

物語の場面抽出と場面連鎖の構造

Structure of scene connections extracted from stories

藤井 美緒† 中山 伸一† 真栄城 哲也†
Mio Fujii Shin-ichi Nakayama Tetsuya Maeshiro

1. 序論

本研究では物語の構造表現およびその構造に基づいた分類ならびに検索を実現するために、物語の構造をプロットで表して解析する。プロットとは物語で起きる出来事の最小単位である「場面」を物語の中で語られる順に並べたものである。場面同士のつながりを見る上では、場面間に現れる連鎖の形態が一つの指標になる。また、場面の内容と前後の連鎖の形態を比較することで、プロットの類似性を見ることができる。

2. 方法論

物語をプロットとして複数の場面に切り分ける際、どこまでを一つの場面と捉えるかについては様々な定義が可能である。本研究では、以下の四つの定義を用いて物語を階層のおよび複数の視点から捉えることにした。

- (1) **動詞単位** 一つの動作を一つの場面とし、文を構成する動詞毎に一つの場面を割り当てる。
- (2) **人物単位** 動作の主体が同じ登場人物である間を一つの場面とする。
- (3) **中間単位** 連続する動作をまとめて一つの場面とする。多くの場合は動詞単位と人物単位の間に近い分け方になる。
- (4) **シーン単位** 人物や動作よりも広い視点から、内容にまとまりのある一連の流れを一つの場面とする。

なお、ここでは表現手法の検討に重点を置いているため、どの手法でも手動で場面を抽出した。

プロットのもう一つの要素である連鎖については、抽出した場面間の関係と定義し、関係の種類によって分類する。分類の際に手がかりとなるのは、場面同士をつなぐ接続語や指示語、時間を表す特定の単語などである。ここでは日本語の接続に関する文献[1-5]を参考に、動詞単位の場面に現れる連鎖を以下の六種類に分類した。

- (1) **因果** 前後の場面が因果関係を持つ連鎖。
- (2) **転換** 場面の内容が前後で途切れ、描写の視点が変わる連鎖。
- (3) **順接** 前の場面に次の場面が追加された連鎖。
- (4) **逆接** 前の場面の出来事に背反あるいは対立する出来事が次の場面で起きる連鎖。
- (5) **修飾** 二つの場面のどちらかがもう一方の場面を修飾あるいは解説する連鎖。
- (6) **並立** 場面同士が完全に並列する関係にあり、前後の

順番を入れ替えても内容が変わらない連鎖。

なお、分類の候補が複数生じて判断が難しい場合には、番号が若いほど優先順位が高いと見なし、該当する種類の中で最上位の種類に分類する。

3. 結果および考察

千一夜物語[6]の類似するストーリー展開を持つ五種類の物語の一部(以後、部分プロットとする)を解析対象とし、実際に四通りの場面の抽出手法を用いて抽出した場面の特徴を調査した。

抽出された場面には、手法毎に異なる特徴がある。動詞単位の抽出手法は定義が明確であり、文章から客観的に場面を抽出することができる。これは最も細かく場面を分割する手法でもあるため、他の手法とも関連付けやすく、すべての手法の基準となる。一方、中間単位の抽出手法は文意を重視して、連続する動作を一つの場面として抽出する。そのため動詞単位では過度に分割されてわかりにくくなる場面群も、中間単位では直感的に、よりわかりやすい形で抽出される。その結果、中間単位で抽出された場面は連続する動作毎に何が起きたのかをつかみやすい。人物単位は動作の主体となる人物を基準に場面を抽出するため、行為者(登場人物)の変遷を追うことができる。シーン単位の抽出手法は一つの場面对応する文章の量が最も多いため、話の流れをより大まかに捉えることができる。したがって、物語の流れを全体的に把握したいときなどに、シーン単位で抽出した場面を利用してあらかじめ知る事ができる。

それぞれの手法の場面の分割点については、シーン単位の分割点では必ず中間単位の場面も分割された。このことから、人物単位を除いた各抽出手法の間には、シーン単位が最上層に位置し、動詞単位が最下層に位置する階層関係が作られているといえる。このとき、一場面あたりに含まれる内容の抽象度は、抽出手法の階層レベルに比例する。シーン単位と中間単位の間で抽出の定義を変えることによって、二手法の間に位置する別の抽出手法を作ることできる。ここで人物単位の抽出手法が階層関係に含まれない理由は、場面の抽出基準が他の手法とは異なるためである。人物単位の手法では動作の主体(動詞の主語)を基準にして場面の抽出を行うため、主体が変わるたびに機械的に場面が分割される。その結果、文意から場面を分割するシーン単位や中間単位の場面とは分割点が重ならないことがあり、抽出された場面そのものも他の手法とは異なる性質を持つ。

四通りの手法で場面を抽出した後、各場面の分割点に現れる連鎖の結果を分析した。ここでも人物単位の手法は他の手法とは異なる特徴を示し、連鎖の種類を明確な基準で分類することができない。これは人物単位の場面が行為者(登場人物)を基準に抽出されるため、その分割点には行為者の変化のみが現れ、前後の場面内容の対比が難しくな

† 筑波大学大学院図書館情報メディア研究科

るためである。しかし、人物単位の場面間で重要なのは動作を行う人物の変遷であることから、連鎖の種類を場面の内容関係に基づいて分類する必要はない。他の手法の連鎖に関しては、動詞単位と同じく場面同士の内容の関係から六種類に分類できる。実際に各手法の連鎖について分類を行うと、手法によって現れる連鎖の傾向が異なることがわかった。表1と図1, 2はすべての部分プロットを対象に、現れた連鎖の種類毎の出現回数と連鎖の総数に占める割合を見たものである。これを見ると、階層関係の上位に位置する手法ほど出現する連鎖の種類が偏る傾向のあることがわかる。

こういった傾向が現れる原因には、各手法で抽出される場面の特徴が挙げられる。ここで定義した連鎖は前後の場面間の関係だけを抽出するようになっているため、間に無関係な場面を挟んで後の場面と強く関連することがあっても、階層関係の下位に位置する動詞単位の場面抽出ではその関係をうまく抽出することができない。一方、中間単位では文章の意味上の接続関係も考慮して場面が分割されるため、抽出される連鎖の種類が動詞単位とは異なるものになる。ここでは連続した動作が一つの場面と見なされるため、修飾や並列に分類される連鎖は動詞単位に比べて減少することが多い。中間単位で一場面にとめられる連鎖には、他にも対比の意味で使用される逆接や連続動作を結びつけた因果関係などがある。さらにシーン単位では、下位の抽出単位で修飾・並列に分類された関係がすべて一つの場面に包括されて、連鎖としては現れない。シーン単位に現れる連鎖の中で他の手法よりも転換の割合が高い理由も、転換は物語中の視点の変化を表す連鎖であり、物語の流れそのものに影響を与える重要な連鎖であるためだと推測できる。

階層関係にある手法間で場面の分割点が共通している場合には、そこに現れる連鎖の種類も同じであることが多かった。しかし、分割点が同じであっても手法によって連鎖の種類が異なる事例もある。これは階層関係の下位に位置する抽出手法で別々に捉えていた場面を、上位の手法でま

とめて捉えたことから生まれた結果である。抽出手法の階層が上がるほど連鎖の表す場面間の関係は話の流れに沿ったものになりやすく、場面の連鎖にもあらずじが反映されやすくなる。

4. 結論

本研究では、物語の構造を場面の連鎖であるプロットによって表現する手法を提案した。また、実際に四通りの抽出手法で場面を抽出し、場面間に現れた連鎖を六種類に分類した。

プロットの類似性を判断する根拠にもなる特徴的な場面の連鎖は、特徴的な場面が連続して現れることもあれば、関係のないノイズを挟んで散発的に現れることもある。このとき場面の抽出手法の階層を上げると、場面内容の抽象度が高くなるためプロットの特徴的な連鎖部分からノイズが減る。特徴的な場面の連鎖だけが抽出された状態で場面と連鎖の双方に重みを与えて比較すると、類似したプロットの検出精度が上がると考えられる。

5. 参考文献

- 山口明穂, 秋本守英編. 日本語文法大辞典. 明治書院, 2001, 974p.
- 佐治圭三. “文章中の接続語の機能”. 国文法講座 6 時代と文法—現代語. 山口明穂編. 明治書院, 1987, p. 127-154.
- 特集 接続詞のすべて. 月刊文法. vol. 2, no. 12, 1970(10), p. 10-124.
- 塚原鉄雄. 接続の論理—接続詞と接続助詞—. 月刊文法. vol. 2, no. 2, 1969(12), p. 68-74.
- 森岡健二ほか編. 口語文法講座 6 用語解説編. 明治書院, 1965, 404p.
- 豊島与志雄ほか訳. 完訳 千一夜物語. 改版. 岩波書店, 1988, 13冊.

表 1. 場面の抽出手法と連鎖の種類別出現頻度

	動詞単位	中間単位	シーン単位
1. 因果	80 (9.33%)	89 (17.7%)	6 (8.22%)
2. 転換	23 (2.68%)	23 (4.56%)	22 (30.1%)
3. 順接	427 (49.8%)	300 (59.5%)	42 (57.5%)
4. 逆接	29 (3.38%)	33 (6.55%)	3 (4.11%)
5. 修飾	272 (31.7%)	51 (10.1%)	0 (0.00%)
6. 並立	26 (3.03%)	8 (1.59%)	0 (0.00%)
連鎖総数	857	504	73

単位：出現回数（割合）

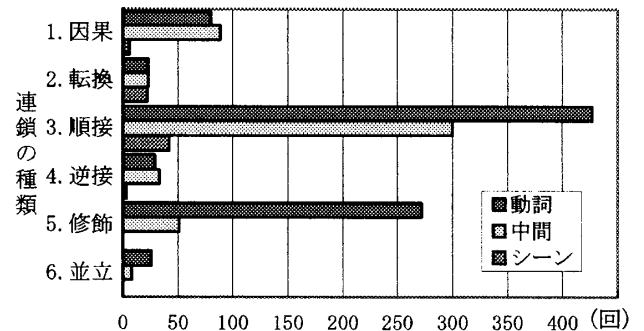


図 1. 場面の抽出手法毎に異なる連鎖の種類別の出現回数

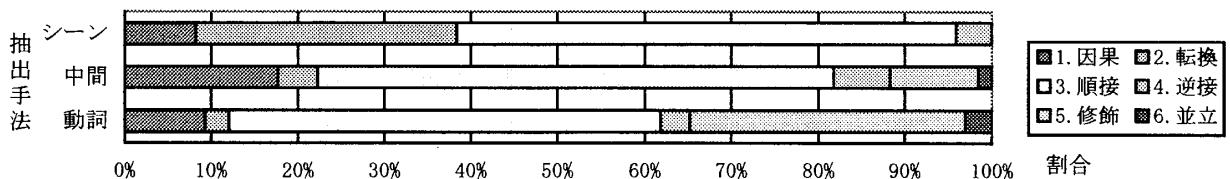


図 2. 場面の抽出手法毎の連鎖の種類別の割合