ISSN 0970-9819

IJT
International Journal of Translation

(*A Half-Yearly Review of Translation Studies*)

Vol. 29 No. 1 Jan-Jun 2017

Special Number on
CRITICISM AND THEORY
IN/VIA TRANSLATION: INDIAN AND WESTERN PERSPECTIVES

Guest Editor
PRABHA SHANKAR DWIVEDI
IIT, Tirupati, Andhra Pradesh

Founder Editor
UJJAL SINGH BAHRI

Editors
HARPREET KAUR BAHRI
DEEPINDER SINGH BAHRI

INTERNATIONAL JOURNAL OF TRANSLATION
VOL. 29, NO. 1, JAN-JUN 2017

The Role of Shallow Parsing/Chunking in Machine Translation

ABHIGYAN DWIVEDI
Dr. H. S. Gour Vishwavidyalaya, Sagar, M. P.

The term "chunk" in computational linguistics was first introduced in (Abney 1991). Chunk literally means "to make block of" particular thing. In computational processing of language this term is used to make blocks of syntactic elements. These syntactic elements may be either phrase or adjacent words making some sense. In literature it has been defined as:

> A textual unit of adjacent word tokens which can be mutually linked through unambiguously identified dependency chains with no recourse to idiosyncratic lexical information.

For example in "a strong body", the article "a" and adjective "strong" both are dependent on the "body." So these are adjacent word tokens and are mutually linked together with dependency chains.

Chunk is also defined as:

> the non-recursive core of an intra-clausal constituent, extending from the beginning of the constituent to its head, but not including post-head dependents. (Abney 1990)

Chunking which is also known as shallow parsing is an intermediate process to parse the input sentence. Machine Translation has three basic steps – morphological analysis, syntactic analysis and semantic analysis. Morphological analysis identifies the information on inflections of the input words/lexemes. Part of speech tagging is one of the major tasks in morphological analysis. The part of speech-tagged and

morphologically-tagged word is then entered for syntactic analysis which parses the input (i.e. tagged) data according to any of the available grammar formalisms. The tagged data can be grouped into phrases/groups of similar category i.e. noun phrase, verb phrase, prepositional phrase etc. For this purpose, shallow parsing or chunking is helpful. It smoothes the task of syntactic parser by providing the phrases to process instead of words.

Chunking involves two basic functions-chunk boundary identification and chunk labeling. There are various computational techniques executed for these two actions. Noun Phrase (NP) chunking is the process of marking chunks of noun phrase. This is well researched area of computational linguistics. Further sections are devoted to the techniques/computational methods for chunking.

A lay out for the machine translation and role of shallow parsing or chunking is as follows: