

論文B

音象徴語と動詞の共起パターンに関する 新聞コーパスの共起頻度と母語話者の産出との類似性の検討

玉岡 賀津雄 (名古屋大学)

要旨

新聞は、記者が一般大衆に情報を伝達するために、簡潔でわかりやすい表現を目指して、標準的な表現で書かれている。そのため、日本語母語話者の言語産出を反映しているのではないかと思われる。特に、音象徴語は、幼少から擬音や擬態の感覚が養われて、共感的な意味拡張とともに、多様な表現へと拡張した表現であるため、成人の日本語母語話者では多様な動詞とともに産出されると仮定される。そこで、28種類の音象徴語と動詞の共起パターンを、1991年から1999年までの9年間の毎日新聞のコーパスとこれらの音象徴語に対する36名の母語話者による30秒での動詞の産出を比較した。その結果、多様性の指標であるエントロピーは、母語話者と新聞コーパスで高い相関 ($r=0.83, p<.001$) を示した。エントロピーと規則性の指標である冗長度には有意な違いはなく、両者の類似性が示された。記述的な考察でみられた例外的な音象徴語は28語中4語であった。

キーワード: 音象徴語, オノマトペ, 共起パターン, エントロピー, 冗長度, 新聞コーパス, 日本語母語話者の産出

1. 研究の目的

新聞は、多くの新聞記者によって広く情報を伝達するために、標準的な表現で書かれている。多数の新聞記者によって書かれた記事は、日本語母語話者の典型的な言語産出を反映しているのではないかと思われる。音象徴語¹と共起する動詞を想定すると、これらを書き言葉として使った表現の集合体が新聞コーパスに反映されていると仮定することができる。そう考えれば、母語話者による音象徴語の動詞の産出は、新聞コーパスでの音象徴語と動詞の共起と類似したパターンを示すのではないかと予想される。そこで、本研究では、音象徴語と共起して産出される動詞の共起パターンについて、母語話者の言語産出と新聞コーパスの共起頻度の類似性をエントロピーと冗長度の指標で検討することにした。

2. 先行研究

擬音語は外界の音声を模倣した語、擬態語は物事の様子や状態を表した語とされる (天沼, 1974; 浅野, 1978; 瀬戸口, 1985; 飛田・浅田, 2002; 角岡, 2007; 田守, 1993, 2001, 2002

¹ 日本語でいう「オノマトペ」の意味で、「音象徴語」という表現を使用する。用語についての詳細の説明は、第2節の「先行研究」を参照。

など). 日本語の「オノマトペ」という表現は, これらの擬音語と擬態語ばかりでなく, 音をともしなわれない様態や心情を表す語までが含まれている. 英語の *onomatopoeia* は, 擬音語のみを指し, 擬態語は *mimesis* と呼ばれる (Chang 1990; Hamano 1998). そのため, これらを総称して「音象徴語 (sound-symbolic words)」と呼ぶほうが適切だとする意見もある (羽佐田 2005; 飯田・玉岡・初, 2012; 大坪 1989; 玉村 1989 など). そこで, 本研究では, より厳密な表現として, オノマトペではなく音象徴語を使用する.

英語では, 笑う様子を表現する動詞は, *grin*, *smirk*, *giggle*, *cackle*, *chuckle* など多様である. 英語と比べて, 日本語には異なる物事の状態や様子を表わす動詞が少なく, 笑う様子を表わす場合, 「笑う」の 1 語だけである. そのため, 音象徴語を副詞として使用して, 「けらけら笑う」「くすくす笑う」「にやにや笑う」など, 笑う状態を表現する. こうした音象徴語は, 動作, 活動, 状態を, 擬音や擬態の類像性に訴えることにより, 躍動感や実状感を醸し出す. 結果的に, 音象徴語は, 日本語の動詞表現の単調さを補うという側面もあり (筧・田守, 2003; Toratani, 2007), 副詞として使用されることが多く (羽佐田, 2005; 李, 1998, 2000; スコウラップ, 1993a, 1993b; 田守, 2002; 田守・スコウラップ, 1999; Toratani, 2007), 頻繁に動詞と共に起る.

日本語を母語とする幼児の発話記録に基づく縦断・横断研究をみると, 幼児が言語獲得をする際に, 擬態語や擬情語よりも擬音語を先に習得するという特徴が観察されている (大久保, 1967). さらに, Akita (2009) は, 音を表現した擬音語は類像性が高く, 抽象事象を描写的に表現した擬態語や擬情語は類像性が低いことに着目して, 音象徴語を擬音語, 擬態語, 擬情語に分類し, それらを一般語彙と比較して, 「臨時的擬音語² > 擬音語 > 擬態語 > 擬情語 > 一般語彙」と分類した. Akita (2009) は, これを「語彙的類像性の階層」と捉えている.

日本語を母語とする幼児は, 擬音を感覚的に活動や状態と結びつけながら多様な表現へと拡張していくと予想される. その過程において, 母語話者のメンタルレキシコンに音象徴語が, 動詞と共に起るかたちで記録される. そして, これにより, 個々の音象徴語に対して多様な動詞を産出することができるようになるかと仮定される. たとえば, 「どんどん」という音象徴語であれば, 特定の動詞と共に起り, 「どんどん進む」「どんどん歩く」「どんどん溜まる」などの多様な動詞句をつくる. もともと, 日本語を母語とする幼児が, 太鼓を叩くようなドンドンという具体的な音の印象をもとにして, 擬態的に「進む」「歩く」「溜まる」などへと, 聴覚から視覚へ, そして心理的・感情的な表現形態へと「語彙的類像性の階層」(Akita, 2009) に示されたように拡張されていくと予想される. これは, 「共感覚³的比喩表現」の意味拡張と考えられている (武田, 2000, 2001⁴; 山梨, 1988, 2007; Williams, 1976).

² 臨時的擬音語とは, 話者が強調したいときに臨時に使うような「と一つもおいしい」の「と一つも」などの音象徴語のことである.

³ 心理学でいう「共感覚 (synesthesia)」とは, ある刺激に対して通常感覚だけでなく異なる種類の感覚をも誘発させる特殊な知覚的な現象を意味している (Asano & Yokosawa, 2013; 岩崎, 2009). たとえば, 聞こえた音に色がついて聞こえる「色聴 (sound-color synesthesia)」である. そのため, 認知言語学で使用されている「共感覚」とは意味が異なっている.

⁴ 武田 (2001) では, 「意味拡張」ではなく, 「派生」と呼んでいる.

共感的な意味拡張 (Akita, 2009; 大久保, 1967) は、幼児の母語話者としての音象徴語の獲得をうまく説明している。しかし、音象徴語は、日本語特有の語彙であり、むしろ日本語に限定された感覚が幼少から養われていなくては、習得が難しいとも考えられる。そうであるなら、外国語として日本語を学ぶ学習者には、日本語母語話者と同じような音象徴語に付随した感覚が幼児期の成長とともに身につけていないので、逆にこの種の語彙の習得は難しいのではないかと予想される。実際、針生・趙 (2007) は、中国の大学で日本語を専攻している中国人大学生の 2 年生 40 名と 4 年生 32 名に、有声音と無声音対比の音象徴語と大小を対応付ける課題を課した。しかし、両グループの得点には差がみられなかった。つまり、日本語の学習年数が長くなっても、音象徴語の感覚には違いがみられなかった (針生・趙 2007) のである。

さらに、飯田・玉岡・初 (2012) は、読解能力で分けた上・中・下位の 3 群の中国語を母語とする日本語学習者に対して、音象徴語が修飾する動詞を選択させるテストを行った。その結果、読解能力が高くなるにつれて、音象徴語の理解も向上した。外国人日本語学習者にとっては、音象徴語は人間に普遍的な感覚を介して自然に習得されるというより、学習によって習得されるべき語彙であると言えよう。そのため、日本語環境において、幼少期の感覚と結びついたかたちで音象徴語が習得され、それがさらに意味拡散するかたちで多様な動詞表現と結びついていった⁵と考えるべきであろう。

本研究では、成人の日本語母語話者を対象に、音象徴語から想起される動詞の共起パターンを測定する。また、複数の新聞記者が書いた表現の集合と考えられる新聞コーパスから、音象徴語と動詞の共起パターンを調べる。そして、母語話者の産出と新聞コーパスの類似性を検討する。

3. 調査方法

3.1 音象徴語の選択

本研究のために、次の 5 つの条件で音象徴語を選択した。条件 1 は、基本的にひらがなで表記される畳語であることである。たとえば、「どんどん」は「どん」を 2 回繰り返す畳語である。もちろん、カタカナで「ドンドン」と書かれることもあるが、通常はひらがなで表記されることが多い。条件 2 は、様態や状態を表す副詞として使える音象徴語であることである。たとえば、「とぼとぼ」であれば、「とぼとぼ歩く」と動詞の「歩く」状態を修飾する副詞として使用される。条件 3 は、形容動詞としては使えない音象徴語であることである。音象徴語は基本的には副詞として使用されるが、形容動詞として使用できるものもある。たとえば、「な」を付けて「ばらばらな動き」あるいは「の」をつけて「よれよれの背広」と表現することができる (浜野, 2014)。このような音象徴語は含まないことにした。条件 4 は、サ変動詞「する」以外の動詞とも用いられることである。もちろん、多くの音象徴語がサ変動詞をともなって、「ころころする」「ぐらぐらする」「きらきらする」と表現することができる。しかし、同時に「ころころ転がる」「ぐらぐら揺れる」「きらきら輝く」と他の動詞とも共起することもできる。こういう多様な動詞と共起しうる音象徴

⁵ ただし、音象徴語にも多義性が認められるので、すべてが意味拡張であると考えるのは一般化しすぎであろう。

語を選んだ。そして条件 5 は、誰もが知っている基本的な音象徴語であることである。その基準としては、1991 年から 1999 年までの 9 年間の毎日新聞の全記事 (273, 514, 662 形態素数) で検索して、少なくとも 10 回以上使用されている比較的使用頻度の高い音象徴語であることとした。

以上の条件で、28 種類の音象徴語を選択した。具体的には、「どんどん」(9 年間の毎日新聞で 4,539 回出現、以下同様)、「きらきら」(528 回)、「ばたばた」(397 回)、「ばりばり」(346 回)、「だらだら」(280 回)、「がんがん」(258 回)、「ころころ」(240 回)、「ゆらゆら」(225 回)、「かんかん」(221 回)、「ぐらぐら」(124 回)、「ばちばち」(103 回)、「とぼとぼ」(74 回)、「ぷんぷん」(72 回)、「すらすら」(71 回)、「じろじろ」(57 回)、「ぴよんぴよん」(57 回)「すやすや」(55 回)、「ぼうぼう」(55 回)、「ちよろちよろ」(48 回)、「げらげら」(34 回)、「ことこと」(28 回)、「ごくごく」(27 回)、「めそめそ」(24 回)、「ぼたぼた」(23 回)、「ぐうぐう」(20 回)、「しくしく」(18 回)、「しとしと」(13 回)「ずきずき」(13 回) である。

3.2 日本語母語話者の音象徴語と動詞の共起産出

日本語母語話者の産出には、36 名の大学生を使った。最年少が 18 歳 7 カ月で、最年長が 21 歳 7 カ月であり、平均が 20 歳 4 カ月、標準偏差が 1 歳 2 カ月であった。男性が 32 名、女性が 4 名である。前節で挙げた 28 種類の音象徴語について、各音象徴語と共起する動詞を 30 秒で思いつく限り挙げてもらった。これを、母語話者の産出とした。たとえば、「ばりばり」を提示すると、日本語母語話者は、この音象徴語と共起する動詞を、「(ばりばり) 食べる」とか「(ばりばり) 働く」など動詞を 30 秒以内に思いつく限り書くという課題である。これにより、動詞の種類と共起頻度を数えた。被験者数が 36 名なので、1 つの動詞に対する最大共起頻度は 36 になり、最大共起頻度が限定される。これくらいのサンプルで、新聞コーパスと類似した共起パターンが現れるかどうかを検討した。仮に、36 名くらいのサンプルで大規模の新聞コーパスの共起パターンが再現できるのであれば、母語話者の産出は今後の研究において、効率的な言語の研究法の 1 つと考えることができよう。

3.3 新聞コーパスによる例文抽出の基準

音象徴語と動詞の共起について、述語が省略されている文、あるいは述語があっても動詞がない文は分析の対象から外した。音象徴語と共起する動詞の分類基準は 8 つである。

(1) 活用形、時制、使役、受身、敬語等は、原形で分類した。この基準では、たとえば、「反射させる」は「反射する」としてカウントされる。(2) 自動詞、他動詞は別々に分類した。たとえば、「利く」と「利かす」は別分類とした。(3) 複合動詞は、それを形成する個々の動詞とは別に分類した。たとえば、「描き出す」は「描く」とは別の分類とした。(4) 「形容詞等+なる・する」など、語幹のとれない動詞は、すべて「する」「なる」として分類した。たとえば、「大きくなる」は「なる」の分類とした。(5) サ変動詞「する」(「できる」を含む) は、すべて「する」として分類した。たとえば、「稽古できる」は「できる」であり「する」として分類した。(6) テ形で動詞がつながる場合は、基本的に前の動詞のみを採った。たとえば、「出てくる」は「出る」とした。(7) テ形で動詞がつながる場合でも、前の動詞が後の動詞を修飾している場合は、後ろの動詞を採った。たとえば、「口をつ

いて出る」は「出る」として分類した。(8)テ形でつながる動詞でも、1語になっているとみなせる場合は、合わせたかたちで分類した。たとえば、「近くにやってきた」は「やってくる(遣って来る)」のかたちで『広辞苑』(第5版)に記載があるので、このかたちで数えた。

以上の基準で、毎日新聞の1991年から1999年までの9年間の新聞のコーパスを使って、28種類の音象徴語と共起する動詞数と共起頻度を数えた。なお、この新聞コーパスの総形態素数は、273, 514, 662である。検索には、アメリカのパデュー大学先端技術言語学習研究所の深田淳が作成した日本語用例・コロケーション抽出システム『茶漉』を使用した(<http://tell.fll.purdue.edu/chakoshi/>)。『茶漉』は、コーパスから用例およびコロケーション情報を抽出するシステムである(『茶漉』についての詳細の情報は、深田, 2007; 玉岡, 2016を参照)。

3.4 共起頻度パターンの指標化

本研究では、エントロピー(entropy)と冗長度(redundancy)の概念(Shannon, 1948)を使って、母語話者の産出と新聞コーパスでの共起パターンを比較した。エントロピーは情報量の尺度の1つであり、あいまいさや乱雑度の増減を示す指標である(有本, 1982; 堀, 1979; 海保, 1989; 玉岡, 2010)。エントロピーと同様に、冗長度は表現の多様性と使用頻度から1つの値を算出して、無駄の程度を表すことができる。エントロピーと冗長度の2つの指標を組み合わせることで、ある表現の多様性と規則性を、絶対的な出現頻度にある程度左右されることなく簡単な数値で表すことができる。

分布の偏りを回避するために対数変換、角変換、平方根変換など他の指標に変換することで、正規分布にしたがう変数を作ることがある。コーパス研究では、出現頻度や共起頻度を他の指標に変換することで、頻度分布の影響を受け難くし、出現パターンを数値化する工夫が模索されてきた(影浦, 2000; 鈴木・影浦, 2011; 中俣, 2015, 2016)。エントロピーも変換指標の1つと考えることができる。袋小路(garden path)文についてエントロピーを用いた研究(Den & Inoue 1997; 井上 2000; Inoue & Den 1999)では、動詞に対して共起する主語と目的語の名詞の種類と共起頻度からエントロピーを算出し、エントロピーの大小と袋小路文の曖昧さへの陥りやすさが関係することを示している。また、内元・関根・井佐原(2001)は、与えられた文を形態素の並びに分解し、各形態素に対し文法的な属性品詞や活用などをつける形態素解析に最大エントロピーモデルを応用している。

エントロピーと冗長度を組み合わせた研究も複数みられる。接辞の生産性を接頭辞と接尾辞で比較した研究(Miyaoka & Tamaoka, 2005)では、「非」など12の接頭辞と「性」など12種類の接尾辞の生産性をエントロピーと冗長度を使って比較した。その結果、「右側主要部の規則(right-hand head rule)」(影山, 1993; Williams, 1981)にしたがって、「性」などの接尾辞は「安全性」「主体性」などの語と結びつき、接頭辞よりもより多様な熟語を作ること示した。また、複合動詞の研究(Tamaoka, Lim & Sakai, 2004)では、「(ビールを)飲み終える」([_{VP} [_V (ビールを) NP-o (飲み) V₁] (終える) V₂])などの統語的複合動詞と「(盛り場を)飲み歩く」([_{VP} (盛り場を) NP-o (飲み歩く) V₁+V₂])などの語彙的複合動詞の2番目の動詞(V₂)がどのような1番目の動詞(V₁)と共起するかをエントロピーと冗長度の指標で比較している。その結果、「ビールを飲む」などの動詞

句について、2 番目の動詞 (V₂) で「終える」と表現できる統語的複合動詞は、多様な動詞と結びついて複合動詞を作る傾向があり、生産性がきわめて高いことを示した。

表 1: 日本語母語話者による音象徴語と共起する動詞の産出と新聞コーパスによる動詞の共起頻度

音象徴語	日本語母語話者				新聞コーパス			
	種類	頻度	エントロピー	冗長度	種類	頻度	エントロピー	冗長度
どんどん	48	118	4.85	13.1%	926	4424	7.81	20.8%
だらだら	40	112	4.60	13.5%	71	220	4.72	23.2%
がんがん	35	83	4.52	11.8%	90	234	5.60	13.7%
ばりばり	28	80	4.05	15.9%	40	152	4.08	23.4%
ばたばた	27	77	4.09	13.9%	89	317	4.52	30.2%
すらすら	27	110	3.75	21.1%	28	62	4.22	12.2%
ゆらゆら	26	99	3.64	22.6%	51	134	4.49	20.8%
ちよろちよろ	19	81	3.55	16.5%	20	34	3.93	9.2%
ことこと	18	52	3.37	19.1%	7	22	2.15	23.3%
ぼたぼた	18	78	3.25	22.1%	7	19	1.88	33.1%
ぱちぱち	15	65	3.24	17.0%	27	60	4.04	15.0%
とぼとぼ	15	62	2.74	29.8%	16	70	2.62	34.6%
ぴよんぴよん	14	91	2.47	35.1%	8	50	3.37	19.2%
しとしと	14	51	3.00	21.1%	3	12	1.04	34.3%
ぐらぐら	13	61	2.66	28.2%	20	87	2.87	33.7%
ころころ	13	63	3.12	15.6%	32	179	3.36	32.8%
きらきら	13	88	2.30	37.8%	42	442	2.58	52.1%
めそめそ	13	82	2.10	43.4%	4	23	1.44	27.8%
かんかん	12	49	2.90	19.1%	11	40	2.50	27.6%
ごくごく	12	52	1.96	45.2%	4	16	1.19	40.7%
すやすや	10	59	2.21	33.4%	9	39	2.10	33.9%
じろじろ	10	61	2.04	38.6%	8	56	1.31	56.5%
しくしく	9	42	1.90	40.0%	8	16	2.78	7.3%
ぐうぐう	9	43	2.16	31.9%	5	13	1.51	35.1%
ずきずき	9	68	1.79	43.5%	5	10	2.05	11.9%
ぷんぷん	8	49	2.05	31.6%	9	41	1.69	46.6%
ぼうぼう	8	43	2.24	25.4%	5	7	2.13	8.3%
げらげら	6	50	1.35	47.9%	8	31	1.96	34.5%
平均			2.93	26.94%			3.00	27.21%
標準偏差			0.93	10.98%			1.52	12.65%

注: 表の「種類」とは各音象徴語について日本語母語話者によって産出されたあるいはコーパスで共起した動詞の種類 (type frequency) を示し、「頻度」は産出されたあるいは共起した動詞の延べ頻度 (token frequency) を示す。

エントロピーと冗長度の 2 つの指標は、敬語表現にも応用されている (玉岡・宮岡・林, 2003)。これは、「行く」「食べる」「見る」などの動詞が尊敬と謙譲で、質問紙条件と面接条件で、多様性がどう異なるかを検討した研究である。その結果、謙譲表現のほうが尊敬

表現よりもエントロピーが高く、冗長度が低くなり、また面接条件の方が質問紙条件よりも多様な敬語表現が使用されていた。さらに、面接場面での謙讓表現は、きわめて多様であることがエントロピーで示された。

4. データ分析

4.1 エントロピーと冗長度の指標による新聞コーパスと母語話者の比較

28 種類の音象徴語に対する動詞との共起頻度パターン（詳細は表 1 を参照）を、エントロピーの指標で表した際の、平均と標準偏差は、表 2 に示した。まず、母語話者の産出と新聞コーパスのエントロピーのピアソンの積率相関係数は、 $r=0.83$ ($p<.001$) となり、非常に高く、この係数は有意であった。母語話者の産出は、新聞コーパスでの音象徴語の動詞との共起パターンと強い関係があることが示された。また、対応のあるサンプルの t 検定⁶で、両者のエントロピーを比較した。その結果、主効果は有意ではなかった [$t(27) = 0.41, p=.68, ns$]。これは、音象徴語と動詞の共起パターンが、母語話者の産出と新聞コーパスにおいて、類似していることを意味している。

表 2 : エントロピーの相関および t 検定の結果

頻度データの出典	1	2
1 母語話者の産出	—	
2 新聞コーパス	-0.90 ***	—
平均	2.93	3.00
標準偏差	0.93	1.52
t 検定の結果	$t(27)=-0.41, p=.68, ns$	

注: N (音象徴語の数)=28. *** $p<.001$. ns は有意でないことを示す。

28 種類の音象徴語と動詞との共起について、冗長度の平均と標準偏差を表 3 に示した。母語話者の産出と新聞コーパスの冗長度のピアソンの積率相関係数は、 $r=0.35$ (ns) であり、この係数は有意ではなかった。そのため、規則性という点では、両者には弱い関係しか認められなかった。さらに、両者の冗長度を対応のあるサンプルの t 検定で比較した。その結果、主効果は有意ではなかった [$t(27) = -0.41, p=.68, ns$]。エントロピーと同様に、母語話者の産出と新聞コーパスは、冗長度で示した共起パターンの規則性からみても、類似していることが示された。

以上のように、28 種類の音象徴語と動詞の共起パターンについて、エントロピーと冗長度の指標を使って、母語話者の産出と新聞コーパスを比較した結果、エントロピーには、強い相関が、冗長度については弱い相関が認められた。両者を比較した結果は、エントロピーと冗長度のいずれの指標でも有意な違いはなかった。そのため、日本語母語話者による音象徴語に対する動詞の産出と新聞コーパスにおける音象徴語と共起する動詞の種類と

⁶ 28 種類の音象徴語に対する母語話者による動詞の産出と新聞コーパスによる動詞の共起頻度は、異なるアプローチではあるが、同じ音象徴語に対して、エントロピーが計算されている。そのため、個々の音象徴語は、母語話者の産出と新聞コーパスにおいて対応があると想定されるので、対応のあるサンプルの t 検定を使って分析した。

共起頻度は、多様性と規則性において類似していることが示された。とりわけ、エントロピーの指標は、母語話者と新聞コーパスの関係が強く、なおかつ両者が類似していることを示したので、きわめて有効な指標であると言える。

表 3：冗長度の相関および *t* 検定の結果

頻度データの出典	1	2
1 母語話者の産出	—	
2 新聞コーパス	-0.47 *	—
平均	26.94%	27.21%
標準偏差	10.98%	12.65%
<i>t</i> 検定の結果	$t(27)=-0.10, p=.92, ns$	

注：N(音象徴語の数)=28. *ns*は有意でないことを示す。

4.2 母語話者の産出と新聞コーパスとのエントロピーと冗長度の散布図

ここまでエントロピーと冗長度の指標を使って、母語話者の産出と新聞コーパスの共起頻度を比較した。確かに、全体的にみれば、音象徴語と動詞の共起パターンは、母語話者の産出と新聞コーパスで類似していると言える。そこで、例外について検討するために、母語話者の産出と新聞コーパスを別々にして、エントロピーと冗長度の散布図を描いて、28種類の音象徴語について個々に検討することにした。

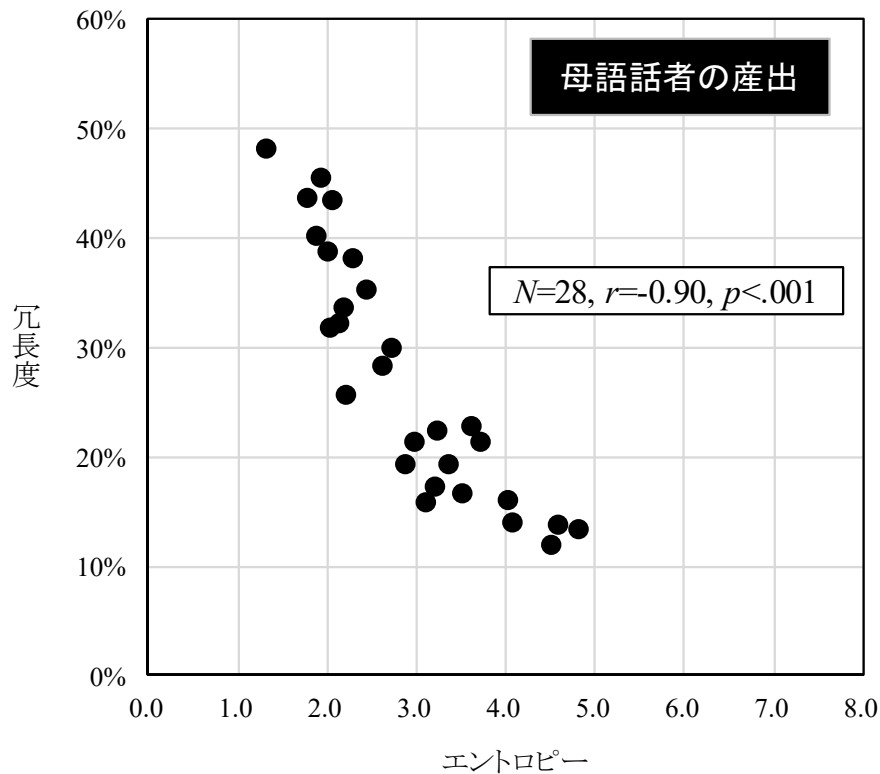


図 1：母語話者のエントロピーと冗長度の散布図

まず、日本語母語話者の産出は、図 1 のような散布図になる。28 種類の音象徴語のエントロピーと冗長さのピアソンの積率相関係数は $r = -0.90$ ($p < .001$) で、強い逆相関であった。これは、図 1 からわかるように、個々の音象徴語と動詞の共起パターンは、多様性という視点ではエントロピーの 1 から 5 の範囲に分布し、規則性という視点では冗長さの 10% から 50% の範囲に分布していた。エントロピーの多様性の指標と冗長さの規則性の指標が類似した共起パターンを示していると考えられる。母語話者の産出は、限られた時間内で産出される動詞についての共起パターンであり、エントロピーと冗長さの 2 つの指標からみても、全体として同じようなパターンとして示されたという見方もできよう。

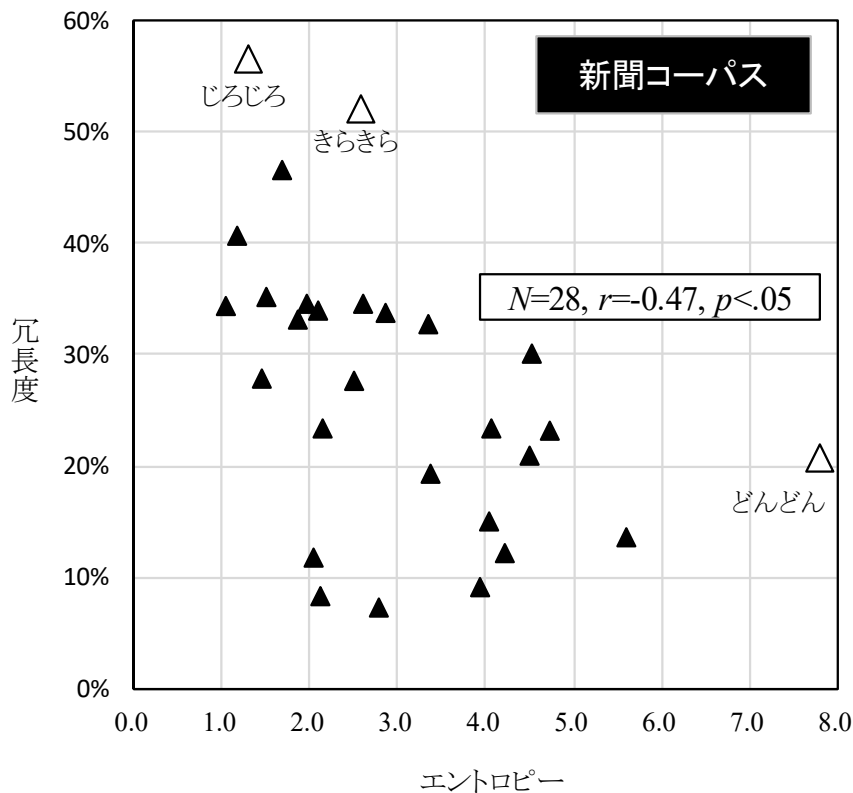


図 2：新聞コーパスのエントロピーと冗長さの散布図

新聞コーパスの共起パターンは、エントロピーは 1 から 8、冗長度は 0% から 60% と、母語話者の産出に比べて、より広い範囲での分布を示した。図 2 からわかるように、「どんだん」はエントロピーが 7.81 となり、音象徴語の中で最も高かった。その意味で、「どんだん」は、新聞コーパスでも例外的に多様性に富んだ音象徴語と言えよう。具体的には、表 1 からわかるように、「どんだん」は、926 種類の異なる動詞と共起し、その総共起頻度は 4,424 回にもなる。80 回以上の共起頻度を示したのは 9 つの動詞で、「X なる」が 451 回、「増える」が 215 回、「出る」が 153 回、「進む」が 118 回、「広がる」が 98 回、「行く」が 93 回、「入る」が 92 回、「やる」が 87 回、「変わる」が 80 回である。さらに、1 回だけの共起頻度の動詞が 256 種類もあった。共起する動詞の種類の高さと複数回出現する動詞との共起頻度の高さが際立っていた。

新聞コーパスで、冗長度が 50%以上の大きさになったのは、「じろじろ」と「きらきら」である。「じろじろ」は、わずか 8 種類の動詞としか共起せず、全体の共起頻度も 56 回である。「見る」が 44 回で全体の 78.6%になり、あとは、「見つめる」が 4 回、「眺める」が 2 回、「見比べる」が 2 回で、1 回の頻度の動詞が 4 種類ある。一方、「きらきら」は、42 種類の動詞と共起し、総共起頻度は 442 回と大きい。しかし、3 つの動詞との共起がきわめて多く、「輝く」が 147 回、「X する」が 121 回、「光る」が 116 回である。この 3 つの動詞だけで共起頻度が 384 回になり、全体の 86.88%になる。続いて、「反射する」が 8 回、「はね返す」が 4 回、「光り輝く」が 4 回となっている。30 種類もの動詞が、わずか 1 回だけの共起である。これら 2 つの音象徴語は、新聞コーパスの絶対的な種類の数と共起頻度ではかなり異なっているものの、冗長度の指標で示した共起パターンをみると、「じろじろ」と「きらきら」は特定の動詞と頻繁に共起し、規則性の高い音象徴語であるという共通の特徴を持っている。9 年間の毎日新聞という大きな新聞コーパスで調べると、母語話者の産出では考察できないような強い規則性をもつ音象徴語が数種類みられるようである。

4.3 新聞コーパスと母語話者の差からみた個別の音象徴語の特性

28 種類の音象徴語と動詞の共起パターンで、新聞コーパスと母語話者の産出を比べると、全体として類似していることはすでに述べた。しかし、新聞コーパスだけをエントロピーと冗長度で記述的に考察すると、図 2 の散布図のように、全体から逸脱した共起パターンの「どんどん」とやや離れたところに「じろじろ」および「きらきら」が隣接してみられた。そこで、さらに、図 3 に示したように、エントロピーと冗長度について、新聞のコーパスから母語話者の産出を引いた散布図を描いた。新聞コーパスを基準としているので、負の領域は母語話者の産出の指標が大きいことを示し、正の領域は新聞コーパスの指標が大きいことを示す。たとえば、エントロピーが高くかつ冗長度が大きい場合は、図 3 の散布図の第 1 象限になるが、これは新聞コーパスのほうが母語話者の産出に比べて、いずれの指標でも大きいことを示す。

図 3 をみると、新聞コーパスと母語話者の産出の差は、全体としてエントロピーも冗長度も 0 に近いところで分布しており、両者が類似した共起パターンを示していることがわかる。あくまで記述的な直観による考察ではあるが、4 つの音象徴語にやや逸脱した違いが見受けられる。すでに触れたように、「どんどん」は新聞コーパスでは多様な動詞と共起する。新聞コーパスのエントロピーは、母語話者の産出と +2.96 の差がある。対照的に、「しとすと」は、母語話者の産出のエントロピーのほうが新聞コーパスよりも大きく、-1.96 の差がみられた。新聞コーパスでは、「しとすと」はわずかに 3 つの動詞としか共起せず、総共起頻度も 12 回であった。その中で、「降る」が 9 回で、「降り続く」が 2 回、「続く」が 1 回であった。母語話者の産出では、「泣く」「する」「落ちる」「話す」「降る」「歩く」「したたる」「しみる」「濡れる」「ぬむる」の 10 種類の動詞と共起して、総共起頻度は 18 回であり、多様性が大きいことがわかる。

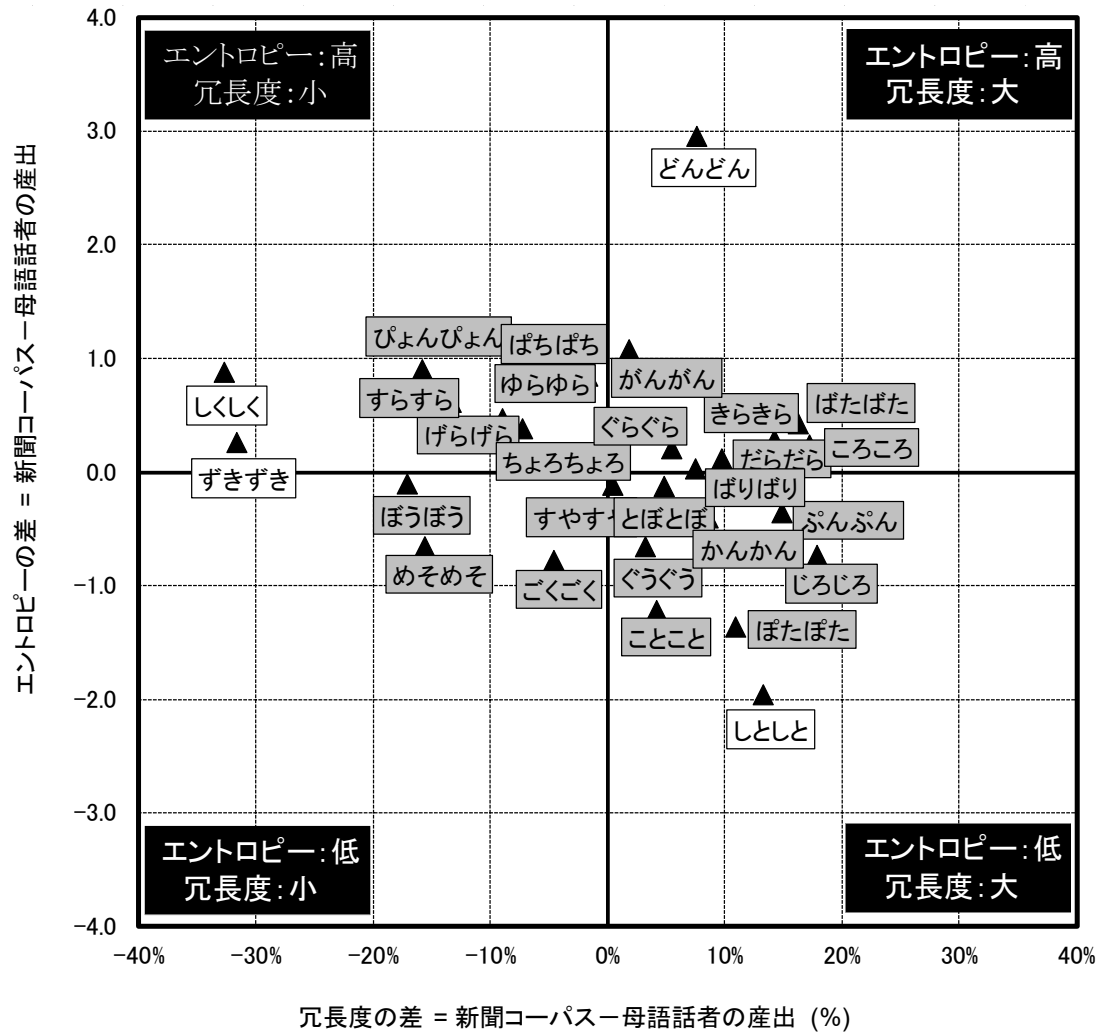


図3：新聞コーパスと母語話者の産出のエントロピーと冗長度の差の散布図

冗長度で、母語話者の産出が新聞コーパスより大きかったのは、「しくしく」の-32.7%と「ずきずき」の-31.6%であった。たとえば、母語話者の産出では、「しくしく」は6種類の動詞と共起したが、「泣く」が圧倒的に多く、総共起頻度の20回中で12回であった。一方、新聞コーパスでは、8種類の動詞が共起したが、「泣く」は、「痛む」と同じ3回であった。ところが、36名の母語話者の産出では、「痛む」は0回であった。このことは、母語話者が「しくしく」から「痛む」を想起していないことを示している。確かに、「しくしく」は、『広辞苑』(第5版)によると多義語であり、2つの意味がある。1つは「声をひそめて弱々しく泣くさま」であり、もう1つは「さほど激しくはないが、絶えず痛むさま」である。母語話者は、「しくしく」を「泣く」状態を示す音象徴語と捉えているのに対して、新聞コーパスでは、「痛む」状態を描く表現としても使用されている。いずれも共起回数そのものは少ないものの、両者に対照的な違いがみられた。また、「ずきずき」の母語話者による産出では、4種類の動詞と共起したが、「痛む」が7回中4回で、「膿みだす」「高潮する」「Xする」がそれぞれ1回ずつであった。一方、新聞コーパスでは、5種類の動詞と共起し、やはり「痛む」が多く、10回中4回であるが、「Xする」も3回あり、「痛み始め

る」「うづく」「締め付ける」がそれぞれ1回ずつであった。『広辞苑』(第5版)によると「ずきずき」は単義語であり「脈打つように連続して痛みが襲うさま」を示す。母語話者も新聞コーパスも基本的に「痛み」についての想起が多かった。

以上のように、記述的な考察では、新聞コーパスと母語話者で、わずかではあるが異なる共起パターンを示す音象徴語がみられることがわかった。

5. 総合考察

音象徴語は、幼少の頃から擬音や擬態の感覚が養われて、共感的な意味拡張 (Akita, 2009; 大久保, 1967) とともに、動作、活動、状態を示す多様な表現へと拡張していくと思われる。この言語習得のプロセスで、音象徴語は、動詞の共起情報とともに、おもに副詞として (羽佐田, 2005; 李, 1998, 2000; スコウラップ, 1993a, 1993b; 田守, 2002; 田守・スコウラップ, 1999; Toratani, 2007), メンタルレキシコンに記録され、成人の日本語母語話者では、多様な動詞とともに産出されるようになると仮定される。一方、新聞は、複数の記者によって一般大衆に正確な情報を伝達するために、簡潔でわかりやすい表現を目指して書かれたものである。その意味で、新聞コーパスは、母語話者による言語表現を標準化したものであると想定することができよう。もしこの想定が正しいとすれば、日本語母語話者による音象徴語に対する動詞の産出における共起パターンは、新聞コーパスの共起パターンと類似したものになると仮定することができる。本研究では、28種類の音象徴語を対象として、共起する動詞のパターンをエントロピーと冗長度で指標化して、新聞コーパスと日本語母語話者を比較した。

本研究の結果、新聞のコーパスから得られた動詞の頻度パターンは母語話者の産出と類似していることがわかった。特に、エントロピーは、母語話者の産出と新聞コーパスとで強い関係を示し、さらに両者の共起パターンに違いが認められなかった (表2を参照)。また、冗長度では、相関は弱かったものの、やはり両者に違いがなかった (表3を参照)。そのため、母語話者の音象徴語に対する動詞の産出と新聞コーパスでの動詞頻度のパターンが類似していることが示された。これには、母語話者の動詞の想起が、新聞コーパスの共起パターンを反映していることが考えられる。コーパスもあくまで言語の産出であり、母語話者のメンタルレキシコンに記録された語彙の想起を具現化したものであると想定できよう。また、逆に、新聞コーパスの共起パターンが、母語話者の動詞の想起に影響することも考えられる。新聞は複数の記者により、一般大衆に情報を伝達する目的で標準的な表現で書かれている。そのため、読者が新聞から音象徴語と動詞の共起パターンを学び、それが母語話者のメンタルレキシコンを構築していくというプロセスも想定される。このように、母語話者の産出と新聞の共起頻度は、お互いに密接な関係をもって成り立っていると思われる。

ただし、新聞コーパスでは、母語話者の産出よりも、分布にもう少し大きな広がりが見られ、記述的な考察で (図2を参照)、いくつかの例外的な音象徴語が観察された。特に、「どんどん」は、新聞コーパスでも例外的に多様性に富んだ音象徴語で、エントロピーが非常に大きく、7.81となった (表1を参照)。冗長度の指標で示した共起パターンをみると、「じろじろ」と「きらきら」は特定の動詞と頻繁に共起し、規則性の高い音象徴語であるという共通の特徴を持っている。

新聞コーパスと母語話者の産出の共起パターンの違いについて、差を散布図に描き、記述的に考察してみると(図3を参照)、ほとんどの音象徴語が、エントロピーと冗長さの差が0に近いところに集まっていた。このことは、*t*検定の結果が示すように、新聞コーパスと母語話者が大枠で類似していることを示している。それでもやや中心から外れていると思われる音象徴語が4つみられた。新聞コーパスでエントロピーがきわめて大きく、多様な動詞と共起する「どんどん」、逆に日本語母語話者のエントロピーのほうが大きかった「しとしと」、新聞コーパスの冗長さが大きく、強い規則性がみられた「しくしく」と「ずきずき」であった。以上のように、例外であると思われる音象徴語が、28語中4語(14.29%)あった。

母語話者の産出は、30秒という限られた時間内での36名の大学生による集計であり、特にテーマがあって産出しているわけでもない。それにもかかわらず、全体として、母語話者の産出と新聞コーパスは類似した共起パターンを示した。36名の30秒という制限でも、新聞コーパスの音象徴語と動詞の共起パターンと類似した結果が出たことを基準にして、今後、産出時間および人数を増やすなどして、コーパスの共起パターンとの類似性を比較して、コーパスが母語話者の産出を真に反映したものであるかどうかを検討していく必要がある。また、音象徴語は、母語話者の習得過程から、動詞との共起パターンが強く予想される例である。そこで、形容詞と名詞の共起、主語・目的語の名詞と動詞の共起など、より一般的な共起パターンでも、音象徴語と同じように、新聞コーパスと母語話者が類似した結果を示すかどうか検討していきたい。

謝辞

本研究は、挑戦的萌芽研究「動詞と共起する名詞群の検索エンジンの構築と読解および聴解に対する共起知識の貢献」(研究課題番号:16K13242, 代表:玉岡賀津雄)の助成を受けた。

文献

- Akita, Kimi (2009) *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics*. Doctoral dissertation, Kobe University.
- Asano, Michiko and Kazuhiro Yokosawa (2013) Determinants of synaesthetic colours for different types of graphemes: Towards a comprehensive model. *Visual Cognition*, 21(6): 674-678.
- Chang, Andrew C. (ed.) (1990) *A thesaurus of Japanese mimesis and onomatopoeia: Usage by category*. Tokyo: Taishukan.
- Den, Yasuharu and Masakatsu Inoue (1997) Disambiguation with verb-predictability: Evidence from Japanese garden-path phenomena. *Proceedings of the Nineteenth Annual Conference on the Cognitive Science Society*. 179-184.
- Hamano, Shoko (1998) *The Sound-Symbolic System of Japanese*. CSLI Publications.
- Inoue, Masakatsu and Yasuharu Den (1999) Influence of verb-predictability on ambiguity resolution in Japanese. *Proceedings of the Second International*

- Conference on Cognitive Science and the Sixteenth Annual Meeting of the Japanese Cognitive Science Society Joint Conference*. 499-502.
- Miyaoka, Yayoi, & Tamaoka, Katsuo (2005) An investigation of the right-hand head rule applied to Japanese affixes. *Glottometrics*, 10: 45-54.
- Shannon, Claude E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, 27: 379-423 (Part I) and 623-656 (Part II).
- Tamaoka, Katsuo, Hyunjung Lim and Hiromu Sakai (2004) Entropy and redundancy of Japanese lexical and syntactic compound verbs. *Journal of Quantitative Linguistics*, 11(3): 33-250.
- Toratani, Kiyoko (2007) An RGG analysis of manner adverbial mimetics. *Language and Linguistics*, 8: 311-342.
- Williams, Edwin. (1981) On the Notions 'Lexically Related' and 'Head of a Word'. *Linguistic Inquiry*, 12: 245-274.
- Williams, Joseph M. (1976) Synaesthetic adjectives. *Language*, 52: 461-478.
- 浅野鶴子 (編) (1978) 『擬音語・擬態語辞典』講談社.
- 天沼寧 (編) (1974) 『擬音語・擬態語辞典』東京堂出版.
- 有本卓 (1982) 『確率・情報・エントロピー』森北出版.
- 飯田香織・玉岡賀津雄・初相娟 (2012) 「中国人日本語学習者の音象徴語の理解」『日中言語研究と日本語教育』5, 46-54.
- 井上雅勝 (2000) 『ガーデンパス現象に基づく日本語文理解過程の実証的研究: 予測的処理の可能性』博士論文, 大阪大学.
- 岩崎純一 (2009) 『音に色が見える世界: 「共感覚」とは何か』PHP 新書.
- 内元清貴・関根聡・井佐原均 (2001) 「最大エントロピーモデルに基づく形態素解析: 未知語の問題の解決策」『自然言語処理』8 (1), 127-141.
- 大久保愛 (1967) 『幼児語の発達』東京堂.
- 大坪併治 (1989) 『擬声語の研究』明治書院.
- 海保博之 (1989) 「第 1 講: 情報をはかる—エントロピー・情報伝達量・冗長度」海保博之 (編) 『心理・教育データの解析法 10 講—応用編』14-26. 福村出版.
- 笈寿雄・田守育啓 (2003) 『オノマトピア: 擬音・擬態語の楽園』勁草書房.
- 影浦峽 (2000) 『計量情報学: 図書館/言語研究への応用』丸善.
- 影山太郎 (1993) 『文法と語形成』ひつじ書房.
- 角岡賢一 (2007) 『日本語オノマトペ語彙における形態的・音韻的体系性について』くろしお出版.
- スコウラップ・ローレンス (1993a) 「日英オノマトペの対照研究」『言語』22 (6), 48-55.
- スコウラップ・ローレンス (1993b) 「日本語の書きことば・話しことばにおけるオノマトペの分布について」笈寿雄・田守田守 (編) 『オノマトペ: 擬音・擬態語の楽園』77-100. 勁草書房.
- 鈴木崇史・影浦峽 (2011) 「名詞の分布特徴量を用いた政治テキスト分析」『行動計量学』38 (1) : 83-92.

- 瀬戸口律子 (1985) 「中国語の“象声詞” 日本語の擬音語との比較を中心に」『大東文化大学紀要』23, 341-357.
- 武田みゆき (2000) 「日・中擬音語の語彙度—人間の活動に関する音と外界音を中心に—」『ことばの科学』13, 171-186.
- 武田みゆき (2001) 「中国語にみる共感覚表現についての一考察」『ことばの科学』14, 107-118.
- 玉岡賀津雄 (2010) 「コーパス分析の研究例 3: 語彙的・統語的複合動詞の特徴についての計量的解析」中本敬子・李在鎬・黒田航 (編) 『新しい認知言語学研究法入門』184-199. ひつじ書房.
- 玉岡賀津雄 (2016) 「共起表現研究のためのコーパス検索入門」『レキシコンフォーラム』7, 239-264. ひつじ書房.
- 玉岡賀津雄・宮岡弥生・林炫情 (2003) 「エントロピーと冗長度で表現の多様性と規則性を表す試み: 韓国語系日本語学習者の敬語表現を例に」『日本語科学』14, 98-112.
- 玉村文郎 (1989) 「日本語の音象徴語の特徴とその教育」『日本語教育』68, 1-12.
- 田守育啓 (1993) 「日本語オノマトペの統語範疇」笈寿雄・田守育啓 (編) 『オノマトペ: 擬音・擬態語の楽園』17-75. 勁草書房.
- 田守育啓 (2001) 「日本語のオノマトペの語形成規則」『言語』30 (9), 42-49.
- 田守育啓 (2002) 『オノマトペ: 擬音・擬態音をたのしむ』東京: 岩波書店.
- 田守育啓・ローレンス・スコウラップ (1999) 『オノマトペ: 形態と意味』くろしお出版.
- 中俣尚己 (2015) 「初級文法項目の生産性の可能性—動詞に接続する文法項目の場合—」『計量国語学』29 (8), 275-295.
- 中俣尚己 (2016) 「「ている」の意味分類と生産性」『計量国語学』30 (7), 417-426.
- 羽佐田理恵 (2005) 「副詞の視点から見た感情を表す音象徴語: その分析過程から導かれた問題点への取り組み」武内道子 (編) 『副詞的表現をめぐって』175-211. ひつじ書房.
- 浜野祥子 (2014) 『日本語のオノマトペ—音象徴と構造—』くろしお出版
- 針生悦子・趙麗華 (2007) 「日本語における有声音/無声音対比の擬音語と大小の対応づけ: 日中大学生による判断の比較」『日本認知科学大会発表論文集』23, 66-67.
- 飛田良文・浅田秀子 (2002) 『現代擬音語・擬態語用法辞典』東京堂出版.
- 深田淳 (2007) 「日本語用例・コロケーション情報抽出システム『茶漉』」『日本語科学』22, 161-172.
- 堀淳一 (1979) 『エントロピーとは何か』講談社ブルーバックス.
- 山梨正明 (1988) 『比喩と理解』東京大学出版会.
- 山梨正明 (2007) 「メタファーと認知のダイナミクス—知のメカニズムの修辭的基盤」, 楠見孝 (編) 『メタファー研究の最前線』3-29. ひつじ書房.
- 李殷娥 (1998) 「日本語と韓国語のオノマトペに関して」『国際開発研究フォーラム』10, 73-88.
- 李殷娥 (2000) 「日本語と韓国語のオノマトペに関する対照研究」名古屋大学博士論文.

(2016年8月22日受付, 2017年1月16日再受付)

Paper B

Investigation of the Similarity between Newspaper Corpus
and Native Speakers' Production in Collocation Patterns
of Sound-symbolic Words and Verbs

TAMAOKA Katsuo (Nagoya University)

Abstract:

Newspapers articles are written by reporters for the general public, so provide accurate information using simple standard expressions. However, it is not clear whether written texts in newspapers reflect typical language production by native speakers. It is assumed that mature native Japanese speakers produce various sound-symbolic words—such as onomatopoeia and mimesis, which are cultivated through childhood experience—with various related verbs. The present study investigated the similarities of collocation patterns for 28 different sound-symbolic words co-occurring with verbs, by comparing nine years of Asahi Newspaper articles (1991-99) with verbal production by 36 native Japanese speakers within 30 seconds. No significant differences were found in either the variation criterion of entropy or the regularity criterion of redundancy for collocational patterns between newspaper corpus and native speakers' production. The result indicated a great similarity between newspaper corpus and native speakers. Exceptional words were only found in 4 out of 28 sound-symbolic words from the descriptive perspective.

Keywords: sound-symbolic words, onomatopoeia, collocation pattern, entropy, redundancy, newspaper corpus, native speaker's production