



Phonology-Augmented Statistical Transliteration for Low-Resource Languages

Hoang Gia Ngo¹, Nancy F. Chen², Nguyen Binh Minh¹, Bin Ma², Haizhou Li²

¹National University of Singapore, Singapore

²Institute for Infocomm Research, Singapore

ngohgia@u.nus.edu, nfychen@i2r.a-star.edu.sg
 nguyen.binh.minh92@u.nus.edu, {mabin, hli}@i2r.a-star.edu.sg

Abstract

Transliteration converts words in a source language (e.g., English) into phonetically equivalent words in a target language (e.g., Vietnamese). This conversion needs to take into account phonology of the target language, which are rules determining how phonemes can be organized. For example, a transliterated word in Vietnamese that begins with a consonant cluster is phonologically invalid. While statistical transliteration approaches have been widely adopted, most do not explicitly model the target language’s phonology, and thus produce invalid outputs. The problem is compounded for low-resource languages where training data is scarce. In this work, we present a phonology-augmented statistical framework suitable for languages with minimal linguistic resources. We propose the concept of pseudo-syllables as structures representing how segments of a foreign word are arranged according to the target language’s phonology. We use Vietnamese, a tonal language with monosyllabic structure as an example. We show that the proposed system outperforms the statistical baseline by up to 70.3% relative, when there are limited training examples (94 word pairs). We also investigate the trade-off between training corpus size and transliteration performance of different methods on two distinct corpora.

Index Terms: machine translation, under-resourced languages

1. Introduction

In every language, new words are constantly being invented or borrowed from foreign languages, especially in colloquial speech. For example, ‘facebook’ is now part of the vocabulary in many languages other than English. These new words present out-of-vocabulary (OOV) challenges to spoken language technologies such as automatic speech recognition [1], keyword search [2], and text-to-speech [3]. Transliteration is a mechanism for modeling OOV’s adopted from foreign languages. Transliteration converts words written in one writing system (source language, e.g., English) into phonetically equivalent words in another writing system (target language, e.g., Vietnamese) [4], and is often used to translate foreign names of people, locations, organizations, and products. For example, the English word *facebook* can be transliterated to Vietnamese phonemes using X-SAMPA notation [5]: f @I _1 . b_< u k _2 (“.” is a syllable delimiter).

Letter-to-sound tools employing statistical approaches are common solutions for OOV’s [6][7]. Transliteration deals with two different language systems, with the input as letters from a source language, and output as phonemes of a target language. Transliteration outputs therefore need to comply with phonological rules of the target language. Such phonological rules of OOV’s adopted from foreign words are often difficult to model

because compared to OOV’s originating from the native language, their numbers are smaller in size. This often results in non-interpretable outputs by statistical transliteration models [8]. The problem is compounded in low-resource languages such as Vietnamese [9]. Given the limited corpus size of such languages, performance of statistical transliteration approaches is suboptimal. On the other hand, rule-based transliteration approaches have been shown to produce phonologically-valid outputs with little training resources [10]. However, rule-based approaches have their performance limited by the complexity of the predefined rules, and therefore, under-perform with larger datasets, as compared to statistical methods [10].

We propose a transliteration framework in which the statistical n-gram language modeling is augmented with phonological knowledge. We propose the concept of pseudo-syllables in statistical models to impose phonological constraints of syllable structure in the target language, yet retain acoustic authenticity of the source language as closely as possible. We assess the transliteration performance of the proposed framework with existing baseline approaches on low-resource languages, using Vietnamese as an example. We further investigate their performance with corpora of different dialects and various training data sizes. Our proposed framework integrates advantages of rule-based approaches on top of classical statistical transliteration models. The proposed approach ensures phonologically-valid outputs, while maintaining strengths of statistical models (e.g., language-independent, performance scales up with increase in training data size). We are working on applying the framework to other languages such as Cantonese and Mandarin.

2. Background

2.1. Phonology

We introduce two phonological concepts essential to transliteration.

i. Syllable: A syllable is considered the smallest phonological unit of a word [11] with the following structure [12]:

$$[O] N [Cd] + [T] \tag{1}$$

where the “[]” specifies an optional unit. *O* denotes the Onset, which is a consonant or a cluster of consonants at the beginning of a syllable. *N* denotes the Nucleus, which contains at least a vowel. *Cd* denotes the Coda, which mostly contains consonants. *T* denotes lexical tone, a feature existing in many languages to distinguish different words. The syllabic structure above is shared across most languages [12]. However, how consonants (\mathcal{C}) and vowels (\mathcal{V}) constitute Onset, Nucleus, Coda differs across languages. For example, in English, an Onset can be a consonant cluster, such as ‘kl’, while no consonant cluster can be the Onset of a syllable in Vietnamese [13][14].

10.21437/Interspeech.2015-728

ii. *Lexical Tones*: In tonal languages, pitch is used to distinguish the meaning of words which are phonetically the same. This distinctive pitch level or contour is referred to as lexical tones [15]. For instance, there are 6 distinct lexical tones in Vietnamese [16], and 4 distinct lexical tones in Mandarin [17]. Each lexical tone is commonly encoded in phonetic representation with a number. For example, consider two different Vietnamese words: b_< 0 _3 (cow) and b_< 0 _6 (bug). The two words are represented in phonetic units using X-SAMPA notation [5], and have the same Onset (b_<) and Nucleus (0), but are distinguished by the two different lexical tones (tone 3 and tone 6). Around 70% of languages are tonal [15], concentrating in Africa, East and Southeast Asia [18].

2.2. Baseline Model: Joint Source-Channel Model

The joint source-channel model for transliteration was introduced in [19]. Similar approaches have also been proposed for grapheme-to-phoneme conversion in [20].

The transliteration problem can be formulated as follows: given a string of graphemes $\mathbf{f} = (f_1, f_2, \dots, f_m)$ from source language F , the objective is to convert this input string into a string of phonemes $\mathbf{e} = (e_1, e_2, \dots, e_n)$ in target language E . The string of phonemes is inferred to:

$$\mathbf{e}^* = \arg \max_{\mathbf{e}} p(\mathbf{e}|\mathbf{f}) = \arg \max_{\mathbf{e}} p(\mathbf{e}, \mathbf{f}). \quad (2)$$

The transliteration process can be seen as finding the alignment for sub-sequences of the input string \mathbf{f} and the output string \mathbf{e} [21]. Let there be K aligned transliteration pairs $\langle x, y \rangle_1, \dots, \langle x, y \rangle_K$, where x takes on sub-sequences of \mathbf{e} and y takes on sub-sequences of \mathbf{f} . The joint probability $p(\mathbf{e}, \mathbf{f})$ is estimated through an N-gram model, where the k -th alignment pair $\langle x, y \rangle_k$ is dependent on its N predecessor pairs [19][20]:

$$p(\mathbf{e}, \mathbf{f}) = \prod_{k=1}^K P(\langle x, y \rangle_k | \langle x, y \rangle_{k-N+1}, \dots, \langle x, y \rangle_{k-1})$$

3. Proposed Phonology-Augmented Statistical Framework

3.1. Motivation

Figure 1 shows how the English word “KLEINHANS” is transliterated into Vietnamese. Units 3, 7, 12 and 16 denote Vietnamese lexical tones. Units 4, 8 and 13 are delimiters between syllables. The arrows approximate the correspondence between the English graphemes and the Vietnamese phonemes.

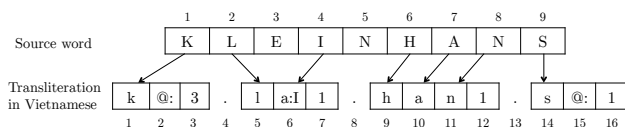


Figure 1: Standard statistical transliteration model (no explicit phonological knowledge)

Some observations made from this example are:

1. Lexical tones are added to the transliteration output (encoded by the numbers at position 3, 7, 12 and 16).
2. The source word and its transliterated version can have different numbers of syllables: the original word has 2 syllables, while the transliterated output has 4.
3. There might not be an obvious correspondence between a segment in the source word and a target phoneme: either segment “EI” or “EIN” of the source word can correspond to phoneme /a:I/ at position 6.

4. New phonemes can be added to the output to retain the phonological structure of the source language: a ‘schwa’ is added to imitate the pronunciation of the consonant cluster “KL” that does not have an equivalent in Vietnamese.

In existing statistical transliteration models, the phonological considerations listed above are implicitly modeled using n-gram language modeling of the source language graphemes [22], or joint sequences of graphemes and phonemes [19][20]. Due to limited training data in low-resource languages, phonological structure in the transliteration output is not well-modeled, resulting in a high rate of invalid syllables in the output [10]. For example, k l E i _1 . J a: n is another transliteration output in Vietnamese phonemes produced by a statistical model for the word “KLEINHANS” in Figure 1. The output is invalid in Vietnamese phonology since the first syllable has a consonant cluster, and the second syllable has no lexical tone. Empirically, at least 21% of transliterated entries lack lexical tones, when we run Sequitur G2P [23] (a statistical system) on 100 Vietnamese transliteration pairs extracted from OpenKWS13 corpus [24]. Previous rule-based approaches have been shown to improve transliteration performance by imposing phonological constraints on their outputs [10]. However, predefined rules are likely to make mistakes with words not observed in training data. Rule-based approaches are thus outperformed by statistical models in larger data sets [10].

Our proposed approach has the strengths of both the rule-based and statistical approaches by integrating the phonology of the target language explicitly with a statistical model. The proposed framework can thus generate transliterated outputs that are phonologically valid yet improves the performance with more data. Figure 2 shows the three steps of the proposed approach: (1) pseudo-syllable formulation, (2) grapheme-to-phoneme mapping, and (3) lexical tone assignment.

3.2. Pseudo-Syllable Formulation

Pseudo-syllable is a representation of how segments of a foreign word are arranged according to the syllable structure specified by the target language’s phonology. The concept is inspired by how native speakers process a foreign loan word by imposing native phonological constraints on the foreign word’s form [25]. The structure of a pseudo-syllable is defined as: $s_i = \{[O_i] N_i [Cd_i]\}$, where O_i, N_i, Cd_i are the Onset, Nucleus and Coda of the i -th pseudo-syllable respectively. Onset, Nucleus and Coda can be seen as sub-syllable roles assigned to different parts of a syllable.

Extra unit representing the ‘schwa’ can be added to form pseudo-syllables to maintain acoustic authenticity of the source language, such as consonant clusters and fricative finals. This is a phenomenon known as epenthesis, commonly observed among loanwords in languages such as Vietnamese [16], Japanese [25], and many more [26][27][28]. For example, given the foreign word $\mathbf{f} = \text{“KLEINHANS”}$, a possible sequence of

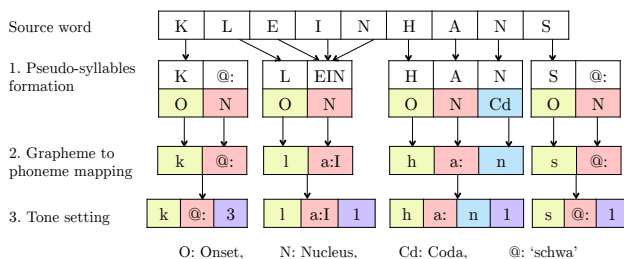


Figure 2: Phonology-augmented statistical framework

pseudo-syllables is: $s_1=\{K, @:\}$ with $O_1=\{K\}$, $N_1=\{@:\}$; $s_2=\{L, EIN\}$ with $O_2=\{L\}$, $N_2=\{EIN\}$; $s_3=\{H, A, N\}$ with $O_3=\{H\}$, $N_3=\{A\}$, $Cd_3=\{N\}$; $s_4=\{S, @:\}$ with $O_4=\{S\}$, $N_4=\{@:\}$, where “@:” is the graphemic representation whose phonemic counterpart is the vowel schwa [ə:].

i. Training: Given a foreign word $\mathbf{f} = [f_1, f_2, \dots, f_m]$ and its transliterated output $\mathbf{e} = [e_1, e_2, \dots, e_n]$, let \mathcal{T}_i be the function to assign a letter f_i a sub-syllable role in a pseudo-syllable. For example, the function to assign the first letter of \mathbf{f} to the onset of the second pseudo-syllable is $\mathcal{T}_1 = O_2$. During training, a beam search is used to find the role-assignment \mathcal{T} for all letters of the source word such that the sequence of resulting pseudo-syllables has the same structures as those of the reference transliterated output. Using the word “KLEINHANS” as an example, the assignment function for all letters of the word is $\mathcal{T} = [ON_1, O_2, N_2, N_2, N_2, O_3, N_3, Cd_3, ON_3]$, which results in 4 pseudo-syllables of structures $[O, N]$, $[O, N]$, $[O, N, Cd]$, $[O, N]$, matching the syllable structures in the reference Vietnamese output. Note that the role ON_i assigned to a consonant indicates that the consonant takes the Onset role of the i -th pseudo-syllable and a ‘schwa’ is added as the nucleus.

ii. Decoding: During decoding, the role assignment function \mathcal{T} for all letters of \mathbf{f} is estimated as:

$$\mathcal{T}^* = \arg \max_{\mathcal{T}} p(\mathcal{T} | \mathbf{f}) \quad (3)$$

We model \mathcal{T} as a Markov chain of n -gram, with n an even number:

$$\mathcal{T}^* = \arg \max_{\mathcal{T}} \prod_{i=1}^n p(\mathcal{T}_i | (f_{i-n/2}, \dots, f_{i+n/2})) \quad (4)$$

iii. Smoothing: \mathcal{T}_i in Eq. (4) is estimated from a weighted score of probability of smoothed n -grams:

$$q(\mathcal{T}_i) = \sum_{l=0}^{n/2} \sum_{k=0}^{n/2} \omega_{k,l} p(\mathcal{T}_i | (f'_{i-n/2}, \dots, f'_{i+n/2})) \quad (5)$$

where $(f'_{i-n/2}, \dots, f'_{i+n/2})$ is a smoothed n -gram such that given $0 \leq k, l \leq n/2$ (n is an even number):

$$f'_j = \begin{cases} f_j & \text{if } i - n/2 + l \leq j \leq i + n/2 - k \\ \mathcal{C} & \text{if } f_j \text{ is a consonant} \\ \mathcal{V} & \text{if } f_j \text{ is a vowel} \end{cases}$$

For example, given the tri-gram of letters (B, E, S), its smoothed n -grams are $\{(B, E, S), (\mathcal{C}, E, S), (\mathcal{C}, E, \mathcal{C}), \dots, (\mathcal{V})\}$.

$\omega_{k,l}$ is the weight corresponding to the smoothed n -gram, with $\omega_{k,l} > \omega_{k',l'}$ for $k+l < k'+l'$; for example: $\omega_{1,0}$ (corresponding to (\mathcal{C}, E, S)) $>$ $\omega_{1,1}$ (corresponding to $(\mathcal{C}, E, \mathcal{C})$).

3.3. Grapheme-to-Phoneme Mapping

Transliteration takes into account (1) graphemes of the source word, (2) how the word is pronounced in the source language, and (3) phonological rules of the target language [16][25]. Such relation can be formalized with the following equation:

$$\mathbf{e}^* = \arg \max_{\mathbf{e}} p(\mathbf{e} | \mathbf{f}, \mathbf{v}, \mathcal{L}_f) \quad (6)$$

where \mathbf{e} is the phonemes of the transliterated word in the target language, \mathbf{f} and \mathbf{v} are the graphemes and phonemes of the original word from the source language, and \mathcal{L}_e is the phonological rules of the target language. This relationship is simplified in the proposed model by introducing the sub-syllable role r :

$$\mathbf{e}_{i,r} = \arg \max_{e_{i,r}} p(e_{i,r} | f_i, v_i) \quad (7)$$

with $\mathbf{e}_{i,r}$ as the phoneme at position $r \in \{\text{Onset, Nucleus, Coda}\}$, of the i^{th} syllable in the transliterated word. f_i, v_i are the orthographic and phonetic form of the i^{th} pseudo-syllable.

$p(e_{i,r} | f_i, v_i)$ is learned from the training data prepared in Section 3.2. Pseudo-syllables formulated in Section 3.2 provides a simplified model of the phonological constraints \mathcal{L}_e by modeling (1) valid syllables structures in the transliteration output, and (2) valid phonemes for each sub-syllable role r .

3.4. Lexical Tone Assignment

In most statistical transliteration models, lexical tones are treated the same as phonetic units of the transliterated output. Thus, lexical tone assignment depends on large amounts of training data to be correct [19][29]. Previous studies using tone assignment to complement output of statistical models have shown improvement in transliteration performance [30][31].

Phonology studies show that in many tonal languages, the assignment of lexical tone to a syllable is influenced by phonemes in that syllable [16][32], and lexical tones assigned to adjacent syllables (tone sandhi) [33][34]. In this work, we attempt to use tonal and phonetic context to model tone assignment to a syllable, with no dependence on the target language.

Let \mathbf{t} be the lexical tone assignment for a transliterated word, t_i be the lexical tone assigned to the i^{th} syllable of the transliterated output.

$$\mathbf{t} = \arg \max_{\mathbf{t}} \prod_i p(t_i | t_{i-1}, (e_{i,O}, e_{i,N}, e_{i,Cd}), t_{i+1}) \quad (8)$$

with $(e_{i,O}, e_{i,N}, e_{i,Cd})$ as phonemes at the Onset, Nucleus, and Coda sub-syllable role of the i -th syllable in the output. The training data to model lexical tones is the reference transliterated output used in Section 3.2 and 3.3.

4. Transliteration Experiments

4.1. Experimental Setup

i. Low-Resource Corpus: NIST OpenKWS13 Evaluation:

The Vietnamese corpus is released from the IARPA Babel program [24] for the NIST OpenKWS13 Evaluation, with a setup similar to that in [10]; all experiments share the same test set of 140 words.

ii. Implementation Details: The Joint source-channel model was implemented with Sequitur G2P [23]. A rule-based framework proposed in [10] was used as another reference. In our proposed model, we use a text-to-phoneme tool for American English [35] to generate the source language’s phonemes in Section 3.3 because the majority of words in the corpus are of English-origin.

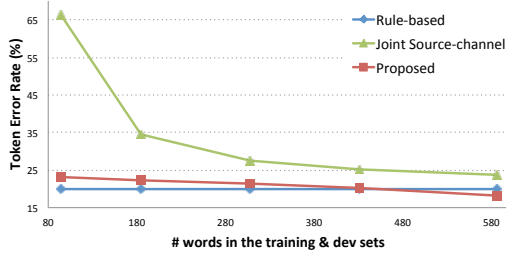
iii. Evaluation Metrics:

- 1) **Token error rate (TER):** tokens include both phones and syllable delimiters.
- 2) **Invalid syllable rate (ISR):** a syllable is invalid if it does not exist among syllables extracted from the lexicon.
- 3) **String error rate (SER):** any error within a string results in string error. TER and SER are computed using SCLITE [36].

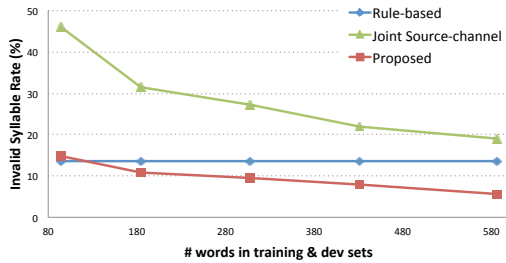
4.2. Results

From Figures 3a, 3b, and 3c, we see that the proposed model consistently outperforms the statistical baseline for all three evaluation metrics, across all the training set sizes.

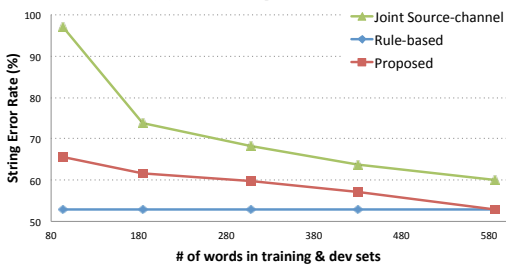
We analyze in more details the performance of the different models at different low-resource setups. Table 1 shows that the proposed system outperforms the baselines at the largest training data set (588 word pairs): it improves the Rule-based model by 8.59% relative in TER, and is on par in ISR and SER; it improves the Joint source-channel model by 22.46% relative in



(a) Token error rate



(b) Invalid syllable rate



(c) String error rate

Figure 3: Low-resource scenario: Performance of different transliteration models as a function of corpus size (NIST OpenKWS13 dataset).

TER, by 70.3% relative in ISR, and by 10.67% in SER. Table 2 shows that the proposed system also outperforms the statistical baseline at the smallest training data set (94 word pairs): it improves Joint Source-channel model by 64.90% relative in TER, by 67.76% relative in ISR, and by 32.34% relative in SER.

The error rates also validate the phonological basis of our proposed approach. As the proposed model better captures syllabic structures and syllables’ boundaries, evident by the much lower ISR compared to the baseline statistical model at low-resource scenarios, the overall SER of the output is also lower.

4.3. Further Analysis: Data size vs. Performance Trade-off

To gain deeper insight into the effectiveness of the proposed approach in comparison with the baselines, we analyzed corpus size and performance trade-off using a larger corpus (HCMUS corpus [9]). The setup is the same as [10].

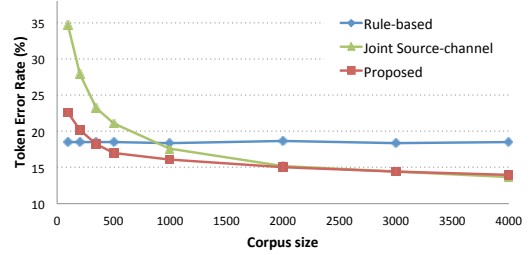
Table 1: Transliteration performance at 588 training word pairs.

Model	TER (%)	ISR (%)	SER (%)
Joint source channel model	23.6	19.0	60.0
Rule-based	19.8	13.6	52.9
Proposed model	18.1	5.63	52.9

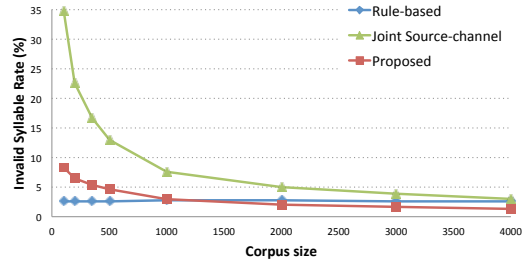
Table 2: Transliteration performance at 94 training word pairs^a.

Model	TER (%)	ISR (%)	SER (%)
Joint source channel model	66.1	46.0	97.1
Rule-based	19.8	13.6	52.9
Proposed model	23.2	14.8	65.7

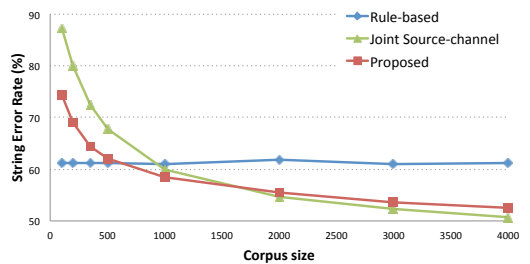
^aSince the rule-based approach is not statistical in nature, its performance remains constant regardless of the training set size in Figure 3a-3c and Table 1-2.



(a) Token error rate



(b) Invalid syllable rate



(c) String error rate

Figure 4: Performance of different transliteration models as a function of training data size (HCMUS corpus [9]).

As shown in Figures 4a-4c, the proposed model performs better than the rule-based model in TER for all corpus sizes larger than 350 words. It is also better than the rule-based model in all three metrics for all corpus sizes larger than 1,000 words. The proposed algorithm performs consistently better than the Joint source-channel model for corpus size up to 1,500 words, and is on par for corpus sizes larger than 1,500 words. In Figure 4b, the consistently low ISR across all data sets produced by our proposed model demonstrates the model’s ability to infer the most common syllables. However, the less typical syllables tend to be missed, evident by the earlier plateauing of the proposed model’s performance compared to the baseline statistical model in Figure 4c. As less typical syllables can still be modeled by the baseline joint source-channel model which permits joint-units longer than one sub-syllabic unit, the baseline model is better at reproducing such syllables during decoding.

5. Conclusion

By augmenting the statistical n-gram language modeling with phonological knowledge, our proposed framework ensures the transliteration outputs are valid, resulting in lower error rates compared with baseline approaches in low-resource scenarios. We are working on generalizing the phonological constraints from syllable level to word level in order to further improve the transliteration accuracy. Furthermore, in this work, we used Vietnamese to demonstrate how often neglected phonological constraints like syllables and lexical tones can be incorporated into our proposed approach. Since such phonological constraints are shared across most languages, we are applying the proposed approach to more languages such as Cantonese and Mandarin.

6. References

- [1] André Mansikkaniemi and Mikko Kurimo, “Unsupervised Vocabulary Adaptation for Morph-based Language Models,” in *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, Stroudsburg, PA, USA, 2012, WLM '12, pp. 37–40, Association for Computational Linguistics.
- [2] Nancy F. Chen, Sunil Sivadas, Boon Pang Lim, Hoang Gia Ngo, Haihua Xu, Van Tung Pham, Bin Ma, and Haizhou Li, “Strategies for Vietnamese Keyword Search,” in *JCASSP*, 2014.
- [3] Robert Eklund and Anders Lindstrom, “How To Handle ‘Foreign Sounds’ in Swedish Text-to-Speech Conversion: Approaching the ‘Xenophone’,” in *Problem. Proc. of the International Conference on Spoken Language Processing*, 1998.
- [4] Kevin Knight and Jonathan Graehl, “Machine transliteration,” *Computational Linguistics*, vol. 24, no. 4, pp. 599–612, Dec. 1998.
- [5] John C. Wells, “Computer-coding the IPA: a proposed extension of SAMPA: <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>,” 2001.
- [6] Wei Gao, Kam-Fai Wong, and Wai Lam, “Phoneme-based transliteration of foreign names for OOV problem,” in *Proceedings of the First International Joint Conference on Natural Language Processing*, Berlin, Heidelberg, 2005, IJCNLP'04, pp. 110–119, Springer-Verlag.
- [7] Paola Virga and Sanjeev Khudanpur, “Transliteration of proper names in cross-language applications,” in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2003, SIGIR '03, pp. 365–366, ACM.
- [8] Kyuchul Yoon and Chris Brew, “A linguistically motivated approach to grapheme-to-phoneme conversion for Korean,” *Computer Speech & Language*, vol. 20, no. 4, pp. 357 – 381, 2006.
- [9] Nam X. Cao, Nhut M. Pham, and Quan H. Vu, “Comparative Analysis of Transliteration Techniques Based on Statistical Machine Translation and Joint-sequence Model,” in *Proceedings of the 2010 Symposium on Information and Communication Technology*, New York, NY, USA, 2010, SoICT '10, pp. 59–63, ACM.
- [10] Hoang Gia Ngo, Nancy F. Chen, Sunil Sivadas, Bin Ma, and Haizhou Li, “A minimal-resource transliteration framework for Vietnamese,” in *INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, 2014*, 2014, pp. 1410–1414.
- [11] Peter Ladefoged and Keith Johnson, *A course in phonetics*, Cengage learning, 2014.
- [12] Brett Kessler and Rebecca Treiman, “Syllable structure and the distribution of phonemes in english syllables,” *Journal of Memory and Language*, vol. 37, no. 3, pp. 295–311, 1997.
- [13] A. C. Gimson, *An Introduction to the Pronunciation of English*, St. Martin's Press, New York, 1970.
- [14] Dinh-Hoa Nguyen, “Vietnamese,” in *The World's Major Languages*, B. Comrie, Ed., pp. 777–796. Oxford University Press, Oxford, 1990.
- [15] M. Yip, *Tone*, Cambridge Textbooks in Linguistics. Cambridge University Press, 2002.
- [16] Thi-Quynh-Hoa Hoang, “A phonological contrastive study of Vietnamese and English,” MA Thesis, Texas Technology College, 1965.
- [17] C Julian Chen, Ramesh A Gopinath, Michael D Monkowski, Michael A Picheny, and Katherine Shen, “New methods in continuous Mandarin speech recognition,” in *Proc. of Eurospeech*, 1997.
- [18] Ian Maddieson, *Tone*, Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.
- [19] Haizhou Li, Min Zhang, and Jian Su, “A Joint Source-channel Model for Machine Transliteration,” in *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, 2004, ACL '04, Association for Computational Linguistics.
- [20] Maximilian Bisani and Hermann Ney, “Joint-sequence models for grapheme-to-phoneme conversion,” *Speech Communication*, vol. 50, no. 5, pp. 434 – 451, 2008.
- [21] Vladimir Pervouchine, Haizhou Li, and Bo Lin, “Transliteration alignment,” in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 2009, pp. 136–144.
- [22] Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer, “The Mathematics of Statistical Machine Translation: Parameter Estimation,” *Comput. Linguist.*, vol. 19, no. 2, pp. 263–311, June 1993.
- [23] Maximilian Bisani, “Sequitur G2P: A trainable Grapheme-to-Phoneme converter: <http://www-i6.informatik.rwth-aachen.de/web/Software/g2p.html>,” 2011.
- [24] NIST, “Open Keyword Search 2013 (OpenKWS13) Evaluation: <http://www.nist.gov/itl/iad/mig/openkws13.cfm>,” 2013.
- [25] Daniel Silverman, “Multiple scansions in loanword phonology: evidence from Cantonese,” *Phonology*, vol. 9, 1992.
- [26] Yvan Rose and Katherine Demuth, “Vowel epenthesis in loanword adaptation: Representational and phonetic considerations,” *Lingua*, vol. 116, no. 7, pp. 1112–1139, 2006.
- [27] C. Uffmann, *Vowel Epenthesis in Loanword Adaptation*, Linguistische Arbeiten. De Gruyter, 2007.
- [28] Suyeon Yun, “Perceptual similarity and epenthesis positioning in loan adaptation,” in *Chicago Linguistic Society*, 2012, vol. 48.
- [29] Paola Virga and Sanjeev Khudanpur, “Transliteration of proper names in cross-lingual information retrieval,” in *Proceedings of the ACL 2003 Workshop on Multilingual and Mixed-language Named Entity Recognition - Volume 15*, Stroudsburg, PA, USA, 2003, MultiNER '03, pp. 57–64, Association for Computational Linguistics.
- [30] Oi Yee Kwong, “Homophones and Tonal Patterns in English-Chinese Transliteration,” in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Stroudsburg, PA, USA, 2009, ACLShort '09, pp. 21–24, Association for Computational Linguistics.
- [31] Yan Song, Chunyu Kit, and Hai Zhao, “Reranking with Multiple Features for Better Transliteration,” in *Proceedings of the 2010 Named Entities Workshop*, Stroudsburg, PA, USA, 2010, NEWS '10, pp. 62–65, Association for Computational Linguistics.
- [32] J. Setter, C.S.P. Wong, and B.H.S. Chan, *Hong Kong English, Dialects of English*. Edinburgh University Press, 2010.
- [33] Kent A Lee, *Chinese Tone Sandhi and Prosody*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1997.
- [34] Takenobu Tokunaga, Dain Kaplan, Chu-Ren Huang, Shu-Kai Hsieh, Nicoletta Calzolari, Monica Monachini, Claudia Soria, Kiyooki Shirai, Virach Sornlertlamvanich, Thatsanee Charoenporn, et al., “Adapting international standard for Asian language technologies,” *Adapting International Standard for Asian Language Technologies*, 2008.
- [35] Kevin Lenzo, “t2p: Text-to-Phoneme Converter Builder. Retrieved from Carnegie Mellon University: <http://www.cs.cmu.edu/afs/cs.cmu.edu/user/lenzo/html/areas/t2p/>,” 1998, December 28.
- [36] NIST, “Evaluation Tools: <http://www.itl.nist.gov/iad/mig/tools/>,” 2007.