

# 『現代日本語書き言葉均衡コーパス』への情報構造 アノテーションの分析

著者	宮内 拓也, 浅原 正幸, 中川 奈津子, 加藤 祥
雑誌名	言語資源活用ワークショップ発表論文集
巻	1
ページ	404-415
発行年	2017
URL	<a href="http://doi.org/10.15084/00001495">http://doi.org/10.15084/00001495</a>

## 『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の分析

宮内 拓也 (国立国語研究所 コーパス開発センター / 東京外国語大学大学院) \*

浅原 正幸 (国立国語研究所 コーパス開発センター)

中川 奈津子 (日本学術振興会 特別研究員 PD / 千葉大学大学院)

加藤 祥 (国立国語研究所 コーパス開発センター)

## Analysis of Annotation of Information Structure on “The Balanced Corpus of Contemporary Written Japanese”

Takuya Miyauchi (National Institute for Japanese Language and Linguistics  
/ Tokyo University of Foreign Studies)

Masayuki Asahara (National Institute for Japanese Language and Linguistics)

Natsuko Nakagawa (Japan Society for the Promotion of Science / Chiba University)

Sachi Kato (National Institute for Japanese Language and Linguistics)

### 要旨

日本語は冠詞のない言語である。ゆえに、日本語から冠詞を持つ言語への翻訳の際には、人によるものでも、機械によるものでも冠詞選択の問題を引き起こすことになる。冠詞選択には、ソースとなる言語における名詞句の情報構造(定性, 特定性など)が影響を与える。本稿では、翻訳における冠詞選択の問題を軽減させるため、『現代日本語書き言葉均衡コーパス』のテキスト(新聞(PN)コアデータ 16 サンプル)内の名詞句に対し、情報構造に関係する文法情報のタグをアノテーションした結果を報告する。特に、本稿ではそのアノテーションの概要と基礎統計について述べる。

### 1. はじめに

冠詞がない言語を母語とする者にとって、冠詞がある言語を習得する際の冠詞選択は難しいものである (Ionin et al. 2004, Tanaka 2013 など)。冠詞選択は、一般に定性 (definiteness) や特定性 (specificity) などの情報構造が大きな影響を与えられられる。英語のように定性により定冠詞と不定冠詞を使い分ける言語もあれば、サモア語のように特定性により冠詞を使い分ける言語もある (Mosel and Hovdhaugen 1992)。さらには、コヴェ語<sup>(1)</sup>のように定性と特定性が共に冠詞選択に影響を与える言語もある (Sato 2013)。

言語処理の分野では英語母語話者が産出した大量のテキストから、英語学習者の冠詞の誤り

---

\* t-miyauchi @ ninjal.ac.jp

(1) コヴェ語は、パプアニューギニアのニューブリテン島で話されるオーストロネシア語族の言語のひとつである。

を検出する手法が提案されている (Nagata et al. 2005). しかし, 日本語母語話者が産出する他言語の冠詞選択を検討する場合, 日本語における名詞句の情報構造を考慮する必要がある. さらに機械翻訳において日本語文を冠詞のある言語に翻訳する際にも, 日本語の情報構造が問題となってくる.

本稿では, 機械翻訳での冠詞選択の問題に関する基礎研究として, 『現代日本語書き言葉均衡コーパス』(Maekawa et al. 2014; 以下, BCCWJ) のテキスト内の名詞句に対して情報構造に関わる文法情報のアノテーションを行った結果を報告する.

日本語の情報構造に関する過去のアノテーションは, 主としてテキスト中に出現する情報が談話中に既出であること(情報状態 (information status)) を共参照情報として付与するものであった. 本研究は, より汎用性を求めるため, 情報状態, 定性, 特定性のみならず, 情報構造関係の様々な項目で名詞句にタグを付与した<sup>(2)</sup>.

## 2. 関連研究

情報構造のアノテーションには, 当該の言語形式の情報構造をどのように決定するかという点で, 二つのタイプがある.

まず初めに, 当該の言語形式に基づいて情報構造を決める研究がある. 例えば, Calhoun et al. (2005) は, Vallduví and Vilkkuna (1998) や Steedman (2000) に言及し, 韻律を採用した.  $L+H^*LH\%$ <sup>(3)</sup> の韻律を持つ形式はテーマ (theme)<sup>(4)</sup> となり,  $H^*L$ ,  $H^*LL\%$ <sup>(5)</sup> の韻律を持つものはレーマ (rheme)<sup>(6)</sup> となる. そして, 当該の NP が以前に言及されたか否か, またそれが以前述べられた個体から言及可能か否かという点をもとに情報構造をアノテーションした. Hajičová et al. (2000) は語順を用いた情報構造のアノテーションを提案した. この研究は情報構造についてプラハ学派の伝統に触発されたものであり, それゆえに動詞より左にある言語形式をトピックとする. これらのアノテーション基準は言語依存であり, 日本語に適用可能なものではない<sup>(7)</sup>.

二つ目のタイプの研究では, 言語学的なテストを採用する. Götze et al. (2007) は言語に依存せず, かつ特定の言語理論にもよらず, 情報状態<sup>(8)</sup> とトピック<sup>(9)</sup>, フォーカス<sup>(10)</sup> をアノ

(2) ラベルの詳細については 3. で述べる.

(3)  $L+H^*LH\%$  の韻律では, 低めに上昇した後, 発話境界の上昇調が現れる (Steedman 2000:645-655).

(4) 概してトピック (topic) に対応する.

(5)  $H^*L$  の韻律では, 急速な高いピッチから始まりそして下がる.  $H^*LL\%$  の韻律は,  $H^*L$  に発話末の低い音調が伴ったものである (Steedman 2000:645-655).

(6) 概してフォーカス (focus) に対応する.

(7) なお, Calhoun et al. (2005) は英語を, Hajičová et al. (2000) はチェコ語を対象にしている.

(8) 情報状態は, 旧情報 (given)/補完可能 (accessible)/新情報 (new) の三つが区別される. ある表現が, それ以前の談話で明示的に言及されてる先行詞があれば旧情報であり, 関連のあるものが言及されていれば補完可能であり, 言及されていなければ新情報である (Götze et al. 2007:§3.2).

(9) トピックについては, アバウトネストピック (aboutness topic)/フレームセッティングトピック (frame setting topic) の二つが区別される. アバウトネストピックは (1) で説明する. フレームセッティングトピックとは, そのトピックを含む文の述語が解釈されるべきフレームであり, 典型的には時間や場所の表現がこれにあたる (Götze et al. 2007:§4.2.3).

(10) フォーカスに関しては, 新情報フォーカス (new-information focus)/対比フォーカス (contrastive focus) の二つが区別される. 新情報フォーカスとは, 談話を進めるために欠けている新しい情報を提供する文の要素であ

テーションするための基準を策定した。例えば、アバウトネストピックは以下(1)の手続きで決定される。

(1) An NP X is the aboutness topic of a sentence S containing X if

「ある名詞句 X が、これを含む文 S のアバウトネストピックであるのは以下のときである:」

a. S would be a natural continuation to the announcement *Let me tell you something about X*

「『X について話させて』という予告の後に、S が自然に続き得るとき。」

b. S would be a good answer to the question *What about X?*

「S が、『X についてはどう?』という質問にふさわしい答えであるとき。」

c. S could be naturally transformed into the sentence *Concerning X, S'*, where S' differs from S only insofar as X has been replaced by a suitable pronoun.

「S が『X に関しては、S'』に自然に変形できるとき。ただし、S' は S における X が適切な代名詞に置き換わっている点のみで異なる。」

(Götze et al. 2007:165)

本研究は Götze et al. (2007) に沿うものであるが、いくつかの点で彼らの研究とは大きく異なっている。まず、本研究ではトピックとフォーカスを直接アノテーションしない。これはトピックやフォーカスがそれぞれに多次元であるためである (Nakagawa 2016)。実際に Götze et al. (2007:163) では、例えば、指示的 (referential) な NP、特定 (specific) 解釈や総称 (generic) 解釈を持つ不定 (indefinite) の NP など様々な種類のアバウトネストピックが区別されている。定性や特定性のような要因はトピックとは独立であると考え、そのようにアノテーションする方がより単純である。第2にトピックやフォーカスと相関すると知られている要因については、定性や特定性以外にも例えば、有生性 (animacy) や動作主性 (agentivity) (Givón 1976, Keenan 1976) など、より多くある。そのため、本研究では情報構造アノテーションの一環としてこれらについてもアノテーションすることとする。

### 3. タグとアノテーション基準

BCCWJ では、長単位と短単位という二つの単位が採用されているが、本研究では、短単位の名詞をアノテーション対象とする。ただし、複合語については、前部要素には指示性 (referentiality) がないということなどを考慮して、前部要素まで含めて一つの名詞と捉える<sup>(11)</sup>。

以下の (2) で示す項目についてラベルを設定した。

る。対比フォーカスとは、他の発話に対して対比を呼び起こす文の要素である (Götze et al. 2007:§5.2)。

(11) つまり、短単位では2語以上の扱いを受けるものに関しても、1語扱いになっている場合があることになる。これは BCCWJ への共参照アノテーションと同様の方針である。

- (2) a. 情報状態 (information status)  
 b. 定性 (definiteness)  
 c. 特定性 (specificity)  
 d. 有生性 (animacy)  
 e. 有情性 (sentience)  
 f. 動作主性 (agentivity)  
 g. 共有性 (commonness)

以下、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。

### 3.1 情報状態

(2a) の情報状態とは、いわゆる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 (discourse-new)」となり、聞き手が知っている情報は「旧情報 (discourse-old)」となる<sup>(12)</sup>。一つのテキスト全体を一つの談話と見なし<sup>(13)</sup>、アノテーションを行った。

- (3) a. 担任だった池田弘子先生は違った。  
 b. スクールカウンセラーでもあった先生の授業は

(読売新聞 [ BCCWJ: PN1c.00001 ])

(3a) の下線部の名詞「池田弘子」はこのテキストで初出の名詞であるため、新情報タグが付与される。一方、(3b) の下線部の名詞「先生」は(3a) の「池田弘子」を指示しているため旧情報タグが付与される。これらの名詞は共参照関係にある。

### 3.2 定性, 特定性

(2b) の定性とは、指示対象を聞き手が同定できるか否かを示すカテゴリーである<sup>(14)</sup>。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 (definite)」であり、同定できないと想定している場合は「不定 (indefinite)」である。本研究では、スコープとして前後3文を見ることとする。

- (4) 「そんな薄い a. かばん じゃ b. 遊び道具 も入らないよ」

(読売新聞 [ BCCWJ: PN1c.00001 ])

(4) の下線部 a. の名詞「かばん」はスコープである(4)の前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。話し手は当然この「かばん」は聞き手により同定しうると考えていると考えられるため、定のタグが与えられる。(4)の下線部 b. の名詞「遊び道具」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ため、不定のタグが付与される。

<sup>(12)</sup> 情報構造自体についての詳細は Kruijff-Korbayová and Steedman (2003) や Hinterwimmer (2011) などを参照のこと。

<sup>(13)</sup> 本研究では、BCCWJ の新聞 (PN) のデータを用いたため、一つの記事が一つの談話であると見なしている。

<sup>(14)</sup> 定性そのものについては、Lyons (1999), Heim (2011) などを参照のこと。

(2c) の特定性は、定性と少々似た概念であるが、話し手が特定の事物を想定しているか否かを示す意味論的カテゴリーである<sup>(15)</sup>。話し手が特定の事物を想定しているならば「特定 (specific)」となり、想定していなければ「不特定 (unspecific)」となる。定性と同様、特定性に関してもスコープとして前後3文を見ることとする。

(5) 行き場を失ったa. 廃タイヤがあぜ道やb. 納屋の横に放置されてきた。

(北海道新聞 [ BCCWJ: PN2e.00001 ])

(5) の下線部 *a.* の名詞「廃タイヤ」は、北海道鷹栖町に放置された約 30,000 本のタイヤを具体的に指しており、これは (5) の前後3文から読み取ることが可能であるため特定のタグが付与される。(5) の下線部 *b.* の名詞「納屋」は特定の納屋が想定されているわけではなく、不特定のタグが与えられる。

### 3.3 有生性、有情性

(2d) の有生性とは、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 (animate)」であり、無生物（植物を含む）は「無生 (inanimate)」である。有生性は名詞句レベルのみで判断し、付与されるものとする。有生性と似た概念として (2e) の有情性がある。これは、情意があるか否かを示すパラメーターである。自由意志による移動が可能な場合は「有情 (sentient)」となり、自由意志による移動はないなら「非情 (insentient)」となる。日本語については、有生/無性の区別よりも有情/無性の区別が重要であるとする先行研究もあり (三上 1953, 山口 1985 など)、また、有生性と有情性の値が異なる場合もあり得ることから、このパラメーターの設定が必要となる<sup>(16)</sup>。情意の有無は名詞句単体では判定できない場合があるため、有情性は述語-項レベルまで見た上で判断し、付与されるものとする。

(6) オオクチバスなどのa. 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b. 湖沼に侵入し、

(読売新聞 [ BCCWJ: PN4c.00001 ])

(6) の下線部 *a.* の名詞「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。また、ブラックバスに情意があるか否かは判断が難しいが、その述語は「侵入する」となっており、これは意志的な動作、行為を表しているため、ここでの「ブラックバス」は有情のタグが付与されることになる。(6) の下線部 *b.* の名詞「湖沼」は無生物であり、情意もないと判断されるため、それぞれ、無生、無情のタグが与えられる。

### 3.4 動作主性

(2f) の動作主性は、事態に関わる人がその事態で果たしている役割を示す。行為を意図的に実現するものは「動作主 (agent)」とし、行為によって変化を被るものを「被動作主 (patient / theme)」とする。このパラメーターについては節レベルまで見て判断し、タグを付与すること

<sup>(15)</sup> 特定性そのものに関しては、Heusinger (2011) などを参照のこと。

<sup>(16)</sup> 例えば、「ゾンビ」と「幽霊」は有生性と有情性の値の差により区別できる。ゾンビは腐敗した人体が自発的意思なく徘徊するため、有生、無情である。一方、幽霊とは死者の魂が未練や遺恨により現れたものであるため、無生、有情となる。

とする。その際、主節と従属節の両方を考慮する。また、「どちらでもよい」「どちらでもない」を許す。

- (7) a. 編み笠をかぶった人なつっこい笑顔を見るだけで、  
 b. もみじの木にとまって仲良く寄り添う二羽のキジバト。  
 c. 独特な雰囲気の写真になりました。

(産経新聞 [ BCCWJ: PN1d.00001 ])

(7a) の下線部の名詞「笑顔」は、主節では被動作主であり従属節では動作主である。このような場合に「どちらでもよい」というタグを付与する。(7b) の下線部の名詞「キジバト」は、それを含む文がこの名詞で終わる体言止めの文であるため、主節では動作主性の判断ができないが、従属節では動作主であるため、「動作主」というタグを付与する。(7c) の下線部の名詞「写真」は動作主でも被動作主でもないため、「どちらでもない」となる。

### 3.5 共有性

(2g) の共有性は、情報を聞き手が既に知っているか話し手が想定しているか否かを示すパラメーターである。聞き手が既に知っているか話し手が想定している情報は「共有 (hearer-old)」であり、知らないか想定している情報は「非共有 (hearer-new)」である。なお、この判断の際はアノテータの世界知識 (world knowledge) を使ってもよいこととし、「想定可能」というラベルも許す。このラベルは、ブリッジング (bridging) を起こしている際に付与される。

- (8) a. キャンティ街道を抜け、b. オリーブ畑に囲まれた田園地帯のc. レストランで、

(読売新聞 [ BCCWJ: PN4c.00001 ])

(8) の下線部 a. の名詞「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。(8) の下線部 b. の名詞「オリーブ畑」は当該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る。(8) の下線部 c. の名詞「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る。

### 3.6 その他

固有名詞については、アノテーションの際、有名の度合いを考慮してよいこととし、アノテータの持つ世界知識を参照してもよいとする。(9a) の形式名詞や (9b) のような慣用表現は対象から外し、それぞれ「形式名詞」「慣用表現」タグを付与する。なお、慣用表現であるか否かについてはアノテータによる揺れを許すこととする。

- (9) a. 様々な人がいるということが  
 b. 聞く耳を持たせてくれるんです。

(読売新聞 [ BCCWJ: PN1c.00001 ])

## 4. 基礎統計

対象は BCCWJ の新聞 (PN) コアデータ 16 サンプルに出現する名詞 2023 件とした。サンプルの選択は BCCWJ-ANNOTATION-ORDER<sup>(17)</sup> に基づく。作業者は BCCWJ-DepParaPAS (植田ほか 2015, 浅原・大村 2016) に付与された共参照情報を確認しながら作業を行う。定性, 特定性, 有生性, 有情性, 動作主性<sup>(18)</sup>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」というタグを認めた。特定性, 動作主性, 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」というタグを認めた。

表 1 にタグの基礎統計を示す。情報状態のラベルは以前アノテーションされた共参照情報に基づいているが, 情報状態の分布と他のラベルの分布は異なっている。よって, この差異より, 他のラベルは日本語からの翻訳の際の冠詞選択に影響を与えらる。

表 1 タグの基礎統計

情報状態	新情報	旧情報	-	-
	1345	678	-	-
定性	定	不定	どちらでもよい	-
	1122	899	2	-
特定性	特定	不特定	どちらでもよい	どちらでもない
	1157	749	116	1
有生性	有生	無生	どちらでもよい	-
	342	1680	1	-
有情性	有情	無情	どちらでもよい	-
	337	1678	8	-
動作主性	動作主	被動作主	どちらでもよい	どちらでもない
	192	338	2	1491
共有性	共有	非共有	想定可能	どちらでもない
	1036	494	489	4

表 2 に情報状態と定性の分割表を示す。不定のラベルは新情報のラベルと共に現れることが多いが, 興味深いことに定のラベルは新情報, 旧情報のどちらのラベルとも現れうるとい傾向がある。これは共参照情報の冠詞選択への貢献が限定的であることを示しているといえる。表 3 は情報状態と特定性の分割表である。これについても情報状態と定性のものと同様の分布を示している。

<sup>(17)</sup> BCCWJ コアデータサンプルにおけるアノテーション優先順序である。以下参照のこと。 <https://github.com/masayu-a/BCCWJ-ANNOTATION-ORDER>

<sup>(18)</sup> ただし, 先に見たように, 動作主性については, 「どちらでもよい」のタグは主節から見た場合と従属節から見た場合で動作主性の値が異なる場合に付与される。



表2 情報状態と定性

	新情報	旧情報
定	497	625
不定	846	53
どちらでもよい	2	0

表3 情報状態と特定性

	新情報	旧情報
特定	531	626
不特定	705	44
どちらでもよい	108	8
どちらでもない	1	0

次に名詞句が含まれる文節の付属語主辞<sup>(19)</sup>と各タグの分布について確認する。付属語主辞は係り受け解析器 CaboCha (Kudo and Matsumoto 2002, 工藤・松本 2002) に含まれる UniDic 主辞規則に基づくものである。付属語主辞と情報状態の分布を表4に示す。付属語主辞が「、」や「。」となっているものは、格表示されない連体止めの表現を表している。付属語主辞と情報状態の関係を見ると、「が」「を」「に」は新情報が多く、「は」は旧情報が多い傾向にあることがわかる。

表4 付属語主辞と情報状態

付属語主辞	新情報	旧情報
、	284	168
。	130	82
が	99	47
は	59	62
も	29	10
を	222	53
に	84	28
の	108	86
で	40	13
と	13	12

付属語主辞と定性の分布を表5に、付属語主辞と特定性の分布表6に示す。表5、表6から、「は」「の」に定性・特定性のものが多い一方、「を」に不定性・不特定性のものが多い傾向にあることがわかる。

## 5. おわりに: まとめと今後の課題

本稿では、BCCWJ に対する情報構造のアノテーションデータについて紹介した。本研究では日本語の名詞句に対し、七つの情報構造に関する概念を導入した。これらのアノテーションラベルの分布は、日本語母語話者の冠詞選択や冠詞誤りの修正に対し共参照情報だけでは十分でないことを示しており、新たに導入されたラベルは翻訳における冠詞選択に役立つと思わ

<sup>(19)</sup> 付属語主辞とは、付属語のうち主要な要素を意味する。

表5 付属語主辞と定性

付属語主辞	定	不定
、	268	184
。	122	90
が	74	72
は	79	42
も	10	29
を	89	186
に	43	69
の	122	72
で	35	19
と	13	12

表6 付属語主辞と特定性

付属語主辞	特定	不特定
、	270	171
。	126	77
が	74	55
は	82	33
も	10	25
を	96	155
に	46	54
の	123	61
で	36	12
と	14	8

れる。

今後の課題は以下に示すとおりである。

まず、情報構造をアノテーションする被験者実験を行う。本稿で示したラベルは言語学者によってアノテーションされたものである。被験者実験のためには、言語学の専門知識を持たない人でも回答が可能のように、わかりやすい質問に落とし込む必要がある。非言語学者により情報構造のラベルを判定する質問を作成し、非言語学者にもわかるような言語学的なテストを設計する。それにより、アノテーションの数だけでなく、標的サンプルをも増やすことができる。

第2に、ラベルの再検討も行っていくことを考える。名詞の飽和/非飽和(西山 2003)や Löbner (1985, 2011) の名詞の4分類 (Sortal タイプ  $\langle e, t \rangle$  / Individual タイプ  $e$  / Relational タイプ  $\langle e, \langle e, t \rangle \rangle$  / Functional タイプ  $\langle e, e \rangle$ ) を付与することで、名詞の指示性と他のラベルの関係性を解明する手掛かりになると考えられる。さらに、動作主性を動作主/被動作主のみでなく、もう少し詳細に意味役割として、受益者 (beneficiary), 経験者 (experiencer), 着点 (goal), 道具 (instrument), 場所 (location) などを付与することで、意味役割と他のラベルの関係性を突き止めることも考えたい。

第3に、機械学習に基づきシソーラスを用いて情報構造の評価モデルを開発する。それにより、日本語テキストに基づき、機械翻訳による冠詞選択の評価を行う。

最後に、本稿で示した情報構造アノテーションを視線計測のデータ (Asahara et al. 2016) と重ね合わせる。これにより、情報構造に対する読み時間のふるまいについて調査を行い (浅原 2017), 情報構造と視線の関係をより深く解明していくことを目指す。

## 謝 辞

本研究は JSPS 科研費 (課題番号: 25284083, 研究代表者: 浅原正幸) の助成を受けている。

## 文 献

- Tania Ionin, Heejeong Ko, and Kenneth Wexler (2004). “Article semantics in L2 acquisition: The role of specificity.” *Language Acquisition*, 12:1, pp. 3–69.
- Junko Tanaka (2013). “A Multivariate Analysis of L2 English Article Use by Articleless L1 Learners.” *Selected Proceedings of the 2011 Second Language Research Forum*, pp. 139–147. Somerville, MA: Cascadilla Press.
- Ulike Mosel, and Even Hovdhaugen (1992). *Samoan Reference Grammar*. Oslo: Scandinavian University Press.
- Hiroko Sato (2013). “Definiteness and specificity in Kove.” *International workshop on information structure of Austronesian languages*, pp. 37–45. Tokyo: The Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Ryo Nagata, Tatsuya Iguchi, Fumito Masui, Atsuo Kawai, and Isu Naoki (2005). “A Statistical Model based on the Three Head Words for Detecting Article Errors.” *IEICE TRANSACTIONS on Information and Systems*, E88-D:7, pp. 1700–1706.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014). “Balanced corpus of contemporary written Japanese.” *Language Resources and Evaluation*, 48:2, pp. 345–371.
- Sasha Calhoun, Malvina Nissim, Mark Steedman, and Jason Brenier (2005). “A framework for annotating information structure in discourse.” *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pp. 45–52. Ann Arbor: The Association for Computational Linguistics.
- Enric Vallduví, and M Vilkuña (1998). “On rheme and kontrast.” P. W. Culicover, and L. McNally (Eds.), *The Limits of Syntax*. San Diego: Academic Press. pp. 79–108.
- Mark Steedman (2000). “Information structure and the syntax-phonology interface.” *Linguistic Inquiry*, 34, pp. 649–689.
- Eva Hajičová, Jarmila Panevová, and Petr Sgall (2000). “A manual for tectogrammatical tagging of the Prague Dependency Treebank.” Technical report, ÚFAL/CKL. (TR-2000-09)
- Michael Götze, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas, and Ruben Stoel (2007). “Information structure.” Stefanie Dipper, Michael Götze, and Stavros Skopeteas (Eds.), *Information structure in cross-linguistic corpora: annotation guidelines for phonology, morphology, syntax, semantics and information structure*. Vol. 7.: Universitätsverlag Potsdam. pp. 147–187.
- Natsuko Nakagawa (2016). “Information structure in spoken Japanese: Particles, word

- order, and intonation.” Unpublished doctoral dissertation, Kyoto University.
- Talmy Givón (1976). “Topic, pronoun, and grammatical agreement.” Charles N. Li (Ed.), *Subject and Topic*. New York: Academic Press. pp. 149–187.
- Edward L. Keenan (1976). “Towards a universal definition of “subject”.” Charles N. Li (Ed.), *Subject and Topic*. New York: Academic Press. pp. 303–334.
- Ivana Kruijff-Korbayová, and Mark Steedman (2003). “Discourse and Information Structure.” *Journal of Logic, Language and Information*, 12:3, pp. 249–259.
- Stefan Hinterwimmer (2011). “Information structure and truth-conditional semantics.” Klaus von Heusinger, Claudia Maienborn, and Paul Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*. Vol. 2.: Mouton de Gruyter. pp. 1875–1908.
- Christopher Lyons (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Irene Heim (2011). “Definiteness and indefiniteness.” Klaus von Heusinger, Claudia Maienborn, and Paul Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*. Vol. 2.: Mouton de Gruyter. pp. 996–1025.
- Klaus von Heusinger (2011). “Specificity.” Klaus von Heusinger, Claudia Maienborn, and Paul Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*. Vol. 2.: Mouton de Gruyter. pp. 1058–1087.
- 三上章 (1953). 『現代語法序説』 刀江書院, 東京.
- 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』 大修館書店, 東京 pp. 198–200.
- 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.
- 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.
- Taku Kudo, and Yuji Matsumoto (2002). “Japanese Dependency Analysis using Cascaded Chunking.” *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pp. 63–69.
- 工藤拓・松本裕治 (2002). 「チャンキングの段階適用による日本語係り受け解析」 情報処理学会論文誌, 43:6, pp. 1834–1842.
- 西山佑司 (2003). 『日本語名詞句の意味論と語用論: 指示的名詞句と非指示的名詞句』 ひつじ書房, 東京.
- Sebastian Löbner (1985). “Definites.” *Journal of Semantics*, 4, pp. 279–326.
- Sebastian Löbner (2011). “Concept types and determination.” *Journal of Semantics*, 28, pp. 279–333.
- Masayuki Asahara, Hajime Ono, and Edson T. Miyamoto (2016). “Reading-Time An-

notations for Balanced Corpus of Contemporary Written Japanese.” *Proceedings of COLING-2016*, pp. 684–694.

浅原正幸 (2017). 「読み時間と情報構造について (ちょっとながめ)」 言語資源活用ワークショップ 2016 発表論文集. M.s.