Abstract

Machine translation is the process by which computer software is used to translate a text from one natural language to another but handling complex sentences by any machine translation system is generally considered to be difficult. In order to boost the translation quality of the machine translation system, simplifying an input sentence becomes mandatory. Many
approaches are available for simplifying the complex sentences. In this paper, Rule based technique is proposed to simplify the complex sentences based on connectives like relative pronouns, coordinating and subordinating conjunction. Sentence simplification is expressed as the list of sub-sentences that are portions of the original sentence. The meaning of the simplified sentence remains unaltered. Characters such as (‘.’,’?’) are used as delimiters. One of the important pre-requisite is the presence of delimiter in the given sentence. Initial splitting is based on delimiters and then the simplification is based on connectives. This method is useful as a preprocessing tool for machine translation.

Reference

- Katrin Tomanek et al. 2003 “Sentence and Token Splitting Based on Conditional Random Fields”, Jena University Language & Information Engineering (JULIE) Lab, Forman, G.
- Satoshi Kamatani, Tetsuro Chino and Kazuo Sumita, “Hybrid Spoken Language Translation Using Sentence Splitting Based on Syntax Structure”, Corporate Research and Development Center, Toshiba Corporation.


Index Terms

Computer Science
Natural Language Processing
Key words

Sentence simplification  Sentence segmentation

POS tag

Machine translation