

4L-6

談話理解におけるチャンキング過程についての一考察

横山哲也 井上雅夫 井村英夫

同志社大学

1. はじめに

談話理解の過程において、chunkingというものは不可欠の働きである。本論では記憶の階層構造と対比させながらchunkingの様々なレベルについて述べ、談話理解のモデルを作成する。

2. chunkingとは

人間は言語を理解するとき、無意識のうちに適当な部分にまとめる。これがchunking(chunk化)である。つまり、chunkingとは入ってきた情報をより高次のレベルでまとめ上げることである。

また、chunkingの副作用として、低次の情報の一部が失われる。例えば、綴りの誤りや前置詞の間違ひは意識して覚えられない限り記憶に残せない。

3. chunkingレベル

chunkingには以下のようなレベルがあると考えられる。レベルが上がるほど、より深い意味処理が行われることになる。

- 1) 語 文字列(音素列)を語としてまとめる。
- 2) 句(格) 語の集まりを句、あるいは格文法でいう格にまとめる。
- 3) 文 格を動詞中心に結びつけ、文としてまとめる。
- 4) 文章 似たような文が入力されたとき、以前の文との融合を試みる。
- 5) 話題 複数の文章が1つの話題としてまとめられる。
- 6) エピソード 話題の流れがエピソードとしてまとめられ長期記憶に固定。
- 7) 一般化 エピソードが一般化され、スクリプトとなる。

本モデルでは1)~5)までを実現した。また、1)~3)は格文法パーザで行う。記憶階層で言えば1)~4)を短期記憶で、4)~5)を中期記憶で行う。5)~7)は長期記憶の領域だがまだ実現できていない。また、短期記憶では意味処理は全くされないという考えもあるが、本モデルではある程度までの意味処理は行われるとした。

4. 記憶の構造

1) 短期記憶 (STM: short-term memory)

よく知られているように、短期記憶の容量は 7 ± 2 である。この数字の単位は文字や語というような固定したものではなく、動的なものである。しかし、本モデルでは簡単のため文を単位にとった。

本モデルでは短期記憶をさらにexplicit部とimplicit部に分けた。explicit部には入力文の内部表現がほぼそのままの形で入り、implicit部には入力文から推論された事柄(入力文と同時に成り立つ事柄)が入る。短期記憶を2つの部分に分解したことによる利点は、1)事実と推論を区別できること、2)短期記憶から、中期記憶あるいは長期記憶へ移行する時にexplicit部とimplicit部が一緒になると考えれば、推論した結果をやがてそう思い込んでしまう「思い込み」の現象も説明できることである。

2) 中期記憶 (MTM: mediate-term memory)

談話の内容を保持するには、短期記憶では容量的に不十分である。しかし、長期記憶は短い時間の間に大きく変化するものではない。そこで両者の中間に中期記憶があると考えられる。

本モデルの中期記憶は、1)短期記憶のようにexplicit部とimplicit部に分かれているのではなく、2)話題ごとに領域が分かれていることに特徴がある。1)は、短期記憶とあわせて「思い込み」の現象を説明することができる。2)は、現在話されている話題に関係した領域のみが活性化していると考えることによって「記憶の検索」の問題を説明することができる。

3) 長期記憶 (LTM: long-term memory)

短期記憶、中期記憶を生成するときに参照される。つまり、本モデルでは記憶の生成はボトムアップで、参照はトップダウンで行われる。

A Study of the Chunking Process in Discourse Understanding

Tetsuya YOKOYAMA, Masao INOUE, Hideo IMURA

DOSHISHA Univ.

5. 処理プロセス

肯定文は以下のように処理される。

1) 格文法パーザによる構文解析

2) 短期記憶の生成

- ・ 構文解析された文（以下、まぎらわしくない限り単に「文」とする）と短期記憶のexplicit部を比較。
- ・ 全ての情報が既知であればなにも行わない。
- ・ 動詞が同じで一部の格に新情報が含まれていればレベル4のchunkingを試みる。
- ・ 全ての情報が新情報、あるいは上述のchunkingに失敗したときはそのままexplicit部に格納。

3) 中期記憶の生成

- ・ 文中に含まれる語から可能性のある話題を得点をつけてリストアップする。得点は長期記憶から得られる。
 - ・ それまでの話題リストに得点を足し合わせる。
 - ・ 話題ごとに領域を作り、explicit部、implicit部の両方に相当するものを格納（レベル5のchunking）。
 - ・ 話題リストは常に短期記憶中の文から構成される。つまり、話題は短期記憶で決定されるといえる。
- 疑問文が入力された場合には、explicit部、implicit

部、中期記憶（その時点での話題の可能性の順）の順に答えが検索される。このとき、話題リストが中期記憶へのリンクポイントの役割を果たす。

図1および図2に短期記憶の容量を2におさえた例を示す。両者は(4)の文のみが異なる。いずれの場合も(5)の質問が発せられた時、短期記憶には(3)と(4)の文しか存在しないので中期記憶がアクセスされる。このとき、図1の場合には「買物」に関係した領域が、図2の場合には「食事」に関係した領域を検索する。そのため同じ質問でも答えが異なる。(*)は現在の話題とその得点を表示させたものである。中期記憶へのリンクが変わっていることがわかる。

6. まとめ

以上のようにchunkingと記憶階層をとらえ、談話理解の過程について考察した。また、実際に計算機上にモデルを作成し、それによって、推論した結果を事実であるかのように思い込む現象、話されている話題によって記憶のリンクが変わることが説明できた。今後の課題としては、省略された部分を補う機能の付加、中期記憶から長期記憶への移行のプロセスの付加がある。

参考文献

田中穂積：談話と記憶構造,サイコロジー,9,35-41,1980

前川肇,井上雅夫,井村英夫:

動詞の意味シソーラスによる談話理解についての考察,情報処理学会第29回全国大会予稿集,1189

```

===== START =====
=> (SETCAP 2)
CAPACITY MAX=2 NOW=0
"SUCCESS"

(1) => JOHN WENT TO THE DEPARTMENT STORE .
      "I SEE."

(2) => HE WENT TO THE RESTAURANT .
      "I SEE."

(3) => HE IS A STUDENT .
      "I SEE."

(4) => HE ATE A FRIED FISH .
      "I SEE."

(5) => WHERE DID HE GO .
      "TO THE RESTAURANT."

(*) => (LISTTOPIC)
      TOPIC(S)...((:EATING 19)
                  (:RESTAURANT 11))
      "SUCCESS"

=> BYE .

===== *BYE* =====
    
```

図1 例1 (買物)

```

===== START =====
=> (SETCAP 2)
CAPACITY MAX=2 NOW=0
"SUCCESS"

(1) => JOHN WENT TO THE DEPARTMENT STORE .
      "I SEE."

(2) => HE WENT TO THE RESTAURANT .
      "I SEE."

(3) => HE IS A STUDENT .
      "I SEE."

(4) => HE BOUGHT A BOOK .
      "I SEE."

(5) => WHERE DID HE GO .
      "TO THE DEPARTMENT STORE."

(*) => (LISTTOPIC)
      TOPIC(S)...((:SHOPPING 7)
                  (:DEPARTMENT_STORE 2))
      "SUCCESS"

=> BYE .

===== *BYE* =====
    
```

図2 例2 (食事)