



# Frustratingly Easy Multilingual Grapheme-to-Phoneme Conversion



Scan  
for  
codebase!

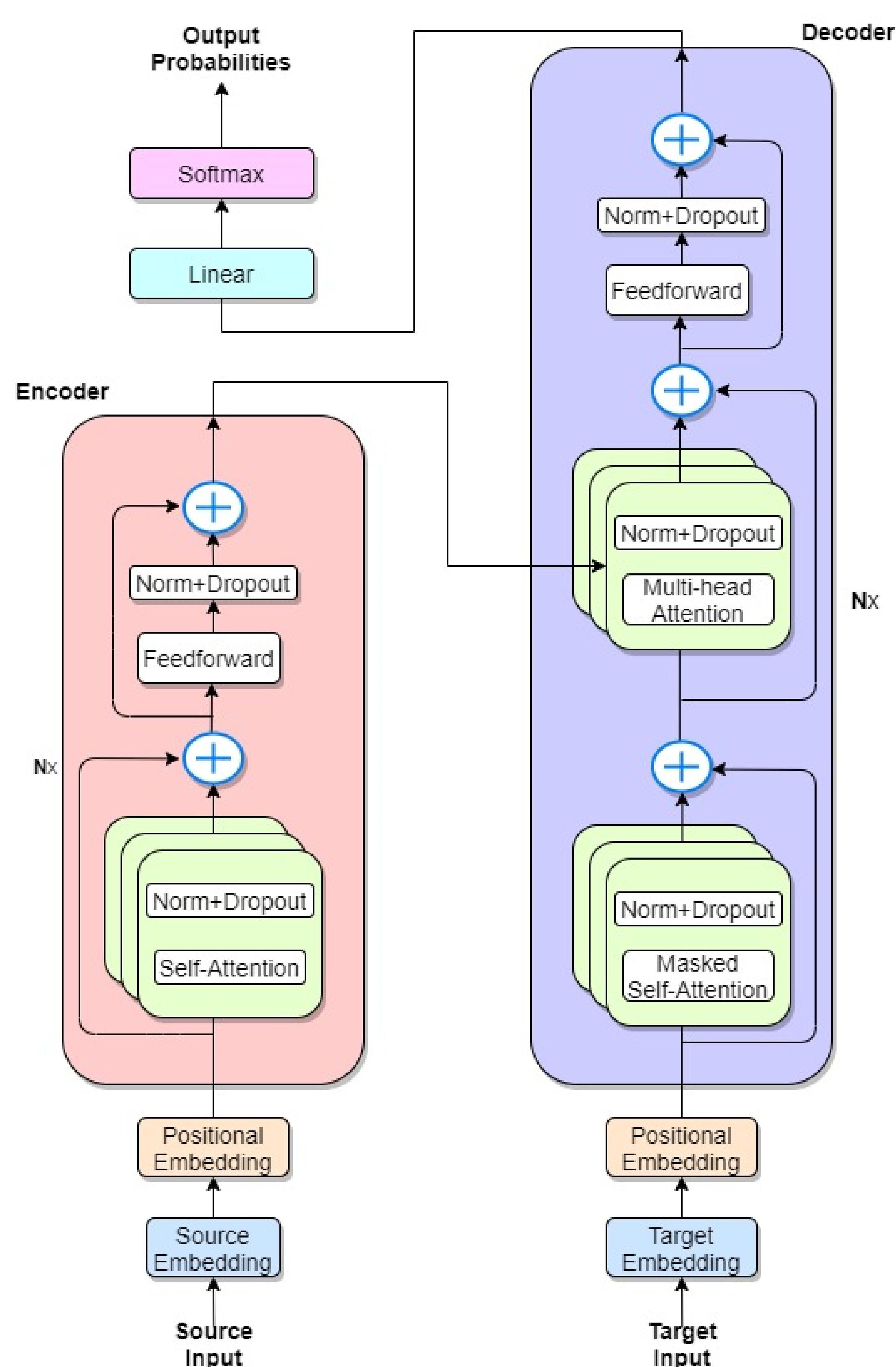
Nikhil Prabhu and Katharina Kann

University of Colorado Boulder

{nikhil.prabhu, katharina.kann}@colorado.edu

## Overview:

- We present two CU Boulder submissions for SIGMORPHON 2020 Task 1 of Multilingual Grapheme-to-Phoneme Conversion (G2P)
- We improve over the high performance of a standard transformer by adding two modifications:
  - multi-task training
  - ensembling



## Multi-task Training:

- We train our model jointly on two tasks:
  - G2P:  $g(w) \rightarrow p(w)$ , with dataset  $D_{g2p}$
  - P2G:  $p(w) \rightarrow g(w)$ , with dataset  $D_{p2g}$
- $\lambda_g$  and  $\lambda_p$  are task embeddings which indicate which task the input belongs to

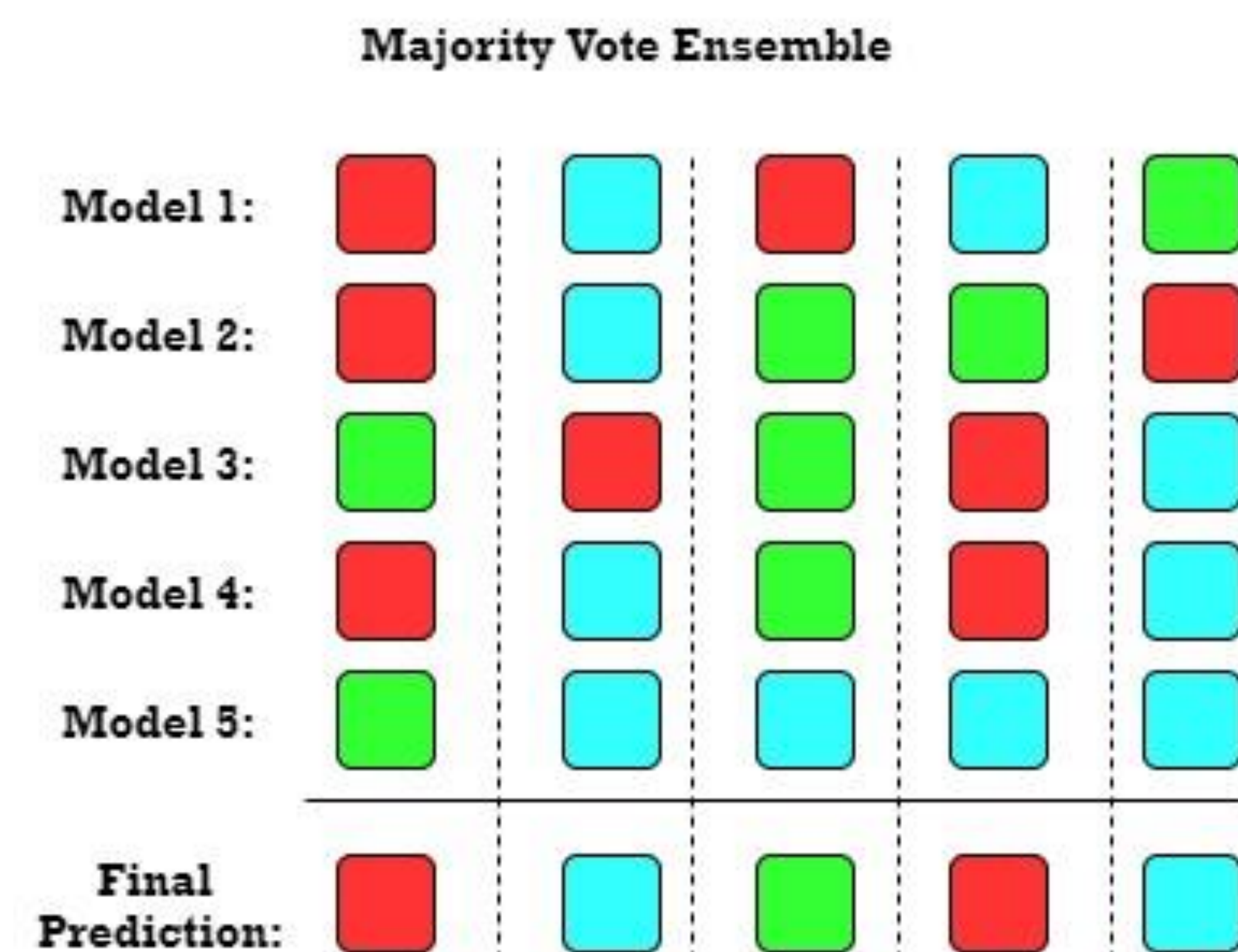
- We then obtain model parameters  $\theta$  that maximize the joint log-likelihood of both datasets:

$$\mathcal{L}(\theta) = \sum_{w \in D_{g2p}} \log p_{\theta}(p[w]|g[w], \lambda_g) + \sum_{w \in D_{p2g}} \log p_{\theta}(g[w]|p[w], \lambda_p)$$

## Ensembling:

Our models are ensembles created via majority voting

- CU-1, is an ensemble of 5 standard G2P transformers and 5 multi-task transformers
- CU-2, is an ensemble of 5 multi-task transformers



## Grapheme-to-Phoneme Conversion:

- Let  $\Sigma_G$  and  $\Sigma_P$  be an alphabet of graphemes and phonemes, respectively
- Consider  $g(w) \in \Sigma_G^*$  and  $p(w) \in \Sigma_P^*$  as the grapheme and phoneme representations of  $w$ , respectively
- G2P then refers to the mapping  $g(w) \rightarrow p(w)$

Grapheme	Phoneme	Phoneme	Grapheme
!aandacht	a:ndax̩t	?a:ndax̩t	aandacht
!basson	bax̩sɔn	?bax̩sɔn	basson
!begint	bəɣ̩ɪnt	?bəɣ̩ɪnt	begint
!gierst	x̩i:r̩st	?x̩i:r̩st	gierst
!heup	h̩ø:p̩	?h̩ø:p̩	heup

## Results:

