

# 修辞ユニット分析における脱文脈化指数の妥当性の 検証

著者	浅原 正幸, 田中 弥生
雑誌名	国立国語研究所論集
号	15
ページ	1-15
発行年	2018-07
URL	<a href="http://doi.org/10.15084/00001593">http://doi.org/10.15084/00001593</a>

## 修辞ユニット分析における脱文脈化指数の妥当性の検証

浅原正幸<sup>a</sup> 田中弥生<sup>b</sup>

<sup>a</sup>国立国語研究所 コーパス開発センター

<sup>b</sup>東京大学大学院 博士課程

### 要旨

本稿では修辞ユニット分析における脱文脈化指数の妥当性について検証する。修辞ユニット分析は、節に相当するメッセージが命題か提言か（発話機能）・主語が空間中どこにあるか（中核要素）・事象が時間軸上どこにあるか（現象定位）を認定して、その組み合わせに基づき「修辞機能」と「脱文脈化指数」を同定するものである。順序尺度である脱文脈化指数は経験的に決められたものではなく恣意的なものであった。この脱文脈化指数が一般にどの程度共有されるものなのかを確認するために、Yahoo! クラウドソーシングを用いた一対比較による被験者実験を行った。一対比較のデータを Bradley-Terry 法に基づく一般化線形モデルを用いて線形順序に変換し、脱文脈化指数の妥当性を検証した。その結果、おおよそ脱文脈化指数の順序は共有される一方、中核要素や現象定位に対する日本語話者の見方を必ずしも反映していないということがわかった\*。

キーワード：選択体系機能理論、修辞ユニット分析、脱文脈化程度、クラウドソーシング

### 1. はじめに

本稿では修辞ユニット分析 (Rhetorical Unit Analysis: RUA) における脱文脈化指数の妥当性について検証する。修辞ユニット分析は選択体系機能言語理論の談話分析手法の一つ (Cloran 1994) であり、当初は英語における母子会話の分析に用いられた。その後、佐野 (2010a)、佐野・小磯 (2011) により日本語への適用が提案されている。これらの談話分析の枠組では、節相当のメッセージが情報を交換するのか品物や行為を交換するのかを規定する「発話機能」、主語要素に対する空間的な距離を表す「中核要素」、副詞や述部に対する時間的な距離を表す「現象定位」を定義し、この組み合わせにより、「脱文脈化指数」と呼ばれる「いま・ここ・わたしからの距離」に相当する観点を割り当てる。脱文脈化指数は、個人的・発話場面に直接かかわる表現や話題を脱文脈化の程度を低いものとし、汎用的・専門的・一般的・発話場面に直接かかわらない表現や話題を脱文脈化の程度を高いものとするように割り当てる。これにより、テキストの個性・一般性・専門性を明らかにする。

日本語のデータとしては田中 (2013) が掲示板データに対するアノテーションを行って検証しているほか、田中・佐野 (2011) が Yahoo! 知恵袋データに対するアノテーションを行っている。

しかしながら、日本語のアノテーション基準については、基準策定者が英語と日本語の差異に

\* 本稿は、言語資源活用ワークショップ 2017 (2017, 於: 国立国語研究所) での発表「修辞ユニット分析における脱文脈化指数の妥当性の検証」をもとにしている。本研究は JSPS 科研費 15K02535 「「修辞機能」と「脱文脈化程度」の観点からのテキスト分析手法確立と自動化の検討」の助成を受けている。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の共同研究プロジェクト「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」(2016–2021 年度) の成果である。

ついて詳しく認識しておらず、Halliday and Matthiessen (2004) による選択体系機能理論を日本語に対してどう適用するかについても一般の作業者が作業できるレベルに適切に規定できていない。

また、特殊な訓練を受けたものが作業できて「発話機能」「中核要素」「現象定位」が規定できたとして、その組み合わせによって割り当てられる「脱文脈化指数」は一部の研究者の内省により作られたもので、その妥当性が問われている。

そこで本稿では「脱文脈化指数」の妥当性を検証するために、「発話機能」「中核要素」「現象定位」がアノテーションされた、対のメッセージを被験者に提示して「いま・ここ・わたしからの距離」の近さについて選択させる実験を行った。実験結果に基づき、既存の日本語に対する修辞ユニット分析手法の問題点を示す。

2節では修辞ユニット分析の詳細について示す。3節では今回行った妥当性の検証手法と結果について述べる。4節にまとめと問題点にどう対処するかについて示す。

## 2. 修辞ユニット分析

### 2.1 分析の単位とメッセージ

修辞ユニット分析の分析対象は「メッセージ」と呼ばれる、Hallidayの機能文法における階層節に相当する単位とされている。佐野(2010b)によると日本語における単位については、述語に基づいて定義されている。しかしながら、述語については何ら定義されていない。格要素を持つ形容詞(形容動詞・連体詞を含む)の扱いや、名詞述語の扱いについては何ら触れられておらず、一般のアノテータに必要な情報は定義されていない。埋め込み節についても定義がされていない。関係節内の関係を埋め込み節とみなし、関係節外の関係を埋め込み節とみなさないのか、もしくは、両方とも埋め込み節とみなすのかなどについて適切な情報が提供されていない。

この分析の単位であるメッセージは以下の三つに分類される。

#### 1) 位置づけ (positioning)

挨拶・定型句・フィラーなど述部を含まない節のみで構成されるメッセージ。これらは修辞ユニット分析において認定の対象外とされる。

#### 2) 自由 (free)

独立して時制やムードなどを表すメッセージ。

#### 3) 拘束 (bound)

発話中に二つ以上のメッセージが出現した場合に従属する側のほうを「拘束」とする。従属性については論理・意味・時制の観点から他のメッセージに依存するかどうかを判定する。この拘束はさらに以下の二つに分類する：

##### ① 意味的従属

時間・場所・因果関係などの節間の意味を陽に形式として表出するもの

##### ② 形式的従属

並列などにより節間の意味(時間)が表出するもの

『現代日本語書き言葉均衡コーパス』(以下 BCCWJ) (Maekawa et al. 2014) の Yahoo! 知恵袋データから例を示す。(1) は質問に対する回答投稿の一部をメッセージ単位で a から d に分割したものである。

- (1) a. カネボウのエフェクト 38 がいいです。  
 b. お湯で洗えば、  
 c. スルっと落ちるのですが、  
 d. 涙や汗では落ちません。

(BCCWJ サンプル ID (以下同様) OC09\_02727)

a と d のメッセージの種類は「自由」である。b は c 「スルっと落ちるのですが」の条件を示しているため「拘束：意味的従属」で、c は d と並列な従属節であるため「拘束：形式的従属」である。

「位置づけ」については日本語向けに導入されたものである。また、分類の名前づけだが、「拘束：意味的従属」が陽に形式的に節間の意味を表出するものであり、「拘束：形式的従属」が陽に形式的に節間の意味を表出しないもの（並列などにより意味が表出するもの）である点に注意が必要である。

これは基準の定義を行ったものが日本語記述文法による、基本的な節の分類に精通していないことに由来する。英語においては格要素が省略されず、かきまぜもあまりおきないために、埋め込みも含めた節のスコープが明示的に定義できるが、日本語においては適切な基準を立てたうえで整理する必要がある。松本ほか (2017) は、Tori-Bank の節分類に基づいた BCCWJ に対する節の意味分類アノテーションについて言及している。このうちのどのレベルをメッセージとみなすのかについて、きちんと検討する必要がある。

## 2.2 発話機能

発話機能は大きく分けて、命題と提言の二つがある。命題は情報の交換に関するメッセージで、提言は品物や行為の交換に関するメッセージと定義されている。これは、場と時間を共有する会話・対話に対する分類であるが、場と時間を共有しない書き言葉については、基本的には「品物や行為の交換」の規定が難しい。明示的に定義できる品物の交換は書き言葉そのものが書かれている媒体（例：「この手紙に同封しているしおり」）以外、どの範囲まで認めるかを規定する必要性がある。田中・浅原 (2017) は Yahoo! 知恵袋に対する発話機能のアノテーションの一致率について分析を行っているが、基本的には出現しないと考える「提言」の範囲を規定できていないがために、アノテータでゆれがあることがわかっている。ウェブサイト上でのやりとりについて「交換」とは何であるかを再定義することが求められる。

(2) は、質問 (a から e) と回答 (f) の一部である。

- (2) a. みなさんはホントにお湯だけで落としていますか？  
 b. それとも一応メイク落としを使っていますか？  
 c. お湯で落とすとマスカラのかす(?)が目に入って  
 d. 大変なんですけど・・・。  
 e. ちなみにオペラマイラッシュ使用です。  
 f. 全体のクレンジングと一緒に落とします。

(OC09\_02515)

(2) の a と b は情報を要求している(交換している)ので「命題」、c から e は、質問する側が自身の状況を補足説明する意味で情報を提供している(交換している)ので「命題」、f は回答する側が情報を提供している(交換している)ので「命題」である。

(3) も、質問(a から c)と回答(d から h)の一部である。

- (3) a. 敏感肌の人にピッタリという化粧水ありますか？  
 b. 美白もできると良いのですが。  
 c. オススメの化粧水を教えてください。  
 d. ファンケル商品をお勧めします。  
 e. 無添加で防腐剤・保存料不使用なので、  
 f. どんな肌質の方にも安心して使えらと思います。  
 g. 値段も手ごろなので、  
 h. 一度試してください。

(OC09\_00105)

(3) の c は「教えてください」と「教える」という行為を相手に要求し、また、h は「試してください」と「試す」という行為を相手に要求しているとも考えられる。行為を要求している発話の機能は「提言」であるが、これらを「提言」とすべきか、これらの表現は情報を要求したり提供したりしていると考え「命題」とすべきか、判断がゆれる。

### 2.3 中核要素

修辞ユニット分析において「中核要素」はメッセージの中心となる要素のことで、「主語によって表現される」と定義されている。この主語相当の要素が談話状況に出現するか否かにより、空間的な距離感を規定する。分析対象の節が指し示す主語相当の人・事象が空間的にどこに位置するかにより分類を行う。

分類として以下の三つがあり、このうち状況内要素については「参加要素」「非参加要素」に細分類される。例を(4)から(10)に示す。太字部分が中核要素である。

- 1) 状況内要素：メッセージの送り手や受け手がいる場に存在する人・事象  
 ① 参加要素：テキストの伝達に参加している人(送り手・受け手) (例(4)(5))

- ② 非参加要素：テキストの伝達に参加していない人・事象 (例 (6) (7))
- 2) 状況外要素：メッセージの送り手や受け手がある場に存在しない人・事象 (例 (8) (9))
- 3) 定言要素：あるカテゴリやクラスに属するメンバーすべてを対象とする要素 (例 (10))
- (4) わたしはけっこうなくせ毛です。 [状況内：参加] (OC09\_00175)
- (5) みなさんはホントにお湯だけで落としていますか？ [状況内：参加] (OC09\_02515)
- (6) 最近顔のしみがどんどん薄くなってきました。 [状況内：非参加] (OC09\_00109)
- (7) わたしのお薦めは資生堂モアリップです。 [状況内：非参加] (OC09\_00058)
- (8) 友達にも、何もしないのに肌きれいな子がいますが、(以下略)  
[状況外] (OC09\_02701)
- (9) 使い捨てコンタクトの使用期限ってありますか？ [状況外] (OC09\_00184)
- (10) 時間が経つにつれ油分は酸化します。 [定言] (OC09\_02500)

田中・浅原 (2016) は Yahoo! 知恵袋の中核要素のアノテーション一致率について分析を行っているが、アノテータ間でゆれがあることを報告している。

まず場と時間を共有していない書き言葉において「状況」とは何であるのかについて規定しなくてはならず、テキストの伝達に参加しているか否かについても不透明である。日本語においては、主語相当を分析対象にするのか、主題相当を分析対象にするのかをまず規定する必要がある。さらに格交代をどう扱うかについても定義する必要がある。主語・主題相当の情報が省略されている場合には、作業者が要素を補完する必要がある。

植田ほか (2015) は、BCCWJ に対する述語を厳密に規定したうえで、ガ格相当のゼロ代名詞を外界応答も含めて補完している。これらを見ることで作業者の負担を減らせる可能性がある。

「定言要素」については、名詞句の定性もしくは特定性に関する項目であると考えられる。冠詞のない日本語において定性・特定性を規定するのは一般の方には難しく、規定も厳密に記述されていない。宮内ほか (2017) は BCCWJ に対して、定性・特定性を含む情報構造アノテーションについて論じているが、書き言葉における状況内外・参加非参加について、情報構造の観点から定義する必要があると考える。

## 2.4 現象定位

現象定位はメッセージに表現される出来事がメッセージの伝達される時間を基準としていつ起きたのかを示す要素である。元基準は concurrent, prior, future など相対時によって表現されている。細分類も含めて 6 種類の分類が準備されている。

- 1) 現在：メッセージの伝達が行われているときに出来事が起きている場合
  - ① 非習慣的・一時的：出来事が非習慣的・一時的なもの
  - ② 習慣的・恒久的：出来事が習慣的・恒久的なもの
- 2) 過去：メッセージの伝達時点より前に出来事が起きたことを示す場合

- 3) 未来：メッセージの伝達時点より後に出来事が起きることを示す場合  
 ① 意図的：出来事が意図的なもの  
 ② 非意図的：出来事が非意図的なもの  
 4) 仮定：ある条件下においてのみ出来事が起きることを示す場合

(11) から (14) はそれぞれ回答の一部の現象定位を示したものである。

- (11) a. 26才女です。 [現在：非習慣的・一時的]  
 b. 私もずっと一重で、 [現在：習慣的・恒久的]  
 c. 21の時と去年と二回埋没法の二重手術を受けました。 [過去]  
 (OC09\_02467)
- (12) 毎日塗っていればハリが甦ります。 [仮定] (OC09\_02501)
- (13) a. 今はまだ化粧でごまかせるかもしれませんが [現在：非習慣的・一時的]  
 b. きっと限界が来ると思います。 [未来：非意図的]  
 (OC09\_02537)
- (14) a. 顔に髪がかからないようにして、 [未来：意図的]  
 b. できるだけ清潔にするよう洗顔をこまめにして、 [未来：意図的]  
 c. 偏食をせず、 [未来：意図的]  
 d. 睡眠をきちんととるといいと思います。 [未来：意図的]  
 (OC09\_00019)

このうち「意図的」は「意志的」とするほうがわかりやすい訳語であると考えられる。

Cloran (1994) の修辞ユニット分析は英語を対象としており、時制の一致が見られる節間の関係も含めて、時制が文法的に陽に表出する絶対時制を対象としている。一方、日本語は相対時制による過去-非過去の対立のみ表出する言語であり、Cloran (1994) の規定にあるとおりに時制や副詞句によっても適切なタグ付けを設定することは難しい。

書き言葉においては、基準時間を、文書作成日時を現在時とするのか、あるいは、テキストが読まれる時間とするのかについて定義する必要があるが、規定されていない。さらに「仮定」「未来」を想定しているが、出来事の不成立については論じられていない。

小西ほか (2013) は BCCWJ に対して時間表現を実時間に写像する情報を付与することにより正規化する方法を提案し、保田ほか (2013) は時間表現と事象表現もしくは二事象表現間の時間的順序関係をアノテーションすることにより相対時を記号化する方法を提案している。

## 2.5 修辞機能・脱文脈化指数

上に述べた「発話機能」「中核要素」「現象定位」の組み合わせにより修辞機能が割り当てられ、これにより「いま・ここ・わたしからの距離」を順序尺度化した脱文脈化指数が定義される。図 1 に割り当てを示す。図中 [n] 内の数字 n が脱文脈化指数を表す。



	提言	命題					
		現在	過去	未来		仮定	現在
		非習慣的・一時的		意図的	非意図的		習慣的・恒久的
状況内：参加	行動 [1]	実況 [2]	状況内回想 [3]	計画 [4]	状況内予想 [5]	状況内推測 [6]	自己記述 [7]
状況内：非参加				状況内予想 [5]			観測 [8]
状況外		報告 [9]	状況外回想 [10]	予想 (予測) [11]	予想 (予測) [11]	推測 [12]	説明 [13]
定言							一般化 [14]

図1 「発話機能」「中核要素」「現象定位」の組み合わせと修辞機能・脱文脈化指数（日本語）

脱文脈化指数順に並べると以下ようになる。

- [1] 行動：品物・行為の交換に関するメッセージ
- [2] 実況：送り手・受け手がいる場に存在する人・事象の現在の行動や状態について表現したメッセージ
- [3] 状況内回想：送り手・受け手がいる場に存在する人・事象の過去の出来事についての情報を交換するメッセージ
- [4] 計画：送り手・受け手が意図して行おうとしている活動について表現したメッセージ
- [5] 状況内予想：意図しないで行われる送り手・受け手の未来の活動，もしくは，送り手・受け手がいる場に存在する送り手・受け手以外の人の未来の活動を表現したメッセージ
- [6] 状況内推測：送り手・受け手がいる場に存在する人・事象が，特定の条件下において起きる出来事について表現したメッセージ
- [7] 自己記述：送り手・受け手の習慣的・恒久的な行動や状態について表現したメッセージ
- [8] 観測：送り手・受け手がいる場に存在する送り手・受け手以外の人・事象の現在の行動や状態について表現したメッセージ
- [9] 報告：送り手・受け手がいる場に存在しない人・事象の非習慣的・一時的な行動や状態を表現したメッセージ
- [10] 状況外回想：送り手・受け手がいる場に存在する送り手・受け手以外の人・事象の過去の出来事についての情報を交換するメッセージ
- [11] 予想（予測）：意図しないで行われる送り手・受け手の未来の活動，もしくは，それ以外の要素による意図的・非意図的な未来の活動を表現したメッセージ
- [12] 推測：特定の条件下においてのみ起きる出来事について表現したメッセージ
- [13] 説明：送り手・受け手がいる場に存在しない人・事象の習慣的・恒久的な行動や状態を表現したメッセージ
- [14] 一般化：特定の要素ではなく，クラス全体について習慣的・恒久的な行動や状態を表現するメッセージ



前掲の(11)について、その発話機能・中核要素・現象定位から修辞機能と脱文脈化指数を確認する。中核要素は太字で、現象定位は斜体で示す。主語が明示されていない場合は、( $\phi =$  )で示した。

(15) a. ( $\phi =$  **私は**) *26才女*です。

[命題&状況内：参加&現在：非習慣的・一時的⇒ [2] 実況]

b. **私も** *ずっと一重*で、

[命題&状況内：参加&現在：非習慣的・一時的⇒ [2] 実況]

c. ( $\phi =$  **私は**) *21の時と去年と二重埋没法の二重手術を受けました*。

[命題&状況内：参加&過去⇒ [3] 状況内回想]

この修辞機能と脱文脈化指数の割り当ては、少数の言語学者の内省に基づくものである。例えば、先行文献でも図2, 3, 4(先行文献での各図をもとに図1にあわせて筆者がまとめた)のように少数の言語学者の内省であっても異なるパターンの割り当てが確認でき、合意が形成されていない。このように理論的な考察がありながらも場当たりの脱文脈化指数が割り当てられており、実証的な研究は進められていない。

	Event Orientation	Proposal	Concurrent		Future		
	Central Entity		non-habitual	habitual	goods/services exchange	information exchange	hypothetical (possible-conditional)
Within material situational setting (MSS)	interactant	Recount [9]	Commentary [2]	Reflection [4]	Action [1]	Plan/Prediction [7]	Conjecture [8]
	other person/object			Observation [3]		Prediction [7]	
Not within MSS	person/object		Report [5]	Account [6]	Prediction [7]		
Class				Generalization [10]			

図2 「発話機能」「中核要素」「現象定位」の組み合わせと修辞機能・脱文脈化指数の割り当てのバリエーション：Cloran (1999: 41)

	Proposal	Proposition						
		concurrent		prior	forecast			
		non-habitual	habitual		non-hypothetical	hypothetical		
				volitional	non-volitional			
interactant	Action [1]	Commentary [2]	Reflection [3]	Recount [6]	Plan [7]	Prediction [8]	Conjecture [10]	
co-present person/object			Observation [4]		Prediction [8]			
absent person/object			Report [5]					Account [9]
generalised person/object								Generalization [11]

図3 「発話機能」「中核要素」「現象定位」の組み合わせと修辞機能・脱文脈化指数の割り当てのバリエーション：Cloran (2010: 52)

	提言	命題					
		現在		過去	未来		仮定
		非習慣的・一時的	習慣的・恒久的		意図的	非意図的	
状況内：参加	行動 [1]	実況 [2]	自己記述 [3]	状況内回想 [6a]	計画 [7]	予想 [8]	推測 [10]
状況内：非参加			観測 [4]		予想 [8]		
状況外		報告 [5]	説明 [9]	回想 [6b]			
定言			一般化 [11]				

図4 「発話機能」「中核要素」「現象定位」の組み合わせと修辞機能・脱文脈化指数の割り当てのバリエーション：佐野・小磯 (2011: 66)

### 3. 検証

#### 3.1 検証方法

本稿では、脱文脈化指数の定義を再検討する。一般の方に共有される脱文脈化指数を目指して、「発話機能」「中核要素」「現象定位」が付与されたデータを対象に、被験者実験により脱文脈化の程度を収集する。対象となるデータはYahoo!知恵袋の中カテゴリ「コスメ、美容」の質問と回答120サンプルである。脱文脈化指数を説明するために「いま・ここ・わたしからの距離」を判定して答えるという課題を設定する。実験協力者はYahoo!クラウドソーシングのアカウントを持つ20歳以上の女性(異なり141人)とする。丁寧な実験協力者を選択するために、チェック設問として厳しめの課題を設定した結果、763人(異なり)の方が落選した。実験協力者は左右に並べた二つのテキスト(Yahoo!知恵袋の2サンプル)と評価対象のメッセージを確認し、左右どちらのメッセージが「いま・ここ・わたしからの距離」に近いかを判定する。図5に実験協力者に呈示した例文対を示す。

左の例は命題であり、中核要素の対象は省略されている「(化粧の濃淡は)」を選択し、「状況外」

とタグ付けされている。現象定位の対象である述語相当として「好みなんでしょうか？」が選択され、「現在：習慣的・恒久的」とタグ付けされている。結果「[13] 説明」となる。右の例は命題であり、中核要素の対象は省略された「私は」とし、「状況内：参加」とタグ付けされている。現象定位の対象である述語は「コンプレックスです。」が選択され、「現在：習慣的・恒久的」とタグ付けされている。結果「[7] 自己記述」となる。

これらを総合して「いま・ここ・わたしからの距離」が近いのは右になる。このデータに対しては24件の有効回答が得られ、左が近いとしたものが11件、右が近いとしたものが13件であった。このようなデータに基づき、次節に述べる統計モデリングを行う。

A OC09_00002_Q_003 ♪	B OC09_00019_Q_005 ♪
<p>彼氏に、「化粧してるとかどうかわからない」と言われたので、濃い目にしたら、満足してみたいんですけど、 普通男性って薄い化粧の方が好きなのかと思ってました。 こういうのはやはり個人の好みなんじゃないかな？ それとも彼は、少し変わってるのでしょうか？ 人によって好みが違うと思います。 一般的には、ナチュラルメイクの方が好印象だと思うのですが・・・・・・・・</p>	<p>私は中一です。 ニキビを潰してしまう、クセがあるのですが、 どうしたら良いのでしょうか？ 気がついたら、引っかいていて、潰れて跡が残ってしまって・・・ とてもこの顔がコンプレックスです。</p>

図5 Yahoo! クラウドソーシングで示した例文対（背景影付きが対象となる文）

### 3.2 モデリング

統計処理はBradley Terryモデルに基づく。Bradley Terryモデルは一对比較データに基づいてオッズ比を一般化線形モデルにより推定することにより、優劣を線形順序にモデル化する手法である。オッズ比の対数を1次元的に表現することで、限られた一对比較データから線形順序を構成することができる。要素  $n$  個中の要素  $i$  と要素  $j$  を一对比較した場合に、要素  $i$  が要素  $j$  よりも脱文脈化程度が高い（順序記号を用いて  $i > j$  と表す）確率を  $P_{i > j}$  として

$$P_{i > j} = \frac{\pi_i}{\pi_i + \pi_j}$$

となるオッズ比  $\pi_i$  を導入する。ここで  $\lambda_i = \log \pi_i$  となる係数を導入すると以下のようになる。

$$\text{logit}(P_{i > j}) = \lambda_i - \lambda_j$$

logit はロジット関数（ロジスティック関数の逆関数）であり、オッズ比の対数を意味する。

すべての一对比較が独立であると仮定して、この係数  $\{\lambda_i\}$  ( $i=1 \dots n$ ) を一般化線形モデルに基づ

く最尤法（切片（定数項）のないロジスティック回帰）によって推定する。

分析は「中核要素」の評価・「現象定位」の評価・「脱文脈化指数」の評価の三つを行った。

分析においては人手によりラベルが同じ対を排除した。ランダムに一対比較をサンプリングした結果、「脱文脈化指数」についてはラベルが出現しなかったものもあるが、出現したラベルについてのみ分析を行う。

### 3.3 結果

表1に中核要素のモデリング結果を示す。係数の大小が実験協力者が答えた「いま・ここ・わたしからの距離」の相対的な大小関係を表す。負の値であっても小さいほうが距離が近いことを表す。実験協力者は状況内：参加<状況外<状況内：非参加<定言の順に「いま・ここ・わたしからの距離」の順序関係を定義している。特に状況内要素については参加要素と非参加要素を区別している傾向が見られた。

表1 中核要素のモデリング結果：Bradley Terry モデルによる係数

状況内：参加	状況外	状況内：非参加	定言
-0.1939	0.0000	0.1083	0.3169

表2に現象定位のモデリング結果を示す。実験協力者は未来：非意図的<現在：非習慣的・一時的<過去<未来：意図的<現在：習慣的・恒久的の順に「いま・ここ・わたしからの距離」の順序関係を定義している。

表2 現象定位のモデリング結果：Bradley Terry モデルによる係数

未来 非意図的	現在 非習慣的・一時的	過去	未来 意図的	現在 習慣的・恒久的
-0.5436	-0.4199	-0.2502	-0.1671	0.0000

表3 脱文脈化指数のモデリング結果：Bradley Terry モデルによる係数

(表中\*は他のラベルの可能性を表す)

[2] 実況 現在 非習慣的・一時的 状況内 {参加, 非参加}	[5] 状況内予想 未来 {意図的*, 非意図的} 状況内 {参加*, 非参加}	[9] 報告 現在 非習慣的・一時的 状況外	[3] 状況内回想 過去 状況内 {参加, 非参加}	[4] 計画 未来 意図的 状況内 参加
0.0000	0.0447	0.0825	0.3082	0.4212
[7] 自己記述 現在 習慣的・恒久的 状況内 参加	[13] 説明 現在 習慣的・恒久的 状況外	[10] 状況外回想 過去 {状況外・定言}	[8] 観測 現在 習慣的・恒久的 状況内 非参加	[14] 一般化 現在 習慣的・恒久的 定言
0.5647	0.5883	0.7137	0.7428	1.0756

表3に脱文脈化指数のモデリング結果を示した。実験協力者は[2] 実況< [5] 状況内予想< [9] 報告< [3] 状況内回想< [4] 計画< [7] 自己記述< [13] 説明< [10] 状況外回想< [8] 観測< [14] 一般化の順に「いま・ここ・わたしからの距離」の順序関係を定義している。

### 3.4 考察

まず「中核要素」について考察する。佐野(2010b)は、状況内：参加<状況内：非参加<状況外<定言の順を提案しているが、被験者実験で得られた結果は、状況内：参加<状況外<状況内：非参加<定言の順であった。日本語母語話者は状況内：参加と状況内：非参加を明示的に区別したうえで、状況内：非参加を状況外よりも遠いものとして扱う傾向が見られた。

次に「現象定位」について考察する。佐野(2010b)は、現在：非習慣的・一時的<過去<未来：意図的<未来：非意図的<仮定<現在：習慣的・恒久的の順を提案しているが、被験者実験で得られた結果は、未来：非意図的<現在：非習慣的・一時的<過去<未来：意図的<現在：習慣的・恒久的の順であった。日本語は過去-非過去の対立のみが陽に表出しない言語ではあるが、「未来：非意図的」をより近いものとして扱う傾向が見られた。

最後に組み合わせより得られる佐野(2010b)に基づく「脱文脈化指数」の順序は、スピーアマンの順位相関係数が0.80と高い相関を持っているが一致しているとまでは言えない。

以上の結果から、佐野(2010b)の問題点として、空間と時間のとらえ方が一般の人の認識と若干異なることがあげられる。

書き言葉において、状況内か状況外か、さらに参加しているかしていないかについての明確な指針がないために「中核要素」において状況内：非参加と状況外間の判別が不定になっているのではないかと考える。それ以前に対象が主語相当か主題相当かについてきちんと定義されておらず、中核要素のアノテーション側の問題があるのではないかと考える。

「現象定位」においては、アノテーション基準が日本語の時制をどのように扱うか、明確化できていないために適切に作業できていない可能性がある。また、被験者実験側で、未来：非意図的が現在：非習慣的・一時的よりもより近いものとして、とらえられる傾向にあることも興味深い。

「脱文脈化指数」については、相関係数が高いながらも、適切な割り当てがされていないように見える。中核要素において、被験者実験側では状況内：参加と状況内：非参加で差があることが確認されている。この差が「脱文脈化指数」の[2] 実況・[3] 状況内回想・[5] 状況内予想においては捨てられており、指標に反映されていない。

### 4. おわりに

本稿では修辞ユニット分析における脱文脈化指数の割り当てについての妥当性を被験者実験により検証した。佐野(2010b)の割り当て方式が、順位相関係数では0.80と高い一方、必ずしも一般の被験者の感覚と一致しているわけではなく、中核要素や現象定位においては順序が一致しない傾向があることがわかった。

問題点として、Cloran(1994)が提案するような「発話機能」「中核要素」「現象定位」が現代

日本語書き言葉に対してタグ付けられるのかという観点と、「発話機能」「中核要素」「現象定位」から「脱文脈化指数」を構造主義的に決定できるのかという2点があげられる。

前者については、時制や主語・主題の基準があいまいであり、特殊な訓練を受けたものであっても安定してタグ付けするのは難しい。日本語の特性に合わせた別の基準を再構築する必要がある。後者については、タグ付けの問題がなかったとしても、単純な組み合わせによる脱文脈化指数の割り当てが一般人の感覚と乖離している。

問題点の解決として、「脱文脈化指数」そのものを被験者実験的に付与すること、「発話機能」「中核要素」「現象定位」について日本語の特性を知ったうえでタグ付け可能な人を育成することとどちらが低いコストであるのかを考える必要がある。

しかしながら、前提としている時空をどのように定義するかを再確認しておきたい。特殊相対性理論においては、ミンコフスキー時空と呼ばれる対称で非退化な双線型形式を持つ4次元の実ベクトル空間で時空が表現され、このミンコフスキー時空上の点（世界点）が事象として定義される。ミンコフスキー時空上ではミンコフスキーノルム（世界距離）と呼ばれる数学的な距離の公理を満たさない（正定値性を満たさない）ノルムが定義される。

ここで、テキストに出現する事象表現を適切な時空上の点に写像することを、リサーチエスチョンとして再定義する（図6）。そのリサーチエスチョンを解くために、アノテーションに基づくテキストが表出する特徴だけでなく、複数の観測者（実験協力者・アノテータ）に基づく相対的な判断の分布を用いる必要がある。さらに、事象表現集合に対して、脱文脈化指数のような全順序に基づく鎖構造が規定できるのかを疑う必要がある。ミンコフスキーノルムのように正定値性を満たさない計量が定義される時空を想定するのであれば、全順序律を外した半順序集合に基づく束構造を考えなければならない。そのうえで、内省によらない、より実証的な手法を用いて適切なモデル化を行う必要がある。

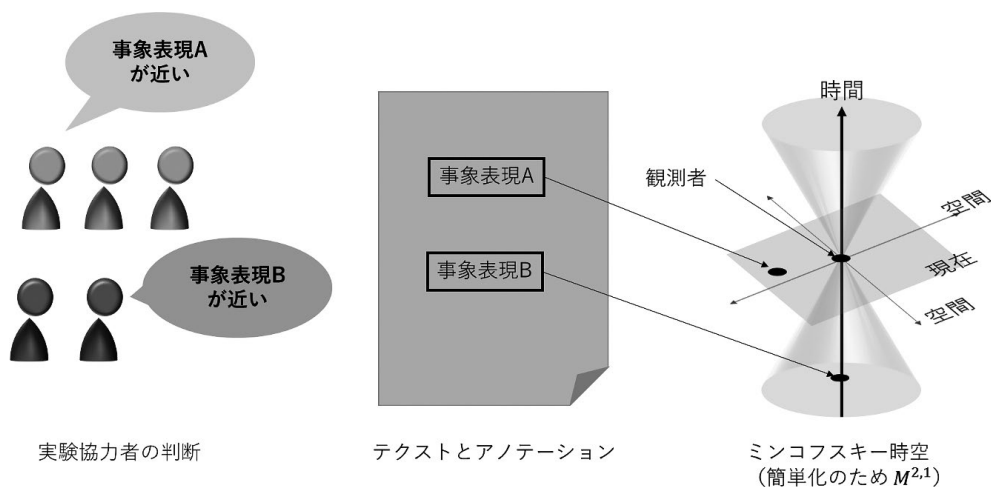


図6 リサーチエスチョンの再定義



## 参考文献

- Cloran, Carmel (1994) *Rhetorical units and decontextualisation: An enquiry into some relations of context, meaning and grammar*. Vol. 6. Monographs in systemic linguistics. Nottingham, United Kingdom: University of Nottingham.
- Cloran, Carmel (1999) Contexts for learning. In: Frances Christie (ed.) *Pedagogy and the shaping of consciousness*, 31-65. London, United Kingdom: Bloomsbury Academic.
- Cloran, Carmel (2010) Rhetorical unit analysis and Bakhtin's chronotype. *Function of language* 17(1): 29-70.
- Halliday, M. A. K. and C. M. I. M. Matthiessen (2004) *An introduction to functional grammar*. London: Arnold.
- 小西光・浅原正幸・前川喜久雄 (2013) 「『現代日本語書き言葉均衡コーパス』に対する時間情報アノテーション」『自然言語処理』20(2): 202-222.
- Maekawa, Kikuo, 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: 345-371.
- 松本理美・浅原正幸・有田節子 (2017) 「『現代日本語書き言葉均衡コーパス』に対する節の意味分類情報アノテーション—基準策定、仕様書作成の必要性について—」『言語資源活用ワークショップ2016 発表論文集』336-346.
- 宮内拓也・浅原正幸・中川奈津子・加藤祥 (2017) 「『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構築」『言語処理学会第23回年次大会発表論文集』402-405.
- 佐野大樹 (2010a) 「選択体系機能言語理論を基底とする特定目的のための作文指導方法について—修辞ユニットの概念から見たテキストの専門性—」『専門日本語教育研究』12: 19-26.
- 佐野大樹 (2010b) 『日本語における修辞ユニット分析の方法と手順 ver.0.1.1—選択体系機能言語理論（システム理論）における談話分析—（修辞機能編）』。 <http://researchmap.jp/systemists/> 資料公開 / (2018年3月31日確認)
- 佐野大樹・小磯花絵 (2011) 「現代日本語書き言葉における修辞ユニット分析の適用性の検証—「書き言葉らしさ・話し言葉らしさ」と脱文脈化言語・文脈化言語の関係—」『機能言語学研究』6: 59-82.
- 田中弥生 (2013) 「評価の高低によるクチコミサイト「アットコスメ」における談話構造の特徴—修辞ユニット分析を用いて—」『神奈川大学言語研究』35: 1-23.
- 田中弥生・浅原正幸 (2016) 「Yahoo! 知恵袋における修辞ユニット分析の中核要素認定に関する諸問題」『言語処理学会第22回年次大会発表論文集』881-884.
- 田中弥生・浅原正幸 (2017) 「Yahoo! 知恵袋における修辞ユニット分析の発話機能認定に関する諸問題」『言語処理学会第23回年次大会発表論文集』831-834.
- 田中弥生・佐野大樹 (2011) 「Yahoo! 知恵袋における質問の修辞ユニット分析—脱文脈化 - 文脈化の程度による分類—」『電子情報通信学会技術研究報告（第2回集合知シンポジウム）』110: 13-18. 電子情報通信学会.
- 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』205-214.
- 保田祥・小西光・浅原正幸・今水穂・前川喜久雄 (2013) 「『現代日本語書き言葉均衡コーパス』に対する時間情報表現・事象表現間の時間的順序関係アノテーション」『自然言語処理』20(5): 657-681.

## 関連 URL

- コーパス検索アプリケーション『中納言』 <https://chunagon.ninjal.ac.jp/> (2018年3月31日確認)
- 『国語研日本語ウェブコーパス』検索系『梵天』 <http://bonten.ninjal.ac.jp/> (2018年3月31日確認)



## Validation of the Degree of Decontextualisation in Rhetorical Unit Analysis

ASAHARA Masayuki<sup>a</sup>      TANAKA Yayoi<sup>b</sup>

<sup>a</sup>Center for Corpus Development, NINJAL

<sup>b</sup>Ph. D. Candidate, The University of Tokyo

### Abstract

This paper empirically evaluates the validity of the degree of decontextualisation in rhetorical unit analysis, a method of analysing messages in terms of: speech function, i.e., whether the message is a proposition or proposal; central entity, i.e., whether the subject appears in the presented situation or not; and event orientation, i.e., where the event occurs on a temporal axis. The rhetorical function and degree of decontextualisation of a message are identified by considering a combination of these features, as the linearly ordered set of the degree of decontextualisation is not defined empirically, but arbitrarily. We conducted a pairwise comparison experiment via Yahoo! crowdsourcing in order to evaluate the degree of decontextualisation that non-linguists find acceptable. The data were converted into a linear order by a generalised linear regression using the Bradley-Terry model, and the degree of decontextualisation was validated. The results revealed that, while the participants mostly shared the linear order of the degree of decontextualisation, the central entity and event orientation features did not always reflect the perspective of native Japanese speakers.

**Key words:** systemic functional linguistics, rhetorical unit analysis, decontextualisation, crowdsourcing