知覚連語の同値性と M 言語による日本語解析システム「ささゆり」における日本語文簡易化の方法

高橋 亘
関西福祉科学大学社会福祉学部

近年我々は、日本語解析システム「ささゆり」の基礎をあたえる知覚連語の言語学の構築を推し進めてきた。[1]

知覚連語の形成規則はある種の語の結合を意図的に禁止している。その代表的なものは知覚連語の形成規則は連体修飾をする動詞文と被修飾名詞の結合の禁止である。この禁止則のため日本語解析システム「ささゆり」は本来的に複文の修飾関係を捌く機能を保持している。複文は、捌かれた結果、修飾する動詞文（修飾子）と修飾される名詞（接合名詞）との対応関係と接合名詞を含んで後続する知覚連語（後続子）の構成関係に分解されるが、これら二つの関係が接合名詞の意味を限定する。[2]

本発表の第一のテーマは接合名詞が形式名詞である複文の単文化の問題である。意味を限定された形式名詞に意味的に最も近い日常的名詞の特定する技術が述べられる。

第二のテーマは知覚連語の二種の同値性についてである。知覚連語が構成要素から成り立ち、純粋な意味を保持するという二面性は二種の同値類の存在を示唆する。

一つは知覚連語を構成する範疇の共通性がもたらす同値性であり、共通な範疇を保持する知覚連語が一つの同値類を構成する。これは、いわば共通のキーワードを保持する知覚連語の探索技術と直接的に関わるものである。

もう一つは、意味的距離の近い知覚連語が構成する同値類である。これは意味空間の構成が前提であり、知覚連語の構成によって始めて定義が可能になるものである。このような同値類は、共通のキーワードを保持しない知覚連語の探索技術を提起するものであり、情報検索の新しい技術である。

二種の同値性は、双方の利便性が相まって情報検索に寄与するものであり、コミュニケーション支援の基本的方法を提示する。

参考文献

[1] 高橋 亘,『コミュニケーション支援の情報科学』, 現代図書 (相模原, 2007, 4 月).

[2] 高橋　亘, “M 言語による日本語解析システム「ささゆり」の意味解析--- 連体修飾のある日本語文の意味解析 ---”, 『Mumps』, Vol. 24 (2008) 27-33.

[3] 宮地絵美, 高橋　亘, “M 言語による聾者のための日本語簡易化機能--- 連体修飾のある日本語文の単文化と形式名詞の意味推定 ---”,『Mumps』, Vol. 24 (2008) 35 ～ 40.