

日本語における名詞句の情報構造と語順の相関についての 統計的検討

宮内 拓也[†]・浅原 正幸^{††}

本稿では、情報構造に関係する文法情報がどのように語順に影響を及ぼすのかについて調査するため、『現代日本語書き言葉均衡コーパス』の名詞句に対して情報構造に関わる文法情報のタグを付与した BCCWJ-InfoStr を利用して、名詞句の係り先からの距離（文節数）をベイジアン線形混合モデルによりモデル化した結果を報告する。その結果、日本語の名詞句の語順は、(I) 情報状態が旧情報であるものが新情報であるものに先行する、(II) 共有性が共有であるものが非共有であるものに先行する、(III) 定名詞句が不定名詞句に先行する、(IV) 有生名詞句が無生名詞句に先行するという傾向が確認された。これは、機能主義言語学の分野で言及されている「伝達のダイナミズム」・「旧から新への情報の流れ」・「名詞句階層」を支持するものである。

キーワード：正規語順、日本語、情報構造、ベイジアン線形混合モデル

Statistical Study of a Correlation between Information Structure and Word Orders of Noun Phrases in Japanese

TAKUYA MIYAUCHI[†] and MASAYUKI ASAHARA^{††}

The aim of this paper is to investigate how grammatical information regarding information structure effects word orders. We report results of modeling distances between NPs and predicates on which they depend by the Bayesian linear mixed model, using “BCCWJ-InfoStr,” where tags regarding information structure are annotated to NPs in the text of “Balanced Corpus of Contemporary Written Japanese.” As a result, estimated word orders of noun phrases in a sentence are as follows: (I) discourse-old NPs precede discourse-new NPs, (II) hearer-old NPs precede hearer-new NPs, (III) definite NPs precede indefinite NPs, and (IV) animate NPs precede inanimate NPs. These results support “Communicative Dynamism,” “From-Old-To-New Principle,” and “Nominal Hierarchy,” referred to in the field of functional linguistics.

Key Words: *Word Order, Japanese Language, Information Structure, Bayesian Linear Mixed Model*

[†] 東京外国語大学, Tokyo University of Foreign Studies

^{††} 人間文化研究機構国立国語研究所, NINJAL, Japan

1 はじめに

日本語は語順が自由な言語であり¹、その基本語順については様々な研究がある。特に、ヲ格要素とニ格要素が共に文に現れる二重目的語構文については、計算言語学の領域においても理論言語学の領域においても多くの研究がある (Yamashita and Kondo 2011; Orita 2017; 笹野, 奥村 2017; Hoji 1985; Miyagawa 1997; Matsuoka 2003)。その中でも、日本語の二重目的語構文について、格要素ではなく情報構造により語順が決まる傾向にあることを指摘した研究として、Asahara, Nambu, and Sano (2018) が挙げられる。

Asahara et al. (2018) は、『現代日本語書き言葉均衡コーパス』(Maekawa, Yamazaki, Ogiso, Maruyama, Ogura, Kashino, Koiso, Yamaguchi, Tanaka, and Den 2014) (以下, BCCWJ) のコアデータに対する述語項構造・共参照情報アノテーション (植田, 飯田, 浅原, 松本, 徳永 2015) を用いて、ガ・ヲ・ニすべての格要素を含む二重目的語構文を抽出し、その共参照情報をもとに格名詞句に情報の新旧を付与した。その結果、名詞句が直接目的語 (ヲ格要素) であるか間接目的語 (ニ格要素) であるかという点よりも、情報の新旧が語順に与える影響が大きいという傾向が明らかになった。

しかしながら、名詞句を特徴付ける情報構造に関わる要素は、共参照に基づく情報の新旧だけではない。例えば、Givón (1976) や Keenan (1976) は有生性 (animacy)、動作主性 (agentivity) がトピックやフォーカス²といった情報構造と関係があることを示している。そのため、情報の新旧以外の情報構造にかかわる要素についても検討する必要がある。また、Asahara et al. (2018) では二重目的語構文の場合のみが対象となっているため、それ以外の場合についてどのような語順になるのかを調査する必要がある。

そこで、本稿では、BCCWJ 内の名詞句に対して情報構造に関わる文法情報をアノテーションした BCCWJ-InfoStr (Miyachi, Asahara, Nakagawa, and Kato 2018; 宮内, 浅原, 中川, 加藤 2018) を利用して、情報の新旧以外のものも含めた情報構造に関わる文法情報がどのように語順に影響を及ぼすのかについて、二重目的語構文以外も対象に調査した結果を報告する。つまり、本研究は Asahara et al. (2018) で示された傾向を踏まえ、より広い観点、および対象から検討するものであると位置づけられる。

具体的には、BCCWJ-InfoStr に含まれる名詞句とその係り先との距離を、情報状態・共有性・定性・有生性という 4 つの情報構造に関わる文法情報の特徴量として、バイジアン線形混合モ

¹ 本稿で言う「語順が自由」とは、名詞句の (述語に対する) 配列に関して、命題的意味を変更させることなく、容認され得る複数の選択肢が存在することを意味している。同じ項や述語を用いて語順のみが異なっているような文を複数想定したとき、それらのニュアンスや語用論的意味解釈の同一性について保証するものではない点に注意されたい。以降についても同様である。

² トピックとは聞き手・読み手の前提となっている部分で、主張 (assertion) を導入する働きをする。フォーカスとは聞き手・読み手の前提にないような新しい情報を伝える部分であり、主張に直接関わってくる。詳細は Erteschik-Shir (1997, 2007) などを見られたい。

デル (Bayesian Linear Mixed Model; (Sorensen, Hohenstein, and Vasishth 2016)) により回帰分析する。その結果, 名詞句は文中で, (I) 情報状態が旧情報であるものが新情報であるものに先行する, (II) 共有性が共有であるものが非共有であるものに先行する, (III) 定名詞句が不定名詞句に先行する, (IV) 有生名詞句が無生名詞句に先行するという傾向が確認された。これらの傾向は, 主に機能主義言語学の分野で言及されている「伝達のダイナミズム」(Firbas 1971)・「旧から新への情報の流れ」(久野 1978)・「名詞句階層」(角田 1991) を支持する結果となった。

以下, 2 節では分析方法として, 使用したデータの概要と統計手法について示す。3 節では統計分析の結果を提示する。4 節では言語学的な考察を行う。5 節でまとめと今後の研究の方向性を示す。

2 方法

本節では分析に用いたデータについて解説する。BCCWJ-DepParaPAS (植田 他 2015; 浅原, 大村 2016) と BCCWJ-Infostr (Miyauchi et al. 2018; 宮内 他 2018) を重ね合わせたデータを用いる。

2.1 BCCWJ-DepParaPAS

BCCWJ-DepParaPAS は係り受け情報 (BCCWJ-DepPara) (浅原, 松本 2018) と述語項構造・共参照情報 (BCCWJ-PAS) を一つのファイルにしたデータである。

図 1 中, 表層形+係り受けの列が, 表層形と係り受け情報を表している。表層形は国語研短単位が規定する単位に分かち書きされており, * ではじまる列が国語研長単位によって規定される文節境界を表し, 各文節の係り先を表す。例えば, * 3 4D 0/1 0.000000 は, この行以下が 0-origin で 3 番目の文節を表し, 当該文節が 4 番目の文節に係ることを表す。0/1 は, 当該文節の末尾から数えて 0-origin で 0 番目の単語が付属語主辞であり, 末尾から数えて 0-origin で 1 番目の単語が自立語主辞であることを表す。0.000000 は係り受け解析器が出力する係り受けに関するパラメータが登録される領域である。

図 1 中, PAS 情報の列が BCCWJ-PAS に含まれる述語項構造である。id は, 述語のガ・ヲ・ニの項名詞句に割り当てられる。また, 述語についてはその項名詞句の情報が付与されている。例えば `ga="36" ga_dep="dep" ni="33" ni_dep="zero" o="35" o_dep="dep" type="pred"` は, ガ格名詞句が `id="36"` の名詞句「先生 (たち)」で係り受け関係 (`ga_dep="dep"`) があり, ニ格名詞句が前の文にある `id="33"` の名詞句であり (同一文内にないために `ni_dep="zero"` とゼロ代名詞であることを明示), ヲ格名詞句が `id="35"` の名詞句「レットル」で係り受け関係があることを表現する。なお, `type="pred"` は, 名詞述語でない述語であることを意味する。他に, 図 1 には記載されていないが, 共参照情報として `id="36"` が事前文脈に同一指示名詞句があるこ

表層形+係り受け	PAS 情報	情報状態	共有性	定性	有生性	係り先	係り先との距離
* 0 1D 0/1 0.000000							
ほか	-						
の	-						
* 1 4D 1/3 0.000000							
先生	id="36"	旧情報	想定可能	定性	有生	4	2
たち	-						
に	-						
は	-						
、	-						
* 2 3D 2/4 0.000000							
「	-						
お	-						
しゃべり	-						
」	-						
の	-						
* 3 4D 0/1 0.000000							
レットル	id="35"	新情報	共有	定性	無生	4	0
を	-						
* 4 -1Z 3/4 0.000000							
張ら	ga="36" ga_dep="dep" ni="33" ni_dep="zero" o="35" o_dep="dep" type="pred"						
れ	-						
て	-						
い	-						
た	-						
。	-						

図 1 調査に用いるアノテーション例

とを含んでいる。

本研究の評価対象は、BCCWJ-DepParaPAS の述語項構造アノテーションで定義される各項名詞句と係り受けアノテーションで定義されるその係り先 (図 1 中、係り先列) との距離 (図 1 中、係り先との距離列) である。係り先との距離は、係り先と当該名詞句との間に入る文節数で評価する。この値が、次に示す名詞句の情報構造 (情報状態、共有性、定性、有生性) により統計的にモデル化できるかを検討する。なお、名詞句の情報構造のうち、情報状態は BCCWJ-DepParaPAS の共参照情報に由来する。

2.2 BCCWJ-Infostr

BCCWJ-Infostr は BCCWJ のテキスト (新聞 (PN) コアデータ 16 サンプル、全 16,657 語 (短単位)、5,195 文節、739 文) 内の名詞句 2,023 に対し、情報構造に関係する文法情報のタグ (情報状態、共有性、定性、特定性、有生性、有情性、動作主性) をアノテーションしたものである³。図 1 では、情報状態、定性、有生性、共有性についてのみ、ゴシック体で示す。ただし、

³ BCCWJ-InfoStr では、情報構造をアノテーションしたいくつかの先行研究 (Hajičová, Panevová, and Sgall 2000; Götze, Weskott, Endriss, Fiedler, Hinterwimmer, Petrova, Schwarz, Skopeteas, and Stoel 2007) とは異なり、トピックやフォーカスを直接アノテーションしていない。例えば、トピックやフォーカスをアノテーションしてい

複合語については、前部要素には指示性 (referentiality) が無いこと等を考慮して、前部要素まで含めて一つの名詞と捉えることにより、実質的に長い名詞句へアノテーションされている。この単位に対する方針については、基本的には BCCWJ-DepParaPAS の作業方針と同じとなっている。上記に示したタグは言語学の専門的な知識を持つものでないとアノテーションができないために、1人の言語学専攻の博士課程の学生がアノテーションした上で、3人によりその結果が確認されている。アノテータは BCCWJ-DepParaPAS に付与された共参照情報を確認しながら作業を行った。

本研究では Miyauchi et al. (2018), 宮内 他 (2018) で付与されたタグの内、情報状態, 共有性, 定性, 有生性の各値を用いた⁴。情報状態とは話し手・書き手⁵の観点に基づく情報の新旧で、共有性は聞き手・読み手⁶の観点に基づく情報の新旧である (Prince 1992)。定性については情報構造との関連が指摘されており (Erteschik-Shir 2014), 情報構造と語順に関係がある (Asahara et al. 2018) ことから定性と語順の関係を調査する必要がある。また、定性を表示する機能を持つ冠詞の習得は、冠詞を持たない言語を母語とする者にとっては難しい (Ionin, Ko, and Wexler 2004; Tanaka 2013) ことから、その解消を目指して冠詞推定の研究が行われている (竹内, 河合, 細田, 永田 2013; 乙武, 永田 2016) が、定性と語順との関連が分かることにより冠詞推定に寄与する可能性がある。有生性は、存在動詞の差異 (いる／ある)⁷のほか、格交代を表出する使役・受動態との関連があり⁸、語順に影響を及ぼす可能性がある。

情報状態, 共有性, 定性, 有生性について、名詞句に付与されたタグの値は以下 (1) の通りである。

- (1) a. 情報状態 (information status):
 「新情報 (discourse-new)」 / 「旧情報 (discourse-old)」
 b. 共有性 (commonness):

る Götze et al. (2007) でも、実際には指示的 (referential) な名詞句、特定 (specific) 解釈や総称 (generic) 解釈を持つ不定 (indefinite) の名詞句など、様々な種類のアバウトネストピック (aboutness topic) が区別されている。これらを考慮し、BCCWJ-InfoStr では、定性や特定性のような要因はトピックと独立にアノテーションされている。このことは、情報構造におけるトピックやフォーカスがそれぞれに多次元的であり (Nakagawa 2016), より小さい要素に分解可能であることを示唆している。

⁴ 本研究で利用しない特定性・有情性・動作主性については Miyauchi et al. (2018), 宮内 他 (2018) を参照のこと。また、本研究でこれらのタグを使用しない理由については 5 節を見られたい。

⁵ これ以降、単に「話し手」と表記する。

⁶ これ以降、単に「聞き手」と表記する。

⁷ 日本語については、有生／無性の区別よりも有情／非情の区別が重要であるとする研究もある (山口 1985) が、この点については本研究に特に大きな影響を及ぼさない。

⁸ 例えば、日本語の受動文においては、有生性制約 (animacy constraint) があることがよく知られている (Kuroda 1979)。動作主が二格で表現される受動文において、以下 (i a) で示すように、無生の主語は容認されづらい。一方、(i b) のように、有生の主語は問題なく容認される。

- (i) a. ??机が花子に叩かれた。
 b. 太郎が花子に叩かれた。

「共有 (hearer-new)」／「非共有 (hearer-old)」／「想定可能 (bridging)」

c. 定性 (definiteness):

「定 (definite)」／「不定 (indefinite)」

d. 有生性 (animacy):

「有生 (animate)」／「無生 (inanimate)」

なお、定性・有生性については、与えられた文脈から判断できない場合に「どちらでもよい」というタグが認められている。共有性については、その概念が認めがたい場合に「どちらでもない」というタグも認められている。これらのタグを含む名詞句については、本研究の分析対象から除外した。

以下、Miyauchi et al. (2018), 宮内 他 (2018) をもとに、実例と共にそれぞれのタグのアノテーション基準を示す。

2.2.1 情報状態・共有性

情報状態とは、いわゆる談話上の旧情報と新情報の区別である。ある談話において、新たな情報は「新情報」となり、既出の情報は「旧情報」となる⁹。一つのテキスト（BCCWJ 新聞サンプルにおける記事単位）全体を一つの談話とみなし、アノテーションされた。

実例として、(2) を見られたい。

(2) a. 担任だった池田弘子^[新情報]先生は違った。

b. スクールカウンセラーでもあった^[旧情報]先生の授業は [...]

(読売新聞 [BCCWJ:PN1c.00001])

(2a) の名詞「池田弘子先生」はこのテキストで初出の実体 (entity) であるために、末尾の短単位名詞「先生」に新情報タグが付与されている。一方、談話中で (2a) より後方に出現する (2b) の「先生」は (2a) の「池田弘子先生」を指示しているため旧情報タグが付与されている。これらの名詞は共参照関係にあり、BCCWJ-DepParaPAS のアノテーションから展開できるが、展開したのち、全数確認を行いタグ付与がなされた。

共有性は、談話成立時点において、情報を聞き手が既知していると話し手が想定しているか否かを示す分類である。聞き手が既知していると話し手が想定している情報は「共有」であり、知らないと想定している情報は「非共有」である¹⁰。なお、この判断の際はアノテータの世界知識を使ってもよいこととされており、「想定可能」というタグも許されている。このタグは、ブリッジングを起こしている際に付与される。

⁹ 情報状態についてのより詳細は Kruijff-Korbyová and Steedman (2003) や Hinterwimmer (2011), Prince (1992) 等を参照のこと。

¹⁰ 共有性についてのより詳細は Prince (1992) などを見られたい。

実例として, (3) を見られたい.

- (3) [共有 キャンティ街道] を抜け、[非共有 オリーブ畑] に囲まれた田園地帯の [想定可能 レストラン] で、[...] (読売新聞 [BCCWJ:PN4c_00001])

(3) において, 名詞「キャンティ街道」は, 世界遺産にも登録されている, ワインで有名な街道であり, アノテータは既にこの街道について知っていたため, 共有のタグが付与された. 「オリーブ畑」については, この記事からどんなオリーブ畑であるのか判断できないため, 非共有のタグが与えられている. 「レストラン」はキャンティ街道のレストランを指しており, ある種のブリッジングを起こしているため, 想定可能のタグが付与されている.

2.2.2 定性

定性とは, 指示対象を聞き手が同定できるか否かを示す分類である. 指示対象を聞き手が同定できると話し手が想定していれば「定」であり, 同定できないと想定していれば「不定」である¹¹. 同データでは, 判定する際に確認する文脈として前後3文を見ることとしている.

実例として, (4) を見られたい.

- (4) 高等部では自由な校風もあって、流行に乗ってかばんを薄くつぶしたり、ピアスをした。呼び出して注意する先生もいたが、二、三年時に担任だった池田弘子先生（七十五）は違った。「そんな薄い [定 かばん] じゃ [不定 遊び道具] も入らないよ」 [...] (読売新聞 [BCCWJ:PN4c_00001])

(4) の3文目の名詞「かばん」はスコープである前3文以内に既出の名詞であり, ここでは具体的に聞き手の持ち物のかばんを指示している. 話し手はこの「かばん」は聞き手により同定しようと想定していると考えられるため, 定のタグが与えられている. 一方, 「遊び道具」については, 特に具体的な何か遊び道具を指示しているわけではないことから, 不定のタグが付与されている.

2.2.3 有生性

有生性とは, 生きているか否かを示すカテゴリーである. 生物 (人間, 動物など) は「有生」であり, 無生物 (植物を含む) は「無生」である. 有生性は名詞句レベルのみで判断し付与されるものとしている.

実例として, (5) を見られたい.

- (5) オオクチバスなどの [有生 ブラックバス類] が、少なくとも四十三都道府県の七百六十一のため池や [無生 湖沼] に侵入し、[...] (読売新聞 [BCCWJ:PN4c_00001])

¹¹ 定性についてのより詳細は Lyons (1999), Heim (2011) などを参照のこと.

(5)における「ブラックバス」は生物であるため、有生のタグが付与される。「湖沼」は無生物であり、無生のタグが与えられる。

2.3 基礎統計

表 1 に、本研究で用いたデータの各特徴量の頻度をピボットテーブル形式で示す。

情報状態が旧情報である場合 (678 件)、共有性が共有 (611 件) になりやすい傾向にある。一方、情報状態が新情報である場合 (1,343 件) は、共有性が共有 (425 件)・想定可能 (455 件)・非共有 (459 件) といずれの可能性もありうる。共有 (旧情報 611 件・新情報 425 件) の場合、定性 (旧情報 604 件・新情報 406 件) である可能性が高い。有生性については、新聞記事という性質上、無生のものが多い傾向にある。その他の特徴量の傾向については、Miyachi et al. (2018), 宮内 他 (2018) を参照されたい。

表 2 に、名詞句の係り先との距離の分布を示す。係り先との距離は、当該名詞句と係り先文節との間に入る文節数で表現する。ゆえに、名詞句と係り先文節が隣接する場合には係り先の距離を 0 とする。例えば、図 1 において「先生 (たち)」の係り先との距離は 2 であり、「レット」の係り先との距離は 0 である。基本的に、約半数の名詞句は隣接文節に係るために係り先との距離は 0 である。但し、係り先の距離が 0 であるもののうち、77 名詞句は新聞記事に頻出する文末名詞句 (体言止め) であった。

2.4 モデル化の方法

本研究では、情報状態、共有性、定性、有生性の各タグをもとに、係り元名詞句とその係り先文節との距離 $dist$ (図 1 中、係り先との距離) をベイジアン線形混合モデルで評価した。アノテーションされた名詞句 2,023 のうち、特徴量のタグに「どちらでもよい」「どちらでもない」が含まれる 7 例と文末に出現した名詞句 77 例を除外した 1,939 を分析対象とした。係り受けの情報は 2.1 節の BCCWJ-DepParaPAS の情報を用いた。

具体的には以下のような線形式でモデル化を行った：

$$dist \sim \text{Normal}(\mu, \sigma),$$

$$\text{但し, } \mu \leftarrow \alpha + \beta_{\text{情報状態}}^* + \beta_{\text{共有性}}^* + \beta_{\text{定性}}^* + \beta_{\text{有生性}}^*.$$

ここで、 $dist$ は係り元名詞句と係り先名詞句の間に入る文節の数 (図 1 中、係り先との距離) とする。Normal は平均 μ 標準偏差 σ の正規分布とし、切片 α と名詞句の各カテゴリパラメータ $\beta_{\text{旧情報情報状態}}, \beta_{\text{新情報情報状態}}, \beta_{\text{共有}}, \beta_{\text{共有性}}$ (図 1 中、情報状態の列に入っている値に応じて 0-1 の値を割り当てる、以下同様), $\beta_{\text{想定可能}}, \beta_{\text{非共有}}, \beta_{\text{共有性}}$ (図 1 中、共有性の列), $\beta_{\text{定性}}, \beta_{\text{不定定性}}$ (図 1 中、定性の列), $\beta_{\text{有生}}, \beta_{\text{有生性}}, \beta_{\text{無生}}, \beta_{\text{無生性}}$ (図 1 中、有生性の列) の取りうるパラメータ割り当ての線形結合で平均 μ を定式化した。各パラメータに割り当てられる値が大きいくほど、名詞句の係り受けの距離が遠いこ

表 1 各特徴量の頻度

情報状態	共有性	定性	有生性	頻度
旧情報	*	*	*	678
旧情報	共有	*	*	611
旧情報	共有	定性	*	604
旧情報	共有	定性	無生	481
旧情報	共有	定性	有生	123
旧情報	共有	不定性	*	7
旧情報	共有	不定性	無生	5
旧情報	共有	不定性	有生	2
旧情報	想定可能	*	*	33
旧情報	想定可能	定性	*	6
旧情報	想定可能	定性	無生	4
旧情報	想定可能	定性	有生	2
旧情報	想定可能	不定性	*	27
旧情報	想定可能	不定性	どちらでもよい	1
旧情報	想定可能	不定性	無生	18
旧情報	想定可能	不定性	有生	8
旧情報	非共有	*	*	34
旧情報	非共有	定性	*	15
旧情報	非共有	定性	有生	15
旧情報	非共有	不定性	*	19
旧情報	非共有	不定性	無生	12
旧情報	非共有	不定性	有生	7
新情報	*	*	*	1,343
新情報	どちらでもない	*	*	4
新情報	どちらでもない	定性	*	1
新情報	どちらでもない	定性	無生	1
新情報	どちらでもない	不定性	*	3
新情報	どちらでもない	不定性	無生	3
新情報	共有	*	*	425
新情報	共有	定性	*	406
新情報	共有	定性	無生	380
新情報	共有	定性	有生	26
新情報	共有	不定性	*	19
新情報	共有	不定性	無生	17
新情報	共有	不定性	有生	2
新情報	想定可能	*	*	455
新情報	想定可能	どちらでもよい	*	2
新情報	想定可能	どちらでもよい	無生	2
新情報	想定可能	定性	*	31
新情報	想定可能	定性	無生	23
新情報	想定可能	定性	有生	8
新情報	想定可能	不定性	*	422
新情報	想定可能	不定性	無生	383
新情報	想定可能	不定性	有生	39
新情報	非共有	*	*	459
新情報	非共有	定性	*	59
新情報	非共有	定性	無生	20
新情報	非共有	定性	有生	39
新情報	非共有	不定性	*	400
新情報	非共有	不定性	無生	329
新情報	非共有	不定性	有生	71
総計				2,021

表 2 係り先との距離の分布

係り先との距離	0	1	2	3	4	5	6	7	8	総計
頻度	1,014	809	140	40	10	5	1	0	2	2,021

とを表す。つまり、他の要素よりも先行する傾向にあることを表す。

ベイズ推定時には、各名詞句の特徴に応じてカテゴリパラメータ割り当てを与える。例えば、図1の「先生(たち)」の事例では、モデル化する係り先との距離を2として、情報状態:旧情報・共有性:想定可能・定性:定性・有生性:有生の特徴量を与える。同様に「レットル」の例では、係り先との距離を0として、情報状態:新情報・共有性:共有・定性:定性・有生性:無生の特徴量を与える。事前分布として平均0分散 σ' である標準分布を各パラメータに与え、ハイパーパラメータ σ' についてもベイズ推定を行った。なお、同等のモデルを頻度主義的な線形混合モデル(lmer)により推定したところ収束しなかった。

推定された線形モデルの線形式(一次方程式)を解釈するには、傾きであるカテゴリパラメータを確認する。簡単化のために出力 y 、切片 a 、傾き b の線形式 $y = a + b$ を用いて説明すると、傾き b が正の場合、 y の値を増やす効果がある。一方、傾き b が負の場合 y の値を減らす効果がある。このように、カテゴリパラメータが正である場合には出力である係り先との距離を遠くする効果があると解釈でき、カテゴリパラメータが負である場合には出力である係り先との距離を近くする効果があると解釈できる。

ベイジアン線形混合モデルにより rstan パッケージ (Stan Development Team 2018) を用いて推定を行った¹²。warm-up 後のイテレーションを 15,000 回に設定し、4 回シミュレーションを実施した。全てのモデルは収束した。

ここで用いたベイジアン線形混合モデルとは、線形混合モデルのベイズ主義的なモデルである。線形混合モデルは、多要因の集約が必要ないという点で、分散分析や対応あり t 検定などのアプローチよりも本研究にとって適切である。帰無仮説に基づく頻度主義的な線形混合モデルでは、本研究のような多要因の回帰が困難であるが、ベイズ主義的なモデルでは柔軟にモデル化が可能となる。特に、ベイジアン線形混合モデルには、線形混合モデルを使用した頻度分析に比べて2つの大きな利点がある。まず、事前分布を使用して情報を組み込むことができる。さらに多数のランダム分散成分を持つ複雑なモデルをマルコフ連鎖モンテカルロ法によって適合させることができる。以上の利点から、本研究ではベイジアン線形混合モデルを用いて、係り元名詞句とその係り先文節との距離 $dist$ を評価した。

¹² 動作環境は Panasonic Let's Note CF-SV 上の R version 3.5.1 (2018-07-02, x86_64-w64-mingw32)。

3 結果

統計処理の結果, 表3の結果が得られた.

Rhat は収束判定パラメータで 1.2 以下を収束とみなす. mean は事後平均であり, 各タグにより係り先との距離が長くなるか (+ 方向) 短くなるか (- 方向) を数値で示す. 日本語は主辞が後置されるために, 係り先との距離が長くなる要素が語順においては先行する要素となる. これをもとに各名詞句の語順を推定すると, 表4の通りとなり, 名詞句は文中で, 「旧情報 → 新情報」, 「共有 → 想定可能 → 非共有」, 「定 → 不定」, 「有生 → 無生」の順で並ぶという推定結果が得られた.

具体的には, 図1の「先生(たち)」は情報状態:旧情報・共有性:想定可能・定性:定・有生性:有生であることから, $\alpha + \beta_{\text{情報状態}}^{\text{旧情報}} + \beta_{\text{共有性}}^{\text{想定可能}} + \beta_{\text{定性}}^{\text{定}} + \beta_{\text{有生性}}^{\text{有生}} = 0.243 + 1.147 + 0.764 + 1.147 + 0.442 = 3.743$

表3 統計処理の結果

Parameter	Rhat	mean	sd
α	1.17	0.243	0.004
$\beta_{\text{情報状態}}^{\text{新情報}}$	1.07	-0.632	0.007
$\beta_{\text{情報状態}}^{\text{旧情報}}$	1.02	1.147	0.008
$\beta_{\text{共有性}}^{\text{共有}}$	1.12	1.332	0.005
$\beta_{\text{共有性}}^{\text{想定可能}}$	1.20	0.764	0.006
$\beta_{\text{共有性}}^{\text{非共有}}$	1.00	-1.975	0.011
$\beta_{\text{定性}}^{\text{定}}$	1.00	1.147	0.003
$\beta_{\text{定性}}^{\text{不定}}$	1.02	0.488	0.011
$\beta_{\text{有生性}}^{\text{有生}}$	1.06	0.442	0.008
$\beta_{\text{有生性}}^{\text{無生}}$	1.15	-0.697	0.012
σ	1.14	0.063	0.008
log-posterior	1.02	28,198.286	300.838

表4 名詞句とその係り先との距離

	推定結果	推定された語順
情報状態	$\beta_{\text{情報状態}}^{\text{新情報}} < \beta_{\text{情報状態}}^{\text{旧情報}}$	旧情報 → 新情報の順
共有性	$\beta_{\text{共有性}}^{\text{非共有}} < \beta_{\text{共有性}}^{\text{想定可能}} < \beta_{\text{共有性}}^{\text{共有}}$	共有 → 想定可能 → 非共有の順
定性	$\beta_{\text{定性}}^{\text{不定}} < \beta_{\text{定性}}^{\text{定}}$	定 → 不定の順
有生性	$\beta_{\text{有生性}}^{\text{無生}} < \beta_{\text{有生性}}^{\text{有生}}$	有生 → 無生の順

表 5 推定結果の差の確認

推定結果の差		
情報状態	$\beta_{\text{旧情報情報状態}} - \beta_{\text{新情報情報状態}}$	1.779
共有性	$\beta_{\text{共有共有性}} - \beta_{\text{非共有共有性}}$	3.307
	$\beta_{\text{共有共有性}} - \beta_{\text{想定可能共有性}}$	0.568
	$\beta_{\text{想定可能共有性}} - \beta_{\text{非共有共有性}}$	2.739
定性	$\beta_{\text{定定性}} - \beta_{\text{不定定性}}$	0.659
有生性	$\beta_{\text{有生有生性}} - \beta_{\text{無生有生性}}$	1.139

と距離がモデル化される。「レットル」は情報状態：新情報，共有性：共有，定性：定，有生性：無生であることから， $\alpha + \beta_{\text{新情報情報状態}} + \beta_{\text{共有共有性}} + \beta_{\text{定定性}} + \beta_{\text{無生有生性}} = 0.243 - 0.632 + 1.332 + 1.147 - 0.697 = 1.393$ とモデル化される。この2つの距離を比較すると「先生（たち）」が「レットル」よりも係り先より遠いことが確認できる。

各特徴量に割り当てられた差を表5で示す。Asahara et al. (2018)の調査では，ヲ・ニ格の格要素よりも，情報状態に相当する共参照情報が語順に影響を与えるという結果であったが，情報状態よりも共有性のほうが顕著な距離の差を表出することがわかる。また，共有性の中でも，共有と想定可能の差が小さく，想定可能と非共有の間に大きな差があることが確認できた。名詞句の共有性が想定可能であることは，日本語においては顕在的な標識がないために言語処理において認識が困難な現象である。反対に，今回得られた結果から，語順が共有性の推定に有効な可能性が示唆された。

なお，この分析手法は帰無仮説に基づかないために，有意差という概念は存在しない。得られた推定値の差が得られた2標準偏差よりも大きい場合に顕著であると判断する。今回得られた結果はいずれも顕著な結果であった。

4 考察

4.1 情報状態，共有性

表4で示したように，情報状態に関しては，旧情報の名詞句が新情報の名詞句に先行する傾向が確認された。また，共有性に関しては，共有の名詞句が想定可能の名詞句に先行し，想定可能の名詞句が非共有の名詞句に先行するという結果であった。より具体的には，旧情報／共有の名詞句が新情報／非共有の名詞句に先行している(6a)のような文は，旧情報／共有の名詞句が新情報／非共有の名詞句に後続する(6b)のような文よりも対象のコーパスにおいて多く出現する。対象としているコーパスはBCCWJであり，BCCWJが現代日本語の書き言葉について

て代表性を有していることを踏まえれば, この結果は現代日本語の書き言葉の特徴を示していること考えることができる¹³.

- (6) a. [旧情報/共有 蓮舫さん] は [...] 自分の [新情報/非共有 意見] を言った。
(読売新聞 [BCCWJ: PN1c.00001])
- b. [新情報/非共有 海外派兵] とごう慢な外交がさらに [旧情報/共有 国] を弱らせていると [...] (毎日新聞 [BCCWJ: PN4b.00001])

これらの結果は, まさに文中で旧情報を前に置き新情報を後ろに置くという「旧から新への情報の流れ」(久野 1978) を支持している。これは (7) のように定義されている。

- (7) 「旧から新への情報の流れ」(From-Old-To-New Principle)

文中の語順は, 旧情報を表す要素から, 新情報を表す要素へ進むのを原則とする。

(久野 1978, p. 59)

この原則により語順の選好性が説明され得る。Kuno (2004) に基づいて, 英語の語順の選好性について説明する。(8) を見られたい。

- (8) a. John gave the girl a book.
b. ?? John gave a girl the book.

(Kuno 2004, p. 325)

Kuno (2004) は, 英語の語順が自由であるとされる部分として, (8) のように直接目的語と間接目的語の間の語順に関して, この原則の有効性を検討している。(8) において, (8a) は容認される一方で, (8b) の容認性は著しく低下する。(8a) では, 間接目的語である *the girl* が談話上旧情報であり, 新情報である直接目的語の *a book* に先行している。これは (7) で示した原則に従う語順である。しかし, (8b) においては, 新情報を表す間接目的語の *a girl* が旧情報を表す直接目的語 *the book* に先行している。これは (7) で示した原則に反する語順であり, (8b) の容認性の低さは「旧から新への情報の流れ」に帰着させることができる。

以上のように, 「旧から新への情報の流れ」は, 英語における自由な語順における選好性のみならず, 日本語における情報状態, および共有性についての語順の推定結果を予測するものである。

さらに, 情報状態, および共有性についての語順の推定結果は, 機能的文眺望 (Functional Sentence Perspective) の研究で議論される, 「伝達のダイナミズム」(Firbas 1971) を支持する。これは (9) で示される概念である。

¹³ ただし, 本稿で検討しているレジスタは新聞のみである。情報構造と語順の相関についてレジスタによって変わるとい報告はないことから, 情報構造と語順の相関がレジスタに (少なくとも大きくは) 依存しないと考えられる。なお, 以降もこれを前提とする。他のレジスタにおける詳細な検討は今後の課題としたい。

(9) 「伝達のダイナミズム」(Communicative Dynamism)

By the degree or amount of CD carried by a linguistic element, I understand the relative extent to which the element contributes to the development of the communication, to which, as it were, it ‘pushes the communication forward’.

「ある言語要素によって運ばれる CD の量によって、その要素がコミュニケーションの発展に寄与する、いわばコミュニケーションを押し進める相対的な大きさがわかる。」

(Firbas 1971, p. 136)

(9) に従うと、文の要素は担う伝達情報の量に応じて少ない情報量のものから多いものへ順に並べられるということになる。Firbas (1979) によれば、CD が小さい要素は文中の左方に配置され、CD が大きい要素は右方に配置される¹⁴。CD の大小は文脈への依存度によって決まってくる。旧情報／共有の名詞句が文脈依存度が相対的に高く、新情報／非共有の名詞句が相対的に文脈依存度が低いことはそれらの定義から明らかであるため、旧情報／共有の名詞句は相対的に CD が大きく、新情報／非共有の名詞句は相対的に CD が小さいことになる。このように考えると、「伝達のダイナミズム」は情報状態、および共有性についての語順の推定結果を予測しうる。

4.2 定性

表 4 の通り、定性については、定の名詞句が不定の名詞句に先行する傾向があるという結果であった。この結果は、(10a) のような文は (10b) よりも対象のコーパスにおいて多く出現することを意味しており、現代日本語の書き言葉の特徴を示していると考えられる。

- (10) a. そんな薄い [定 かばん] じゃ [不定 遊び道具] も入らないよ

(読売新聞 [BCCWJ: PN1c.00001])

- b. [不定 看護師 = 当時 (五三) =] が、[...] [定 病院駐車場] から車で連れ去られ [...]

(中日新聞 [BCCWJ: PN4f.00001])

前述の「伝達のダイナミズム」(Firbas 1971, 1979) は、定の名詞句と不定の名詞句の文中での分布についても説明を与えることができる。定、および不定の定義から、定の名詞句は相対的に文脈依存度が高く、不定の名詞句は相対的に文脈依存度が低いことがわかる。よって、前述のように文脈依存度により CD の大小が決まり、CD の大小に応じて (自然な) 語順が決まることを考えれば、文脈的依存度が相対的に高い定の名詞句が前方に現れやすく、逆に文脈的独立度が高い不定の名詞句が後方に現れやすいことが予測される。

実際に、語順が自由であり、かつ冠詞による定性の表示を行わないという点で日本語と類似

¹⁴ なお、Firbas (1979) はこのような語順を “ordo naturalis” 「自然な語順」と呼んでいる。

しているスラヴ系の言語では、かつてより定性と語順の関係が指摘されており (Krámský 1972; Chvany 1973; Szwedek 1974), それは「伝達のダイナミズム」に沿うものである。例として, (11) のロシア語の例文を見られたい。

- (11) a. Na stole stojala lampa.
on table stood lamp-NOM.INDEF
- b. Lampa stojala na stole.
lamp-NOM.DEF stood on table
「ランプがテーブルの上にあった。」 (Chvany 1973, p. 266)

(11a) の主語 *lampa* 「ランプ」は不定で解釈されやすく, (11b) の主語 *lampa* は定で解釈されやすいとされる。これら 2 つの主語の違いは, 文の前方に位置するか後方に位置するかという文中における位置のみである。

定の名詞句が文中の前方の位置を占める傾向があり, 不定の名詞句が後方の位置を占める傾向が見られる¹⁵ という表 4 で示した結果から, 日本語もスラヴ諸語と同様の傾向が見られることがわかる。さらに, それは「伝達のダイナミズム」によって説明され得るものである。

4.3 有生性

有生性に関しては, 有生の名詞が無生の名詞に先行する傾向があるという結果であった。この結果は, (12a) のような文が (12b) よりも対象のコーパスにおいて多く出現することを意味しており, 現代日本語の書き言葉の特徴であると考えられる。

- (12) a. [有生 蓮舫さん] は [...] 自分の [無生 意見] を言った。
(読売新聞 [BCCWJ: PN1c.00001])
- b. [無生 共同通信社] は今回審査を受ける九人の [有生 裁判官] にアンケートをした。
(西日本新聞 [BCCWJ: PN3g.00001])

¹⁵ なお, 英語においても同様の傾向が見られると考えることもできる。例えば, (8) の例文では, 直接目的語と間接目的語の冠詞による定性の表示から定の名詞句が相対的に前方に位置し, 不定の名詞句が相対的に後方に位置することが確認できる。また, 以下 (ii) のような存在構文に見られる, 定の名詞が出現できないという定性制限 (definiteness restriction) についてもこの傾向と類似したものであると考えられる。

- (ii) a. There is a wolf at the door.
b. *There is the wolf at the door.
(Milsark 1977)

(ii b) では定の名詞句を存在構文の主語にした際に非文になることが示されているが, (iii) で示すように, 虚辞の *there* を用いた存在構文でない形であれば容認される。

- (iii) The wolf exists at the door.

しかしながら, (ii b) と (iii) は, 例えば (11) などとは異なり, 相互に異なる文構造であるため, この点について結論を出すにはより詳細な検討が必要となる。

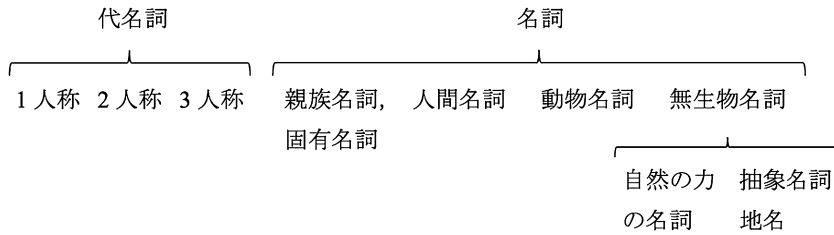


図 2 名詞句階層

これは、有生の名詞が無生の名詞より階層が高いとする有生性階層に従う形となっている。例えば、Silverstein (1976) は、文の成立において名詞句の階層でより高い位置にあるものが主語となる文構造が優先されるとする。Silverstein の階層をもとに、Dixon (1979) や Zubin (1979) の研究などを受けて角田 (1991) が修正した名詞句階層を図 2 に示す。図中の左方が階層の上位を、右方が下位を表す。図 2 においては、代名詞を除き、親族名詞、(有生の) 固有名詞¹⁶、人間名詞、動物名詞といった有生の名詞が、無生の名詞である無生物名詞 (自然の力の名詞、抽象名詞、地名) よりも階層上で高い位置を占めることがわかる。

日本語の基本語順は SOV であると考えられており、スクランプリング等による語順の変更がない限り¹⁷、主語は他の要素より前方に現れる。そのため、有生の名詞句が無生の名詞句に先行するという本研究の結果は、実質的に有生の名詞句が主語に立ちやすいという Silverstein の名詞句階層と一致すると考えることができる。また、角田 (1991) は名詞句階層において、親族名詞、固有名詞、人間名詞、動物名、自然の力の名詞、抽象名詞、地名といった多様な名詞の階層を想定しているが、少なくとも有生と無生という 2 項の対立に限定すれば、この階層の妥当性が統計的に確認されたといえる。

5 おわりに

本稿では、情報構造に関する文法情報がどのように語順に影響を及ぼすのかについて調査するため、BCCWJ 内の名詞句に対して付与された情報構造に関わる文法情報のタグ (Miyachi et al. 2018; 宮内 他 2018) の各値を利用し、値が付与された名詞句の文末からの距離をベイジアン線形混合モデル (Sorensen et al. 2016) によりモデル化した結果を報告した。

本研究では、名詞句は文中で、「旧情報 → 新情報」、「共有 → 想定可能 → 非共有」、「定 → 不定」、「有生 → 無生」の順で並ぶという推定結果が得られた。これらの結果は、先行研究の

¹⁶ ここで、固有名詞は「太郎」「花子」などの人間を指すもののみを指しており、無生の固有名詞は想定されていない (角田 1991, p.39)。

¹⁷ スクランプリングなどが起こる場合についての詳細な検討は今後の課題である。

「伝達のダイナミズム」(Firbas 1971) や「旧から新への情報の流れ」(久野 1978), 「名詞句階層」(Silverstein 1976) を支持するものである。従来の機能主義的言語学の領域の成果を, データから統計的に検討したという点に本研究の貢献がある。

なお, BCCWJ-InfoStr には, 有情性, 特定性, 動作主性のタグも付与されている。有情性は有生性と, 特定性は定性と強い相関¹⁸があるために同時に入れると線形モデルにおいて多重共線性の問題があるために今回のモデル化から除外した。しかしながら, それぞれ相関がある情報構造のタグと同じ傾向があることが言える。また, 予備調査においては BCCWJ-InfoStr に含まれる動作主性についても統計分析を行ったが, 件数が少なく統計的に顕著な傾向が見られなかった。今回選択したパラメータの単純な交互作用を含むモデルについても分析を行ったが, いずれのモデルも収束しなかった。このことから, 各特微量間の関係は不明である。今後, データ量を増やしたうえで, 各特微量間の関係を階層的にモデル化することで各特微量間の関係を検討したい。

笹野, 奥村 (2017) は大規模ウェブコーパスを用いて日本語二重目的語構文の基本語順に関してさまざまな調査を行った。しかしながら, 彼らの調査においては, 情報の新旧の観点が入り入れられていないという問題点がある。Asahara et al. (2018) の共参照情報に基づく分析が示す通り, 日本語の二重目的語構文の基本語順は既出の名詞句が先行する顕著な傾向があり, 分類語彙表に基づく動詞や名詞の分類やヲ・ニの表層格においては特段の傾向が見られなかった。本研究は, 同調査を任意の名詞句に拡張するとともに, 情報構造を情報状態, 共有性, 定性, 有生性の観点で整理したうえで, 再分析したものである。笹野, 奥村 (2017) で示されている「二格名詞が着点を表す場合は有生性を持つ名詞の方が「にを」語順をとりやすい」については, 本研究の結果の有生名詞句が無生名詞句に先行するという結果に一致する。これらの結果から, 日本語の格要素の語順については, ガ格以外においては本研究で取り扱った情報構造 (情報状態・共有性・定性・有生性) がより強い影響を与える特徴であり, 笹野, 奥村 (2017) の調査においても情報構造を含めた再調査が必要であることを示唆する。

謝 辞

本研究は, The 16th International Conference of the Pacific Association for Computational Linguistics (PACLING 2019) の発表 “Statistical Approaches to a Correlation between Information Structure and Word Orders of Noun Phrases in Japanese” (Miyachi and Asahara 2019) に基づいている。また, 本研究の一部は国立国語研究所コーパス開発センター共同研究プロジェクト

¹⁸ 具体的には, 有情性と有生性の間の κ 係数は 0.98 であり, 特定性と定性の間の κ 係数は 0.96 である (Miyachi et al. 2018; 宮内 他 2018)。なお, Landis and Koch (1977) の示した κ 値を評価する基準では, 0.81–1.00 の値は「ほとんど完璧な一致」(Almost Perfect) であるとされる。

ト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」, JSPS 科研費 17J07534, 19K23073, 18K18519, 17H00917, 18H05521 の助成を受けている。

参考文献

- Asahara, M., Nambu, S., and Sano, S.-I. (2018). “Predicting Japanese Word Order in Double Object Constructions.” In *Proceedings of the 8th Workshop on Cognitive Aspects of Computational Language Learning and Processing*, pp. 36–40, Melbourne. Association for Computational Linguistics.
- 浅原正幸, 松本裕治 (2018). 『現代日本語書き言葉均衡コーパス』に対する文節係り受け・並列構造アノテーション. 自然言語処理, **25** (4), pp. 331–356.
- 浅原正幸, 大村舞 (2016). BCCWJ-DepParaPAS: 『現代日本語書き言葉均衡コーパス』係り受け・並列構造と述語項構造・共参照アノテーションの重ね合わせと可視化. 言語処理学会第 22 回年次大会発表論文集, pp. 489–492.
- Chvany, C. V. (1973). “Notes on ‘Root’ and ‘Structure-preserving’ in Russian.” In Corum, C., Smith-Stark, T., and Weiser, A. (Eds.), *You Take the High Node and I Will Take the Low Node*, pp. 252–290. Chicago Linguistic Society, Chicago, IL.
- Dixon, R. M. W. (1979). “Ergativity.” *Language*, pp. 59–138.
- Erteschik-Shir, N. (1997). *The Dynamics of Focus Structure*. Cambridge University Press.
- Erteschik-Shir, N. (2007). *Information Structure: The Syntax-Discourse Interface*. Oxford University Press.
- Erteschik-Shir, N. (2014). “Information Structure and (in) Definiteness.” In *Crosslinguistic Studies on Noun Phrase Structure and Reference*, pp. 23–51. BRILL.
- Firbas, J. (1971). “On the Concept of Communicative Dynamism in the Theory of Functional Sentence Perspective.” In *Sborník Prací Filozofické Fakulty Brněnské Univerzity A19*, pp. 135–144.
- Firbas, J. (1979). “A Functional View of Ordo Naturalis.” *Brno Studies in English*, **13**, pp. 29–59.
- Givón, T. (1976). “Topic, Pronoun, and Grammatical Agreement.” In Li, C. N. (Ed.), *Subject and Topic*, pp. 149–187. Academic Press, New York.
- Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., Schwarz, A., Skopeteas, S., and Stoel, R. (2007). “Information Structure.” In Dipper, S., Götze, M., and Skopeteas, S. (Eds.), *Information Structure in Cross-linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*, Vol. 7, pp. 147–187. Universitätsverlag Potsdam.

- Hajičová, E., Panevová, J., and Sgall, P. (2000). “A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank.” Tech. rep., ÚFAL/CKL. (TR-2000-09).
- Heim, I. (2011). “Definiteness and Indefiniteness.” In von Stechow, K., Maienborn, C., and Portner, P. (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Vol. 2, pp. 996–1025. Mouton de Gruyter.
- Hinterwimmer, S. (2011). “Information Structure and Truth-Conditional Semantics.” In von Stechow, K., Maienborn, C., and Portner, P. (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Vol. 2, pp. 1875–1908. Mouton de Gruyter.
- Hoji, H. (1985). *Logical Form Constraints and Configurational Structures in Japanese*. Ph.D. thesis, University of Washington.
- Ionin, T., Ko, H., and Wexler, K. (2004). “Article Semantics in L2 Acquisition: The Role of Specificity.” *Language Acquisition*, **12** (1), pp. 3–69.
- Keenan, E. L. (1976). “Towards a Universal Definition of “Subject”.” In Li, C. N. (Ed.), *Subject and Topic*, pp. 303–334. Academic Press, New York.
- Krámský, J. (1972). *The Article and the Concept of Definiteness in Language*. Mouton de Gruyter, Hague.
- Kruijff-Korbayová, I. and Steedman, M. (2003). “Discourse and Information Structure.” *Journal of Logic, Language and Information*, **12** (3), pp. 249–259.
- Kuno, S. (2004). “Empathy and Direct Discourse Perspectives.” In Horn, L. and Ward, G. (Eds.), *The Handbook of Pragmatics*, pp. 315–343. Blackwell, Malden.
- 久野すすむ (1978). 談話の文法. 大修館書店, 東京.
- Kuroda, S.-Y. (1979). “On Japanese Passives.” In Bedell, G., Kobayashi, E., and Muraki, M. (Eds.), *Explorations in Linguistics: Papers in Honor of Kazuko Inoue*, pp. 305–347. Kenkyusha, Tokyo.
- Landis, J. R. and Koch, G. G. (1977). “The Measurement of Observer Agreement for Categorical Data.” *Biometrics*, **33**, pp. 159–174.
- Lyons, C. (1999). *Definiteness*. Cambridge University Press, Cambridge.
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., and Den, Y. (2014). “Balanced Corpus of Contemporary Written Japanese.” *Language Resources and Evaluation*, **48** (2), pp. 345–371.
- Matsuoka, M. (2003). “Two Types of Ditransitive Constructions in Japanese.” *Journal of East Asian Linguistics*, **12**, pp. 171–203.
- Milsark, G. (1977). “Toward an Explanation of Certain Peculiarities of The Existential Construction in English.” *Linguistic analysis*, **3**, pp. 1–29.

- Miyagawa, S. (1997). “Against Optional Scrambling.” *Linguistic Inquiry*, **28**, pp. 1–25.
- Miyauchi, T. and Asahara, M. (2019). “Statistical Approaches to a Correlation between Information Structure and Word Orders of Noun Phrases in Japanese.” In *Proceedings of the 16th International Conference of the Pacific Association for Computational Linguistics*. (#58).
- Miyauchi, T., Asahara, M., Nakagawa, N., and Kato, S. (2018). “Information-Structure Annotation of the “Balanced Corpus of Contemporary Written Japanese”.” In Hasida, K. and Pa, W. P. (Eds.), *Computational Linguistics*, Vol. 781 of *Communications in Computer and Information Science*, pp. 155–165, Singapore. Springer.
- 宮内拓也, 浅原正幸, 中川奈津子, 加藤祥 (2018). 『現代日本語書き言葉均衡コーパス』への情報構造アノテーションとその分析. 国立国語研究所論集, **16**, pp. 19–33.
- Nakagawa, N. (2016). *Information Structure in Spoken Japanese: Particles, Word Order, and Intonation*. Ph.D. thesis, Kyoto University.
- Orita, N. (2017). “Predicting Japanese Scrambling in the Wild.” In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2017)*, pp. 41–45, Valencia, Spain. Association for Computational Linguistics.
- 乙武北斗, 永田亮 (2016). 冠詞推定のための情報構造仮説の検討. 言語処理学会第 22 回年次大会発表論文集, pp. 493–496.
- Prince, E. F. (1992). *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, Chap. The ZPG letter: Subjects, Definiteness, and Information Status, pp. 295–325. Benjamins, Amsterdam.
- 笹野遼平, 奥村学 (2017). 大規模コーパスに基づく日本語二重目的語構文の基本語順の分析. 自然言語処理, **24** (5), pp. 687–703.
- Silverstein, M. (1976). “Hierarchy of Features and Ergativity.” In Dixon, R. M. W. (Ed.), *Grammatical Categories in Australian Languages*, pp. 112–171. Australian National University, Canberra.
- Sorensen, T., Hohenstein, S., and Vasishth, S. (2016). “Bayesian Linear Mixed Models using Stan: A Tutorial for Psychologists, Linguists, and Cognitive Scientists.” *The Quantitative Methods for Psychology*, **12** (3), pp. 175–200.
- Stan Development Team (2018). “RStan: The R interface to Stan.” R package version 2.18.2.
- Szwedek, A. (1974). “A Note on the Relation between the Article in English and Word Order in Polish.” In Jacek, F. (Ed.), *Papers and Studies in Contrastive Linguistics*, Vol. 2, pp. 213–225. Adam Mickiewicz University Press, Poznań.
- 竹内裕己, 河合敦夫, 細田直見, 永田亮 (2013). 前方文脈を考慮した冠詞の推定. 言語処理学会第 19 回年次大会発表論文集, pp. 717–720.

- Tanaka, J. (2013). “A Multivariate Analysis of L2 English Article Use by Article-less L1 Learners.” In Voss, E., Tai, S.-J. D., and Li, Z. (Eds.), *Selected Proceedings of the 2011 Second Language Research Forum*, pp. 139–147, Somerville, MA. Cascadilla Press.
- 角田太作 (1991). 世界の言語と日本語. くろしお出版, 東京.
- 植田禎子, 飯田龍, 浅原正幸, 松本裕治, 徳永健伸 (2015). 『現代日本語書き言葉均衡コーパス』に対する述語項構造・共参照関係アノテーション. 第8回コーパス日本語学ワークショップ予稿集, pp. 205–214.
- 山口光 (1985). 存在文と所有文. 金田一春彦, 林大, 柴田武 (編), 日本語大辞典, pp. 198–200. 大修館書店, 東京.
- Yamashita, H. and Kondo, T. (2011). “Linguistic Constraints and Long-before-short Tendency.” In *IEICE Technical report (TL)*, No. TL2011-19, pp. 61–65.
- Zubin, D. A. (1979). “Discourse Function of Morphology: The Focus System in German.” In Givón, T. (Ed.), *Syntax and Semantics: Discourse and Syntax*, pp. 409–504. Academic Press, New York.

略歴

宮内 拓也：2019年東京外国語大学総合国際学研究所博士後期課程単位取得満期退学。国立国語研究所非常勤研究員，日本学術振興会特別研究員等を経て，2019年より東京大学大学院教育学研究科特任助教。修士（言語学）。言語処理学会，日本言語学会，Slavic Linguistics Society 各会員。

浅原 正幸：2003年奈良先端科学技術大学院大学情報科学研究博士後期課程修了。2004年より同大学助教。2012年より人間文化研究機構国立国語研究所コーパス開発センター特任准教授。2019年より同教授。博士（工学）。言語処理学会，日本言語学会，日本語学会各会員。

(2019年11月1日 受付)

(2020年2月8日 再受付)

(2020年3月12日 採録)