4H-02　接続助詞の意味解析

沢田康子　足立守啓　池原悟　村上仁一
鳥取大学工学部知能情報工学科

1　はじめに

機械翻訳などの自然言語処理において、重文の接続の意味を解析することは重要な課題の一つとなっている。重文の接続の意味関係については、述語の意味に着目した方法[1]、「用言の連用中止＋(助詞)て」で接続する文を対象とする方法[2]、主節と従属節の「は/が格」の意味関係に着目する方法[3]などがある。

これに対して、本研究では、使用頻度の高い接続助詞「ば」、「と」、「ても」によって接続される重文に着目し、その接続の意味を、主節及び従属節の持つ構造的特徴から決定する方法を提案する。各々の接続助詞の意味分類は、英語との対応付けにより行う。また、判定条件においては、重文の構造上の特徴を調べ、計算機上で判定を行うことを考慮して定める。

2　重文に含まれる接続助詞の意味分類

2.1　対象とする接続助詞の種類

新聞記事経済欄 1 年分 (93,075 文) を解析した結果、接続助詞を持つ重文は 17,242 文存在した。そのうち出現頻度が 1,000 を越える接続助詞は「て」、「が」、「ば」、「と」、「ても」である。この中で接続助詞「て」[2],[3]、「が」[4] については既に研究されている。そこで、本研究では接続助詞「ば」、「と」、「ても」を持つ重文を対象とする。

2.2　接続助詞「ば」、「と」、「ても」の意味分類

接続助詞「ば」、「と」、「ても」を持つ重文について、英文との対応付けにより意味分類を行う。分類結果を表 1 に示す。なお、分類は基礎日本語学習辞典[5]による。

表 1: 接続助詞「ば」、「と」、「ても」の意味分類

「ば」の意味分類	接続の意味	「と」の意味分類	接続の意味	「ても」の意味分類	接続の意味
if	仮定条件	if	仮定条件	although	逆接の既定条件
when	時間、発見	when	時間、発見	even if	逆接の仮定条件1
must	必要	must	必要	no matter	逆接の仮定条件2
according to	話題の出所	according to	話題の出所	whether	逆接の仮定条件3
the 比較級	比例	even if	逆接の仮定条件1	it is all right	許可、許容
as well as	並列	no matter	逆接の仮定条件2	not have to	不要
should	提案	whether	逆接の仮定条件3	anyhow	ー
in terms of	立場、観点、換言、換算	in terms of	立場、観点、換言、換算	at (the) 最上級	ー
副詞 + speaking	比喩	副詞 + speaking	比喩	ー	ー
talking of	言及	talking of	言及	ー	ー
considering	考慮	considering	考慮	ー	ー
as compared	比較	as compared	比較	ー	ー
what 等	疑問	ー	ー	ー	ー

3　接続助詞「ば」「と」「ても」の意味の判定法

3.1　接続の意味の判定条件

各接続助詞に対して、各標本データ 500 文から主節と従属節の構造上の特徴を各意味分類ごとに分析する。その結果、意味分類は、接続助詞「ば」については 13 種類、「と」については 12 種類、「ても」については 8 種類に分類した。「ば」の意味判定方法の一部を表 2 に示す。

3.2　判定条件の適用順序

各判定条件は、全体の正解率が最大となるよう、$P_x = n_x p_x / q_x$ の大きい順に適用する。ここで、P_x は適用順序の優先度を、n_x は判定条件 x の対象とする重文の割合を、p_x は判定率を、そして q_x は誤判定率を表す。

Semantic analysis of conjunctive particle
Yasuko SAWADA,Morihiro ADACHI,Satoru IKEHARA,Jin'ichi MURAKAMI
Tottori University

表 2: 接続助詞「ば」の意味判定例

意味分類	判定条件	例文
if (仮定条件)	*1*. 主節に推量、可能性、懸念可能、完了を表す表現が存在	株式市場が安定すれば、為替も落ち着くだろう。
	2. 主節及び従属節に「~化」という表現が存在	企業活動が活発化すれば、税収が増える。
	:	:
when (時間, 発見)	*1*. 主節の述部が過去形	父はお酒を飲めば、歌を歌いました。
	2. 主節に「」に囲まれた「文」が存在	コピーを頼めば「私、そんな仕事をするために入ったんじゃありません」ときっぱり断られた。
		:

4 思考実験方法と結果

新聞記事経済欄 1 年分中の「ば」、「と」、「ても」を持つ重文各 500 文 (標本データ) より定めた判定条件から p_x、q_x の値を求めた。その後、3.2 節の適用順序に従って、実験データ各 200 文を対象に実験を行った。その結果を人手で判定して、判定率を求めた。なお、最終的に判定に漏れたデータは、標本データで各接続助詞の接続の意味として最も割合の高かったものに分類した。判定率は以下の式 (1) から求めた。実験結果を表 3 に示す。

$$\text{判定率} = \frac{\text{正解文数}}{\text{全原文数 (各 200 文)}} \quad (1)$$

表 3: 思考実験結果

	「ば」	「と」	「ても」
判定率	92%	85%	82%
(正解文数)	(183 件)	(170 件)	(163 件)

また、デフォルト値として、原文における「ば」、「と」、「ても」を最も出現頻度の高い用法と仮定して判定した結果を表 4 に示す。その設定方法を表 5 に示す。

「ば」、「と」、「ても」の判定精度はデフォルト値と比べて、判定率にして 8% 、22% 、19% 上昇した。この結果より、接続助詞「ば」、「と」、「ても」を持つ重文の接続の意味を、主節と従属節の構造的特徴を基に作成した判定条件によって解析する手法の有効性が示された。

表 4: デフォルト値

	「ば」	「と」	「ても」
判定率	84%	63%	63%
(正解文数)	(168 件)	(126 件)	(126 件)

表 5: デフォルト値の設定方法

接続助詞	デフォルト値に用いる用法
ば	if (仮定条件)
と	if (仮定条件)
ても	even if (逆接の仮定条件)

5 おわりに

本研究では、接続助詞「ば」、「と」、「ても」を持つ重文の接続の意味を、主節と従属節の構造的特徴より作成した判定条件から解析する手法を提案し、その有効性が示された。

誤判定、及び判定漏れを起こした対象文について調べたところ、第 2 章で述べた英文との対応付けによる分類では対処できない表現も多くみられた。これに関しては、さらに分類を追加、あるいは細分化することで対処することが可能と思われる。

今後は、さらなる解析精度の向上を目指すべく、用法分類、及び判定条件を再検討する必要がある。また、本手法を実際に計算機上で行う場合の検証、及び本手法を「ば」、「と」、「ても」以外の接続助詞を持つ重文の接続の意味解析に有効かどうかの検証も行っていく必要がある。

参考文献

[1] 木村啓一、西沢信一郎、中川裕志 : 述語の意味を用いた日本語順接複文の理解システム、情報処理学会自然言語処理研究会、96-NL-111,pp,47-54 (1996)

[2] 仁田義雄 : 複文の研究 (上) シテ形接続をめぐって、くろしお出版 (1995)

[3] 向仲こう : 動詞の主体の属性を用いた複文の連接関係の解析、自然言語処理 4 (4) ,3-164 (1997)

[4] 井上慎一、池原悟、足立守啓 : 接続助詞「が」を持つ重文の意味解析 、H10 後期電子情報通信学会、全国大会、D-5-2, pp.31 (1998)

[5] 国際交流基金　日本語国際センター : 基礎日本語学習辞典、凡人社 (1986)