現代日本語書き言葉均衡コーパス（BCCWJ）のコア・データに基づく関係節付加曖昧名詞句と先行文脈内の結束連鎖の分析

中野 陽子（関西学院大学）†

Cohesive Chains Formed between Noun Phrases Including Ambiguous Relative-Clause Attachments and the Preceding Context—Analyses of the Core Data of the Balanced Corpus of Contemporary Written Japanese

Yoko Nakano (Kwansei Gakuin University)

要旨

「黄色い服を着た少女の母親」のように関係節（下線部）が２つの名詞句（少女、少女の母親）のうち、どちらを修飾するのか曖昧な名詞句を関係節付加曖昧名詞句（関係節＋名詞句１の名詞句２）と呼ぶ。関係節付加曖昧名詞句とその先行文脈とのあいだの関係について、英語の関係節の非制限用法に基づいた想定はできるがコーパスに基づいた研究はされていない。そこで現代日本語書き言葉均衡コーパスのコア・データから関係節付加曖昧名詞句を含む分を抽出し、個々の事例毎に日本語母語話者２名に名詞句１、２と先行する談話とのあいだに形成される語彙的結束について、その種類を判定してもらった。判定結果を集計して関係節付加曖昧名詞句と先行文脈の関係を分類した。その結果、従来の先行研究では理論に基づき一種類しか仮定されていなかったが、この分析によって日本語では先行する談話と関係節付加曖昧名詞句との関係のパターンには数種あることが分かった。

１．はじめに

心理言語学の実験では文が単独で提示されることが多いが、日常生活で使われている文はテキストを構成している複数の文の１つとなっており、先行する他の文からの情報を参考に理解される。関係節付加曖昧名詞句を含んだ文の処理に関する心理言語学的研究も同じことが言える。関係節付加曖昧構文は実験の中では単独で提示されることが多い。関係節付加曖昧名詞句には構造的に曖昧な部分があるが、もし先行する他の文の情報があれば、その曖昧性を解消することができる。例えば、英語の関係節の制限用法は先行文脈内に関係節が修飾している名詞句の指示物と同じ種類のものが複数あることが前提となっているとき、その中のどれを指すのか特定するときに使われる。関係節付加曖昧名詞句で使われている関係節の用法は制限用法なので、テキストの中にあれば、先行する文の中に関係節の先行詞となっている名詞句と同じものまたは同等の語句があり、それが関係節の付加に関する曖昧性を解消すると考えられる。

下記の例（1）では下線部が関係節付加曖昧名詞句となっており、下線部のみを単独で読んでも、関係節の that liked swimming in the river が dog と boy のどちらを修飾しているのか曖昧である。しかし先行する文脈に 2 匹の犬がおり、その内の１匹が川で泳ぐのが好きであることが述べられている。先行する文と関係節付加曖昧名詞句との照応関係に整合性を持たせるために、関係節の that liked swimming in the river は boy ではなく、dog を修飾しているという解釈の方が自然である。

† y-k.nakanoATkwansei.ac.jp

(1) A boy had two dogs[a-1&b-1]. One dog liked swimming in the river and the other dog[b-2] liked running along the river bank. The boy's father walked the dog[b-3] of the boy that liked swimming in the river.

例 1 では先行する文内の名詞句(dog)が関係節付加曖昧名詞句内に繰り返し現れることで曖昧性が解消したのである。

１つのテキストの中に同じ名詞句、あるいは同等の語句が繰り返し現れると、それらの語句を含む文がお互いに関連付けられ、複数の文からなるテキストができる。このような関連付けを結束(cohesion)と呼び、語句の連なりは文の繋ぎの役割を果たしており結束連鎖(cohesive chain)と呼ばれる（Halliday & Hassan, 1976）。例（1）では同じ dog という語が繰り返されて文同士が関連付けられテキストを構成している。また dogs[a-1&b-1] と dog[b-2] と dog[b–3] とで結束連鎖が形成されている。結束連鎖を形成する語句と語句の関係は、同じ語句同士の関係に限らず、複数の種類に分類される（詳細は２．２を見てください）。

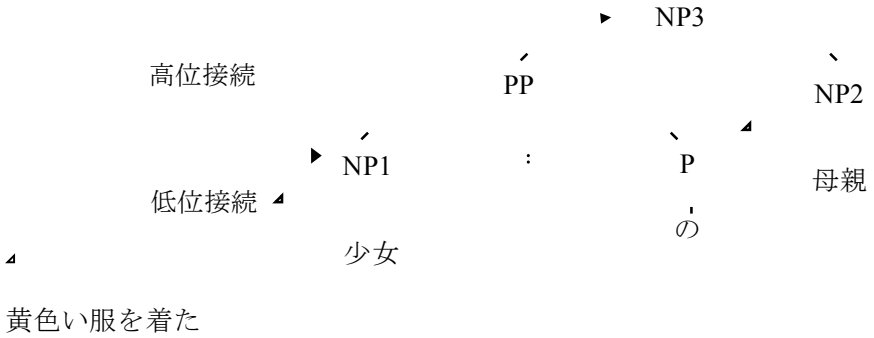
例 1 で見たように、関係節付加曖昧名詞句を含むテキストでは、関係節の先行詞である名詞句(dog)と、先行する文内に現れた同じ名詞句(dog)または同等の語句が含まれている。そこで本研究では BCCWJ のコア・データから関係節付加曖昧名詞句を含むテキストを抽出し、関係節付加曖昧名詞句がテキスト内で先行する文とどのような関係を結んでいるのか、またそれが関係節付加の曖昧性の解消に役立っているのかについて、結束連鎖の種類を分析することによって検討した。

２．背景

２．１　関係節付加曖昧名詞句

「黄色い服を着た少女の母親」のように関係節（下線部）が２つの名詞句（少女、少女の母親）のうち、どちらを修飾するのか曖昧な名詞句を関係節付加曖昧名詞句（関係節＋NP2 の NP1）と呼ぶ。関係節付加曖昧名詞句を樹形図にすると下記の例 2 のようになる。

(2) 黄色い服を着た少女の母親



階層的な樹形図における NP1 と NP2 の高さが異なることから、位置の低い方の NP1 への接続を低位接続（low attachment, LA）、高い方の NP2 への接続を高位接続（high attachment, HA）と呼ぶ。

関係節の接続に関する好みは言語によって異なるという先行研究がある。スペイン語と英語の関係節付加曖昧名詞句に関する調査ではスペイン語母語話者は高位接続を好む傾向が見られ、英語母語話者では低位接続を好む傾向が見られたと報告されている（Cuetos &

Mitchel, 1988）。この研究をきっかけにさまざまな言語で関係節付加曖昧名詞句における関係節の接続に関する好みの調査が行われた。日本語は高位接続が好まれることが示唆されている（Kamide & Mitchell, 1997；中野、早野、西内、井本, 2007）。日本語では関係節のあとに NP1 と NP2 が現れる。それと同じように中国語でも関係節のあとに NP1 と NP2 が出現するが中国語母語話者は高位接続を好むことが報告されている（Shen, 2006）。先行文脈の影響を調べた研究もいくつかある（フランス語：Zagar, et al. 2010; オランダ語：Desmet et al. 2002; ギリシャ語：Papadopoulou & Clahsen, 2006）。どの研究も文処理中の様子を調べる実験と関係節の接続に関する最終判断を調べる課題を実施している。先行文脈の影響があるかどうか文処理中の様子を調べる実験の結果は一致していない。これらの研究はさまざまな点で異なっており、オンラインの文処理の研究結果が異なる理由を特定するのは難しい。一方、関係節の接続に関する最終判断を調べる課題の結果は一致している。どの研究でも先行文脈の影響を受けて関係節の接続が選択される結果となっている。

(3) 低位接続文脈（複数のNP1、単数のNP2）

"L'audience allait débuter et on attendait le juge. Le public nombreux bavardait bruyamment et commentait l'affaire. La chanteuse[a-1] et ses avocats[b-1] se tenaient dans un coin du prétoire. Un journaliste a borda l'avocat[b-2] [N1] de la chanteuse[a-1] [N2] qui paraissait plus confiant(e) que les autres." (The hearing was about to begin and everyone was waiting for the judge. The audience was chatting noisily and talking about the case. The singer [female] and her barristers [male] were standing in a corner of the courtroom. A journalist approached the barrister [male N1] of the singer [female N2] who seemed more confident [feminine or masculine gender] than the others.)

(Zagar et al. 2010; p. 427)

Zagar らの実験で使われた例を見てみると、複数の弁護士（avocats[b-1])が先行文脈に登場するが、歌手は(La chanteuse)1 人だけである。一方、関係節付加曖昧名詞句（二重下線部）では歌手（la chanteuse[a-1]）と弁護士(l'avocat[b-2])が 1 人ずつ登場している。歌手は 1 人しかいないので関係節で特定しなくても指示対象が明確であるが、弁護士は複数いるので関係節の制限用法を用いて特定するとどの弁護士について言及しているのか明確になるため、文脈は低位接続を支持する文脈となっている。実際の実験では歌手を複数形にして弁護士を単数形にすることによって高位接続を支持する文脈条件も作られた。視線計測の実験では二重下線部のような完全な関係節付加曖昧構文が提示されたが、文完成課題では関係節の部分が空欄となっており、被験者が文を完成させるようになっていた。

上記の例 3 では先行文脈内の名詞句と関係節の先行詞が同じ名詞句であり、同じ名詞句の繰り返しで結束性連鎖が形成されている。ただし、文脈内の名詞の複数形であるのに対し、関係節の先行詞は同じ名詞の単数形であり、関係節の先行詞は意味上、文脈内の名詞の複数形に含まれる。

２．２　結束の種類と結束連鎖

結束には２つの種類がある—文法的結束と語彙的結束である。文法的結束は照応、置換、省略、接続などによって形成される。語彙的結束は繰り返しやコロケーションによって形成される（Halliday & Hassan, 1976）。本研究では関係節付加曖昧名詞句内の関係節の先行詞

と、先行するテキスト内にある先行詞と同じまたは同等の語句との間の関係を調べる。同じ語句または同等の語句の繰り返しを扱うため、本研究では語彙的結束の中の繰り返しを扱う。下記の例 4 a～d のそれぞれで下線部の語が if 節の主語の he と同じものを指している。このように繰り返される語句は結束連鎖（cohesive chain）を形成する。4 つの例は結束連鎖を形成している語彙の種類という点で異なっている—同じ語の繰り返し（4a）、同意語または同意語に近いもの（4b）、上位語（4c）、一般的な用語（4d）。

(4) There is a boy climbing that tree.

a. The boy's going to fall if he doesn't take care.　（同一語）

b. The lad's going to fall if he doesn't take care.　（類義語）

c. The child's going to fall if he doesn't take care.　（上位語）

d. The idiot's going to fall if he doesn't take care.　（一般的な語）

(Halliday and Hassan, 1976; pp. 280-281)

本研究ではコーパスから「関係節＋NP1 の NP2」の名詞句を含むテキストを抽出し、テキスト内で NP1 と NP2 と結束連鎖を形成する語句が、どのような種類の結束連鎖を形成しているのか分類することによって、先行文脈の関係節の接続の曖昧性の解消への影響を調べた。

３．本研究

３．１　材料のサンプリング

現代日本語書き言葉均衡コーパス（BCCWJ）のコア・データから検索エンジンの中納言を用いて、「関係節＋NP1 の NP2」の名詞句を含むテキストを抽出し、各ジャンル毎の数を算出した。そのあと、分析に必要な適正サンプル数を計算し[1]、総数における各ジャンルの比率を変えないようにランダムに抽出した（表１）。

表１：抽出されたテキスト数と分析対象にしたテキスト数

ジャンル	新聞	雑誌	書籍	白書	Yahoo 知恵袋	Yahoo ブログ	合計
抽出数	572	133	133	89	173	169	1269
比率(%)	45	10	10	7	14	14	100
分析対象のテキスト数	134	30	30	21	42	42	299

本研究は「関係節＋NP1 の NP2」の名詞句とその前にあるテキストとの関係を調査対象としているため、「関係節＋NP1 の NP2」の名詞句の前にテキストが無い事例は分析の対象外として、その数のテキストを、残りのテキストからランダムに抽出した。また同じテキストが複数回サンプルに入った場合は１回と数え、分析適正数を満たせるように残りのテキストからランダムに抽出した。

[1] 下記の計算式が９５％信頼区間内に入るテキスト数の計算に用いられた。

$$n \geq \frac{N}{\frac{N-1}{P(1-P)} \times \left(\frac{e}{Z}\right)^2 + 1}$$

N=the number of samples, P=0.5, e=0.05, Z=1.96

Yahoo 知恵袋と Yahoo ブログにも「関係節＋NP1 の NP2」の名詞句を含むテキストが入っていたが、テキストとして意味を成さない事例もあり、本研究では分析しないことにした。従って、新聞、雑誌、書籍、白書から抽出した事例を分析対象とした。

３．２　分析方法

日本語母語話者２名の判定者に、抽出されたテキストについて、関係節の接続傾向のほか、関係節付加曖昧名詞句「関係節+NP1 の NP2」に先行するテキストの中に、NP1、NP2、関係節の内容が記述されているかについて分野ごとに判定してもらった。また、先行するテキストにこれらの要素が記述されている場合は、これらの要素と関係節付加曖昧名詞句とのあいだの関係についても分類してもらった。判定者間の信頼度は各分野ごとに Cohen's Kappa が 0.8 以上であった（新聞：0.823、雑誌：0.871, 書籍：0.830, 白書：0.937）。表 4 以降の結束連鎖を形成する語の種類の分類には統計ソフトのエクセルを用いた。例えば、NP1 または NP2 の名詞と先行するテキストの繰り返されている語が同じかどうかコマンドを入力して検出し同一語を抽出して数を算出するようにした。

４　結果

４．１　NP1 及び NP2 に関する先行するテキスト内での言及

判定者に下記の３点について分析してもらった。

(1)「関係節＋NP1 の NP2」の名詞句に先行するテキスト内で NP1 についての言及があるか。
(2)「関係節＋NP1 の NP2」の名詞句に先行するテキスト内で NP2 についての言及があるか。
(3)関係節の先行詞は NP1(低位接続)または NP2(高位接続)のどちらか。

表 2：先行するテキスト内での NP1 と NP2 の言及と関係節の接続の比率（数）

NP1 と NP2 に関する言及	関係節の接続の選択		
	低位接続	高位接続	合計
どちらについても言及がない。	43(25)	57(33)	100(58)
NP1 についてのみ	44(27)	56(35)	100(62)
NP2 についてのみ	37(11)	63(19)	100(30)
NP1 と NP2 の両方	48(32)	52(35)	100(67)
合計	44(95)	56(122)	100(217)

日本語では限定用法または非限定用法であるかどうかは表記から判断することが難しく、NP1 と NP2 のどちらにも言及がなかった事例は非限定用法に該当する可能性がある。また中納言では先行文脈の語数が 500 字と限られている。この範囲外で言及があった可能性もある。「関係節＋NP1 の NP2」の名詞句に先行するテキスト内で NP1 にも NP2 にも言及がなかった事例では高位接続の方が低位接続よりもやや多かったが、上記のような点を考慮すると接続の傾向について断定することはできない。NP1 と NP2 のどちらか、または両方について言及がある事例では高位接続を選択する事例が多くなっているがどの場合もあまり大きな差はない。

４．２　関係節に関する先行するテキスト内での言及

判定者に「関係節＋NP1 の NP2」の名詞句に先行するテキスト内で関係節についての言及があるかどうかについて判定してもらい、その合計を算出した（表 3)。関係節についての言及がない場合の方が言及がある場合よりも多かった。

表 3：先行テキスト内での関係節に関する言及の比率（数）

関係節に関する言及なし		関係節に関する言及あり		合計
低位接続	高位接続	低位接続	高位接続	
27(59)	48 (104)	17(36)	8(18)	100(217)

４．３　繰り返しによる語彙結束を形成する NP1、NP2 及び先行するテキスト内名詞句

語彙的結束の繰り返しを同一語、類義語、上位語、一般的な語に分類した。「関係節＋NP1 の NP2」の名詞句内（例 5 の下線部：[黒潮が育てた$_{関係節}$ [漁船$^{a\text{-}8}{}_{NP1}$] の[民俗文化$_{NP2}$]]）で、NP1 または NP2 と同じ語が先行する文脈内にある場合は**同一語**（NP1=漁船$^{a\text{-}8}$と漁船$^{a\text{-}1}$)、NP1 または NP2 の類義語が先行する文脈内にある場合は**類義語**（NP1=漁船$^{a\text{-}8}$と船$^{a\text{-}2}$)、NP1 または NP2 の上位語が先行する文脈内にある場合は**上位語**（NP1=漁船$^{a\text{-}8}$と船舶（例 6 には含まれていない)）、NP1 または NP2 の一般的な語が先行する文脈内にある場合は**一般的な語**（NP1=漁船$^{a\text{-}8}$と海生丸$^{a\text{-}3}$、漁生丸$^{a\text{-}4}$、正丸$^{a\text{-}5}$、直美丸$^{a\text{-}6}$、美衣丸$^{a\text{-}7}$）とした。

(5) 岸壁につながれた漁船$^{a\text{-}1}$は、よく見ると、どれもこれも「眼のある船$^{a\text{-}2}$」だった。海生丸$^{a\text{-}3}$、漁生丸$^{a\text{-}4}$、正丸$^{a\text{-}5}$、直美丸$^{a\text{-}6}$、美衣丸$^{a\text{-}7}$…、みんな舳に可愛い眼が付いていた。　種子島にはこれまで何度も訪れていたが、気が付かなかった。[黒潮が育てた$_{関係節}$]漁船$^{a\text{-}8}$ (NP1) の民俗文化(NP2)が、

語彙的結束連鎖が形成されている事例について、その種類を分類したところ（表 4)、同じ語、類義語、上位語、一般的な語の 4 種類は、それぞれ 38.39%, 4.52%, 15.81%, 41.29%の比率となり、NP1 と NP2 の同じ語を繰り返す、または一般的な語に言い換える比率が高いことが分かった。更に種類毎に関係節の接続が高位接続か低位接続かについて分類した(表 5)。

表 4：繰り返しの語彙の種類の比率（数）

繰り返しの語	NP1	NP2	合計
同一語	32(38)	68(81)	100(119)
類義語	43(6)	57(8)	100(14)
上位語	45(22)	55(27)	100(49)
一般的な語	56(72)	44(56)	100(128)
合計	45(138)	55(172)	100(310)

表 5：繰り返しの語彙の種類と関係節の接続の比率（数）

繰り返しの語	繰り返されている語句	関係節の接続		合計
		低位接続	高位接続	
同一語	NP1	46(13)	54(15)	100(28)
	NP2	44(35)	56(45)	100(80)
類義語	NP1	39(7)	61(11)	100(18)

	NP2	100(2)	0(0)	100(2)
上位語	NP1	43(9)	57(12)	100(21)
	NP2	37(10)	63(17)	100(27)
一般的な語	NP1	43(20)	57(26)	100(46)
	NP2	45(37)	55(45)	100(82)
合計		37(113)	44(133)	56(171)

先行研究では関係節の制限用法は先行文脈に同じ種類のものが２つ以上あり、そのうちどれを指しているのか明示するために使われることが前提となっているが、先行文脈内の語と NP1 または NP2 の関係を分析したところ当てはまらない事例も多くあった。例えば、下記の例 6 では、先行文脈は過去から現在の日本の農業の様子を記述しており、農業界 $^{b-19}$ は上位語として先行文脈全体を指し、[[逆風の吹く $^{a-6}$ $_{関係節}$] [[日本 $^{b-15}$ $_{NP1}$]の[農業界 $^{c-19}$ $_{NP2}$] $_{NP3}$] $_{NP4}$]は現在の日本の農業の様子を総括している。農業界の一部を指すのではなく、全体を総括する表現として関係節付加曖昧名詞句が使われている。このような例から関係節の先行詞が上位語、先行文脈内の語が下位語の事例もあり、先行研究で想定されている以外の語彙的結束性の連鎖が形成されていることがわかった。そこで表 6 のように先行文脈の語句が一般的な語を、NP1 または NP2 の下位語となっている場合と同じレベルの語である場合とに分類した。

(6) 社説　二千一・２・二十一　【中日 $^{b-1}$ 農業 $^{c-1}$ 賞】危機 $^{a-1}$ 突破に若者の力　中日 $^{b-2}$ 農業 $^{c-2}$ 賞が第六十回を機に衣替えし、若い農家 $^{c-3}$ に絞って顕彰することになった。日本 $^{b-3}$ の農業 $^{c-4}$ 危機 $^{a-2}$ 突破の力となることを期待する。　三十数年前、ちゃぶ台にこぼれたわずかなご飯粒 $^{c-5}$ を「もったいない」と言いつつ口に運んだ時代、農業 $^{c-6}$ はまだ国 $^{b-4}$ の基幹的な産業 $^{c-7}$ であった。が「飽食の時代」と呼ばれる今、その $^{c-8}$ 存在は、とかく軽く見られがちである。　そんな時代に、中部地方 $^{b-5}$ の農業者 $^{c-9}$ を顕彰する中日 $^{b-6}$ 農業 $^{c-10}$ 賞は審査対象年齢を四十歳以下に絞り、二十一世紀を担う若い農家 $^{c-11}$ を励ますことになった。　背景に、日本 $^{b-7}$ の農業 $^{c-12}$ に対する危機 $^{a-3}$ 感がある。何よりも、国際 $^{b-8}$ 競争の激化 $^{a-4}$ が日本 $^{b-9}$ の農業 $^{c-13}$ を揺さぶっている。安い労賃や、広大で安価な土地で生み出される海外 $^{b-10}$ の農作物 $^{c-14}$ が輸入解禁となり、宿命的な悪条件下 $^{a-5}$ で作られる国産 $^{b-11}$ 農作物 $^{c-15}$ を駆逐しつつある。　とくに国際 $^{b-12}$ 分業論を信奉する人々は、生産性の低い日本 $^{b-13}$ の農業 $^{c-16}$ そのもの $^{a-17}$ を経済発展の足手まといととらえ $^{a-6}$「日本 $^{b-14}$ に農業 $^{c-18}$ はいらない」とまで述べている。まさに[[逆風の吹く $^{a-6}$ $_{関係節}$] [[日本 $^{b-15}$ $_{NP1}$]の[農業界 $^{c-19}$ $_{NP2}$] $_{NP3}$] $_{NP4}$]であり、。。。

NP1 または NP2 が先行する文脈内で繰り返されている語にとって、どのような関係にあたるかを分類し、その数を算出した（表 6）。先行文脈の語が例 6 の海生丸 $^{a-3}$ で NP1 がその総称で「船」や「漁船」なら**下位語**とした。

表 6：結束連鎖を形成する繰り返される名詞句の種類の数

繰り返しの語	語彙の種類	関係節の接続	低位接続	高位接続	合計	合計
一般的な語	同レベルの語	NP1	18	13	31	67
		NP2	7	29	36	

	下位語	NP1	19	22	41	61
		NP2	17	3	20	
上位語		NP1	9	12	21	48
		NP2	10	17	27	
合計			80	96	176	176

５．まとめ

本研究は現代日本語書き言葉均衡コーパスのコア・データから関係節付加曖昧名詞句を含むテキストを抽出し、関係節付加曖昧名詞句とそれに先行するテキストの部分とで形成されている結束連鎖を分析した。その結果、心理言語学の先行研究で想定していた結束は同一語の繰り返しで成立されるもののみだったが、多くの種類の結束連鎖があることが分かった。表 5 を見ると同一語では高位接続の方が低位接続より多くなっており、関係節付加曖昧構文を単独で提示している研究の結果と一致する。一方、表 5 や表 6 で NP1 や NP2 と結束連鎖を形成している他の種類の語を見ると、必ずしも高位接続が低位接続より多くなってはいない。したがって、文脈情報が関係節の接続の選択にどのように影響するか、心理言語学的研究を行った場合、従来よりも複雑な仕組みが明らかになる可能性がある。コーパスから得られるデータに基づいた研究と心理言語学的な実験から得られるデータに基づいた研究の成果を合わせていくとより発展的な研究ができる可能性がある。

謝　辞

本研究は、喜田桃世さん、近藤眞樹さん、西本優さんにご協力をいただきました。また、科学研究費補助金基盤（C）（代表者：中野陽子　No. 24520484）による補助を得ています。ここに記して感謝の意を表します。

文　献

Cuetos, F., and Mitchell, D.C. (1985). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30*, 73-105.

Desmet, T. Baecke, C. D., and Brysbert, M. (2002). The influence of referential discourse context on modifier attachment in Dutch *Memory & Cognition, 30*, 150-157.

Halliday, M. A. K., and Ruqaiya Hasan. 1976.*Cohesion in English.* London: Longman.

Kamide, Y., & Mitchell, D.C. (1997). Relative clause attachment: Non-determinism in Japanese parsing. *Journal of Psycholinguistic Research, 26*, 247-254.

Papadopoulou, D., and Clahsen, H. (2006). Ambiguity resolution in sentence processing: the role of lexical and contextual information. *Journal of Linguistics, 42*, 109−138.

Zagar, D., Pynte, J., and Rativeau, S. (1997). Evidence for Early closure Attachment on First pass Reading Times in French. T*he Quarterly Journal of Experimental Psychology Section A, 50*, 421-438.

中野陽子、早野賢譲、西内万貴、井本智子（2007）中国人留学生の第二言語としての日本語における関係節付加曖昧構文の処理について　国際社会文化研究第 8 号 109-126.