

放送ニュースの動詞連用形名詞の平易化

美野 秀弥 田中 英輝

NHK 放送技術研究所 〒157-8510 東京都世田谷区砧 1-10-11

E-mail: {mino.h-gq, tanaka-h.ja}@nhk.or.jp

1. はじめに

著者らはNHKのニュースを外国語話者にとって「やさしい日本語」に書き換えることを検討しており[1,2]、その一環として、日本語能力試験(以下、JLPT)の出題基準¹に基づいた語彙の平易化に取り組んでいる。JLPTは、1~4級の級別に語彙リストを公表している他、数詞や派生語などの扱いについても説明している。著者らは、派生語の中で、動詞連用形名詞(以下、連用形名詞)が難易度の高い1,2級の表現であることに着目した。連用形名詞とは動詞の連用形が名詞へと派生した形態素である(「見通し」「悩み」など)。連用形名詞はニュースでもよく使用されており、平易化による効果は大きい。また、動詞からの派生語という性質上、連用形名詞は用言(動詞,サ変名詞)へと平易化できることが多く、その場合、係っている連体修飾句を連用修飾句へと変換する必要がある。この変換は、サ変名詞や他の用言(形容詞、形容動詞など)からの派生語の言い換えにも応用することができ、汎用性は高い。

そこで、本稿では、放送ニュース中に出現する連用形名詞を、用言へと平易化することに取り組んだ。特に、ニュース中の出現頻度が高かったことから、連体修飾句を1つ持つ連用形名詞を平易化の対象とした。

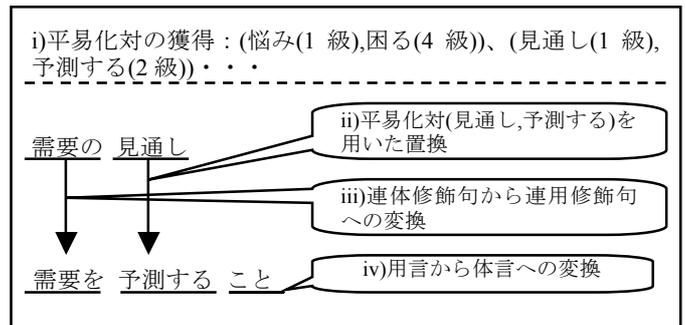
2. 提案手法

2.1. 概要

本稿では、JLPTの級を下げる操作を「平易化」、単語と平易化した単語との関係を「平易化の関係」、単語と平易化した単語との対を「平易化対」と呼ぶ。

そして、連用形名詞を用言へと平易化し、体言相当句に変換するまでの流れを図1に示す。本稿では、i)~iii)の処理について扱っており、特に、i)平易化対の獲得、とiii)連体修飾句から連用修飾句への変換、が本提案手法の重要な処理である。連用形名詞から用言へと平易化するためには、i)で獲得した平易化対を用いて置き換えるだけでは実現できず、iii)の処理で連用形名詞に係っている連体修飾句を連用修飾句へと変換する必要がある。

¹ JLPTは2010年に新しくなったが、出題基準(級別に、語彙,漢字,文法項目のリストが掲載されたもの)が公表されていないため、本稿では旧基準を用いている



- i) 平易化対の獲得(2.2で説明)
(連用形名詞,平易化した用言)の平易化対を獲得する。
- ii) 平易化対を用いた置換
獲得した平易化対(例:(見通し,予測する))を用いて、ニュース文中の連用形名詞を用言に置換する。
- iii)連体修飾句から連用修飾句への変換(2.3で説明)
連用形名詞に係っている連体修飾句(例:「需要の」)を、連用修飾句(例:「需要を」)へと変換する。
- iv)用言から体言への変換
用言へと平易化した句を体言相当句にするために「こと」「様子」「具合」などを付与して、用言から体言へと変換する(本稿ではこの処理について扱わない)。

図1: 連用形名詞から用言への平易化の流れ

2.2. 平易化対の獲得

本節では、以下の3つの手法で平易化対を獲得する。国語辞典を用いた手法とシソーラスを用いた手法では、獲得できる平易化対が少ないため、連用形名詞の特徴を用いた手法で獲得する平易化対を増やしている。

2.2.1. 国語辞典を用いた手法

一般に、国語辞典の見出し語を説明したものが語釈文であり、その語釈文は平易な表現になっている場合が多いと考えられる。そこで、論文[2]で報告した国語辞典[3]を用いた手法を利用して、図2の手順で平易化対を獲得する。

- i) 国語辞典からの抽出
論文[2]で報告した手法を用い、国語辞典の連用形名詞である見出し語とその語釈文中に含まれている言い換え可能性のある用言とを、対として抽出する。
例)見出し語:救い 語釈文:救うこと。助けること。
- ii) 平易化対(連用形名詞,平易化した用言)の獲得
i)で抽出した対の中から平易化の関係にある対を平易化対として獲得する。
例)(救い(1級),救う(2級))(救い(1級),助ける(2級))

図2: 国語辞典を用いた手法

2.2.2. シソーラスを用いた手法

単語間の上位下位関係が記述してある日本語ワードネット[4]などのシソーラスは、言い換えをするためのリソースとして使われる。例えば、「怒り」と「激怒」は上位下位関係にある（「怒り」が上位語、「激怒」が下位語である）。下位語は上位語を意味的に包含しているので、下位語「激怒」から上位語「怒り」への言い換えの可能性がある。そこで、上位下位関係を用いて、図3の手順で平易化対を獲得する。

- i) 日本語ワードネットからの抽出
(下位語,上位語)の対を獲得する。
例) (備え,準備)(備え,仕込み)(備え,用意)(備え,警備)
- ii) 類似度を用いた手法
論文[2]で報告した、国語辞典の語釈文間の類似度を見出し語間の類似度とする手法を用いて、i)で獲得した対の中から類似度の高い対のみを抽出する。
例) (備え,準備)(備え,仕込み)(備え,用意) (~~備え,警備~~)
- iii) 平易化対(連用形名詞,平易化した用言)の獲得
ii)で抽出した対の中から、平易化の関係にある対を平易化対として獲得する。
例) (備え(級外),準備(3級))(備え(級外),用意(3級))

図3：シソーラスを用いた手法

2.2.3. 連用形名詞の特徴を用いた手法

連用形名詞と派生元である動詞は、ほぼ同義であると考えられる。例えば、連用形名詞「見込み」の派生元である動詞「見込む」は、品詞が異なるが、ほぼ同義である。同義であると仮定すれば、動詞「見込む」を平易化した用言「思う」「考える」などは、「見込み」を平易化した用言であるといえる。この特徴を用いて、図4のように、連用形名詞と、派生元の動詞を平易化した用言の対とを平易化対として獲得する。

- i) (連用形名詞,派生元の動詞)の対の獲得
連用形名詞と派生元の動詞とを対として獲得する。
例) (見込み,見込む)
- ii) (動詞,平易な動詞)の平易化対の獲得
2.2.1 と 2.2.2 の手法を用いて(動詞,平易な動詞)の対を新たに獲得する。
例) (見込む,思う)(見込む,考える)
- iii) 平易化対(連用形名詞,平易化した用言)の獲得
i)と ii)を組み合わせ、(連用形名詞,平易な動詞)を平易化対として獲得する。
例) i)(見込み,見込む) + ii)(見込む,思う/考える)
→(見込み(1級),思う(3級))(見込み(1),考える(3級))

図4：連用形名詞の特徴を用いた手法

2.3. 連体修飾句から連用修飾句への変換

連用形名詞を用言へと平易化することで、連用形名詞に修飾していた連体修飾句を連用修飾句へと変換する必要がある。表1に変換した例を示す。連体修飾句はどちらも「名詞(需要,異議)+の」の形であるのに対して、変換後の連用修飾句は「名詞(需要)+が」「名詞(異議)+を」と異なる形になっている。

連体修飾句	連用形名詞	連用修飾句	用言
需要 <u>の</u>	落ち込み	需要 <u>が</u>	減る
異議 <u>の</u>	申し立て	異議 <u>を</u>	主張する

表1：連体修飾句(左)から連用修飾句(右)への変換例

これは、連体修飾句の助詞「の」が連用修飾句の助詞「が」「を」のどちらにも対応できるからであり、変換するためには対応先を同定しなければならない。表1の上部の例では、連体修飾句「需要+の」に対応している連用修飾句の候補「需要+が」や「需要+を」などの中から、用言「減る」に対して適切な連用修飾句「需要+が」を決定して変換する必要がある。

そこで、本節では、図5のような流れで連体修飾句を連用修飾句へ変換することを提案する。

まず、連体修飾句変換規則を用いて、連体修飾句と対応している連用修飾句の候補を出力する。候補が1つの場合は、それを連用修飾句とする。候補が複数ある場合は、格フレームとの照合を行い、候補の中から連用修飾句を決定する。

以下、2.3.1では連体修飾句変換規則の適用について、2.3.2では格フレームとの照合について説明する。

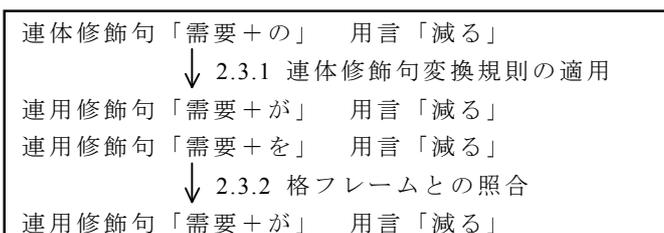


図5：連体修飾句から連用修飾句への変換の流れ

2.3.1. 手順1：連体修飾句変換規則の適用

連体修飾句と対応している連用修飾句の候補を決めるための連体修飾句変換規則を、山本ら[5]を参考にし、表2のように作成した。山本らは、本節の処理とは逆に、連用修飾句を連体修飾句に変換することに取り組んでいる。表2の型名は、片岡ら[6]の分類を参考に作成した。

形容詞型は、形容詞の活用形を連体形にすることで連体修飾句となる場合であり、活用形を連体形から連用形にすることで連用修飾句に変換する。

名詞型は、「名詞+助詞/助詞の連続(以下、助詞相当句)」の形で連体修飾句となる場合であり、助詞相当句の置換を行うことで連用修飾句に変換する。表2の例では、助詞相当句「との」を助詞「と」に置換することで連用修飾句に変換している。置換先の助詞相当句は連体修飾句が持つ助詞相当句ごとに異なる。

そこで、連体修飾句が持つ助詞相当句ごとに、置換先の連用修飾句が持つ助詞相当句の候補を対応付けたものを助詞相当句置換規則と呼び、以下のように作成した。

型名	名詞型 (名詞+助詞相当句)	形容詞型 (形容詞連体形)
変換方法	助詞相当句置換規則を適用	形容詞を連用形に変換
変換例	「社会 <u>との</u> 」+「かかわり」→「社会 <u>と</u> 」+「かわる」	「 <u>激しい</u> 」+「当たり」→「 <u>激しく</u> 」+「当たる」

表 2：連体修飾句の変換規則

- i) 「の」、「からの/への/での/との」、「による」
表 3 に示すように、置換先の助詞相当句の候補が複数存在する。助詞相当句の決定には、2.3.2 の格フレームとの照合が必要になる。
- ii) 「としての/についての/にかけての/にあたっての/に対しての/を通じての/」
後ろの「の」を除く。
例) 保障についての考え→保障について考える
- iii) 「に対する」、「に関する」など助詞相当句の中に用言相当語が含まれている場合
助詞相当句中の用言相当語の活用形を連用形にする。
例) 健康に関する悩み→健康に関して悩む
- iv) 副助詞「など」「だけ」などを伴う場合
副助詞を除いて、助詞相当句置換規則を適用する(副助詞は必ず後ろに i)から iii)の助詞相当句を伴う)。
例) 議会など:に対する働きかけ→議会など:に対して働きかける

置換前	の	からの	への	での	との	による
置換先	が/を/ で/に	から/ が/に	へ/に/ を	で/に	と/に	が/で

表 3：助詞相当句置換規則(置換先の候補が複数ある場合)

2.3.2. 手順 2：格フレームとの照合

手順 1 で複数の候補が出力された場合、格フレームと照合を行い、変換先を決定する。格フレームとは、用言とそれが持ちうる格要素(名詞+助詞)を記述したものである。本稿では、格要素ごとに出現頻度が併記されている京都大学の格フレーム[7]を利用した。特に、用言と格要素の二項関係として(名詞,助詞,用言,頻度)の対を抽出し、それを用いた。ただし、格フレームの中で、格要素が用言の直前にある場合をメイン格フレーム対、直前にない場合をサブ格フレーム対と呼び、別々に抽出した。手順 1 で出力された複数の候補と格フレームとの照合は、次の手順で行う。

- i) 連用修飾句の変換候補の抽出
手順 1 で出力された連用修飾句の変換候補を(名詞,助詞相当句,用言)の形で抽出
例) 連体修飾句「復興+への」 用言「努める」
→ 連用修飾句の変換候補：(復興,へ,努める)
(復興,に,努める)(復興,を,努める)

- ii) メイン格フレーム対との照合
i)の候補とメイン格フレーム対との照合を行い、照合できた対の中から頻度が最も高い対を用いて連用修飾句を決定する。
例) メイン格フレーム対(復興,に,努める,326)
→ 連用修飾句「復興+に」 用言「努める」
- iii) サブ格フレーム対との照合
メイン格フレーム対と照合できなかった場合、サブ格フレーム対を代わりに用いて同様の照合を行い、連用修飾句を決定する。照合できなかった場合は、変換不可とする。

3. 評価実験

3.1. 実験概要

提案手法を 2000 年から 2009 年までの NHK ニュース記事中の文節に適用して、評価実験を行った。平易化の対象は、連体修飾句を 1 つ持つ連用形名詞の難語(1,2 級,級外)とした。そして、提案手法を用いて得られた平易化結果を、連体修飾句の分類ごとに以下の 2 つに分けて別々に評価を行った。

- 評価 1) 手順 1 のみで変換が行われる場合
手順 1 のみで連用修飾句への変換が可能な連体修飾句(形容詞型、格フレームとの照合を必要としない助詞相当句を持つ名詞型)を持つ連用形名詞を、20 例ずつ無作為に抽出し、著者が評価を行った。
- 評価 2) 手順 1 に加えて、手順 2 が必要な場合
手順 2 を用いなければ連用修飾句への変換ができない連体修飾句(格フレームとの照合を必要とする助詞相当句 (表 3)を持つ名詞型)を持つ連用形名詞を、20 例ずつ無作為に抽出し、著者を含めた 3 名の被験者が評価を行った(ただし、助詞相当句「による」を持つ連体修飾句は対象が 20 例に満たなかったため評価対象から除外した)。

また、「通り」のように、名詞として自立度が高く、動詞「通る」に戻すと意味的な関連が希薄になる連用形名詞は平易化の対象から除いた。これは、著者が国語辞典の語釈文を参考にして行った。

評価基準は、表 4 に示すように、平易化対で置換した用言が意味的に言い換え可能かどうかの「用言の評価」と、連体修飾句から連用修飾句への変換が適切かどうかの「修飾句の評価」を設定し、それぞれ 4 段階で評価した。ただし、修飾句の評価は、用言の評価が意味の類似性が認められる 0,1,2 となった対象に対してのみ行った。

評価値	評価基準(用言の評価/修飾句の評価)
0	言い換え可能かつ、違和感を全く感じない
1	言い換え可能だが、若干違和感を感じる
2	文として不適切だが、意味の類似性がある
3	言い換え不可

表 4：評価基準

連体修飾句	用言の評価	修飾句の評価
-------	-------	--------

		0	1	2	3	0	1	2	3
形容詞型	A	10	2	2	25	14	0	0	0
名詞型	A	15	1	0	25	10	4	2	0
の	A	29	5	2	4	21	1	1	13
	B	24	5	6	5	18	0	1	16
	C	28	5	2	5	20	1	2	13
からの	A	24	3	3	4	8	2	4	16
	B	25	2	0	7	6	0	0	21
	C	22	5	0	7	9	1	1	16
への	A	30	4	1	5	29	0	0	6
	B	21	5	2	12	20	1	1	6
	C	26	1	6	7	26	0	0	4
での	A	32	1	0	3	23	1	0	9
	B	24	5	3	3	15	3	3	11
	C	29	4	0	3	17	8	0	8
との	A	24	0	3	12	20	1	0	6
	B	24	2	1	12	20	4	1	2
	C	25	1	1	14	20	3	1	3

表 5：評価結果(A:著者、B,C:著者以外の被験者)

3.2. 実験結果

評価結果を表 5 に示す。表 5 の上部(形容詞型,名詞型)が評価 1 の結果、表 5 の下部(の,からの,への,での,との)が評価 2 の結果である。以下、評価 0,1,2 を正解として議論する。

- ・ 用言の評価について

用言の評価の精度は評価 1,2 全体で 76.4%となった。ただし、平易化対の中には、(訴え,訴える)といった、ほぼ同義語である(連用形名詞,派生元の動詞)の形の平易化対も含んでいる。この形の平易化対を除いて平易化を行った場合の精度は 56.0%となった。

- ・ 評価 1 の修飾句の評価について

形容詞型、格フレームとの照合を必要としない名詞型ともに、修飾句の評価の精度は 100%であった。

- ・ 評価 2 の修飾句の評価について

格フレームとの照合を行う必要のある名詞型の場合の精度は全体で 69.6%であった。表 6 には、連体修飾句が持つ助詞相当句別の精度を示す。ただし、Baseline は表 3 の置換先の候補を無作為に選んだ場合の精度である。

助詞相当句	の	からの	への	での	との
精度(%)	60.7	36.9	82.8	71.4	86.4
Baseline	25.0	33.3	33.3	50.0	50.0

表 6：助詞相当句別の精度

3.3. 考察

評価実験で誤りと評価されたデータを分析し、誤りの原因を考察した。

- ・ 用言の評価で誤りとなった原因は、連用形名詞と平易化した用言との間の情報の欠落であった。誤り例では、連用形名詞「訴え」を用言「言う」に平易化しており、語釈文中の「不平や痛みを」の情報が欠落していることが誤りの原因である。この誤りを改善するためには、情報の欠落部分と平易化可能性と

の関係性を明らかにする必要があるが、データの中には被験者によって評価が分かれた場合も少なくなく、誤りと感じる境界線に個人差があることを考慮して検討を行う必要がある。

誤り例) 住民の訴え → 住民が言うこと

「訴え」の語釈文: 不平や痛みなどを人に言う。

- ・ 修飾句の評価は、正しい助詞相当句が付与されなかった場合に誤りとなる。誤り例では、連体修飾句「職員からの」が連用修飾句「職員に」に変換されているが、正解は「職員が」である。誤りの原因は、格フレームの用例不足と、「職員に申し出ること」が非文にならず格フレームに出現することにある(表 6 で、助詞相当句「からの」の精度が悪いのは主にこれが原因である)。誤りを改善するためには、格フレームの用例の増加の他に、2.3.1 の連体修飾句変換規則の変更が必要となる。例えば、連体修飾句「名詞+からの」の名詞が、時間を表している場合は連用修飾句「名詞+に」に、主格を表している場合は連用修飾句「名詞+が」に、それぞれ変換するという変換規則を加えれば、誤り例は改善できる。
誤り例) 職員からの申し出 →×職員に申し出ること (○職員が申し出ること)

4. おわりに

本稿では、NHK のニュース中に出現する連用形名詞を平易な用言へと言い換える手法を提案し、検討を行った。連用形名詞を用言へと平易化した場合、連用形名詞に係っている連体修飾句を連用修飾句へと変換する必要があり、連体修飾句変換規則の適用と格フレームとの照合を組み合わせたことが有効であることが分かった。しかし、変換後の接尾辞の付与や活用形の修正などの用言から体言への変換処理については本稿では取り扱わなかった。文に埋め込むことを考えた場合、体言への変換は必須であり、今後の課題となる。

文 献

- [1] 田中, 美野, やさしい日本語によるニュースの書き換え実験, 自然言語処理研究会, Vol.2010-NL-199 No.11, 2010
- [2] 美野, 田中, 国語辞典を使った放送ニュースの名詞の平易化, 言語処理学会第 16 回年次大会 pp.760-763, 2010
- [3] 三省堂, 例解小学国語辞典 第 3 版, 2005
- [4] NICT, 日本語ワードネット(V1.1), 2009-2010
- [5] 山本, 大橋, 「サ変動詞+名詞」の複合名詞への換言, 自然言語処理, Vol.12 No.3 pp.19-42, 2005
- [6] 片岡, 増山, 山本, 動詞型連体修飾表現の”N1 の N2”への言い換え, 自然言語処理, Vol.7 No.4 pp.79-98, 2000
- [7] 京都大学大学院情報学研究科 黒橋研究室, 京都大学格フレーム(Ver 1.0), 2009