

文脈制約によるタイ語会話文法モデルの構築

Chumphol Krootkaew 永井 秀利 中村 貞吾 野村 浩郷

九州工業大学 情報工学部

1 はじめに

本研究では、タイ語会話におけるいくつかの会話文脈制約やタイ語の表層的な言語表現などを分析・整理してタイ語会話構造の基本的な要素を分析する。その会話構造要素によってタイ語会話の文章レベルと文句レベルをつなぐ言語モデルを構築する。

2 タイ語の特徴と問題点

タイ語処理の研究においては、曖昧な文法のため、さまざまな問題が現れる。例として、タイ語の特徴とともに問題点を以下に示す。

2.1 形態素解析と品詞の決定

タイ語で記述された文章は、日本語や中国語と同様に語と語の間にスペースが存在しない。さらに句読点も存在しないため句や文の切れ目が曖昧である。したがって、辞書を参照しながら語の切れ目を確認することが必要である。話し言葉の場合は相手が理解しやすいためポーズ (pause) をいれることがある。

品詞においては、タイ語の語形が活用しないため、語形情報は存在しない。品詞を決定するには、まわりの単語の意味と語順が重要になる。

2.2 語順と構文の曖昧性

タイ語では、他動詞文はSVOの語順をとる。したがって英語の文法に近いと言われる。言語の語順研究 [3] に関して、タイ語と日本語と英語は、三つの異なるタイプに属す。日本語とタイ語は語順に関してかなり一貫して逆さまである。英語の語順は一貫性がない。日本語と同じになったり、タイ語と同じになったりする。

タイ語の句構造の語順は、日本語と違って修飾語は被修飾語よりも後ろにおかれる。例えば、タイ語の [機械][計算] は、日本語の「計算機」に訳すことができる。さらに、「計算した機械」や、および「機械が計算した」のようにも訳すことができる。

このように、構文解析においては、文句の構造の曖昧性が大きな問題になってくる。しかし、言語表現の

解析を詳細にし、さらに談話の文脈情報を用いることによって、この曖昧性を減少することができる。

2.3 文章解析と文脈制約の必要性

タイ語は、英語などの節効率言語に対して、段落効率言語であると言われる。英語は、節および文がそれぞれ自立した情報の単位として機能するが、一方では、主語、目的語、時制などの同じ情報が複数の節中に何度も出現するため、段落全体としてみた場合は冗長である。逆に、タイ語は一般に各々の節単位では完全な情報を持っていないが、不完全な情報は文脈的に補われるので段落全体としてみた場合は冗長性が少く効率的に情報伝達が行なわれる。

タイ語の文は、談話の先頭に現れることができるかどうかと言う観点から、開始文 (initial sentence) と非開始文 (non-initial sentence) とする二つのタイプに分類することができる [2]。開始文は、それ単独で事象を記述するだけの十分な情報を担っている完成された文であるが、非開始文は単独では十分な情報を持っておらず、文解釈においては、不足する情報を文脈から補ってやる必要がある。このように、文章の解析においては文間の文脈解析が必要である。

3 文脈制約による会話の構造

従来の研究 [1] では、文脈制約の照応結束性と文要素の省略制約に着目してタイ語の文章を開始文と非開始文に分類し、これらによりタイ語の文脈モデルを設定した。開始文と非開始文の文章レベルから文句構造までの文法はまだ明らかになっていない。ここでは、タイ語会話の言語モデルを設定するため、タイ語会話構造の要素を分析し、より明確な文脈の境界 (boundary) を設定してタイ語会話の言語モデルを構築する。

3.1 談話の分割

会話中に、二人の関与者が話し手と聞き手の役割を交替しながら、交互に談話を発する。ここでは、その談話の文章を会話構造の要素として談話ユニット (discourse unit) と呼ぶ。したがって、会話の談話ユニットは、一文だけではなく複数文のこともある。逆

に句または語のこともある。本研究は談話ユニットをより単純な会話構造要素(「節」(clause))に分割するため以下のような制約で「節」の区切りを決定する。

- 句読法 (punctuation)
書き言葉では句読点がないが、話し言葉では文の構造や名詞句の論理的関係を明らかにするためはっきりしたポーズ (pause) を用いる。
- 構造を示す表現
構文構造の曖昧性を減少するため、手がかりとなる表層的な言語表現がある。これらの表現は、「節」の先頭または「節」の終りに位置する。

3.2 表層表現による節 (clause) の分類

タイ語では、日本語の文末表現に対応して、文章の機能、情報、構造などを定義する言語表現がある。これらの表現によって、前節で分割された談話ユニットの「節」をさまざまな文脈情報を持つ会話の要素として分類する。分類された節の一部を例として以下のように示す。

- 話題節 (topic clause)
話題部の役割を持つ節で、場所表現、時間表現、強調表現などがある。
- 述語節 (predicate clause)
主要部 (head) が動詞または形容詞である。動詞の分類によって、省略のパターンが分類できる。
- 補語節 (complementary clause)
英語の前置詞と日本語の助詞に対応する標識によって特徴づけられる。
- その他 (句の並び、応答節など)
会話の決まり文句的なもの。

分割されたこれらの節は談話ユニット内部構造の基本構造要素となる。これらの節は、句構造文法の文句として分析できる。

3.3 談話ユニット内部構成

談話ユニットの内部構成を確定するため、文脈情報によって結びつけられている節の集合を「発話セグメント」(utterance segment) と定義する。これと会話における話者の発言交替規則 (turn-taking) や隣接発話対 (adjacency pair) 制約などによって、前節で分類した「節」を以下のように発話セグメントの構成要素である「開始節」と「展開節」に分割する。

<発話セグメント> ::= <開始節><展開節の並び>

- 話しを導入する「開始節」(initial clause)
開始節は、発話セグメントの先頭に位置し、話し手が何かについて話す時の、その何かのことで、既知情報 (theme) を表す。話題節のほかに発話の先頭の応答節や隣接発話対制約の述語節なども開始節と判定される。この場合、既知情報は、前の発話の発話セグメントから補われる。
- 話しを展開する「展開節」(expansion clause)
展開節は、開始節の既知情報に対して新しい情報を加える。上の述語節と補語節が展開節と判定される。

発話セグメントは、開始節と展開節間及び展開節と展開節間の不完全な情報を互いに補う境界 (boundary) となる。

4 タイ語会話の言語モデル

タイ語会話の言語モデルを先に述べたような会話構造の要素などを用いて設定する。そのモデルの一部は以下のように示される。

会話の談話	::=	<談話ユニットの並び>
<談話ユニット>	::=	<発話セグメントの並び>
<発話セグメント>	::=	<開始節><展開節の並び>
	::=	<開始節>
<開始節>	::=	<話題節> <応答節>
	::=	<(隣接発話対) 述語節>
	::=	...
<展開節>	::=	<述語節> <補語節>
	::=	...

5 おわりに

タイ語の文章構造を文脈的な情報という観点から述べた。表層的な言語表現制約などによって、文句レベルの会話の基本要素を分析した。その会話の基本要素の集合を文脈的な観点から文脈の構成を定義し、タイ語会話の言語モデルを設定した。

参考文献

- [1] Chumphol, Nagai, Teigo, Nomura : "A model for contextual constraints and its implementation on the CAT2 formalism", The 2nd symposium on Natural language processing, 2-4 August 1995, Thailand.
- [2] Vichin Panupong : "The structure of Thai: Grammatical System", Ramkhamhaeng University Press, 1991.
- [3] 角田 太作 : "世界の言語と日本語", しろくお出版, 1991.