

# Neural Borrowing Detection with Monolingual Lexical Models

John E. Miller

jemiller@pucp.edu.pe

Emanuel Pariasca

emanuel.pariasca@pucp.edu.pe

César A. Beltrán Castañón

cbeltran@pucp.edu.pe

Artificial Intelligence/Engineering  
Pontificia Universidad Católica del Perú  
San Miguel, Lima, Perú

## Abstract

Identification of lexical borrowings, transfer of words between languages, is an essential practice of historical linguistics and a vital tool in analysis of language contact and cultural events in general. We seek to improve tools for automatic detection of lexical borrowings, focusing here on detecting borrowed words from monolingual wordlists. Starting with a recurrent neural network lexical model and competing entropies approach, we incorporate a more current Transformer based lexical model. From there we experiment with several different models and approaches including a lexical donor model with augmented wordlist. The Transformer model reduces execution time and minimally improves borrowing detection, and the augmented donor model shows some promise. A substantive change in approach or model seems necessary for significant gains in detection of lexical borrowings.

## 1 Introduction

When words shows striking similarity between languages, they may be inherited from a common ancestor, borrowed from one contact language into the other, or less frequently, similar by chance or common creative process, e.g., onomatopoeia.

Identification of lexical borrowings, transfer of words between languages, is an integral step in the comparative method of historical linguistics (Campbell, 2013) and a vital tool for investigations of language contact, and cultural influences. Automated or assisted detection of lexical borrowings is one of many important and difficult tasks remaining in historical linguistics (List, 2019a).

We use neural network lexical language models of the phonology and phonotactics of a language to identify lexical borrowings. While language models refer generally to “any system trained only on the task of string prediction” (Bender and Koller,

2020), our lexical language models operate on monolingual wordlists limited to common or essential concepts, with vocabularies limited to language sound segments (phonemes, allophones). A fixed vocabulary of typically less than 100 coded sound segments eliminates the out-of-vocabulary problem, and a limited wordlist size of at most a few thousand lexical entries, reduces the practical model complexity.

We use the recurrent neural network model and competing lexical entropies approach of Miller et al. (2020) as a jumping off point, and incorporate a light-weight forward direction Transformer model from Vaswani et al. (2017) into this approach. We demonstrate our light-weight Transformer model over 41 different languages from the World Loan Database (Haspelmath and Tadmor, 2009) and contrast performance versus the recurrent model of Miller et al. (2020).

We perform several experiments expanding both the technical and application scope of our light-weight Transformer model: 1. develop a direct model approach that discriminates between inherited and borrowed words directly without the need to model inherited and borrowed word entropies separately, 2. enhance competing entropies and direct approaches to account for borrowed words by individual donor language, 3. train multiple borrowed and inherited models, and base decisions on majority vote of model combinations, 4. perform search for best model parameters using experimental sequences of randomly assigned parameter values, 5. explore a limited multi-lingual approach of using an additional donor language table to supplement borrowed words from the recipient language.

Results are evaluated for borrowing detection and, secondarily, execution time. Our light-weight Transformer minimally improves borrowing detection and reduces execution time versus the baseline. Direct detection of borrowed words performs less

well than the competing entropies approach. Multiple model combinations and parameter search show little improvement over the Transformer model, each with increased execution time. Adding a donor language wordlist minimally improves borrowing detection.

Our contributions are:

- Incorporation of Transformer model and direct approach into the borrowing detection problem,
- Demonstration that the competing entropies approach offers better borrowing detection than the direct approach,
- Demonstration that the multiple models approach and parameter search do not improve borrowing detection,
- Prototyped a way to improve detection performance via added donor wordlists.

## 2 Related Work

### Approaches to Borrowed Word Detection

When words are borrowed into a language, initially, they may retain the phonology, phonotactics, or other features of the donor language. But over time the borrowed words will adapt to the recipient language – making borrowing detection more difficult (Kiparsky, 2014).

Our monolingual wordlist approach gambles that some remnant of the phonology or phonotactics of the donor language will still be perceptible long after the word is borrowed, and that competing entropies lexical models can be used to detect borrowed words. This is the approach taken by Miller et al. (2020) where curated monolingual wordlists of inherited and borrowed words are modeled with Markov chain and recurrent neural network lexical models, and then inherited versus borrowed word decisions are made based on which model estimates the minimum entropy. The approach is modestly successful, yet not sufficient as a tool to determine lexical borrowing. Professional linguists also take into account substantial multilingual and cross-linguistic lexical and contextual information.

An ambitious and complex borrowed word detection approach by Mi et al. (2020, 2018) identifies borrowed words by Uyghur from Turkish, Arabic, Russian and Chinese donors. Word embeddings are calculated using monolingual language models from corpora of Uyghur and donor languages, and

then cross-lingual embeddings are constructed between Uyghur and donor languages with the help of bilingual dictionaries. The cross-lingual embeddings are used to identify candidate borrowed words between Uyghur and donor languages. Candidate words are tested as borrowed or inherited based on sound similarity and other language features. This approach is not limited to curated wordlists indexed by concepts. Future research planning will consider (Mi et al., 2020, 2018).

List and Forkel (2021) and List (2019b) repurpose lexical cognate identification methods from multilingual wordlists to the task of identifying borrowed words on a newly developed dataset of South-East Asian languages. Cognate detection, by searching for full or partial cognates based on sound similarity, is applied to identify language family internal cognate sets, and then “cognate sets are compared across language families and clustered into sets of potentially borrowed words” (List and Forkel, 2021, pg 4). We hope to integrate this approach into future research.

Another approach to borrowed word detection is to construct phylogenetic models of language families based on wordlists, including also intruder languages that are not necessarily part of the language family. Observed discrepancies in the model, in particular lexical items that detract from hierarchical family relations and contribute instead to lateral transfers, are likely due to borrowed words (List et al., 2014; Delz, 2014).

### Neural Network Language Models

Use of recurrent neural networks to model textual language was popularized by Bengio et al. (2003). The language model is relatively sparse and still useful today with the incorporation of long short-term memory (LSTM) or gated recurrent unit (GRU) cells. Miller et al. (2020) employed a lexical language version of this model.

The attention module by Bahdanau et al. (2015), paired with a recurrent layer, explicitly takes into account the differential contribution of language elements on the position or element being modeled. Vaswani et al. (2017) realized that attention doesn’t require a recurrent layer, “all you need is attention”, resulting in reduced complexity and cost. They constructed a standard Transformer module of attention, add and normalization, and feed forward layers, and incorporated it into their state of the art translation system. Our lexical language models

incorporate a light-weight Transformer module.

Recent advances include Bi-directional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) and many others such as GPT, ELMo, and RoBERT which use millions or billions of parameter language models. This seems overkill since the number of parameters is orders of magnitude greater than a typical wordlist size. However, the method used by BERT of bi-directional training with a masked language model avoids the problem of model dependencies, and similarly the model of Wada et al. (2020) with independent bi-directional models. Both merit further consideration.

### 3 Materials and Methods

#### 3.1 Materials

Wordlists extracted from all 41 language tables of the World Loan Database (Haspelmath and Tadmor, 2009) are used in our experiments. Tables document concept, value, borrowed status, language donor, etc., and consist of 800 to 2,000 words representing some 1,400 concepts from the Intercontinental Dictionary Series (Key and Comrie, 2015). Tables have recently been updated to add International Phonetic Alphabet (IPA) (Moran and Cysouw, 2018) representations of sound segments for all lexical entries (Tresoldi et al., 2019). A snippet from the Imbabura Quechua language table is shown in Table 1.

Concept	Value	Segments	Borrowed	Donor
World	pacha mama	p a tʃ a + m a m a	False	
Valley	yunga	j u ŋ g a	False	
Foam	putsuju	p u t s u x u	False	
Knife	kuchillo	k u tʃ i ʒ u	True	Spanish

Table 1: Snippet of Imbabura Quechua table.

In the experiment, where an additional Spanish donor wordlist is used in training, the Spanish word table from IDS is used as source, since Spanish is not among the 41 WOLD languages. However, the Spanish language word table from IDS was orthographic, so we transcribed the Spanish to segmented IPA, using the Easy Pronunciation portal (Baytukalov, 2019) to produce a phonetic representation of the Spanish wordlist, and then LingPy (List et al., 2018) with cross-linguistic transcription (List et al., 2021) module and a transcription profile to construct segmented IPA.

### 3.2 Methods

#### 3.2.1 Neural models and experiments

**Recurrent model** Miller et al. (2020)’s lexical language model, based on Bengio et al. (2003), was adapted to work with sound segments and estimate lexical entropies using appropriate dropout and regulation (Figure 1). An innovative aspect of their borrowed word detection approach was to train separate models for inherited and borrowed words and have the models compete for which word is inherited or borrowed. Words belong to the model which estimates the lesser entropy for them.

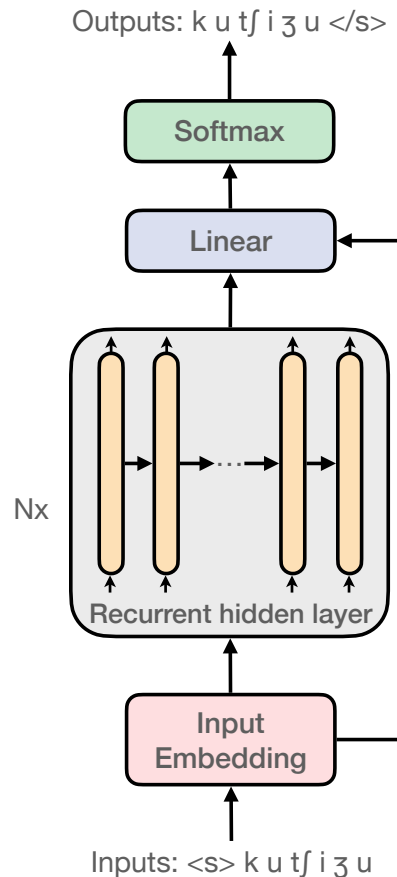


Figure 1: Recurrent Lexical Model

**Light-weight Transformer** Our technology improvement over the baseline model replaces the recurrent layer with a light-weight Transformer module (Vaswani et al., 2017) which includes Attention (Bahdanau et al., 2015) and Transformer features of adding and normalization layers, and a feed forward layer (Figure 2). The model uses a forward only (left-to-right) causal model — this reduces model complexity and avoids unintended dependencies between inputs and outputs.

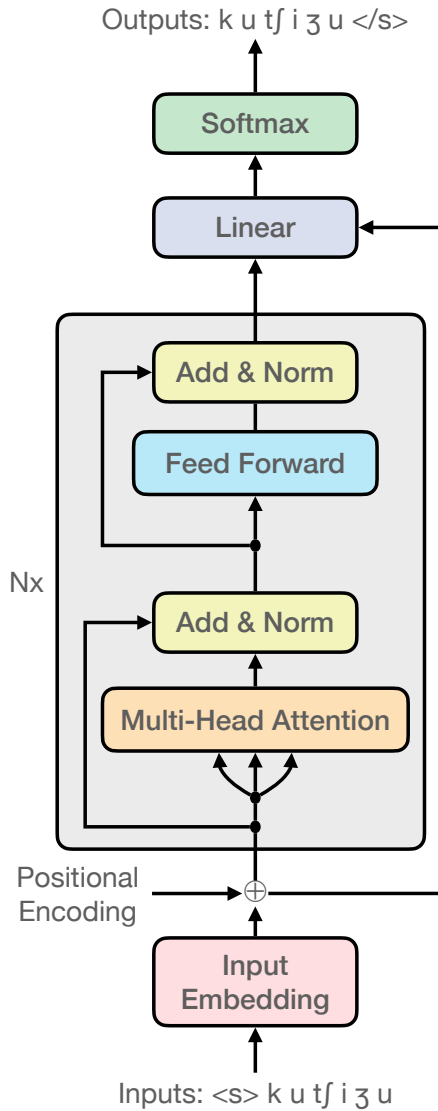


Figure 2: Light-weight Transformer Lexical Model

**Direct model - experiment** We developed a light-weight Transformer to predict inherited versus borrowed words directly from wordlists without the intermediate calculation of inherited and borrowed word entropies. This model flattens Transformer outputs and optionally merges embeddings (see connector on right of Figure 2) before connecting to the softmax layer.

**Word donor - experiment** We broadened the problem definition to include donor source of the borrowed words. Inputs include indication of word donor instead of simply inherited versus borrowed, where  $d = 0$  designates an inherited word, and  $d \in [1, D]$  designates which is the borrowed word donor. A minimum donor wordlist size of 75 words was used to assure enough data to fit donor models; less than 75 word donor sources were combined.

For competing entropy models this results in  $D + 1$  individual entropy models per language — one for inherited words and one for each donor source. In the decision procedure, all models compete for which has the lowest entropy to select the word. For the direct model, only one model is created and it directly discriminates between donors.

**Multiple word models - experiment** With competing entropy models, inherited word models are paired with borrowed word models to compete for which estimates the lower entropy for a word and so claims that word. With  $M$  inherited and borrowed word models each, there are  $M \times M$  model pairs competing to claim words as inherited or borrowed. We developed the multiple models capability, which determines whether a word is inherited or borrowed based on majority vote of the model combinations. Model training and calculation of entropies is much more costly than entropy comparisons,  $C \gg c$ , so cost of training multiple models and taking majority vote on all entropy comparisons is approximately  $(M + M) \times C$ .

**Parameter search - experiment** Each language is unique whether considering the entire language or just phonology and phonotactics in a wordlist. Each lexical model for computing entropy will be different and maybe each configuration of light-weight Transformer parameter settings should be different as well. We developed the capability to run experiments on each training dataset, to determine optimal training parameters based on minimum validation cross-entropy for a lexical model, and then train multiple instances  $M$  of each model at the optimal training parameters. Majority vote from combinations of models trained at optimal conditions determine inherited versus borrowed word status.

This experiment is costly. For each training set, there are  $E$  experiments; for each experiment, there are  $I$  iterations of samples from the training set; for each iteration there are  $R$  replicate runs which return validation cross-entropies. Optimal training parameters for inherited and for borrowed words are used to train  $R$  model replicates on inherited and borrowed words training data. The resulting  $R \times R$  model combinations predict inherited versus borrowed word status based on majority vote, as for multiple word models. So the cost per train-test partition or fold is  $2 \times (E \times I \times R + R)$ .

**Additional donor wordlist - experiment** Lack of sufficient borrowed words for training was a detractor so we enhanced our data methods to permit an additional donor wordlist source. This was a limited experiment where we include a Spanish wordlist from IDS (Key and Comrie, 2015), transcribe it to IPA (§3.1), and combine it with Spanish donor borrowed words for target languages where Spanish is the primary donor language: Imbabura Quechua, Mapudungun, Otomi, Q’eqchi’, Wichí, Yaqui, and Zinacantán Tzotzil. We applied our enhanced data methods with both competing entropy and direct models for these seven languages.

### 3.2.2 Decision and Evaluation procedures

Test F1 score for the borrowed word category is the primary evaluation metric, with precision and recall reported as well. Total time duration of execution, combined training and testing, is a secondary measure.

Experiments ran 10-fold cross-validations over all 41 WOLD language tables except for the additional donor wordlist experiment which used just seven language tables. In the 10-fold cross-validation, wordlists are partitioned into 10 different folds of 9/10 training and 1/10 testing data. Models train on training data, predict borrowing on testing data, and summarize detection results from testing as F1 score, precision and recall over all folds. Results are reported here as the average of F1 score, precision, and recall, plus the execution time, over cross-validations and language tables.

Experiments were run under Windows 10 on an Intel portable with 6 cores (12 logical) and 16 GB of memory. TensorFlow 2.4 was configured to use 8 parallel threads; other user tasks were curtailed.

## 4 Experiments

### 4.1 Competing Entropies and Direct Models

Parameter settings specific to the recurrent model and constant parameter settings for the light-weight Transformer are shown in Table 2. Typical number of parameters per neural model are also reported. Results of the recurrent neural network competing entropies approach from (Miller et al., 2020) and our replication of that analysis are shown in Table 3. Note that Table 3 also reports Transformer parameter values that vary over experiments.

F1 score measures are similar for original and replicate recurrent neural studies. Execution time for the recurrent model is 1 hr and 57 minutes.

Parameter	Recurrent	Transformer
cell_type	GRU	LSTM
N_layers	1	1
embedding_len	32	32
hidden_internal_len	32	32
merge_embedding	True	True
learning_rate_schedule	Decay	Transformer
decay_factor	0.95	
recurrent_l2	0.001	
recurrent_output_dropout	0.2	
merge_embedding_dropout	0.2	
number of parameters	≈ 13,000	≈ 10,000

Table 2: Recurrent and Transformer Model Parameters

### Light-weight Transformer

Results for the light-weight Transformer with the competing entropies approach on inherited versus borrowed words are shown in Table 3. Trials at 50 and 80 training epochs are reported with learning rate and dropouts held constant. A split settings for inherited and borrowed models is also shown with inherited, 50, and borrowed, 80, epochs. F1 scores are 1 percentage point greater for the greater number of epochs. Performance of the light-weight Transformer is on par or minimally better than the recurrent neural model. Adoption of the light-weight Transformer model reduced execution time.

**Inherited versus donor** Since borrowed words can vary substantially by word donor, we speculated that fitting an entropy model on borrowed words by donor language could result in better fitting donor models with a commensurate improvement in overall borrowed word prediction. Results for this experiment at 50 and 80 training epochs, and the split 50, 80 epochs settings, are reported for the competing entropies approach in Table 3. Detection performance remains essentially unchanged or slightly reduced for donor models while execution time increased slightly.

### Direct neural model

The approach of competing entropy models, may be more complicated than necessary. All the data is available for a direct neural model to use in discriminating between inherited and borrowed words.

Results of our experiment with a light-weight Transformer model to directly discriminate between inherited and borrowed words are reported in Table 3. All trials used 120 training epochs and the same learning rate and dropout parameters. For the competing entropies approach, embeddings are merged with recurrent or light-weight Transformer

output before calculating segment probabilities and word entropies. This improved model detection performance. We tested this explicitly with our flattened direct model with merge and no-merge of embeddings.

The direct model scores 5 percentage points lower F1 score than the competing entropies model and takes half again as much time to execute. There is no difference between merge and no-merge of embeddings with the Transformer output.

The direct model for inherited versus donor shows no difference in borrowing detection between inherited versus borrowed and inherited versus donor approaches. Execution times are similar between all experiments on the direct model, and are a little longer than those for the competing entropies model when training for 80 epochs.

## 4.2 Multiple Models and Search

An advantage of the competing entropy models approach is that when there are several entropy models, they can be combined with any of their opposite inherited or borrowed word models. In particular, with three models each of inherited and word models, nine competing entropy model pairs can be formed to make borrowed word decisions.

Table 4 reports results for two experiments on competing entropy models where three models each for inherited and borrowed were trained at the same set of conditions and then a majority vote taken over the nine model combinations for borrowed word decisions. The experiment with 80 training epochs, used different learning rates and embedding dropouts for inherited versus borrowed word models, while the experiment at 50 epochs used constant learning rates and dropouts. Detection results are on par with those reported previously for the single competing entropies approach (Table 3). Execution times are increased by a little more than three times that for the single competing entropies approach.

Since the phonology, phonotactics, language culture and contact situation are different for each language, maybe different training parameters are required for each language wordlist. Table 4 reports results of an experiment on competing entropy models where eight experiments on three parameters are performed. Per the experimental method in §3.2.1 one iteration with three replicates

---

<sup>1</sup>Precision and Recall, in error in the Original "Baseline - recurrent" model, is corrected here.

each are performed for each randomly chosen experimental condition for each fold of the cross-validation. Parameter settings were randomly sampled from the parameter ranges shown in Table 4. Resulting optimal parameter settings are used to perform three replicate trials on their corresponding inherited or borrowed word model. The nine combinations of inherited and borrowed models take a majority vote on borrowed word decisions.

Parameter search does no better than using multiple inherited and borrowed word models trained at fixed parameter settings, which does no better than a single pair of inherited and borrowed models trained at fixed settings. Execution times have exploded to about a day due to the quantity of model training for search models.

## 4.3 Added Spanish Donor Table

Several of the Latin American languages in WOLD (Imbabura Quechua, Mapudungun, Otomi, Q'qechi', Wichí, Yaqui, Zinacantán Tzotzil) have Spanish as the primary and only significant language donor. For each of these languages we added the Spanish wordlist in segmented IPA to the existing borrowed words of the training set, and then trained the models and evaluated test performance for these seven languages in a 10-fold cross-validation. This was a crude attempt to take advantage of data quantity with the hope that Spanish phonotactics would translate sufficiently into Spanish borrowed word phonotactics via our light-weight Transformer model.

Table 5 shows detection results using this method. Also shown are the average of detection results over these same seven languages from the light-weight Transformer with competing entropies and direct approaches. We find at best a slight improvement in F1 score by using the Spanish language wordlist. Recall suffers and precision improves as a result of using the Spanish wordlist. Direct model results remain noncompetitive.

## 5 Discussion

The purpose of this study was to develop and improve tools for automatic or assisted identification of lexical borrowings in monolingual wordlists. We replicated the recurrent lexical model and competing entropies approach of Miller et al. (2020) for detecting borrowed words, producing similar detection results and capturing a baseline execution time.

Experiment	Epochs	Learning rate	Embed dropout	Attention dropout	Transform dropout	F1 score	Precision	Recall	Time hr:min
<b>Baseline - recurrent</b>									
Original <sup>1</sup>	45	0.01				0.603	0.546	0.697	
Replicated	45	0.01				0.604	0.546	0.699	1:57
<b>Competing entropies - Transformer</b>									
Borrowed	50/80	.0075/.0035	0.4/0.2	0.2	0.1	0.616	0.559	0.713	1:12
Borrowed	50	0.0055	0.3	0.2	0.1	0.606	0.548	0.701	1:08
Borrowed	80	0.0055	0.3	0.2	0.1	0.614	0.557	0.706	1:33
Donor	50/80	.0075/.0035	0.4/0.2	0.2	0.1	0.599	0.527	0.724	1:29
Donor	50	0.0055	0.3	0.2	0.1	0.605	0.536	0.720	1:21
Donor	80	0.0055	0.3	0.2	0.1	0.606	0.533	0.730	1:35
<b>Direct - Transformer - flattened</b>									
Borrowed(nomerge)	120	0.0025	0.3	0.3	0.3	0.559	0.480	0.726	1:49
Borrowed(merge)	120	0.0025	0.3	0.3	0.3	0.557	0.506	0.659	1:53
Donor(nomerge)	120	0.0025	0.3	0.3	0.3	0.544	0.447	0.766	1:47
Donor(merge)	120	0.0025	0.3	0.3	0.3	0.552	0.479	0.694	1:52

Table 3: Competing Entropies and Direct Model Experiments - 10 Fold Cross-validation

Epochs	Learning rate	Embed dropout	Attention dropout	Transform dropout	F1 score	Precision	Recall	Time d:hr:min
<b>Multiple entropy models</b>								
80	.0075/.0035	0.4/0.2	0.2	0.1	0.615	0.559	0.710	4:22
50	0.0055	0.3	0.2	0.1	0.617	0.561	0.708	4:08
<b>Parameter search</b>								
50-100	.0025-.0075	0.1-0.5	0.2	0.1	0.621	0.563	0.715	23:53

Table 4: Parameter Search and Multiple Models Experiments - 10 Fold Cross-validation

We developed a light-weight Transformer model (Bahdanau et al., 2015; Vaswani et al., 2017) and observed that it performed minimally better than par versus the recurrent model in borrowed word detection, and was more responsive with reduced execution times. The light-weight Transformer offers a viable base for exploring different lexical borrowing detection approaches.

### 5.1 Competing Entropies and Direct Models

Our meta-level design and analysis, contrasts Transformer with recurrent model results, and forms a 2-factor design of: competing entropies versus direct approaches, and inherited and borrowed versus inherited and word donor approaches.

The light-weight Transformer with competing entropies approach performed five percentage points better than the light-weight Transformer with a direct approach. Competing entropies seems a useful approach to test for lexical borrowings, more so than fitting a larger but less meaningful direct neural model.

With inherited versus donor models, we tested whether modeling donors separately would result in better prediction performance, thinking that treating donors individually might give more coherent training and test subsets. Resulting performance was just on par with the corresponding borrowed word approach; no benefit was conveyed by modeling donors separately. This suggests that any benefit due to modeling more coherent language subsets is offset by the reduced sample size for such subsets. This result seems more likely a problem of insufficient data rather than a dismissal of the utility of modeling donors separately.

### 5.2 Multiple Models and Search

With multiple models and the competing entropies approach, we took the cross product of inherited with borrowed word models and used majority vote to indicate borrowed words. The multiple models approach conferred no advantage. Modeling inherited versus borrowed words seems sufficiently consistent that voting has nothing to add.

Experiment	Epochs	Learning rate	Embed dropout	Attention dropout	Transform dropout	F1 score	Precision	Recall	Time min
<b>Results for Languages with Additional Donor Language Table</b>									
Entropies	100/150	0.0055	0.1	0.1	0.1	0.789	0.841	0.752	52
Direct	120	0.0035	0.3	0.3	0.3	0.735	0.743	0.742	41
<b>Results for Languages without Donor Language Table</b>									
Replicated Baseline - Recurrent						0.777	0.719	0.855	
Competing Entropies Borrowed - Transformer						0.775	0.713	0.855	
Direct Borrowed - Transformer						0.704	0.632	0.830	

Table 5: Additional Spanish Donor Language Table Experiments - 10 Fold Cross-validation - Over Languages: Imbabura Quechua, Mapudungun, Otomí, Q’qechi’, Wichí, Yaqui, Zinacantán Tzotzil

Performing parameter search for training of inherited and borrowed word models, was ineffective and costly. Detection results were no better than for multiple models, and execution time over all 41 WOLD languages was about a day. The single competing entropies approach with light-weight Transformer gives almost as good detection results with just 1.5 hours execution time. Again modeling seems sufficiently consistent within the given parameter sampling ranges that no benefit is gained.

### 5.3 Added Spanish Donor Table

Lack of sufficient data is a major detractor in obtaining good model fits and reproducibility on test cases; this is especially true for highly parameterized models where the parameter counts equal or exceed the data counts. We added training data from a Spanish language wordlist as though they were borrowed words, fake borrowed words, for each of the seven WOLD language tables where Spanish is the primary donor language.

Transformer models learned from actual and fake borrowed words, and produced F1 scores slightly better than par versus the competing entropies approach without data augmentation. Recall decreased and precision increased. This indicates that the Transformer model learned the Spanish wordlist so well that it no longer detected borrowed words that were better adapted to the recipient language. Similarly the better learning prevented inherited words from being confused with Spanish borrowed words. This suggests that with a table of fake borrowed words that conforms more faithfully to word adaption to each language, we could see a meaningful improvement in detection of borrowed words. We observed an improvement for direct detection too, but still with poorer performance versus the competing entropies approach.

## 6 Conclusion

We constructed an effective and responsive light-weight Transformer model as a lexical language model for use in the competing entropies approach for detection of lexical borrowings. Detection performance was on par to minimally better than that using a recurrent neural network model; execution times were reduced. A direct detection model using a light-weight Transformer showed poorer detection of lexical borrowings than the competing entropy models approach. Entropy seems to capture important evidence about lexical items that our direct model does not.

Incorporation of language donor into detection models performed on par with the the corresponding borrowed word models. Lack of sufficient data detracts more from the model than within language coherence contributes. Majority vote over the cross product of inherited and borrowed word models, and parameter search via random experimental trial design, offer no benefit. This vouches for the relative consistency of model results for current parameter ranges and model architecture.

Addition of a donor language wordlist, as though borrowed words, reduced recall and increased precision. With better simulated borrowed words, rather than simple adoption of the donor wordlist, recall and precision should improve. This offers a promising avenue of investigation.

## Acknowledgments

We sincerely thank the reviewers for their helpful and detailed comments. The first author has received funding and encouragement from the Graduate School of the Pontificia Universidad Católica del Perú (PUCP) through the Huiracocha-2019 scholarship program.



## References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). *ICLR*.
- Timur Baytupalov. [Easypronunciation.com](#) [online]. 2019.
- Emily Bender and Alexander Koller. 2020. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 5185–5198, Stroudsburg. Association for Computational Linguistics.
- Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. [A neural probabilistic language model](#). *J. Mach. Learn. Res.*, 3:1137–1155.
- L. Campbell. 2013. *Historical Linguistics: An Introduction*. Edinburgh University Press.
- Marisa Delz. 2014. Mismatches between phylogenetic trees in historical linguistics. Master’s thesis, Eberhard-Karls-Universität, Tübingen, Germany.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [BERT: pre-training of deep bidirectional transformers for language understanding](#). *CoRR*, abs/1810.04805.
- Martin Haspelmath and Uri Tadmor, editors. 2009. *World Loanword Database (WOLD)*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Mary Ritchie Key and Bernard Comrie, editors. 2015. *Intercontinental Dictionary Series (IDS)*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Paul Kiparsky. 2014. [New perspectives in historical linguistics](#). In Claire Bower and Bethwyn Evans, editors, *The Routledge Handbook of Historical Linguistics*, pages 64–102. Routledge.
- Johann-Mattis List. 2019a. [Automatic detection of borrowing \(open problems in computational diversity linguistics 2\)](#). Web blog at: <http://phylonetworks.blogspot.com/2019/03/automatic-detection-of-borrowing-open.html>.
- Johann-Mattis List. 2019b. Automated methods for the investigation of language contact, with a focus on lexical borrowing. *Language and Linguistics Compass*, 12:1–16.
- Johann-Mattis List, Cormac Anderson, Tiago Tresoldi, and Robert Forkel. 2021. [Clts. cross-linguistic transcription systems](#).
- Johann-Mattis List and Robert Forkel. 2021. [Automated identification of borrowings in multilingual wordlists - version 1; peer review: 2 approved, 1 approved with reservations](#). *Open Research Europe*, 1:79.
- Johann-Mattis List, Simon Greenhill, Tiago Tresoldi, and Robert Forkel. 2018. [Lingpy. a python library for historical linguistics. version 2.6.4](#). Web page. With contributions by Christoph Rzymiski, Gereon Kaiping, Steven Moran, Peter Bouda, Johannes Dellert, Taraka Rama, Frank Nagel.
- Johann-Mattis List, Shijulal Nelson-Sathi, Hans Geisler, and William Martin. 2014. Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *Bioessays*, 36(2):141–150.
- Chenggang Mi, Lei Xie, and Yanning Zhang. 2020. [Loanword identification in low-resource languages with minimal supervision](#). *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 19(3).
- Chenggang Mi, Yating Yang, Lei Wang, Xi Zhou, and Tonghai Jiang. 2018. [A neural network based model for loanword identification in Uyghur](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- John E. Miller, Tiago Tresoldi, Roberto Zariquiey, César A. Beltrán Castañón, Natalia Morozova, and Johann-Mattis List. 2020. [Using lexical language models to detect borrowings in monolingual wordlists](#). *PLOS ONE*, 15(12):1–23.
- Steven Moran and Michael Cysouw. 2018. *The Unicode Cookbook for Linguists: Managing writing systems using orthography profiles*. Zenodo. This is the Language Science Press publication version.
- Tiago Tresoldi, Robert Forkel, and Natalia Morozova. 2019. *CLDF dataset derived from Haspelmath and Tadmor’s “World Loanword Database” from 2009*. Zenodo, Geneva.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, volume abs/1706.03762.
- Takashi Wada, Tomoharu Iwata, Yuji Matsumoto, Timothy Baldwin, and Jey Han Lau. 2020. [Learning contextualised cross-lingual word embeddings for extremely low-resource languages using parallel corpora](#).