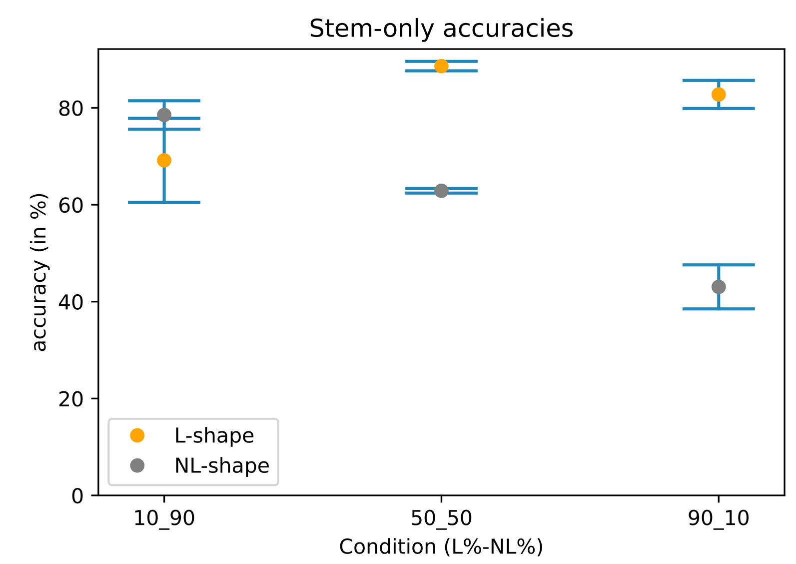


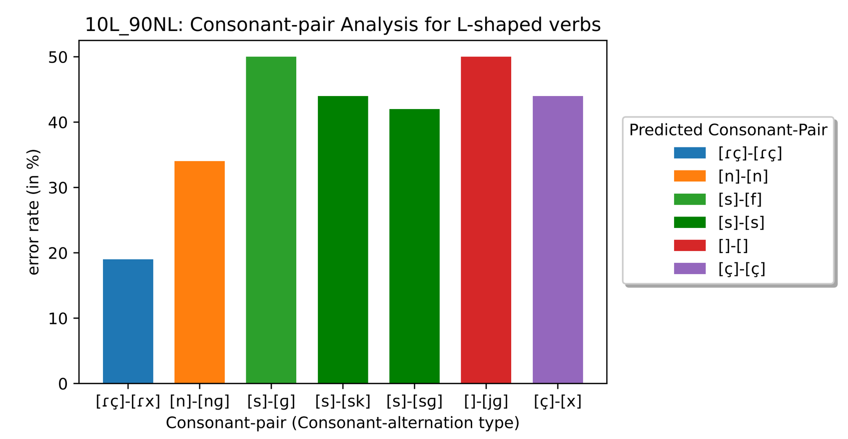

## References

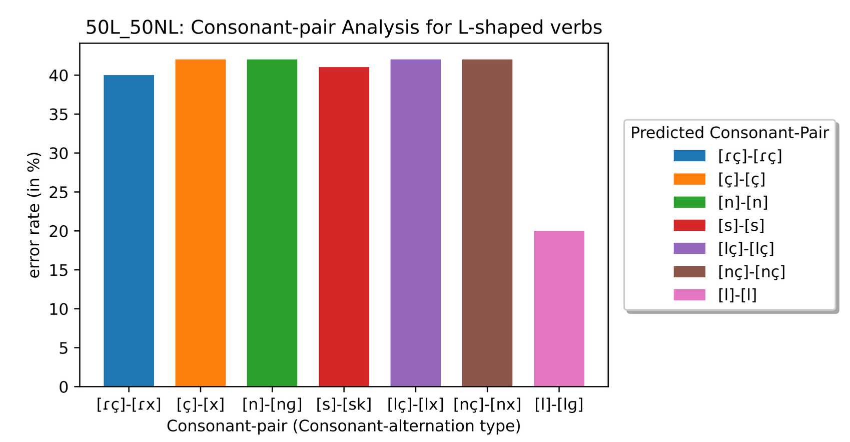
[1] Andrew Nevins, Cilene Rodrigues, and Kevin Tang. The rise and fall of the l-shaped morphome: diachronic and experimental studies. *Probus*, 27:101 – 155, 2015.
[2] Borja Herce. A typological approach to the morphome. 2020.
[3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
[4] Dinah Baer-Henney and Ruben Van de Vijver. On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, 3:221–249, 2012.
[5] Elliott Moreton and Joe Pater. Structure and substance in artificial-phonology learning, part i: Structure. *Language and linguistics compass*, 6(11):686–701, 2012.

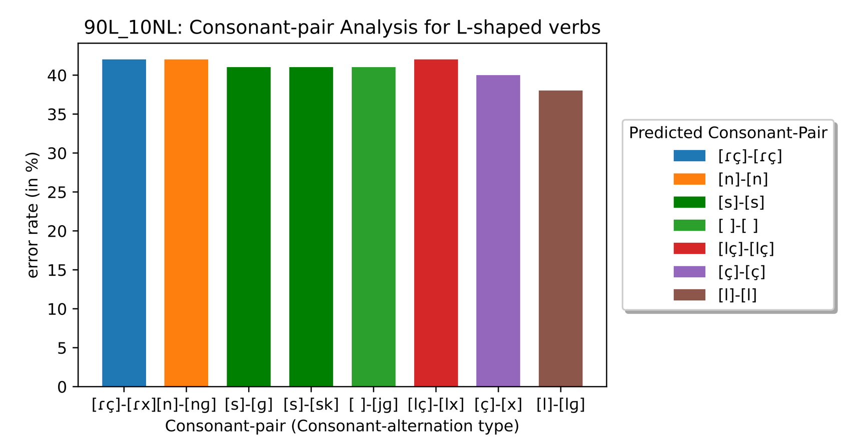
## Results




## Conclusion

- Under a naturalistic distribution in the lexicon (the 10-90 condition), L-shaped verbs are more difficult to learn than non-L-shaped verbs, mirroring the unproductivity of the L-shaped morphome found in human wug experiments [1].
- Surprisingly, L-shaped verbs require less lexical support than Non-L-shaped verbs. Training with 10% of L-shaped verbs yielded a stem accuracy of 70%, while training with 10% of Non-L-shaped verbs yielded only a stem accuracy of 40%.
- Error analyses show that the irregular stem consonant (1sg IND and SUBJ) of an l-shaped consonant pair is more error-prone than the regular stem consonant.
- **Potential future work** includes: 1) Probing for the influence of other cognitive factors such as phonological complexity [4, 5] and morphological complexity [2]. 2) Comparing data with human data from [1]. 3) Run artificial language learning experiments for a comparison of human and machine learning architectures.