

# テ形接続節の接続関係の推定

大石 亨

北陸先端科学技術大学院大学  
情報科学研究科

松本 裕治

奈良先端科学技術大学院大学  
情報科学研究科

## 1 はじめに

接続関係 (coherence relation) とは、文章を構成する要素である節と節、あるいは文や文群の間の意味的なつながり関係をいう。接続関係の適切な理解は、文章の全体構造の正しい理解へとつながり、最終的には文章全体のより深い理解をもたらす。われわれは接続関係の適切な理解を通して、文章における首尾一貫性をとらえているといえる。[9]

本稿は、連用形接続とともに最も広く用いられる節間の接続形式である「テ形」による接続の接続関係を推定する手法を提案する。

## 2 テ形接続の表す接続関係

本稿では、いわゆる接続助詞の「テ」を考察対象とする。以下のように、テ形接続によって表される意味は、非常に多岐にわたる。これは、これらの意味が「テ」そのものではなく、接続される節の意味によって分化するものであることを物語っている<sup>1</sup>。

### (1) Circumstance

痛みをこらえて、走り続けた。

### (2) Additive

太郎は、明るくて、勤勉だ。

### (3) Temporal Sequence

午後は、手紙を書いて、論文を読んだ。

### (4) Cause-Effect

台風が来て、家が破壊された。

### (5) Means-End

お金を借りて、新しい車を買う。

### (6) Contrast

太郎は結婚して、次郎は就職した。

### (7) Concession

彼は、お金があつて、貸さない。

以下では、これらの関係がどのように認識されるかというモデルを示し、実験によってこのモデルを検証する。

<sup>1</sup>用例は、Hasegawa 1996[2] による。

## 3 接続関係の階層構造

テ形接続の意味が多岐にわたるといっても、全く関連のない事態はテ形で接続することはできない。テ形で接続すると、聞き手、あるいは読み手はその事態の間に何らかの関連を推測してしまう。これを仮に「共立性 (togetherness)」と呼ぶ。共立性は、その時間的な性質によって、二つの下位カテゴリーに分類できる。継起的 (sequential) なものと並立的 (parallel) なものである。前者は二つの事態が連続して起こることを、後者は同時並列的に起こることを表している。それぞれのカテゴリーは、事態の内容によって、さらに細分化される。図1にテ形接続によって表される接続関係の階層構造を示す。詳細な説明は、Oishi 1998 [4, 6] を参照されたい。

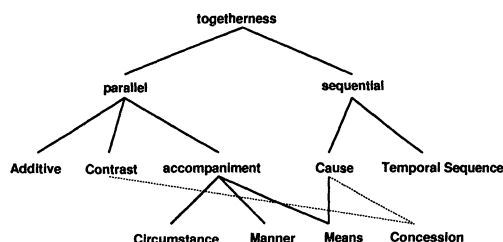


図1: 接続関係のカテゴリー構造

## 4 接続関係の推定

接続関係の推定に用いたアルゴリズムは、以下のとおりである。

**STEP1** Assume part of semantic structures of the conjuncts by reverse linking

**STEP2** Unify them with a verb's semantic structures

**STEP3** Infer the most feasible relation between them

STEP1 では、テ形で接続される節のそれぞれの構成要素から、リバースリンキングによって意味構造の断片が生成される。リンキングとは、概念構造を言語構造

$$\begin{aligned}
& (\forall w, e, g, p) go(e, y, p) \wedge locational(e) \wedge goal(g) \\
& \quad \supset pp("w - ni", g) \wedge place(g) \\
& (\forall w, e, g, p) go(e, y, p) \wedge possessional(e) \wedge goal(g) \\
& \quad \supset pp("w - ni", g) \wedge thing(g) \\
& (\forall w, s, y, p) be(s, y, l) \wedge locational(e) \wedge at(l, p) \\
& \quad \supset pp("w - ni", p) \wedge place(p) \\
& (\forall w, e, x, y) act(e, x, y) \supset pp("w - ga", x) \wedge animate(x) \\
& (\forall w, e, y, s) become(e, y, s) \supset pp("w - ga", y) \\
& (\forall w, s, y, l) be(s, y, l) \supset pp("w - ga", y) \\
& (\forall w, e, x, y) act(e, x, y) \supset pp("w - o", y) \\
& (\forall w, e, x, y, s) act(e, x, y) \wedge become(e, y, s) \\
& \quad \supset pp("w - o", y)
\end{aligned}$$

図 2: リンキングルールの例

に変換するための制約であり、概念構造の項の位置にしたがって、文法的役割が付与される。リバースリンキングはこの過程を逆にたどるものであり、言語構造から概念構造を復元するものである。日本語の場合、格助詞によってこの処理が呼び出される。STEP2 では、動詞の概念構造が惹起されて、STEP1 で出力された構造の断片と単一化 (unification) される。この時点で、単一化されなかった概念の断片は放棄され、それぞれの節の概念構造が一意に決まる。リンキングルールと動詞の概念構造の例を図 2 と図 3 にそれぞれ示す。

STEP3 では、これらの構造の組合せによって接続関係を推測する。

#### 4.1 接続関係認識に関与する属性

接続関係認識に関与する属性のうち、重要なものとして次の三つを考える。

- 1) 節間での主語の一致および動作主性
- 2) 各節で記述される事態の項構造およびアスペクト属性
- 3) 両方の節に関与する名詞句に伴う規範的事態

以上の属性に依存せず、我々の接続関係の認識に影響を及ぼすと考えられるものを、あらかじめ、表 1 にまとめておく。これらの表現は、辞書に登録されており、これにマッチしたものは、以下で述べる処理を行わず、接続関係を決定する。

$$\begin{aligned}
& (\forall s, y, z, l) be(s, y, l) \wedge at(l, z) \supset State(s) \\
& (\forall e, x, y) act(e, x, y) \supset TransAct(e) \\
& (\forall e, x) act(e, x) \supset IntransAct(e) \\
& (\forall e, y, p) go(e, y, p) \wedge path(p) \supset Move(e) \\
& (\forall e, y, s, l, z) become(e, y, s) \wedge be(s, y, l) \wedge at(l, z) \\
& \quad \supset Achievement(e) \\
& (\forall e, e_1, e_2, x, y) act(e_1, x, y) \wedge cause(e, e_1, e_2) \\
& \quad \wedge become(e_2, y, s) \wedge be(s, y, l) \wedge at(l, z) \\
& \quad \supset Accomplishment(e) \\
& (\forall s) State(s) \wedge thing(y) \wedge place(z) \supset verb("aru", e) \\
& (\forall s) State(s) \wedge animate(y) \wedge place(z) \supset verb("iru", e) \\
& (\forall e) Move(e) \wedge manner1 \supset verb("hashiru", e) \\
& (\forall e) Move(e) \wedge manner2 \supset verb("aruku", e) \\
& (\forall e) Accomplishment(e) \wedge manner3 \supset verb("nuru", e) \\
& (\forall e) Accomplishment(e) \wedge manner4 \wedge locational(e) \\
& \quad \supset verb("sosogu", e) \\
& (\forall e) Accomplishment(e) \wedge state1 \wedge \\
& \quad identificational(e) \supset verb("mitasu", e)
\end{aligned}$$

図 3: 動詞の意味構造

#### 4.2 プロトタイプと拡張

われわれは、以前の研究で、コーパスに現れた表層表現によって日本語の動詞を約 30 の意味クラスに分類し、それぞれのクラスに語彙概念構造 (LCS) を与えた [5]。これを利用することにより、比較的一貫性のある方法で接続関係を推定することが可能となる。

われわれは、接続関係を一種のカテゴリーと考えている。カテゴリーは、必要十分条件によって規定されるものではなく、プロトタイプ効果をもたらすような構造を持つ [7, 3, 8]。それぞれの接続関係のプロトタイプを次

表 1: 接続関係を規定する表現

接続関係	表現の種類	例
Temporal Sequence	経過動詞	過ぎる, 経過する ...
	終了動詞	終わる, 終える ...
	継続動詞	続く, 引き続く ...
	副詞句 固定表現	その後, 今では ... [TE] になって, ほどして
Means-End	使用動詞 固定表現	使う, 使用する ... によって
Cause-Effect	固定表現	だけあって, をうけて ...
Circumstance	静的関係動詞	沿う, 基づく ...

のように定義する.

- Circumstance  
[x ACT]<sub>2</sub> WITH [x BE z]<sub>1</sub>
- Additive  
[x BE z<sub>1</sub>]<sub>1</sub> AND [x BE z<sub>2</sub>]<sub>2</sub>
- Temporal Sequence  
[x GO TO z<sub>1</sub>]<sub>1</sub> THEN [x GO (FROM z<sub>1</sub>) TO z<sub>2</sub>]<sub>2</sub>
- Cause-Effect  
[x ACT ON y]<sub>1</sub> CAUSE [y BECOME z]<sub>2</sub>
- Means-End  
[x ACT]<sub>2</sub> BY [x ACT]<sub>1</sub>
- Contrast  
[x ACT]<sub>1</sub> WHILE [y ACT]<sub>2</sub>
- Concession  
[x ACT ON y]<sub>1</sub> BUT [y NOT BECOME z]<sub>2</sub>

ここで, WITH, AND, THEN, などは, それぞれの  
 接続関係に対応するものであり, 二つの事態を引数にとり, 一つの一貫した事態を返す関数である. 角括弧 “[ ] ”  
 が個々の節に対応し, 添字は表層における順序, 内部は意味構造を表す. ACT, BE, GO, BECOME は, それぞ  
 れ行為, 状態, 移動, 変化に相当する大まかな事態の分  
 類を表している. これらは, STEP2 で構成される (図  
 3 参照). 変数 x, y が主語の一致不一致を表し, 動作主  
 性は, 事態のタイプによって暗に表現されている (ACT  
 > GO > BECOME > BE).

テ形接続の広い範囲の表現をカバーするために, プロ  
 トタイプを拡張した結果を表 2 (同一主語の場合) と  
 表 3 (主語が異なる場合) に示す. プロトタイプは太字  
 で示され, 様々な方向に拡張されている. それぞれの拡  
 張には理由があるが, 紙面の都合で記述できない. 詳細  
 な説明は, Oishi 1998 [4, 6] を参照されたい.

拡張の結果, 多くの事態の組合せが複数の接続関係を  
 割り当てられてしまう. ここで, 先の階層構造で近く  
 ある接続関係が同一のセルを占める傾向があることに注  
 意されたい. 接続関係は, カテゴリーであるので, その  
 境界は明確に区分されるものではない. しかし, モデル  
 を評価するためには, ある程度の識別を行なわざるをえ  
 ない. そこで, 以下のようなヒューリスティックを用意  
 して, 接続関係を一意に決定した. (I は同一主語を, D  
 は別主語を表す. 引数は各節のイベントタイプである.)

- I(ACT, ACT), I(ACT, GO)  
If 時間設定表現が含まれていれば, *Temporal Sequence*;  
else if 先行節の動詞が様態を含んでいれば, *Circumstance*;  
otherwise, *Means-End*.
- I(ACT, BECOME)  
If 後続節の事態が心理的なものであれば, *Cause-Effect*;  
else if 先行節の動詞が様態を含んでいれば, *Circumstance*;  
otherwise, *Means-End*.

表 2: 事態構造の組合せ (同一主語)

2nd clause	1st clause			
	ACT	GO	BECOME	BE
ACT	Means Cir(manner) TempSeq	TempSeq		Circum
GO	TempSeq Cir(manner) Means	TempSeq	TempSeq Cause	TempSeq Circum
BECOME	Cause Means Cir(manner)	Cause TempSeq	Cause	Circum Cause
BE	Cir(manner)	TempSeq		Additive Cause Circum

表 3: 事態構造の組合せ (別主語)

2nd clause	1st clause			
	ACT	GO	BECOME	BE
ACT	Contrast			Circum
GO		Contrast		Circum
BECOME	Cause	Cause	Contrast Cause	Cause Circum
BE	Concession			Contrast Circum

- I(GO, BECOME)  
If 後続節の事態が心理的なものであれば, *Cause-Effect*;  
otherwise, *Temporal Sequence*.
- I(BECOME, GO)  
If 先行節の事態が感覚的なものであれば, *Cause-Effect*;  
otherwise, *Temporal Sequence*.
- I(BE, GO)  
If 時間設定表現が含まれていれば, *Temporal Sequence*;  
otherwise, *Circumstance*.
- I(BE, BECOME)  
If 後続節の事態が心理的なものであれば, *Cause-Effect*;  
otherwise, *Circumstance*.
- I(BE, BE)  
If 後続節の事態が心理的なものであれば, *Cause-Effect*;  
else if 両方の述語が属性記述的な形容詞または名詞であれば,  
*Additive*;  
otherwise, *Circumstance*.
- D(BECOME, BECOME)  
If 両方の節の主語が「は」でマークされていれば, *Contrast*;  
otherwise, *Cause-Effect*.
- I(BE, BECOME)  
If 先行節の状態が関係的なものであれば, *Circumstance*;  
otherwise, *Cause-Effect*.
- D(BE, BE)  
If 両方の節の主語が「は」でマークされていれば, *Contrast*;  
otherwise, *Circumstance*.

一方, いくつかの組合せで, 関係が指定されていないも  
 のがある. これらは三番目の属性, すなわち両方の節に  
 関与する名詞に結び付いた規範的イベント等を用いて決  
 定されるべきものであろうが, 現在のところ, 名詞辞書  
 が十分ではないので, 既定値として Circumstance 関係  
 を与えた.

## 5 実験結果

実験は、EDR コーパス [1] からランダムに抽出した 280 文に対して行なった。結果を表 4 に示す。(括弧内の数字は、固定表現とのマッチによって決定されたもの)

全体の正解率は 82% であった。このことは、節のイベントタイプによって、(少なくともテ形接続に関しては) かなりの程度、接続関係が認識できることを示している。多くのエラーは接続関係自体の曖昧さによるものであるが、これは、一文を抜き出したために、文脈が考慮できないためである。人間でさえ、前後の文脈なしに接続関係を判断するのは困難であった。これは、接続関係が明確な境界を持たないことの反映でもある。

一方、本手法の限界を示す例として、次のようなものもあった。

(8) あの人は、82 歳になって、あんなに好奇心一杯だ。この例では、イベントタイプの組合せは I(BECOME, BE) であるので、Circumstance 関係がデフォルトで割り当てられる。しかし、われわれは、一般に 82 歳の人々がそれほど好奇心旺盛でないことを常識として知っている。したがって、Concession 関係が認識されることになる。このように、接続関係の認識は、われわれの一般常識と深く関わっている。ただ、それらの常識の基礎を形成する事態の認識には、本稿で述べた接続関係の創発機構が用いられると考えられる。

## 6 おわりに

本稿は、テ形接続の表す接続関係の認識について述べた。テ形接続は、非常に広い範囲の接続関係を表すので、この手法は、無標の接続関係、すなわち接続が明示されない場合にも十分適用できる。一方、有標の接続関係は、ここで述べた無標の関係創発ではない関係を明示するために場合に用いられると考えられる。

## 参考文献

- [1] EDR, *The EDR Electronic Dictionary Technical Guide*, Japan Electronic Dictionary Research Institute Ltd. (1995), (in Japanese).
- [2] HASEGAWA, Y. *Toward a Description of Te-linkage in Japanese*, Grammatical Constructions (eds. Shibatani, M. and Thompson, S. A.), Oxford University Press (1996), chapter 3, 55–75.
- [3] LAKOFF, G. *Women Fire, and Dangerous Things : What Categories Reveal about the Mind*, The University of Chicago press (1987).
- [4] OISHI, A. *Semantic Structures of Japanese Verb Phrases: — Acquisition, Representation and Use —*, PhD thesis, Nara Institute of Science and Technology (1998), Feb.
- [5] OISHI, A. and MATSUMOTO, Y. Detecting the Organization of Semantic Subclasses of Japanese Verbs, *International Journal of Corpus Linguistics*, 2, 1 (1997), 65–89.
- [6] OISHI, A. and MATSUMOTO, Y. Recognition of the Coherence Relations between Te-linked Clauses, *Proceedings of the 17th COLING and the 36th Annual Meeting of ACL*, 2 (1998), 990–996.
- [7] ROSCH, E. Natural categories, *Cognitive Psychology*, 4 (1973), 328–350.
- [8] TAYLOR, J. R. *Linguistic Categorization \*Prototypes in Linguistic Theory*, Clarendon Paperbacks (1989).
- [9] 阿部純一, 桃内佳雄, 金子康朗, 李光五人間の言語情報処理—言語理解の認知科学, *Cognitive Science & Information Processing* 12, サイエンス社 (1994).

表 4: 実験結果

接続関係	人間による 判断 (a)	システムの 出力 (b)	一致 件数 (c)	recall(%) $c/a \times 100$	precision(%) $c/b \times 100$
Temporal Sequence	89	81(46)	79	89	98
Circumstance	75	83(22)	63	84	76
Cause-Effect	64	58(13)	48	75	82
Means-End	45	48(12)	34	76	71
Additive	3	3	3	100	100
Concession	3	5	1	33	20
Contrast	1	2	1	100	50
Total	280	280(92)	229	82	82