

一方向的な発話を実時間で理解する聞き手の認知モデル

6H-2

竹内 勇剛 三輪 和久

名古屋大学 人間情報学研究科

{ytakeuti, miwa}@cog.human.nagoya-u.ac.jp

1 はじめに

これまでの談話理解の研究では、主に時間経過に伴う状態の変化が少ない静的な世界についての談話を対象としてきた^[1]。しかし日常的な談話には時間経過にともなう状態の変化が活発な対象について言及するものが多く存在する。例えばラジオによる野球の実況中継などがある^[2]。このような対象について、その状態を言語によって伝達するとしたとき、聞き手は伝達された対象の状態をすみやかに理解し、次の発話に備えなくてはならない。

本稿では、言語による一方向の伝達情報を実時間で理解する聞き手側の認知モデルを提案する。連続的に送り込まれてくるのみの情報を実時間で理解するためには効率のよい処理が必要である。特に話し手の発話によって操作される聞き手の記憶している知識の操作が重要となる。このモデルでは、心理学的に長期記憶に相当する領域がプロトタイプ的な知識を記憶し、短期記憶に相当する領域がインスタンス的な知識を記憶する。そして伝達された情報をトップダウンに適切なインスタンスである知識フレームに埋め込むことで聞き手は伝達情報を理解できると仮定する。そしてこの認知モデルを計算機に実装することによって、モデルの妥当性を検証する。また、この計算機上でのモデルから、活性化しているインスタンスフレーム数とそれらが存在する領域の大きさの関係から、上記の制約をもった情報理解における効率性について検討する。

2 認知モデル

人間は実生活の中で、指示対象の同定や新旧の情報操作、焦点化の過程などの文脈による解析といった処理を実時間的に行うことでコミュニケーションを成立させている^[3]。Griceは、話し手と聞き手は、制限された時間内であつ部分情報に基づき、意図の伝達あるいは理解に向けて最善と信じる処理を行う^[4]と述べている。

A Cognitive model of listener in one-way and real-time communication.

Yugo TAKEUCHI, Kazuhisa MIWA
Graduate School of Human Informatics, Nagoya University
Nagoya, Aichi 464-01, Japan

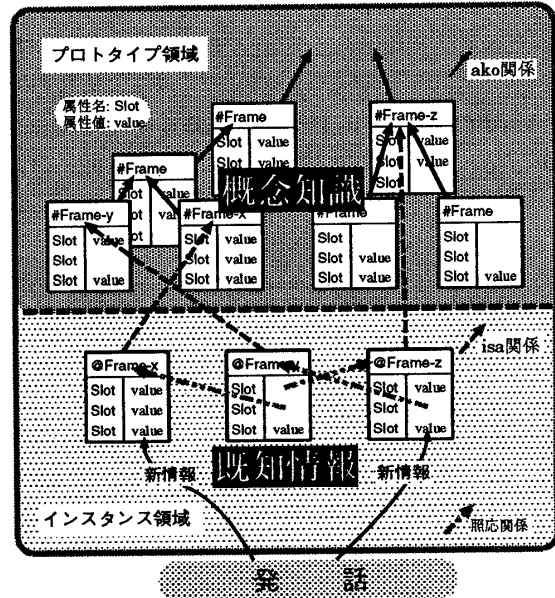


図1: プロトタイプ領域とインスタンス領域

ここで、動的な世界に関する状況を一方向の伝達であつそれが実時間で理解される過程における聞き手側の心的な情報操作の認知モデルを提案する。本モデルでは、プロトタイプ知識とインスタンス知識の二つの知識領域を設定することでこの処理を効率良く達成できると仮定している。

2.1 プロトタイプ知識とインスタンス知識

動的な世界における対象の変化を伝達するには、静的な知識に対する変化の差分だけを伝達する方が効率的である。このとき、土台となるその静的な知識はその対象に対して一般的な概念知識である。そして差分情報（新情報）を与えられた知識は具体的な知識になる。なぜならその差分情報は動的な世界におけるある対象（インスタンス）についての情報だからである。ここで前者の知識の記憶領域をプロトタイプ領域、後者の記憶領域をインスタンス領域と呼ぶ（図1）。

インスタンス領域での知識をここでは既知情報と呼ぶが、これは動的な世界についての言及では、伝達された新情報がその後のある短い時間内に参照される既知の情報として積極的に利用されることを反映しているからである。

2.2 伝達情報の理解の構造

発話による新情報の伝達はプロトタイプ知識からトップダウンに既知情報として活性化されているインスタンス知識の適切なスロットに埋め込まれる。一方的な情報の伝達は、話し手と聞き手が相互に共有していると想定できるプロトタイプ領域の概念知識とそのインスタンスによって行われる。聞き手は話し手の発話によって活性化されたインスタンスフレームが存在するインスタンス領域を新情報（スロット値）の埋め込みのために優先的に検索する。そして適切なスロットが存在したときに、そのスロットに値を埋め込む。

- i. インスタンスフレームの生成と活性化
- ii. 活性化しているインスタンスフレームへの値の埋め込み

活性化しているインスタンスフレームはある期間を経過すると沈黙しインスタンス領域への優先的な検索から除外される。このようにして前述した条件下で、効率のよいコミュニケーションを実現することができる。

3 計算機上での実現

3.1 実験

前節で述べた認知モデルをフレームシステムを用いて計算機上でシミュレートした。ここでは、入力された発話データがどの程度の比率で2.2のiとiiの処理を実行していることを検証する。

実験システムに入力した発話データは、ラジオによるプロ野球の実況放送である。なお、アナウンサー（発話者）と解説者の会話は入力データから除外した。

3.2 結果

以下に全活性化動作インスタンスフレームとその中の最頻出動作インスタンス *pitching*, *batting* についての実験結果を示す。

$$\begin{aligned}
 & \text{iの処理による活性化インスタンスフレーム数:} \\
 & \qquad T(\text{act}) = 54 \\
 & \text{iiの処理による値を埋め込まれたフレーム数:} \\
 & \qquad T_f(\text{act}) = 39 \\
 & \qquad T_f(\text{act})/T(\text{act}) \times 100 = 72.2\%
 \end{aligned}$$

上記のようにi,iiの処理によって72.2%の活性化動作インスタンスフレームが生成され値が埋め込まれた。

特に、最頻出動作インスタンス *pitching*, *batting* については次のようになる。

$$\begin{aligned}
 & \text{iの処理による活性化pitchingインスタンスフ} \\
 & \text{レーム数:} \qquad T(\text{ac:pit}) = 17 \\
 & \text{iiの処理による値を埋め込まれたフレーム数:} \\
 & \qquad T_f(\text{act:pit}) = 16 \\
 & \qquad T_f(\text{act:pit})/T(\text{act:pit}) \times 100 = 94.1\% \\
 & \text{iの処理による活性化battingインスタンスフ} \\
 & \text{レーム数:} \qquad T(\text{ac:bat}) = 9 \\
 & \text{iiの処理による値を埋め込まれたフレーム数:} \\
 & \qquad T_f(\text{act:bat}) = 8 \\
 & \qquad T_f(\text{act:bat})/T(\text{act:bat}) \times 100 = 88.9\%
 \end{aligned}$$

これらの動作に対してはiに対してiiの処理が高い割合でなされることがわかる。

4 今後の課題

今回の実験ではインスタンスフレームが活性化している期間を1イベントとした。このイベントという単位は、発話データにおいて話題の主題の転換によって客観的に区切ったものである。

さらに今後は、活性化しているインスタンスフレームの数（インスタンス領域の広さ）と値の埋め込みの割合の相関関係について検討してゆく。効率的な伝達は、少ないインスタンスフレームにより多くの値を埋め込めたときに達成できると考える。これは計算機システムにおけるCPUとメモリの間に位置するキャッシュの役割と同じように見てとれる。またさらにこの伝達の効率性の問題から、インスタンス領域の大きさの限界を求めることによって、心理学的な短期記憶の容量と対応させた考察が得られると考えている。

参考文献

- [1] Ishizaki, S.: Generating Japanese Text from Conceptual Representation, in McDonald, D. and Bolc, L. (Eds.), *Natural Language Generation Systems*, Springer-Verlag, pp.256-279. (1988)
- [2] 竹内, 三輪: ラジオでの野球の実況中継を理解する認知モデルの検討, 人工知能学会 第9回全国大会論文集 (1995)
- [3] 田窪行則: 対話管理の言語学モデル 対話管理標識としての提題形式, 音声・言語・概念の統合的処理による対話の理解と生成に関する研究 第2年次研究成果報告 pp.237-244 (1994)
- [4] Grice, H.G.: Logic and conversation, in Cole, P. and Morgan, J.L. (eds.), *Syntax and Semantics*, Vol.3, Academic Press. (1975)