Automatic construction of a phonics curriculum for reading education using the Transformer neural network

Cassandra Potier Watkins[1[0000-0002-6588-0614]], Olivier Dehaene[2] and Stanislas Dehaene[1,3[0000-0002-7418-8457]]

[1] INSERM, UMR992, CEA, Neurospin Center, University Paris Saclay, Gif-sur-Yvette, France
[2] Owkin France, 75 rue de Turbigo, 75003, Paris
[3]College de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France

cassandra.potier-watkins@cea.fr

Abstract. Key to effective phonics instruction is the teaching of grapheme-phoneme (GP) correspondences in a systematic progression that starts with the most frequent and consistent pronunciation rules. However, discovering the relevant rules is not a an easy task and usually requires subjective analysis by a native speaker and/or expert linguist. We describe GPA4.0, a submodule to the Transformer neural network model that automatizes the task of grapheme-to-phoneme (g2p) transcription and alignment. The network is trained with four different languages of decreasing orthographic transparency (Spanish<Portuguese<French<English). Our results show that the Transformer model improves on the current state-of-the-art in g2p transcription and that the attention mechanism allows for the alignment of graphemes to their corresponding phonemes. From the g2p aligned words, our software provides an optimally ordered phonics progression based on frequency and consistency in the target language, as well as an ordered list of words that teachers can use. This work exemplifies a practical way that neural networks can be used to develop educational materials for research and teachers. Submodules and phonics output are available at, https://github.com/OlivierDehaene/GPA4.0.

Keywords: phonics instruction, g2p, attention.

1 Introduction

Early phonics introduction is endorsed as the foundation of successful reading instruction in both education research (meta-analysis by the National Reading Panel [1], [2]) and cognitive neuroscience [3, 4]. However, phonics instruction is not universally used. One factor for its relative disaffection could be that knowing what grapheme-phoneme (GP) correspondences to teach, and in what order to teach them, can be a difficult task, given that letter-sound relationships do not all have a one-to-one relationship. Take for example Spanish, a highly *transparent* language, meaning that a given letter is nearly always pronounced the same. In stark contrast is English, which can have many different sounds for a single grapheme (e.g. the ‘a’ in ‘cat’, ‘mate’,

'what' or 'about'). Cross-language research demonstrates that orthographic transparency influences the time and difficulty children have in learning to read[5–9] .

Orthographic transparency is also a conundrum in neural network text-to-speech applications that rely on grapheme-to-phoneme (g2p) transcription. G2p refers to converting words to their phonemes. The current state-of-the art applies long short-term memory (LSTM) networks and recurrent neural networks using sequence-to-sequence (seq2seq) modeling combined with an attention-mechanism [10]. More recently, the Transformer model has brought notable improvements in neural machine language transcription and language parsing [11]. These tasks that are fairly analogous to g2p transcription (both depend on long range dependencies and contextual influences). Improvement made by the Transformer model is in part due to parallel position encoding that curtails the need for recurrence and a self-attention field that enables the concatenation of information between sequences, regardless of their distance. The goal of the current project, GPA4.0 (Fig. 1), is to test for g2p transcription improvements, for the first time to our knowledge, using the Transformer model. With this achieved, we take advantage of the Transformer's attention mechanism to align grapheme input to phoneme output, thus permitting the construction of a phonics progression based on the frequency and consistency of all found GP correspondences for any alphabetic language word list.

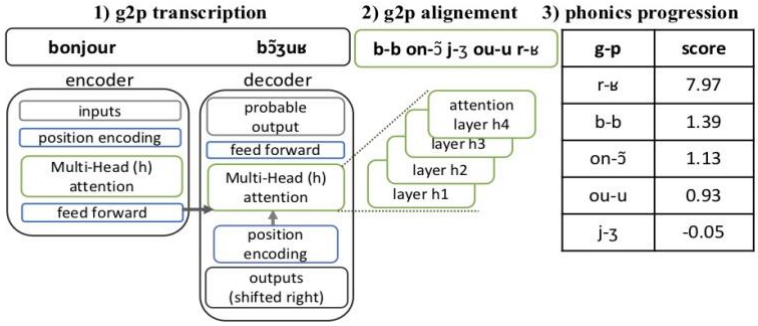

Fig. 1. GPA4.0 steps to constructing a phonics progression. 1) g2p transcription is done using the Transformer neural network. 2) g2p alignment uses attention weights to align the 'grapheme inputs' to their 'phoneme outputs'. 3) a phonics progression is built by according each g2p alignment an aggregated z-score based on frequency and consistency in the word corpus.

2 Experiment

We tested the Transformer model for improved g2p transcription compared to the current-state-of-the-art results [10, 12] on the CMUDict database [13] while also comparing, for the first time to our knowledge, the results of five different languages of varying orthographic transparency: Spanish<Portuguese<French<English. Training was done using one 1080TI NVIDIA GPU on the base models for a total of 10,000 steps. We use Tensor2Tensor (T2T) [14] an open-source system for training deep learning models in TensorFlow [15]. G2p alignment in our model is made possible

using the attention weights of the Transformer model. G2p alignment accuracy was analyzed in French, the only language for which we had a reference for comparison. Table 1 describes the word lists used and provides the minor adjustments made to accommodate the small amount of training data. Training was conducted on 80% of the data. The model's performance was tested on the remaining 20% of data.

To generate a language's phonics progression, we extract all the GP correspondences in the list of g2p aligned words. For each GP correspondence found, we measure its frequency, g2p consistency and phoneme-to-grapheme consistency. The GP correspondences are then sorted by an aggregate weight of the prementioned measures' z-scores (we apply weights of 0.7, 0.25 and 0.05 respectively, but these can be adjusted in the code). The weights are designed to 1) give priority to the most frequent GP correspondences when a pair is particularly consistent and less frequent but highly consistent correspondences.

Table 1. Language wordlists used and adjustments made to the Transformer architecture

language	**number of words used for training**	**number of words used for testing**	**number of hidden layers**	3
Spanish [16]	10,400	2,600	hidden size, number of neurons per layer	256
Portuguese [17]	31,200	7,800	filter size	512
French [18]	8,000	2,000	h, number of attention heads	4
English [19]	8,000	2,000	attention dropout rate	0.2
English [13]	95,069	23,767	dropout rate	0.3

3 Results

3.1 g2p Transcription

The standard measures of word error rate (WER and phoneme error rate (PER) are reported in Table 2. WER is the total number of output errors in which there is at least one phoneme error / total number of words. PER is the Levenshtein distance [20] (the minimum number of single-character edits needed to change one word to the other) of the predicted phoneme sequence to the reference from the original database / the number of phonemes in the reference. Language WER and PER scores reflect, as expected, decreasing orthographic transparency. We report a slight gain over Toshniwal and Livescu's best prior score on the CMUdict database.

Table 2. Word error rate (**WER**) and Phoneme error rate (**PER**) in four languages of decreasing orthographic transparency

	Spanish<	**Portuguese<**	**French<**	**English**	**CMUDict**
WER	0.38%	2.77%	3.18%	15.04%	20.87%
PER	0.07%	0.55%	0.89%	4.50%	4.59%
			Previous best results using the CMUDict database:		WER=21.69% PER=5.04%

3.2 g2p Alignment

Fig. 2 provides an example of encoder-decoder attention (taken from layer-4 multi-head attention in the decoder, see Fig. 1). As the network reads the word “bonjour” or “banane”, it attends to distant information required to know if a vowel followed by the letter ‘n’ will make a single nasal sound (e.g. ‘on’) or two distinct phonemes (e.g. a+ n). GPA4.0 aligns graphemes to phonemes based on the attention carrying the most weight. G2p alignment error rate was assessed for French using the sequence error rate, a correct or incorrect score for each word and the g2p alignment error rate (Levenshtein distance [20]). We report scores of 27.76% and 10.20% respectively. The relatively high sequence error rate compared to the low g2p alignment score is due to the difficulty in parsing silent letters not coded in the phonology of the trained wordlist. 56% of words in the list contain silent letters.

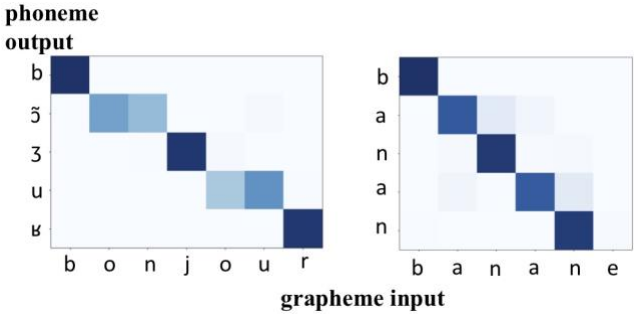

Fig. 2. Encoder-decoder attention in g2p transcription

4 Conclusion

Our results demonstrate improved g2p transcription by the Transformer model. Our submodule, GPA4.0, takes a novel approach to developing applicable phonics tools for the classroom by taking advantage of neural network performance in g2p transcription and, in particular, the attention field for g2p alignment. This work highlights the difficulties for neural networks to learn the GP correspondences in decreasingly transparent languages. The phonics progressions for the four languages analyzed and their ordered wordlists are freely available. These datafiles can be used as a ‘paper’ support to guide reading instruction, or as stimuli for game-based reading applications (e.g. the GraphoGame software [3, 21]). We hope that the GPA4.0 submodule will be taken up as a tool for researchers and educators to generate their own phonics lessons with 100% decodable reading materials. GPA4.0 combines cognitive science and neural network technology for evidence-based reading education. Phonics progressions and word lists for the four different languages analyzed in this paper, as well as the GPA4.0 submodule code, can be downloaded at https://github.com/OlivierDehaene/GPA4.0.

5 Bibliography

1. Cunningham, J.W.: The National Reading Panel Report. Reading Research Quarterly. 36, 326–335 (2001). https://doi.org/10.1598/RRQ.36.3.5.
2. Castles, A., Rastle, K., Nation, K.: Ending the Reading Wars: Reading Acquisition From Novice to Expert. Psychol Sci Public Interest. 19, 5–51 (2018). https://doi.org/10.1177/1529100618772271.
3. Brem, S., Bach, S., Kucian, K., Kujala, J.V., Guttorm, T.K., Martin, E., Lyytinen, H., Brandeis, D., Richardson, U.: Brain sensitivity to print emerges when children learn letter–speech sound correspondences. Proceedings of the National Academy of Sciences. 107, 7939–7944 (2010). https://doi.org/10.1073/pnas.0904402107.
4. Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Filho, G.N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L.: How Learning to Read Changes the Cortical Networks for Vision and Language. Science. 1194140 (2010). https://doi.org/10.1126/science.1194140.
5. Seymour, P.H.K., Aro, M., Erskine, J.M.: Foundation literacy acquisition in European orthographies. British Journal of Psychology. 94, 143–174 (2003). https://doi.org/10.1348/000712603321661859.
6. Goswami, U., Gombert, J.E., Barrera, L.F. de: Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. Applied Psycholinguistics. 19, 19–52 (1998). https://doi.org/10.1017/S0142716400010560.
7. Landerl, K.: Influences of orthographic consistency and reading instruction on the development of nonword reading skills. Eur J Psychol Educ. 15, 239 (2000). https://doi.org/10.1007/BF03173177.
8. Serrano, F., Genard, N., Sucena, A., Defior, S., Alegria, J., Mousty, P., Leybaert, J., Castro, S.L., Seymour, P.H.K.: Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. Journal of Portuguese Linguistics. 10, 183–204 (2011). https://doi.org/10.5334/jpl.106.
9. Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., Blomert, L.: Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. Psychological Science. 21, 551–559 (2010). https://doi.org/10.1177/0956797610363406.
10. Toshniwal, S., Livescu, K.: Jointly Learning to Align and Convert Graphemes to Phonemes with Neural Attention Models. arXiv:1610.06540 [cs]. (2016).
11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need. arXiv:1706.03762 [cs]. (2017).
12. Rao, K., Peng, F., Sak, H., Beaufays, F.: Grapheme-to-phoneme conversion using Long Short-Term Memory recurrent neural networks. Presented at the April (2015). https://doi.org/10.1109/ICASSP.2015.7178767.

13. Weid, R.L.: The CMU Pronouncing Dictionary, http://www.speech.cs.cmu.edu/cgi-bin/cmudict.
14. Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A.N., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., Sepassi, R., Shazeer, N., Uszkoreit, J.: Tensor2Tensor for Neural Machine Translation. arXiv:1803.07416 [cs, stat]. (2018).
15. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Software available from tensorflow.org. 19 (2015).
16. Corral, S., Ferrero, M., Goikoetxea, E.: LEXIN: A lexical database from Spanish kindergarten and first-grade readers. Behavior Research Methods. 41, 1009–1017 (2009). https://doi.org/10.3758/BRM.41.4.1009.
17. Derived Corpora and Counts, https://childes.talkbank.org/derived/.
18. Lété, B., Sprenger-Charolles, L., Colé, P.: MANULEX: a grade-level lexical database from French elementary school readers. Behav Res Methods Instrum Comput. 36, 156–166 (2004).
19. Masterson, J., Stuart, M., Dixon, M., Lovejoy, S.: Children's printed word database: Continuities and changes over time in children's early reading vocabulary. British Journal of Psychology. 101, 221–242 (2010). https://doi.org/10.1348/000712608X371744.
20. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady. 10, 707–710 (1966).
21. Richardson, U., Lyytinen, H.: The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. Human Technology. 10, (2014). https://doi.org/10.17011/ht/urn.201405281859.