

5N-5 チェック・訂正問題を対象とした  
日本語記述知識の曖昧性除去手法

田添 清 渡辺 敏

NTT情報通信網研究所

1. はじめに

ES開発において、専門家による知識の追加・修正を困難にしている要因の一つに、専門家の持つ概念的な知識と現状ツールによって提供される知識表現とのギャップが挙げられる。我々はこの問題を解決するために、日本語で記述された知識をそのまま知識表現と見なす段階的知識ベース生成方式の検討を進めている。<sup>[1]</sup>

本稿では、曖昧で多様な表現が可能な日本語文を一意に解釈可能な日本語文に変換する曖昧性除去手法について報告する。

2. 問題解決へのアプローチ

日本語記述知識の曖昧性を排除するためには、記述用語の厳密な定義と表現の多様性を回避することが重要である。用語の定義方法（辞書化）については別途報告することとし、ここでは辞書の存在を前提に議論を進める。

日本語記述知識の解釈時における曖昧性は、①用語の省略や多義性、②係受け関係の解釈、③助詞・助動詞の使用による表現の揺らぎ、等に起因する。これらの曖昧性を自動的に除去することが理想であるが現状の技術では限界がある。そのため、入力された知識を計算機に解釈可能な範囲で勝手に解析し、解析結果を元に再構成された日本語文（限定表現と呼ぶ）を生成し、利用者に原知識との同義性を確認させることを考えた。

3. 限定表現とは

限定表現は利用者と計算機間の文章理解のギャッ

プを埋めるものであり、一種の計算機用言語の役割を果たす。しかし、本方式では利用者は限定表現文法を詳しく知る必要はなく、自由に表現すれば良い。

限定表現は計算機の内部表現と対応づけが可能な用語と構文情報から構成された辞書に基づく自然語表現である。その結果、①使用用語を辞書用語で限定、②処理と直接関係ない表現を全て削除、③画一的な表現への統一、等の特徴を持つ。そのため、人から見た場合、ごちない表現となる部分がある。

4. 基本処理方式

限定表現を生成するため、ドメイン用語と用語の属性、文型パターン、用語と用語の関係が記述された辞書を用意する。それぞれ用語情報辞書、文型情報辞書、関連情報辞書と呼ぶ。これらの辞書を利用して日本語記述知識を洗練する。（図1参照）

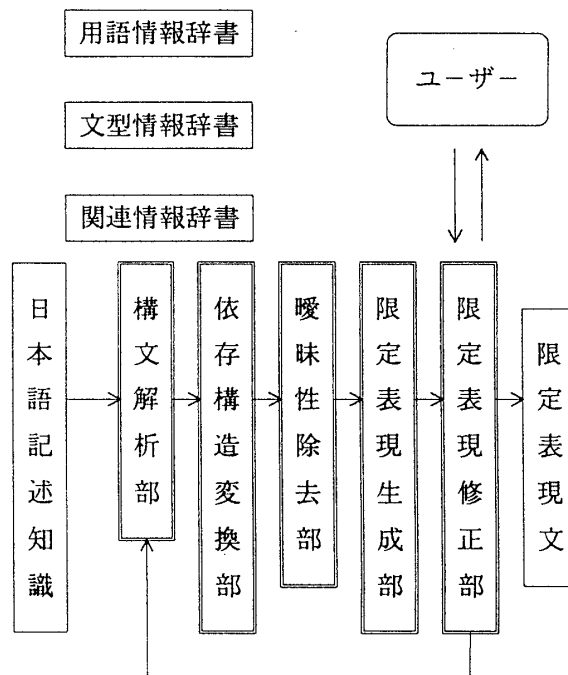


図1：基本処理方式

各部の処理概要を以下に示す。

**構文解析部**

日本語記述知識の解析は自然言語解析パッケージを利用して、形態素解析、係り先解析、係り受け解析を行い、依存構造<sup>[2]</sup>の候補を抽出する。

**依存構造変換部**

構文解析で得られた解析結果を基に、専門用語の利用法に着目して文構造を一つに絞り込むと共に、文関係の把握を容易とするため、接続詞を主体とする依存構造に変換する。

**曖昧性除去部**

変換された依存構造を基に用言の文型情報や用語の属性や修飾関係を分析し、分析ができない表現や自動的に変換できない部分を抽出する。

**限定表現生成部**

変換された依存構造と曖昧性除去部の解析結果を基に、修正案を含む依存構造に再変換し、これを元に日本語化する。

**限定表現修正部**

原文、依存構造から生成された日本語文及び解析結果を表示すると共に、不具合な表現については利用者との会話により、修正を実施する。この作業を繰り返し、最終的に原文の意味との整合性を確認させる。

**5. 限定表現の修正支援**

本支援方式では、計算機との会話で、計算機の認

識できる表現に速やかに集束させることができなければならない。そのため、提示候補の選択を主体に会話が進められ、利用者による再入力を最小とするのが望ましい。そのため、用言用語とその主語や目的語になり得る用語との関係、用語間の修飾関係、用語の属性、等に着目して、置換或いは挿入すべき用語候補抽出すると共に選択メニューを表示し、選択された用語に基づき表現が自動的に訂正されるように工夫した。主な修正支援項目は、未知語（未定義語）の使用、用語の欠落、構文誤り（意味誤りも含む）、等がある。

図2に具体的な限定表現案への変換過程を示す。

**6. おわりに**

本稿では日本語記述知識を対象に曖昧性を除去した限定表現に変換する手法について報告した。プロトタイプによる開発を進めており、より高度な支援の可能性や限定表現の読解性向上について、更に検討を進めていく予定である。

**参考文献**

- [1]山本, 渡辺: チェック・訂正問題を対象とした段階的知識ベース生成支援方式, 情報処理学会第45回全国大会, 4H-5, 1992.
- [2]上野: パーソナル・コンピュータによる日英双方向機械翻訳システムの制作, ラッセル社, 1989

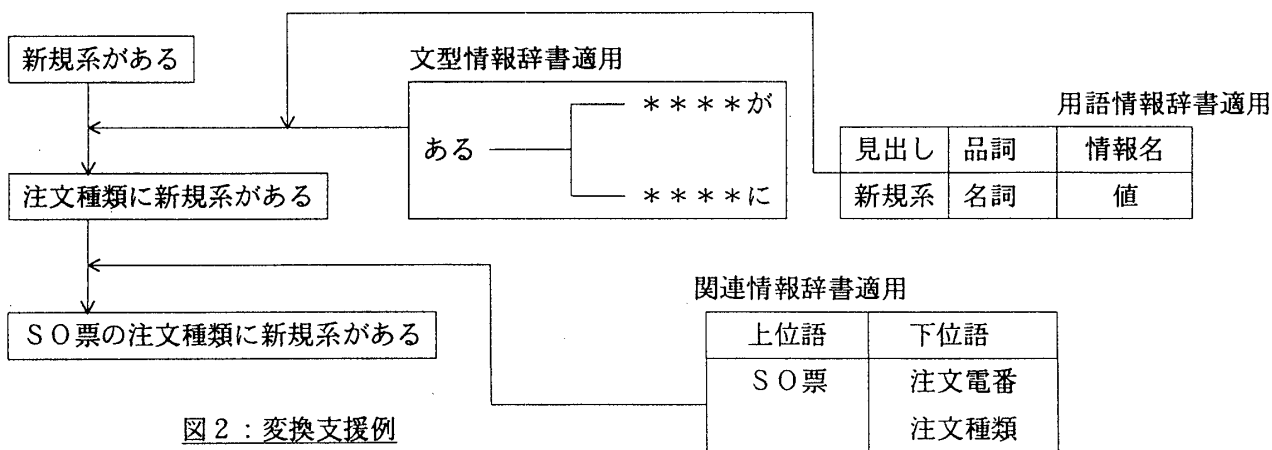


図2: 変換支援例