John Richardson （黒橋教授）

「Improving Statistical Machine Translation with Target-Side Dependency Syntax」

（目的言語側の依存構文による統計的機械翻訳の改善）

平成 28 年 9 月 23 日授与

Machine Translation (MT) is the application of Natural Language Processing that focuses on the automatic translation between languages. Translation is particularly challenging for language pairs with widely different grammatical structures, such as English and Japanese. Syntax-based MT is a translation paradigm based on the principle of generalizing language with grammatical rules. This additional layer of abstraction enables the design of more robust and flexible translation rules. However, the majority of previous approaches to syntax-based MT have employed only source-side grammar (known as 'tree-to-string MT'). This is mainly because syntactic analysis is difficult, prone to error and resulting systems can become overly complicated.

While there have been previous studies on exploiting target-side syntax ('tree-to-tree MT'), results have not been promising. Our aim is to analyze in detail the effectiveness of target-side syntax in the modern world of machine translation. We ask whether the potential improvement in translation quality is able to outweigh the increased complexity of employing a structured target-side representation (in particular, dependency parses).

This thesis begins with an overview of machine translation, outlining the major paradigms and methods of evaluation. We continue by outlining the case study of a state-of-the-art dependency tree-to-tree system, KyotoEBMT (see Figure 1), which we have been developing as a core component of our research on syntax-based MT. The design and extraction of dependency tree-to-tree translation rules are discussed. Analysis of the system gives empirical evidence of the advantages and disadvantages of syntax-based approaches and provides a starting point for our investigation.

We proceed to analyze two major aspects of translation where target-side syntax can be effective: word order and translation fluency. We discuss our approaches to each of these areas, describing experiments assessing the effectiveness of our proposed approaches and discussing the potential impact of each method.

While this thesis concentrates on statistical syntax-based approaches, the field has recently seen a surge in interest in translation methods based on neural networks. The final chapter presents an overview of future work that could incorporate ideas from this paradigm. We conclude by discussing the potential impact and future directions of our work.

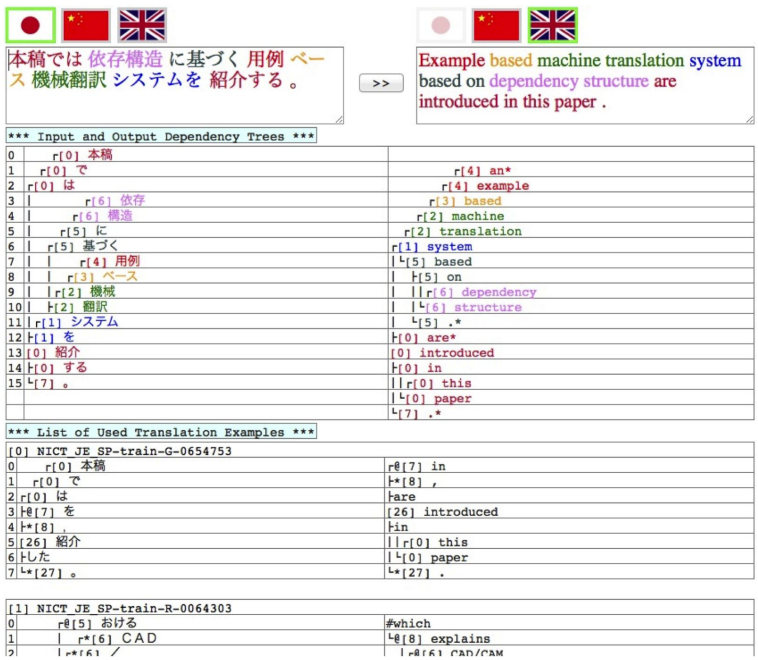

Figure 1: An example translation with KyotoEBMT showing translation examples with dependency trees.