8N-2

SGLR-plus による話者の対象認識構造を抽出する中国語文パーザの試作

孟 莉　柴田 悟之　宮崎 正弘
新潟大学大学院自然科学研究科

1 はじめに

人間の言語活動における「対象 ⇔ 認識 ⇔ 表現」という過程的構造に着目した、コンピュータによる中国語文の解析の実現を目指し、一般化 LR 法を Prolog 上に実現した SGLR の拡張版である SGLR-plus パーザを用いて、言語の統語構造にあたる表現構造と認識された対象世界を把握する認識構造を抽出する中国語文パーザを試作し、その有効性を検証した結果について述べる。

2 中国語表現の構造

2.1 中国語の特徴

英語は名詞の数、人称代名詞の格、動詞の人称や格やテンスなどによる語尾変化がある。日本語は助詞、助動詞などのような文法上の関係を表す語を他の語に密着させて文を作る。これに対して、中国語は語尾の変化がなく、文を作るとき一定の語順にあてはめて単語をそのまま連ねる。つまり、中国語は「語彙的なことば」である。

2.2 中国語の表現構造と認識構造

時枝誠記によって提唱され、三浦つとむ [1] によって発展的に継承された言語過程説によれば、言語の本質は、認識の対象と認識する主体との統一を反映したものである。

言語を構成する表現には以下の 2 種類ある。

- 客体的表現：話者が対象を概念化して捉えた表現で、客体認識を表す。
- 主体的表現：話者の主観的な感情、要求、意志、判断などを直接表現したもので、主体認識を表す。

言語の対象世界は、話者 (主体) とそれ以外の対象 (客体) とから構成される。対象は実体、属性、関係の 3 つの要素から構成される。話者が対象を認識する際の認識構造は着目する対象とその属性の認識である。その構造は客体認識とそれらに対する話者の主体的な判断の認識である主体認識から構成される。

中国語では、常に最も重要な対象を文 (表現) の先頭に位置づけ、話し手、書き手の主体的な表現をその直後に置く「主謂表現」(主述賓) 形式であり、主体的な判断や感情などが主部と述部の間に位置して、英語のような天秤型の形式をしている。図 1 に中国語の基本的構造を示す。

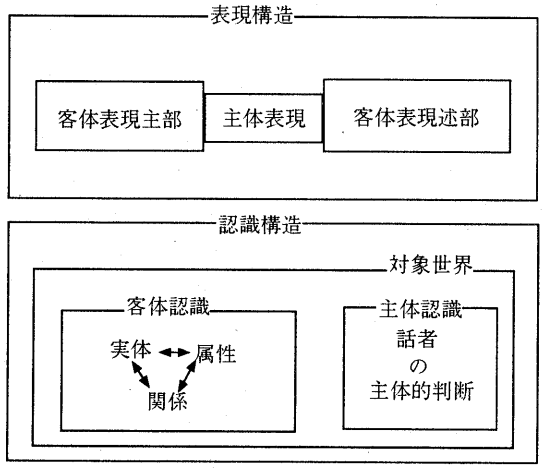

図 1　中国語の基本的な構造

An Experimental Chinese Parser for Extracting Speaker's Recognition Structure
Li Meng, Noriyuki Shibata, Masahiro Miyazaki
Graduate School of Science and Technology,
Niigata University

2.3 中国語の主体的表現

中国語の主体的表現には、肯定・否定の判断、「語気辞」、「主体判断詞」（助動詞）などがある。以下に例を示す。

表 1 中国語の主体的表現の例

肯定	是，有	他是学生。
否定	不，没	他不是学生。
疑问	吗	他是学生吗?
请求	吧	你来吧。
感叹	啊	他来啊!
可能	能，会，可以	他能唱歌。
推量	可能	他可能来。

2.4 中国語の客体的表現

中国語の客体的表現には、話者が注目している、また動作を行う対象の実体である動作主や、実体の動作を表す動的属性、状態を表す静的属性、そして動的属性の効果が及ぼす目標対象などがある。

3 中国語文パーザの試作

SGLR(A Sequential Generalized LR Parser; 逐次型一般化 LR) パーザ [2] は、横型探索をベースとした富田法を Prolog 上に記述したもので、先読み情報を利用することにより無駄のない高速な統語解析を実現する。ここではSGLR パーザの拡張版であるSGLR-plus パーザ [3] 上で中国語の文法、辞書を記述し、中国語文パーザを構築した。

現段階で、中国語の単文レベルの文を完全に解析できるよう、文法と辞書の整備を進めている。なお、認識構造は英語の認識構造を参考に将来的な翻訳を見越して英語文と親和性があるように設計した。図 2、図 3 に文法・辞書記述例と解析結果例を示す。

```
主谓句(Dec) --> 客体表现_主部(Subj), 主体表现1(Aff), 客体表现_述部(Pred),
{sp1(Subj, Aff, Pred, Dec)}.
主体表现1(Aff0) --> 主体判断辞(Aff), {aa(Aff, Aff0)}.
主体表现1(Aff0) --> 否定判断辞(Neg), 主体判断辞(Aff), {na(Neg, Aff, Aff0)}.
肯定判断辞(Aff0) --> 肯定判断辞0(Aff), {aff(Aff, Aff0)}.
否定判断辞(Neg0) --> 否定判断辞0(Neg), {neg(Neg, Neg0)}.
主体判断辞(Aff0) --> 主体判断辞0(Aff), {aff(Aff, Aff0)}.

人称代名词([实体, [把握, [关系, [人称, 一人称]], [属性侧面, [数, 单数]],
[标识, 我]]]) --> [ 我 ].
人称代名词([实体, [把握, [关系, [人称, 二人称]], [属性侧面, [数, 单数]],
[标识, 你]]]) --> [ 你 ].
肯定判断辞0([主体判断, [把握, 肯定]]) --> [ 是 ].
否定判断辞0([主体判断, [把握, 否定]]) --> [ 不 ].
名词([实体, [把握, [概念表示, 歌]]]) --> [ 歌 ].
动词([动态的属性, [把握, [核心结构, 唱], [属性侧面, 动作表示]]]) --> [ 唱 ].
动词([动态的属性, [把握, [核心结构, 走], [属性侧面, 动作表示]]]) --> [ 走 ].
```

図 2 文法記述例と辞書記述例

```
例：他不能唱歌。(彼は歌うことができない。)
|- 表现构造
   |- 陈述句表现
      |- 句子
         |- 主谓句
            |- 客体表现_主部
            |  |- 体词
            |     |- 人称代名词 — < 他 >
            |- 主体表现1
            |  |- 否定判断辞
            |  |  |- 否定判断辞0 — < 不 >
            |  |- 主体判断辞
            |     |- 主体判断辞0 — < 能 >
            |- 客体表现_述部
               |- 动态表现1
                  |- 动词短语
                     |- 述宾式动词短语
                        |- 谓词
                        |  |- 动词 — < 唱 >
                        |- 体词
                           |- 名词 — < 歌 >
|- 认识构造
   |- 客体表现
   |  |- 动作主 — < 他 >
   |  |- 动的属性 — < 唱 >
   |  |- 对象 — < 歌 >
   |- 主体表现
      |- 时制 — < 现在 >
      |- 判断 — < 否定 >
      |- 样相 — < 可能 >
```

図 3 中国語文の表現構造と認識構造

4 おわりに

人間の言語活動である「対象 ⇔ 認識 ⇔ 表現」という過程的構造に着目した時枝誠記による言語過程説を発展的に継承した三浦つとむの説に基づいて、SGLR-plus パーザ上に、表現構造と認識構造を抽出する中国語パーザを試作した。今後、中国語表現の認識構造の詳細化を行い、複文や重文などもっと複雑な文を解析できる、中国語文パーザの構築を進めると共にその解析能力の定量的の評価を行う予定である。

参考文献

[1] 三浦つとむ：日本語はどういう言語か、講談社学術文庫 (1976)

[2] 沼崎・田中：SGLR：逐次型一般化 LR パーザの Prolog による実現, 情報処理学会論文誌 Vol 32 No.3 pp 396-409(1991)

[3] 五百川・宮崎：痕跡処理のための逐次型一般化 LR パーザ SGLR の拡張, 言語処理学会第 4 回年次大会発表論文集 pp.314-317(1998)