

結末構造および一貫性に基づく場面構造の解析*

2G-3

角田 達彦[†]、田中 英彦[‡]

{tsunoda,tanaka}@mtl.t.u-tokyo.ac.jp

東京大学 工学部

1 はじめに

テキストや対話を理解するための自然言語理解のプロセスである談話理解の研究の問題点の一つは、談話セグメントの設定の困難さがあげられる [1]。その結果処理の定義の曖昧さの問題が発生する。ここでは文脈情報の一つとしての物理的な場面を談話セグメントとして定義し、段落間の構造、段落内の場面構造の解析を行なう。自然言語処理において多義性などを解消する手段として、場面などの情報を文脈としてトップダウンに与えることが考えられる [2]。場面の切り出しに関しては頑健性のためにボトムアップの方法をとることを考えると、有効な場面を同定する際の問題は、(1) 場面が明示的に与えられていないときの対象物による同定、(2) 前の場面の継続の判断、(3) 場面の構造化、とまとめられる。(2)と(3)は談話の中の場面の構造を切り出し、一つの場面に焦点を当てるという点で(1)の処理 [3, 4] を支えるプロセスである。今回はこれらに絞って解析を行なう。またこの結果に従い、英文における場面の移動、切替えが行なわれる過程を工学的に実現する方法を考察する。

2 物理的背景としての場面

場面という言葉は複数の概念の組合せであり、以下のような構造を持つ。

抽象的場面

感覚的場面 --+ 視覚以外の感覚に基づく場面
+ 視覚的場面 --+ 行為、体験
+ 物理的背景

抽象的場面は会話内容や心理描写、抽象的概念、話題の扱いがあげられるが、ここでは一つの別の場面として扱う。視覚以外の感覚に基づく場面は登場人物の体験あるいは背景、対象物の属性とみなすことができるとし、明示的にはとりあげない。ここでは対象物や登場人物が作用を起こす場としての物理的背景を独立した場面としてテキストから切り出すことを考える。

3 結末性および一貫性に基づく場面の構造

談話構造は意味的には一貫性 (coherence) と呼ばれるまとまりによって成り立っている。それは意義の連続性を示し、一貫性のある一つのまとまりは談話セグメントに対応する。一般には因果関係などの抽象的關係がその連続性を形成する要

素の例としてとりあげられることが多いが、対象、状況、出来事、行為などに注意を当て、関係づけるすべてが考えられる。その一つが対象物や登場人物が共通に作用を起こす物理的背景としての場面である。

場面という談話セグメントを切り出すために、表層のテキストに現れる構造上のパターンである結末構造を利用して推測を行なう。結末構造は、指示、代用、省略、接続、語彙的結束関係 (世界知識と一般常識) の他に、反復生起、部分的反復生起、並行表現、パラフレーズ、時制と相、機能的文構成 (焦点化と再活性化)、イントネーションなどがあげられる [5]。それは人間の短期記憶を考慮し、情報性と簡潔性のトレードオフをとったパターンであり、一貫性を検出するための有効な手がかりとなる。

結末構造から一貫性を検出するには

1. 照応関係 (anaphoric relation) の理解: 文章中で陰に、あるいは陽に言及されている対象の理解
2. 接続関係 (coherence relation) の理解: 文章中で言及されている事象や状態の間の関係の理解

が必要となるが、すべての一貫性を考慮すると処理が爆発し、また定義が曖昧になるので、前述のように物理的背景としての場面に絞って次章のように解析を行なった。

4 場面对象とした照応・接続関係の解析

場面对象とした場合、着目すべき照応、接続関係は以下の2点に絞り込むことができる。

1. 登場人物、対象物などの焦点対象の特定
2. 焦点対象の存在する場面の特定

実際の解析では

1. 焦点となっているものの位置の特定
 - (a) 場所が明示的に与えられる場合 ... A
 - (b) 場所が明示的に与えられない場合
 - i. 焦点、潜在焦点リストを探索 ... B
 - ii. 語彙的結束性を利用して特定 ... C
 - iii. 世界知識、一般常識が必要 ... D
 - iv. 全く新しい場所の場合 ... E
2. 場面の切り替わり
 - (a) 場面の切り替わりなし ... 0
 - (b) 場所の明示的移動 ... 1
 - (c) 焦点対象の一致が見られないとき
 - i. 前の場面と一致せず = 新規 ... 2

*Analysis of Scene Structure Based on Cohesion and Coherence

[†]日本学術振興会特別研究員

[‡]Tatsuhiko TSUNODA, Hidehiko TANAKA
Faculty of Engineering, University of Tokyo

表 1: 対象文データ

章	0	1	2	3	4	総数
文の数	389	175	169	123	97	953
段落数	120	38	51	29	23	261
大段落数	4	3	1	1	2	11

表 2: 解析結果

	0	1	2	3	4	5	6	7	総数
A	2/1	20/1	5	0	0	0	2	0	29/2
B	165/65	0	0	9/1	0	1	7/2	9/1	191/69
C	11/8	0	0	0	0	0	0	1	12/8
D	12/1	1	0	1	0	0	5	0	19/1
E	0	0	6	0	0	0	3	1	10
計	190/75	22/1	13	13/1	4	6	23/2	18/1	261/80

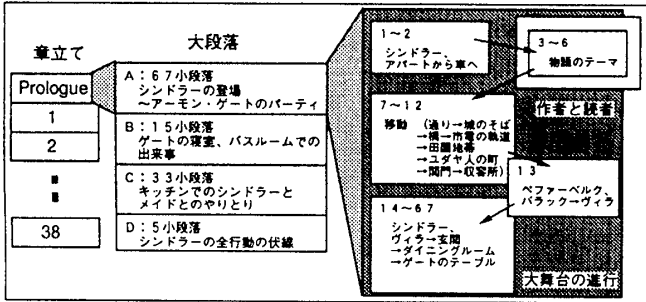


図 1: Schindler's List の全体像と場面構造例

ii. 昔の場面と一致 = 復帰 ... 3

- (d) 手がかり語句 ('by now(そのころ)' など): 時間は一致するが場所は一致しないというニュアンスをもつので新たな場面とする ... 4
- (e) 焦点が人物や移動物の場合、空白時間が大きいと、異なる場面であるとする ... 5
- (f) 時制や相が異なると、回想や心理描写、語りかけ、一般的事実の陳述として新たな場面であるとする ... 6
- (g) 場面がとらえられない場合 ... 7

という分類方法を用いる。

文レベルの解析をすべへ行なうと冗長な場合が多く効率が悪いため、場面単位の近似として段落分けを利用し、段落間の関係に重点を置いて解析を行なう。また場面単位が段落単位よりも小さく、より詳細な解析が必要となった場合のみ段落内の解析を行なうこととする。

5 解析結果と考察

解析対象は Thomas Keneally の Schindler's List のうちのプロローグ (0) と 1 章から 4 章までの 953 文、261 段落である (表 1)。この本の全体像と構造例を図 1 に示す。前章の分類方法に従った解析結果は表 2 に示される。ただし焦点は各段落の第一文の主語とし、'/' の右側は the などの間接照応が存在する事例数である。解析における知見をまとめる。

- 場面を特定するために、目標文の焦点だけでなく、潜在焦点に対する照応を行なう必要がある文例が 3 つ (B-0) あり、この方法では解決できない。
- 世界知識、一般常識として仮定しなくてはならない知識は、(1) 頭文字が略号として使われること (1 例)、(2) 会話は交互に行なわれること (7 例)、(3) 'there was' は前の場面を仮定する (1 例)、(4) 作者 ↔ 読者という特殊な世界の存在、(5) 'it' などで一般的事実の陳述 (4 例) (6) 回想 (3 例)、(7) 時候の 'it' (3 例)、(8) 地名が複合名詞に組み込まれる (1 例)、(9) より深い推論 (2 例)、(10) 'know' が状態を表すこと (2 例)、(11) イタリックなどテキスト表層情報の利用 (1 例) などであった。

- 語彙的結束性を利用するものは 8 例あり、すべて場面の部分-全体関係を要求するものであった。
- 手がかり語句としては、however などの接続詞、by now や always などの時間を表す句、everyone などの限量的指示代名詞などがあげられる。
- 他はすべて通常の照応の問題に帰着した。
- 表より、間接参照を利用すると、場面の切替えを行わない場合の検出に有効ではあるものの、場面の復帰などでも使われる場合もあるので、注意を要することがわかる。
- この手法は以上のような知識の他に、場面フレームとして location と actors だけを検出していけば良いので、工学的実装が容易であるという特長を持っている。
- 段落内で場面が構造を持つ場合は解析したうち 24 段落あった。その内容は (1) 回想、心理描写、事実陳述の混入 (9 例)、(2) 細かい移動 (14 例)、(3) 後で因果関係をつくる伏線の混入 (1 例) であった。各場面の依存関係の詳細な解析は今後の課題とする。

6 おわりに

談話セグメントとして物理的場面を設定し、新規場面への切替えや移動、復帰などの処理を結束構造から一貫性を検出する仮定ととらえ、照応関係、接続関係を解析することによって実現する手法を提案した。照応が完全ならば 84% の正確さで切り分け処理を行なえる。解析を行なった実際の文においては段落レベルにおいてはほぼこの枠組で処理できるという結果が示されたが、照応関係の解決、世界知識などが十分にあることが前提であるので、これらの手法や知識の整備が必要であり、また段落内の細かな場面構造の解析が今後の課題としてあげられる。なおこの研究の一部は、文部省科学研究費の助成による。

参考文献

- [1] 桃内佳雄. 文章における接続関係の解析のための基礎的考察. 情処研報, Vol. 90, No. 64, pp. 97-104, 1990.
- [2] T. Tsunoda and H. Tanaka. Semantic ambiguity resolution by parallel distributed associative inference and contradiction detection. In *Proceedings of IJCNN, Nagoya-93, vol. 1*, pp. 163-166, 10 1993.
- [3] 角田達彦, 田中英彦. 辞書ベース連想による場面同定に必要な文脈情報量の推定. 情報処理学会第 48 回全国大会, Vol. 3, pp. 171-172, 1994.
- [4] T. Tsunoda and H. Tanaka. Analysis of scene identification ability of associative memory with pictorial dictionary. In *Proceedings of COLING-94*, 8 1994.
- [5] R. ボウグラント, W. ドレスラー. テクスト言語学入門. 紀伊国屋書店, 1984.