

文型分析ツール開発に向けた接続助詞的形式名詞の研究
—用例抽出ルールの設定と運用実験—

A STUDY ON CONJUNCTIVE PARTICLE-LIKE FORMAL NOUN TOWARD
DEVELOPMENT OF A FUNCTIONAL-EXPRESSION ANALYZING TOOL:
FORMULATING AND TESTING EXTRACTION RULES

渡辺 文生	武田 知子	川村 よし子
山形大学	サンフランシスコ州立大学	東京国際大学
Fumio Watanabe	Tomoko Takeda	Yoshiko Kawamura
Yamagata University	San Francisco State University	Tokyo International University

1. はじめに

本研究の目的は、日本語教師の読解教材分析や作文指導を支援する文型分析ツールを開発するために、接続助詞的に用いられる形式名詞のリストを作成し、用例を抽出するルールを設定すること、および、それらのルールの運用実験を行ってその検証結果を検討することである。文型分析ツールとして、表現文型抽出器というツールを開発中であるが、そのツールへ接続助詞的形式名詞の抽出機能を組み込む試みを報告する。

文型分析ツールの開発の発端は、「『リーディング・チュウ太』の難易度判定ツールは日本語学習者の作文評価に利用可能か？」という疑問であった。日本語読解学習支援システム『リーディング・チュウ太』には、語彙チェッカーと文章の難易度判定ツールという2種類の判定ツールがあり、読もうとする日本語の文章の語彙それぞれが日本語能力試験の何級に当たるかを分析し、文章全体としてどのくらいの難易度であるかを判定することができる(川村・北村2013)。ツールが分析するテキストとして日本語学習者が書いた作文を入力すると、その作文で使われている語彙の難易度が判定できて、さらにその結果をもとに作文の習熟度評価につなげることができないだろうかと考えた。そこで、渡辺ほか(2013)では、『リーディング・チュウ太』の難易度判定ツールを日本語学習者の作文評価に利用した場合、どこに着目し、何を評価指標とすればよいのか、などの点に関して研究を行った。その結果、作文評価に利用可能かどうかという点については、語彙の難易度だけでもある程度は評価に役立つが、もともと読解支援のために設計されたツールであるため問題点が多いということが分かった。

上記の結果を受けて、日本語学習者の作文習熟度の測定に役立つ機能表現や文型を自動で抽出できるツールの開発を目指すことになった。その端緒として、渡辺ほか(2014)では、接続詞・接続助詞、名詞修飾節、複合動詞に着目し、日本語母語話者と日本語学習者それぞれの作文を分析し、名詞修飾節を中心としたツールの運用実験を行った。その結果浮かんできた課題の一つが、形式名詞に関わる文型の処理である。形式名詞は、(1)のように文末表現の一部として用いられるが、それらの例と、(2)のような名詞修飾節の場合とを、いかに区別して

文型を正しく抽出できるようツールを設計するかということが問題になり、抽出条件を加えることで対応した。

- (1) 子どもの頃はよくケンカしたものだ。
- (2) 小麦粉を牛乳で溶いたものを加える。
- (3) 電話で問い合わせたものの疑問は解けなかった。

本研究では、(3)のように形式名詞が接続助詞的に用いられている場合を区別して抽出できるような機能を文型分析ツールに組み込むことが課題である。複雑な構造の文を読み解いたり、複数の節を組み合わせて長い文を書くためには、狭い意味での接続助詞や名詞修飾節などと並んで、「もの」や「こと」「ところ」などの形式名詞の習得が重要である。日本語教師が読解教材や学習者の作文などをツールにかけることによって、その文章で用いられている接続助詞的形式名詞を自動的に抽出・表示させ、教材分析や作文評価に役立ててもらおうことを目指している。

2. 接続助詞的形式名詞

接続助詞は、複文に含まれる節同士のつながり方を表し、文脈の展開を明示的に表す言語表現である。節をつなげて長い文を作るという点で、副詞節の使い方、言い換えれば接続助詞の使い方の習得は重要だと言える。たとえば、南

(2010)は、日本語学習者による日本語の物語作文を母語話者に評定させ、テ形接続を多く含んだ作文は評定が低いのに対し、論理関係を明確にする接続表現を多く含む作文は評定が高いと述べている。

渡辺(2014)では、物語作文において日本語母語話者と日本語学習者とで、接続助詞使用に関して異なる傾向があることを指摘した。ただ、その分析ではいわゆる接続助詞のみを対象にしており、総合的な分析を目指すうえでは、接続助詞的形式名詞による用例も取り上げる必要がある。

- (4) 何の手がかりもないなかで、いつまでも町中を探しまわる気力は私にはなかった。
- (5) 民宿に泊まってみたいと答えたところ、おばちゃんたちは民宿がどこにあるか知らなかった。

接続助詞的形式名詞とは、形式名詞に先行する節と被修飾名詞である形式名詞を含んだ名詞句が、文の中で副詞的に機能しているものと定義する。具体例としては、先に挙げた(3)のほか、(4)(5)が挙げられる。特に(4)の場合、形式名詞「なか」の後に格助詞の「で」が現れているが、「何の手がかりもないなか」という名詞句が《場所》という意味を持たず、デ格が述語「探しまわる」あるいは「なかった」の格要素として結びつくものとは言えないため、接続助詞的と判断する。

3. 接続助詞的形式名詞のリスト作成手続き

どんな形式が接続助詞的形式名詞として機能しているかを調べるため、白書およびニュースのデータをもとに、表現文型抽出器を使って形式名詞が被修飾名詞となっている名詞修飾節の用例を抽出した。これは、形式名詞を含む例文 (1) (2) (3) のタイプすべてを含むさまざまな用例リストと言うべきもので、これらの用例から接続助詞的に機能している形式名詞として 34 種類を選び出した。

白書 (延べ語数 425 異なり語数 81)
 ニュース (延べ語数 1,174 異なり語数 107)

接続助詞的に機能する形式名詞

あと・ため・もの・こと・よう・とおりに・ほか・うえ・場合・なか・さい (際) ・まえ・とき・うち・かぎり・すえ (末) ・ところ・まま・かわり・くせ・わり・たび・以上・途中・おり (折) ・のち・現在・ま (間) ・最中・はず・せい・おかげ・結果・あいだ

そして、34 種類の接続助詞的形式名詞それぞれに抽出ルールとして、前接条件および後接条件を設定した。たとえば、ほとんどの形式名詞の前接条件は、「。「」または「、」または「の」でないこと」と設定したが、これは、明らかに形式名詞の前が節ではない場合を排除するものである。後接条件は、形式名詞の種類によって様々であるが、たとえば「もの」の場合、「もの」だけでは接続助詞的には用いられないため、「の、」または「として、」という形式が後接すること」という条件を設定した。その結果、節のあとに「ものの、」あるいは「ものとして、」という形式が続く場合は形式名詞が接続助詞的に働いている例として認定することになる。

4. コラムデータによる表現文型抽出器の運用実験

4.1 表現文型抽出器の概要

文型分析ツールとして、現在開発している表現文型抽出器について簡単に触れておく。図 1 は、表現文型抽出器の画面である。

「入力ウィンドウ」に分析したい文章を直接入力またはコピー&ペーストして、分析ボタンをクリックすると、抽出結果が表示される。ほかにも、抽出結果の表現をクリックすると、入力テキスト中のそれらの表現をハイライトするなど、さまざまな工夫が施されているが、機会を改めて紹介することにする。

図 1 表現文型抽出器 (http://yasashii.cloudapp.net/phrase_analyzer/) の画面

抽出結果

表現文型抽出器

もし私はたくさんお金を持ったら、一番したいことはいろんな国々に旅行したいです。いい経験がとられるし、見たことがな所に行けます。またはピザ何枚でも注文することができます。私はピザが大好きなので、毎日ピザを食べることはすごくうれしいことです。一人で食べても、または友達と一緒に食べても最高です。たくさんお金を持つことはとてもいいことです。

もう一つは自分の体の全体に筋肉がついてほしいです。それはやはり、難しいことです。実際に毎日練習をしないと、筋がつくことができません。それならば、毎日練習しなければなりません。でも、私はそんなことはめんどくさいですから、一回だけ練習してすぐにきれいな体と筋肉がついてほしいです。他のほしいことはもっと元気で明るくなることです。

もし私はたくさんお金を持ったら、一番したいことはいろんな国々に旅行したいです。いい経験がとられるし、見たことがな所に行けます。またはピザ何枚でも注文することができます。私はピザが大好きなので、毎日ピザを食べることはすごくうれしいことです。一人で食べても、または友達と一緒に食べても最高です。たくさんお金を持つことはとてもいいことです。

もう一つは自分の体の全体に筋肉がついてほしいです。それはやはり、難しいことです。実際に毎日練習をしないと、筋がつくことができません。それならば、毎日練習しなければなりません。でも、私はそんなことはめんどくさいですから、一回だけ練習してすぐにきれいな体と筋肉がついてほしいです。他のほしいことはもっと元気で明るくなることです。その通りにできたら、完璧です。

分析ボタン

Analyze **Clear**

文字数	単語数	異なり語数	文数	改行数
346	183	86	13	4

	述べ語数	異なり語数	一文当たり平均語数
接続詞	3	2	0.1538
または	2		0.1538
でも	1		0.0769
接続助詞	1	1	0.0769
ので	1		0.0769
機能表現	6	4	0.3077
て欲しい	2		0.1538
ことが出来る	2		0.1538
所に	1		0.0769
なければ	1		0.0769
形式名詞	11	2	0.1538
こと	10		0.7692
とおり	1		0.0769

[表現リストCSV出力](#)

入力ウィンドウ

4.2 コラムデータによる運用実験の手続き

前節で説明した接続助詞的形式名詞のリストとその抽出ルールでうまく表現文型抽出器が稼働するかを確かめるために、運用実験を行った。データには、ウェブから収集可能なコラムの文章、18種類を用いた。

コラムデータ：ウェブ上のコラムの文章

(出典は 47 NEWS、朝日新聞、Yomiuri Online、WEB ちくま)

18種類 886文 34,625文字

運用実験の手続きは次のとおりである。まず、このデータから34種類の形式名詞が含まれるすべての文を、接続助詞的であるかどうかに関わらず抽出し、その後、それらひとつひとつについて、接続助詞的な用例であるかどうか、本研究の執筆者3名がそれぞれ判定を行った。そして、接続助詞的形式名詞の用例として選定された文を、表現文型抽出器で分析し、どの程度正確に接続助詞的形式名詞として認定されるか検証した。

4.3 接続助詞的であるかどうかの判断基準

形式名詞が接続助詞的であるかどうかの判断基準としては、形式名詞の修飾要素が節としての独立性を持つかどうかという点がまず挙げられる。国立国語研究所(1981)は、名詞修飾節の認定要因として、「誰(何)かが」という主体が具体的にその文(脈)から分かるということと、テンスの意識があるということの2点を挙げている(国立国語研究所1981:89)。そのほかにも、述語につながる格要素や連用修飾要素があるかどうかも要因として挙げられる。たとえば、(6)の「10秒もしないうちに」であれば、「10秒も」という要素の存在をもとに節として認定できるが、(7)の「若いうちは」は、特定の主体が想定されない点や述語以外の要素がない点などから、節とは認定できない。

- (6) ハンドルをひねってみると、10秒もしないうちにお湯が出た。
 (7) 人は、若いうちは、多少なりとも他人と違う旅をしなければならない。

もう一つの基準は、形式名詞が後続の述語と格関係を結ぶかどうかという点である。(8)の、「このテーゼにはとくに瑕疵はないように」は述語「思われる」の格要素となっているため接続助詞的とは言えない。一方、(9)の「いつか長い旅を振り返るように、」は述語「残す」の格要素とは判断できないので接続助詞的と判断する。(10)の「年輪を重ねたことで、」は、格助詞「で」が《手段》を表す格として判断する余地もあるが、因果関係を表す副詞節とみなし接続助詞的と判断した。

- (8) このテーゼにはとくに瑕疵はないように思われる。
 (9) いつか長い旅を振り返るように、言葉で残せたら良いと思う。
 (10) 年輪を重ねたことで、呪いの説得力がパワーアップしていた。

4.4 運用実験の結果

コラムデータによる運用実験では、執筆者3名のうち2名以上が接続助詞的と判定した69の用例(Aデータ)と、判定者が1名だけであっても4.3節の基準による総合的な判断によって実験に加えた21の用例(Bデータ)の合計80例が、表現文型抽出器による検証の対象とした。たとえば、(9)はAデータの用例であるが、(10)はBデータの用例で見方によっては単なる名詞修飾節とする可能性もある非典型的な用例と言える。

表現文型抽出器による検証の対象

2名以上が接続助詞的と判定した用例	69 (Aデータ)
判定者が1名であるが総合的判断で加えた用例	21 (Bデータ)

これらの用例を表現文型抽出器の検証した当初の結果は、形式名詞に読点が後接する用例の場合、約9割の抽出率で接続助詞的形式名詞と認定できたのに対し、形式名詞の直後に読点が後接しない用例の場合は、抽出率が約2割と課題を残すものとなった。(11)のように形式名詞「ところ」のすぐ後に「、」がある場合

は、接続助詞的形式名詞として認定されたのに対し、(12)のように格助詞の「で」が後接する場合は認定できないという結果になった。

- (11) 《前略》それに泊まってみたくいと答えたところ、おばちゃんたちは民宿がどこにあるか知らなかった。
- (12) だが、考えまい考えまいと、考えてみたところで、それはもうすでに敵の術中にはまっているので、《後略》

その後、上記の問題点を改善するために、さらに表現文型抽出器の動作を検証・修正した。現段階においては、検索の見出し語そのものを形式名詞のあとの読点や助詞も含めた形にした結果、抽出率が非常に改善した。今後さらに抽出条件の吟味や条件のかけ方の検討を経て更なる改善を図りたい。

形式名詞に読点の後接する用例の抽出率

Aデータの抽出率 100% (21例中の21例)

Aデータ+Bデータの抽出率 96.2% (26例中の25例)

形式名詞の直後に読点の後接しない用例の抽出率

Aデータの抽出率 94.7% (38例中の36例)

Aデータ+Bデータの抽出率 87.0% (54例中の47例)

いまだに接続助詞的形式名詞として抽出できない用例の多くは、Bデータ、つまり非典型的な用例であった。(13)の「～することなく、」は、形式名詞「こと」のあとに助詞ではなく否定の形が来るという点で特殊なものであり、いまのところうまく抽出できていない。(14)は、「ために」だったら抽出できるのだが、「ため」と「に」のあいだに取り立て助詞の「だけ」が入っているために抽出できなかった例である。取り立て助詞が介在する場合にどう対応するか、検討する必要がある。

- (13) そうして病名も告げることなく、ただりボトリールという薬をくれた。
- (14) あのときの、ただただ非日常を味わうためだけに出かけ、そして非日常を味わえただけで満足していた旅。

上記の検証方法は、接続助詞的形式名詞として抽出すべきデータのみをツールに入力して、それらが正しく抽出されるか検証しているものだが、接続助詞的形式名詞として抽出すべきでない用例を適切に排除しているかを検証する必要もある。そこで、コラムデータの中から、接続助詞的ではないと判断した形式名詞の用例を表現文型抽出器で分析して検証することにした。

抽出すべきでない用例を適切に排除しているかの検証対象

形式名詞が含まれる用例 432例 (接続助詞な用例は除く)

その検証の結果、432 例中 17 例が接続助詞的形式名詞として抽出された。その内訳は以下のとおりである。

排除できなかつた用例数	(432 例中の 17 例	3.9%)	
「こと」	3 例	「前」	2 例
「よう」	1 例	「とき」	5 例
「なか」	1 例	「うち」	1 例
「ところ」	1 例	「まま」	2 例
「結果」	1 例		

排除できなかつた具体例として、(15) (16) が挙げられる。(15) は、「中に、」のニ格が述語「見つけた」と結びついているため接続助詞的とは判定できない例であるが、述語の結合価まで使って処理することは難しいので、対応に困る例である。(16) については、形式名詞「まま」の後接条件を改善することによって対応可能だろうと思われる。

(15) オフィスビルがひしめく中に、《中略》その店を見つけた。

(16) 《前略》水は冷たいままなのだった。

5. おわりに

以上、文型分析ツールを開発するために、接続助詞的に用いられる形式名詞の分析を中心に、ツールの運用実験とその検証結果について報告した。表現文型抽出器は、文型分析ツールとして、文章にどのような表現が使われているのかを示す支援ツールである。今後とも抽出率の改善を図っていくが、人が判断に迷う例はコンピュータでも判断できないので、ある程度割り切った結果を表示させることしかできない。(13) から (16) のような用例については、利用者の目で確認しながら活用していただくということが必要である。どこまでを接続助詞的形式名詞と判断し、その抽出条件は何かと考察する作業は、単にツールの開発という目的のみならず、日本語の文法記述という点でも意義があると言えよう。

参考文献

- 川村よし子・北村達也 (2013) 「日本語学習者のための文章の難易度判定システムの構築と運用実験」『Journal CAJLE』14, 18-30
- 国立国語研究所 (寺村秀夫執筆) (1981) 『日本語の文法 (下)』大蔵省印刷局
- 南雅彦 (2010) 「接続表現」南雅彦 (編) 『言語学と日本語教育 VI』65-66 ころしお出版
- 渡辺文生 (2014, March) 「日本語学習者の物語作文における接続詞・接続助詞使用をもとに習熟度を探る」Paper presented at AATJ 2014 Spring Annual Conference
- 渡辺文生・川村よし子・下條光明・藤原美保 (2013, March) 「『リーディング・チュウ太』のレベル判定ツールを用いた日本語学習者の作文の熟達度評価」Panel presented at AATJ2013 Spring Annual Conference

渡辺文生・下條光明・藤原美保・川村よし子・武田知子（2014, March）「日本語習熟度を測る指標の分析および表現文型抽出ツールの開発」 Panel presented at AATJ2014 Spring Annual Conference