

日本語解析システム「ささゆり」における知覚連語の同義性同値類を用いた検索技術と日本語の言い換え技術

高橋 亘

関西福祉科学大学社会福祉学部・大学院社会福祉学研究科

日本語解析システム「ささゆり」の基礎となる日本語の切断についての考え方がスタートしたのは 1998 年である [1]。この技術の日本 M テクノロジー学会での発表は 1999 年に出雲市で行われた第 26 回日本 M テクノロジー学会大会であった [2]。その後、品詞解析機能・知覚連語の学習機能・意味解析機能の強化がなされて行ったが、近年に至って、複文を単文化する技術・形式名詞の意味推定の技術・難解語の言い換え技術などが整っていった。

こうした近年の技術の底流にあるものは意味解析の基礎技術としての知覚連語の同義性同値類を用いた新しい検索技術である。この発表で我々は、知覚連語の同義性同値類に関連する日本語解析システム「ささゆり」の基礎を与える言語学と M 言語の技術の総合報告を試みたい。

我々の日本語解析の基礎にある原理は知覚連語の言語学である。知覚連語は日本語解析システムによって機械学習される。学習の基礎を与えるのは知覚連語の形成規則である。知覚連語の形成規則はチョムスキーの生成文法に似ているが、普遍文法を前提としていない。連語形成が意味的に純粋な文の構成単位を作り出し、意味的に純粋な文の単位が知覚を誘発し、知覚が誘発されることが言語習得の原動力を与えること

を前提としている [3]。これは幼児の言語習得の動機を提示しているだけでなく、意味解析からいえば、知覚連語を機械学習させ、知覚連語と意味要素を関連づけることが意味解析のアーキテクチャーを与えることを意味している。

したがって、日本語解析システム「ささゆり」の意味解析機能には、その開発由来に従って、5 つの特徴がある [4 - 7]。

- (1) 機械学習された知覚連語を基礎として意味空間を張る。
- (2) 日本語解析システムに複文の構文理解機能が本来的に備わっている。
- (3) 知覚連語の間に共通語同値類と同義性同値類の二種の同値類を定義できる。
- (4) (2) を基礎に単文分解が可能である。
- (5) (3) を基礎に共通語同値類を同義性同値類で類別することが出来る。

以上の特徴を持つ意味解析は、情報検索の技術に、キーワード検索には無い、新しい技術を提供することになる。

- (a) 知覚連語を基礎に取るので連語構成要素間の意味要素の重複勘定がない。
- (b) 複文の構文理解を活用し、修飾関係における意味評価を正確にでき

る。これは、修飾を受ける名詞や形式名詞などの意味規定に有効である。

(c) 知覚連語の共通語同値類を同義性同値類で類別し、難解語や専門用の総合的解説機能の基礎を与える。

(d) 単文分解機能に被修飾名詞の意

味推定機能を組み合わせると、構文の複雑な日本語文を簡易表現に言い換える一般的技術を提供出来る。

(e) 同義性同値類の把握は今までキーワード検索の技術で検索しきれなかった情報の検索を可能にする。



図 1 同義性同値類を活用した検索の例;

「強い相互作用」という物理学の専門用語についての例

以上は、これまで折に触れて発表してきた意味解析技術を情報検索のアーキテクチャーとして総合報告するものであるが、今までに触れて来なかった技術として、日本語に埋め込まれた英数字表現の言葉の情報を日本語ベースの意味解析システムに組み入れる方法についても触れていきたい。

引用文献

1. 高橋 亘, “視覚障害者のためのヒューマンインターフェイスにおけるユニバーサルデザインと人工知能” 『関西福祉科学大学紀要』 No. 1, 41-49 (1998).
2. 高橋 亘, “大域変数の階層構造と日本語文切断のアルゴリズム”, 『Proceedings '99 M Technology Association of Japan』, 7-1 ~ 7-4 (1999).

3. 高橋 亘, 『コミュニケーション支援の情報科学』, 現代図書 (相模腹, 2007).
4. 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.
5. 高橋 亘, “日本語解析システム「ささゆり」における連体修飾のある日本語文の意味解析”, 『関西福祉科学大学紀要』, Vol. 12, 81 ~ 90 (2009).
6. 高橋 亘, “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と知覚連語間の意味的距離” 『総合福祉科学研究』, Vol. 1 (2010) 91 ~ 100.
7. 高橋 亘, “知覚連語の同値性と日本語文簡易化の方法 ---M 言語による日本語解析システム「ささゆり」の意味解析---” 『Mumps』, Vol. 25 (2010) 9 ~ 21.