

Conférence internationale:

**The 32nd Pacific Asia Conference on Language, Information and Computation
(PACLIC 32)**

Dec 1-3, 2018 @ The Hong Kong Polytechnic University, Hong Kong SAR

Site Web: <http://www.cbs.polyu.edu.hk/2018pacific/call-for-papers.php>

Étudiant: **Ngoc Tan Le**

Formation: **Doctorat en Informatique Cognitive**

Membre junior de l'Institut des Sciences Cognitives

Université du Québec à Montréal

Objectif de la conférence

The long tradition of PACLIC conferences emphasizes the synergy of theoretical analysis and processing of language - from theoretical frameworks to cognitive accounts, from lexical processing to language understanding, and from computational modelling to multi-lingual applications.

The most important purpose of PACLIC conferences is to provide a forum where researchers in different fields of language study in different areas in the Pacific-Asia region working on issues pertaining to different languages can come together and talk, get to know each other, learn old wisdom, be enlightened by new insights and generally get entertained intellectually, and come home ready to initiate a new research program with new research partners in a new state of mind.

They value hybrid talks on linguistic principles and implementation details, massive data collection and extraction of abstract rules, automated proficiency evaluation and philosophical contemplation on language learning, although pure theory and pure technology will also be appreciated. They welcome bi-lingual or multi-lingual research, while due respects will be paid to mono-lingual research.

Résumé du contenu de la communication

Grapheme-to-phoneme models are key components in automatic speech recognition and text-to-speech systems. With low-resourced language pairs that do not have available and well-developed pronunciation lexicons, grapheme-to-phoneme models are particularly useful. The current work presents an approach that applies an alignment representation for input sequences and pre-trained source and target embeddings to overcome the transliteration challenge for a low-resourced languages pair.

In this paper, we presented an approach for machine transliteration in low resource settings, based on the alignment representation for input sequences and pre-trained source and target embeddings. This approach was applied experimentally to the French-Vietnamese low-resource language pair in a transliteration task. The results showed promising improvement compared to the state-of-the-art approaches, with a large increase of +7.30 BLEU and a reduction in translation error rate (TER) of -8.16 and phoneme error rate (PER) of -14.17.

Eventually, this approach could be extended to any low-resources language pair when a bilingual pronunciation dictionary is available. Therefore, this method is extremely useful for under-resourced languages for which training data is difficult to find.