

正規化を指向した機能動詞表現の述部言い換え

泉 朋子[†] 今村 賢治[†] 菊井 玄一郎[†] 藤田 篤[‡] 佐藤 理史[‡]

[†]日本電信電話株式会社 NTT サイバースペース研究所

{izumi.tomoko, imamura.kenji, kikui.genichiro}@lab.ntt.co.jp

[‡]名古屋大学大学院工学研究科

{fujita, ssato}@nuee.nagoya-u.ac.jp

1. はじめに

テキストマイニングや情報検索など、言語表現の意味を扱う際には、表層的に異なるが意味的には同等である表現を一つに集約する必要がある。特に、日本語においては、後述する機能動詞構文の存在により述部の表現が多様であること、また最近、述語項構造など述部を含むデータを用いる応用が増えていることから、述部の表層的多様性を縮退させることは極めて有用である。

そこで本稿では、「反発をくろう」といった「動作性名詞+助詞+機能動詞」の構造をもつ機能動詞構文に焦点を当て、機能動詞が表す機能語の意味役割を日本語の助動詞や時制が持つ意味に対応づけ、「反発される」といった最も単純な言語表現にまとめる正規化手法を提案する。

意味的に等価な言語表現を一つに「正規化」する場合、その結果は必ずしも自然言語である必要はない。しかし、意味を表現するための汎用的な人工言語を設計することは困難であり、また、正規化処理の妥当性を直観によって判定することもできない。そこで本稿では、自然言語に言い換えるという方針を取る。

なお、ある言語表現を意味的に等価な別の言語表現に変換する手法は「(自動) 言い換え」として多くの研究がある。これらと本研究との違いは、前者が言い換えの前後で意味がほぼ保存されているという厳しい制約を受けているのに対して、後者は言い換えの前後で事実関係が変らない限り極力単純な形に言い換えるという方針を取ることにある。

2. 機能動詞構文

2.1. 機能動詞構文(Light Verb Construction; LVC)とは

機能動詞構文(LVC)とは、「動作性名詞+助詞+機能動詞」の構造を持ち、動作性名詞が動詞の意味役割を果たし、機能動詞そのものはアスペクト/ヴォイス/ムードといった機能語の役割を果たす。そして日本語では、典型的な「する」に加え、「続ける」や「目指す」、「くろう」といった動詞も機能動詞として作用する(Matsumoto, 1996; 村

木, 1991)。

- (1) 開発 [サ変名詞] を 続け [継続] ている。
- (2) 締め出し[動詞連用] を くろう [受動]。

(1)の場合、機能動詞「続ける」はアスペクト表現の一つである「(動作の) 継続」を表わしている。(2)の「くろう」はヴォイス表現の「受動」を表している。これらのLVCはしばしば、「動詞+助動詞」という単純な形で言い換えることができる。

- (3) 開発を続けている → 開発している。
- (4) 締め出しをくろう → 締め出される。

このように、日本語の多様なLVCを単純な形へと言い換えることによりマイニング等の再現率向上が期待できる。

2.2. 対象とするLVC

本稿では、対象とする機能動詞として、村木(1991)に記載されている143組の「格助詞+機能動詞」と藤田 et al. (2006)で新聞記事19年分から得られた出現頻度上位40組の「格助詞+機能動詞」のペアを合わせ、両者の重複を除いたものを使用する。ただし、5種類はMatsumoto(1996)に記載されているLVCの統語的特徴に適合しなかったため、それらを除いた計160種類を使用することとした。これらをもとに、下記の構造を正規化対象とした。

- (5) 正規化対象とするLVC
動作性名詞 + 格助詞 + 機能動詞
e.g., 拡充をはかる, 注意を促す

3. 正規化を指向した制限言語の設計

本稿でいう正規化とは、複数の異なる表現を単一の表現に言い換えることである。つまり、言い換え先の表現は使用する語彙が制限された、一種の制限言語となる。本節では、前節で対象としたLVCの正規化用制限言語について述べる。

3.1. 制限言語の設計基本方針

制限言語の設計に先立ち、(5)の構造を毎日新聞12年分(1991年~2002年)とブログ824万文から抽出した。そして、抽出結果を参考に下記の方針で制限言語の設計を試みた。

- 制限言語は、最も単純な「動詞+助動詞」の表現にする。
- 正規化後の表現パターンを最小限にする。
- 正規化前後で、表現する事態の事実関係が変化しないようにする。

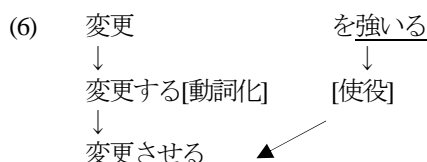
上記の条件を満たすために、正規化では次の2点が必要となる。第一に、動作性名詞を動詞化させる(e.g., 変更→変更する)。次に、機能動詞が表す機能語の意味を助動詞に言い換える。そのために、160種類の機能動詞に対して、それぞれの機能表現に対応した助動詞を割り当てる。対応する助動詞が見つからない場合は機能動詞間で同義のものをまとめ、上記の新聞とブログから抽出された中で出現頻度の一番高い機能動詞を制限言語とする。

3.2. 機能動詞の意味役割と言い換え表現

機能動詞が表す機能表現として、ヴォイス/ムード/アスペクトがある。各機能表現に対して、本手法では次のように表現パターンを設計した。

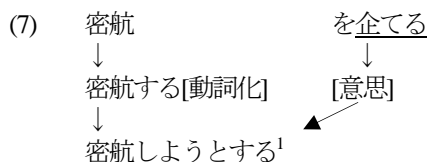
3.2.1. ヴォイス

使役/受動といったヴォイスを表す機能動詞には、「させる」と「される」という2種類の助動詞に対応付けた。



3.2.2. ムード

「意思」や「可能性」を表すムード表現も、同義の助動詞を対応させた。



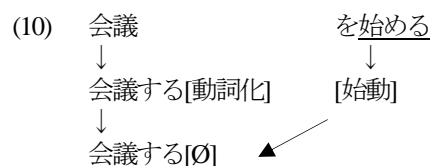
しかし、対象とする機能動詞の中で次の2点のムード表現においては対応する助動詞が見つからなかった。そのため3.1の結果から出現頻度の一番高い「なる」と「ある」と

いう機能動詞を言い換え先とした。

- (8) 原因・結果→「なる」
e.g., 混乱を招く, 反乱を引き起こす, 戦争になる
- (9) 動作主を消失させる→「ある」
e.g., 変化がある, 変化が見られる

3.2.3. アスペクト

「始める」「続ける」「繰り返す」といった機能動詞が表すアスペクト表現に関しては、本稿ではアスペクト情報を欠落させ単純時制にまとめる手法を取った。



本手法で、機能動詞のアスペクト表現を省略させた理由は下記の2点である。第一に、機能動詞で表現されるアスペクト表現は、機能動詞に後続する助動詞によってさらに同義のアスペクト情報が与えられることにより、アスペクト情報が冗長になっている場合が多々あるからである。

- (11) 黙秘 を 続け[継続] ている[継続].

この場合、「継続」のアスペクト情報は、機能動詞「続ける」に加え、助動詞の「ている」でも表現されている。また、日本語の単純時制そのものに、アスペクト情報が含まれており、動詞の現在形には「未来」の意味が、過去形には「完了」が含意される(Tsujimura, 2007, pp.364-369).²

- (12) (今から) 準備に取りかかる[始動].
= (今から) 準備する[未来].

(12)がほぼ同じ意味を表しているように、後続する助動詞や日本語の単純時制に随伴するアスペクト表現により、機能動詞が表すアスペクト情報は他の要素で補うことが可能である。

第二に、助動詞や時制形態素などで機能動詞が表すアスペクト情報を完全に補えない場合でも、cの制限言語設計条件を満たせるからである。アスペクト表現とは、事態が起った時間に、始動や継続といった話の参照となる詳細な時間表現を着色する役割を果たす(Portner, 2006, Ch. 8)。これは、言い換えれば、詳細なアスペクト情報を持つ述部表現は、それらが簡略化された表現を含意しているといえる。

¹ 本稿では「～(し)ようとする」を一語の助動詞として扱う。

² 日本語の「た」や「る」といった過去・現在を表す形態素を時制ではなくアスペクトを表す形態素とする議論もあり、日本語の単純時制の役割に関しては議論の余地がある(e.g., Ogiwara, 1999).

Table 1:LVC 正規化のための制限言語

制限言語	保持される機能表現	対象の LV	例
1. する	n.a.	106 種類	供給を実施する → 供給する
2. される	ヴォイス (受動)	14 種類	足止めをくう → 足止めされる
3. させる	ヴォイス (使役)	9 種類	混乱をもたらす → 混乱させる
4. させられる	ヴォイス (受動+使役)	1 種類	惨敗を喫した → 惨敗させられた
5. できる	ムード (可能)	4 種類	判断ができた → 判断できた
6. しようとする	ムード (意思)	5 種類	拡大を目指す → 拡大しようとする
7. させようとする	ヴォイス+ムード (使役+意思)	1 種類	強化を促す → 強化させようとする
8. をやめる	アスペクト (終結相)	2 種類	調査を打ち切る → 調査をやめる
9. となる	ムード (原因・結果)	5 種類	削減につながった → 削減となった
10. がある	ムード (非動作主性)	6 種類	不足が見られた → 不足があった

(13) 議論を繰り返している→議論している

藤田 et al. (2009)での同義言い換えでは、(13)は「繰り返して議論している」というように、副詞により詳細なアスペクト情報が加えられている。しかし、「繰り返す」は「議論する」という事態に「反復」というアスペクト情報を加えているだけで、これは「議論する」という事態を含意しているといえる。上記の理由から、本技術では機能動詞が表す詳細なアスペクト情報は省略する手法を取る。

ただし、アスペクト表現でも「やめる」といった終結を表す表現は対応する助動詞や時制形態素もなく、欠落させると c の条件から外れるため、出現頻度の一番高い「やめる」を制限言語とした。

(14) 終結相→「やめる」

e.g., 会議を打ち切る, 会話をやめる

このように、一部例外があるが、a～c の条件を満たすような制限言語として Table1 の 10 種類を設計した。³

4. 自動正規化

4.1. 対象機能動詞と正規化ルール

制限言語をもとに、LVC の自動正規化を試みた。本稿では機能動詞の意味機能にのみ焦点を当て自動正規化を試みているため、(15)のような動作性名詞の意味に依存して、機能動詞/本動詞と変化するものは対象外とした。⁴

(15) a. 信頼を置く(= 信頼する) 機能動詞

³ 対応する助動詞がない機能動詞のうち7種類は機能動詞間にも同義のものがなかったため今回の正規化の対象から外した。

⁴ サ変名詞の意味に焦点を置いた LVC の言い換えは、藤田 et al. (2009)を参照。

b. 電話を置く(≠ 電話する) 本動詞

そこで、3.1 の毎日新聞とブログから抽出された最大 10 種類の異なる動作性名詞と機能動詞を組み合わせ、常に機能動詞として作用し一意に言い換えができた 100 種類の機能動詞を正規化対象とした。

また、機能動詞が表すアスペクト表現と単純時制が表すアスペクト情報が対立する場合、含意性が失われる。

(16) a. A 氏の殺害に取り掛かった [始動+過去]

b. A 氏を殺害した [過去(完了)]

(16a)ではA氏はまだ生きているが、(16b)ではA氏は死んでしまった事を表している。この場合、(16a)の機能動詞が表すアスペクト情報を欠落させ、(16b)のように正規化すると事実関係が変わってしまうので、[始動を表す機能動詞]+[過去時制の形態素]の組み合わせが現れた場合は、正規化を適用しないこととした。⁵

4.2. 自動正規化と評価

毎日新聞2003年1年分(392,865文)をテストデータとし、文末から「サ変名詞+格助詞+機能動詞」のLVCを抽出し、3節で述べた制限言語への正規化を行った。⁶Table 1の1～7の制限言語に対応するLVCは、サ変名詞を動詞化させた後、動詞を制限言語の形に語形変化させた(e.g., 足止めをくっている→足止めされている)。一方、8～10ではサ変名詞は動詞化せずに名詞のまま機能動詞部分を制

⁵ アスペクトの対立として[完了]+[現在時制]の組み合わせもあげられるが、今回対象とする機能動詞に[完了]を表すものは含まれていなかったため、[始動]+[過去時制]の組み合わせのみを正規化の例外とした。

⁶ サ変名詞の判定は、形態素解析機 Chasen で「名詞-サ変接続」と解析され、さらに毎日新聞12年分に「サ変名詞+する」の動詞形が1回以上出現しているものを対象とした。

限言語に変換した(e.g., 拡大につながった→拡大となった). なお, 機能動詞に後続する助動詞に関しては, 正規化後もそのまま残した.

評価として, 正規化された 10,005 文からランダムに 300 文選び, 対象とする述部に焦点を当て, 正規化条件の a~c が守られているか, また統語的に誤りがないかを検討した.

評価対象の 300 文中 23 文は「始動」のAspectと過去形が組み合わさった例であり, 過去時制と始動のAspect対立から制限言語へは正規化されなかった. 下記は評価用に選出された 300 文中, 実際に言い換えられた 277 文の精度を表している.

Table 2: 述部自動正規化の精度

評価結果	割合 (頻度)	例
正解	94.2%(261)	結論に達した→結論した
複合名詞エラー	3.9%(11)	*国交正常化実現しようとする
ヴォイスエラー	0.7%(2)	*お祭りされる
IDIOM エラー	0.4%(1)	布石を打つ→*布石している
生成統語エラー	0.4%(1)	*加速させようとしている
意味エラー	0.4%(1)	偽造・変造に当たる≠偽造・変造する

評価の結果, 94.2%の精度で 10 種類の制限言語へと正しく正規化された. 正規化により意味の同義性が損なわれた例は 300 文中 1 例であった(Table 2, 意味エラー). また, 誤りの多くは, サ変名詞を含む複合名詞が動詞化されてしまったものである(Table 2, 複合名詞エラー). これらは, 正規化に伴いサ変名詞を含む複合名詞を分解する必要があることを示しており(e.g., 国交正常化を実現しようとする), 今後検討すべき課題であると考えられる.

また, 正規化した述部を元の文に戻し一文としての精度を検討した結果, 精度が 53.0%と低下した. この問題を解決するには, 格交替や副詞化などの統語変換や(38.2%), 動作性名詞の等位構造変換(2.2%)が必要である(格交替に関しては (藤田 et al., 2006) 参照).

(17) 動作性名詞の等位構造と格交替

条約批准と国内法の整備に取り組む。

*条約批准と国内法の整備する。

条約批准し, 国内法を整備する。

5. 考察

本稿では, 「議論を始める」⇔「議論し始める」や「議論を繰り返す」⇔「繰り返し議論する」という双方向の言い換え技術とは異なり, 冗長なAspect表現を簡略化さ

せた一方の言い換え手法を提案した.

(18) 本手法による正規化

議論を始める	[始動]	} →議論する
議論を続ける	[継続]	
議論を繰り返す	[反復]	

これにより, より多くの述部をまとめることができ, マイニング等の再現率向上につながると考えられる.

しかし, 一言に正規化といっても双方向の言い換えるような精度の高い正規化を求めるものもあれば, 「会議を行う」が「会議できる」を含意するといった, ムードを含めたより広義の含意表現をまとめた正規化が求められる場合も考えられる. 本技術では, (16)のような事態の事実関係が変化するような言い換えを避けた正規化技術を設計したが, 今後はアプリケーションが求めるレベルに合わせ, 正規化の粒度を変更させる技術の構築が必要になると考えられる.

6. 結論

本稿では, 機能動詞表現の言い換え技術を正規化に応用するための手法を提案した. 機能動詞が表すAspectやヴォイスといった機能語の意味役割と, 日本語の助動詞や時制形態素が持つ意味役割を対応づけて考察することにより, 意味の冗長性を省いた「最も単純な言い換え」を構築することができた. これをもとに, 述部を自動正規化した結果, 約 94%の精度で述部を単純な形に言い換えることができた. これは, マイニング等の精度を下げずに再現率向上を図る技術につながると考えられる. 今後はさらに, アプリケーションに応じたレベルでの, 言い換えから正規化の自動変換ができるような技術の構築を目指したい.

参考文献

- 藤田 篤, 降幡 健太郎, 乾 健太郎, 松本 裕治(2006). 語彙概念構造に基づく言い換え生成: 機能動詞構文の言い換えを例題に. *情報処理学会論文誌* 47.6, 1963-1975.
- 藤田 篤, 佐藤 理史, 泉 朋子, 今村 賢治, 菊井 玄一郎(2009). 言い換え用例に基づく機能動詞構文と動詞句の同義性計算. *言語処理学会第15回年次大会発表論文集*.
- Matsumoto, Y. (1996). A syntactic account of light verb phenomena in Japanese. *Journal of East Asian Linguistics* 5, 107-149.
- 村木 新次郎 (1991). *日本語動詞の諸相*. 東京: ひつじ書房.
- Ogihara, T. (1999). Tense and aspect. In N. Tsujimura (ed.) *The Handbook of Japanese Linguistics* (pp.321-348). Malden, MA: Blackwell.
- Portner P. H. (2006). *What is meaning? Fundamentals of formal semantics*. Malden, MA: Blackwell.
- Tsujimura, N. (2007). *An introduction to Japanese Linguistics*. Malden, MA: Blackwell.