**Prasanna Raj Noel Dabre (黒橋教授)**
**「Exploiting Multilingualism and Transfer Learning for Low Resource Machine Translation」**
**(低リソース機械翻訳における多言語性と転移学習の活用)**
**平成 30 年 3 月 26 日授与**

Machine Translation (MT) is an application of Natural Language Processing (NLP) that focuses on the automatic translation between languages. Even if a language pair is resource rich, the amount of data, known as parallel corpora, for a specific domain is rather limited. One solution is to develop techniques that can leverage the data for several different language pairs or domains to improve the translation quality for a given language pair or domain by transfer learning. Another solution is to develop techniques to obtain better translation system using no additional data and instead rely on multilingualism as a form of redundancy.

While most research works claim that multilingualism is important, they do not explore more than two or three languages at a time. Although, it is important to design and develop techniques that improve upon existing models, it has been observed that black-box approaches that rely on simple pre and post processing tend to work as well as other sophisticated techniques. In this thesis, we focus on leveraging the power of multiple source and pivot languages, multiple domains, monolingual corpora and linguistic redundancy in Neural Machine Translation (NMT) and Phrase Based Statistical Machine Translation (PBSMT) settings to improve the translation quality in a resource poor scenario.

We begin the thesis with a history of MT followed by a literature survey of PBSMT and NMT. We then perform a case study of Japanese-Hindi translation and show how using multiple intermediate languages in a N-lingual corpus setting can help improve the quality of translation. We then address the problem of noise in the pivot language setting by statistical significance pruning followed by an efficient post processing technique that relies on features obtained using a deep neural network.

We proceed to first develop simple but effective transfer learning techniques to develop a single translation model for multiple domains for a particular language pair. We then attempt to leverage a resource rich language pair to improve the translation quality of a resource poor language pair. We exhaustively experiment with over 6 resource rich languages and 7 resource poor languages in a number of settings which include monolingual corpora and quantitatively show how using related languages is important. We then show that multi-source MT, a special case of multimodal NLP, can be done by simply concatenating the multiple sentences instead of modifying the NMT architecture. We also show how the multi-source model can be used for transfer learning.

Because NMT architectures are still evolving, our approaches will act as a guideline for studying the relationships between languages as well as promote the development of multilingual resources. Following a listing of the conclusions of our work, the final chapter gives an overview of the future work taking this point into consideration.

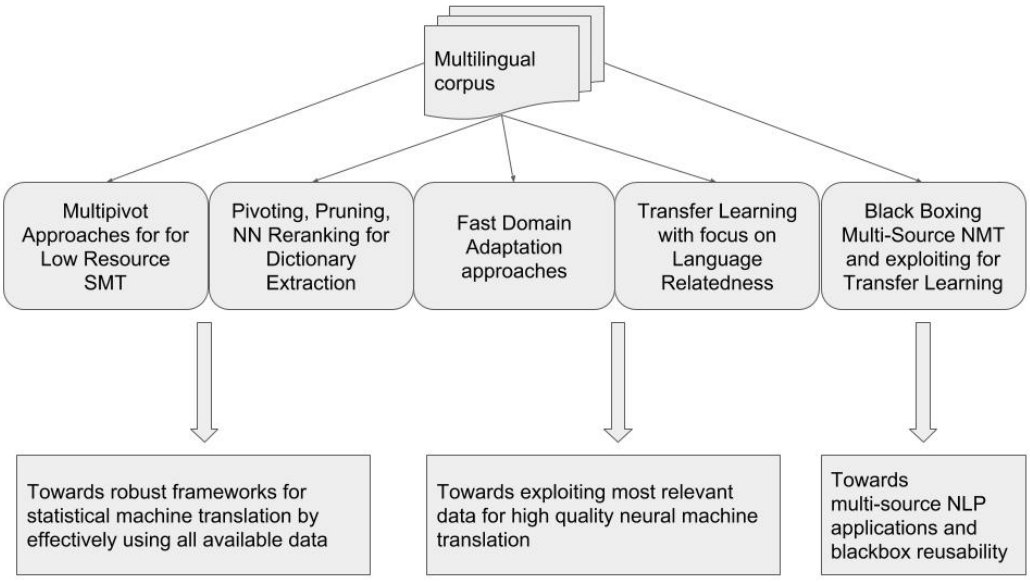


**Figure 1: An overview of this thesis with its implications on the future of MT research**