

情報の濃度に基づく参照表現の生成

5N-5

関洋平 飯島正 原田賢一

慶應義塾大学理工学部

e-mail: yohei@hara.cs.keio.ac.jp

1 はじめに

文を生成する際には、解析過程のように文の構成要素の共通の性質を導き出すよりも、語彙の細かなニュアンスの違いを用いて表現を行うことが不可欠である。特に、同じ個体を指示する際に、立場や見方を変えてみると表現ががらっと変わったり、個体への指示の曖昧性を排除するために、様々な属性を修飾要素として与えることで、個体の特定性を増すとといったことが行われている。

本研究は、客観的事実・関係を入力とした文の生成過程において、生成する事象に関連した個体に関してどこまでの情報を付加して、どこからを省略するかを判断する計算機上での実現の枠組についての提案である。本研究における参照表現は、実体を参照する際に、付加する属性や部分・呼称を元に指示する表現を指す。

2 文を生成するためのモジュール分類

客観的事象を入力として文を生成するためのタスクは、命題を文法的に構成する過程を除いて、認知的にはおおよそ2種類のモジュールに分類できると考えられる。

工学的実現を考える上でこのモジュール分けが妥当かどうかは議論の余地が残るが、本研究ではこれに基づいた生成を考える。入力としては事象系列を与える。

1. 一文単位の生成モジュール

(a) 各観点と信念に基づいた、事象に対応する表層表現の生成モジュール

- 文が生成される場合、その文は必ず何らかの立場に基づいていると考える。立場には、ユーザなどが持つ主観的立場や、定義文のような客観的立場がある。各立場ごとに、各事象および事象に関連した個体の情報を表が持っており、それを利用して生成を行う。

(b) 事象に対する役割に基づいた、文の変形モジュール

- 文の役割(依頼・命令等)に応じて、主語や目的語の省略、モダリティ情報の付加などが行われる。

2. ドメイン内知識を利用した、曖昧性解消モジュール

- 各個体には、その参照表現の特定性を高めるために属性を付加していく過程と、ドメイン内でわかりやすさを示すために、属性を用いて直接対象を指示する場合の二種類の参照方法がある。前者の過程における対象の特定性の高さを、情報の濃度と呼ぶことにする。一般には、この過程における属性の特定性の高さが高い(含む集合が縮まる)ほど濃密な情報となる。

3 参照表現の生成

本章では、上記文生成の戦略を用いて、参照表現の生成を行う。上記二段階のステップは参照表現の選択においても用いられる。すなわち、まず各ユーザの立場から事象の個体の参照表現を選択し、次に、個体に文脈的曖昧性がある場合には、それを取り除く。

本研究における参照表現とは、実体を参照する際に、付加する属性や部分・呼称を元に指示する表現を構成するものである。この際、属性(形容詞・動詞の連体修飾など)が生成対象の入力である事象命題に含まれる場合には、そのまま文としての生成が行われる。しかし、本章で対象とする参照表現の生成方法は、そのような場合を除き、生成事象命題が個体識別子として名前を含み、そこには属性などが含まれないにも関わらず、指示対象を的確な形で生成するケースについての手順を示す。

1. 個体の属性項目をピックアップし、各項目を全て集めた集合のトッ集合を使って、空集合をトップに、
X をボトムに置く形で束構造を構成する。
2. 束構造のトップから始めて、他の個体が同様の属性を持たないところまで、上から下に向けて降ろす

3. 他の個体が属性を持つ集合に含まれなければ、その属性を付加した参照表現を生成する

名前 (ID)	性別	職業	年齢
太郎 (a)	男性	先生	22 歳
花子	女性	会社員	28 歳
太郎 (b)	男性	フリーター	22 歳
次郎	男性	銀行員	38 歳

図 1: 個体の属性の表

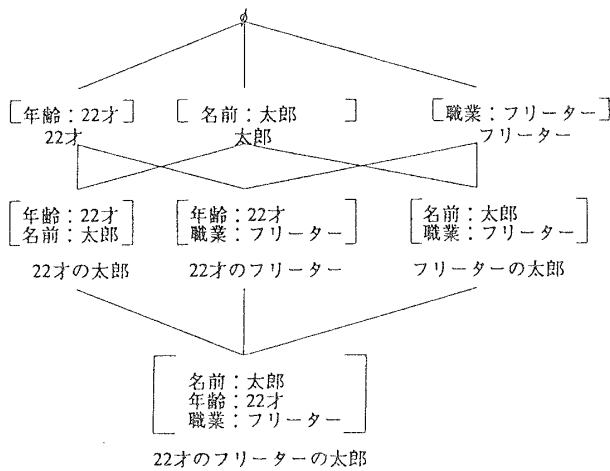


図 2: ある男性の属性の束構造に基づく参照表現

名前 (ID)	観点 1	観点 2	観点 3
太郎 (a)	他人	先生	自分
花子	知人	他人	妻
太郎 (b)	友人	友人	知人
次郎	自分	他人	知人

図 3: 各観点の状態の表

事象	個体 1	個体 2
ドライブに誘う	太郎 (a)	花子
断わる	花子	太郎 (a)
ドライブに行く	太郎 (b)	次郎

図 4: 文として生成したい事象系列の表

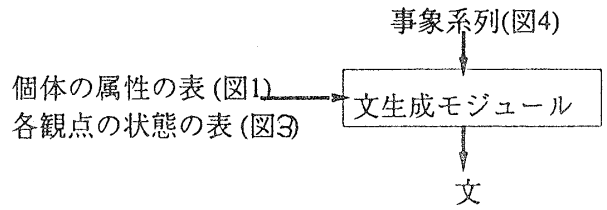


図 5: 全体の構造

図 1 ~ 図 4 に基づいた文の生成を以下に示す。

1. 観点 1 から見た生成文

- 花子さんは旦那さんにドライブに誘われた
- 彼女は断わった
- 太郎君は僕とドライブに行った

2. 観点 2 から見た生成文

- 先生は奥さんをドライブに誘った
- 奥さんは断った
- 太郎君は誰かとドライブに行った

3. 観点 3 から見た生成文

- 僕は妻をドライブに誘った
- 断わられた
- 太郎さんは次郎さんとドライブに行った

4. 客観的に見た生成文

- 先生の方の太郎は花子をドライブに誘った
- 花子は断わった
- フリータの方の太郎は次郎とドライブに行った

4 おわりに

本研究で提案した文生成手法は、客観的事実と各自の信念に基づいた個体の参照情報、および個体固有の属性情報をためこんだ事象系列を入力とする。本研究の応用としては、事象系列として何らかの知識をため込んでおき、それをユーザにわかりやすいかたちで提示するタスクとして、要約などが考えられる。