ID	KK300702
種別	研究ノート
タイトル	「ている」の意味分類と生産性
Title	Meaning Classification and the Productivity of -*teiru*
著者	中俣　尚己
Author	NAKAMATA Naoki
掲載号	30巻7号
発行日	2016年12月20日
開始ページ	417
終了ページ	426
著作権者	計量国語学会

計量国語学 30 巻 7 号（2016 年 12 月）pp.417-426.

研究ノート

「ている」の意味分類と生産性

中俣　尚己（京都教育大学）

要旨

本稿は中俣（2015a）で提案した生産性指数の計算方法を改良し，複数の意味を持つ語に対して各意味ごとの生産性を計算することを提案するものである．ケーススタディーとして，「ている」20,000 例を【継続】【結果存続】【経験】【状態】の 4 つの意味に分類し，生産性指数を計算したところ，上記の順番で生産性が低くなることが確認された．この結果はこれまでに行われた「ている」の意味ごとの習得研究の結果とも一致するものである．また，研究の副産物として，各動詞の「ている」形がどのぐらいの割合でそれぞれの意味になるのかというリストを作ることができた．その結果，多くの動詞が，いずれかの意味に偏りを見せることがわかった．

キーワード：生産性指数，標準化 TTR，BCCWJ，動詞の「ている」形の意味リスト，習得順序

1．はじめに

本稿は中俣（2015a）で提案した生産性の計算方法について，複数の意味を持つ形式の場合に各意味ごとの生産性が計算できなかったという弱点を克服することを目的とする．また，その目的を実現するにあたり，コーパスに出現する多数の「ている」に意味分類のタグ付けを行った．これまで「ている」の意味分類についての研究は数多く行われているものの，日本語の動詞に対して網羅的にどの動詞が「継続」でどの動詞が「結果存続」になるかタグ付けを行った試みはなく，動詞を動作動詞と変化動詞に分類したデータ自体も有益なものとなると考えられる．

2．生産性指数とその問題点

本節では中俣（2015a）の生産性に関する研究を紹介する．2.1 で生産性について，2.2 で生産性指数について説明し，2.3 で生産性指数の計算方法の弱点とその解決策について述べる．

2.1　生産性とは

生産性とは以下のように定義される．

生産性の定義

ある形式 X が一定の関係 R で結びつく要素の多寡の度合い P を生産性と呼ぶ．

（中俣 2015a:275）

「生産性」は長らく接辞研究など形態論の分野で議論されてきた概念であるが，例えば機能語とそれに前接する動詞といった他の関係にも拡張しうるものである．機能語の生産性を考えることは，その項目を応用性の高い文型スキーマとして教えるのがよいのか，あるいは，よく使われる限られた組み合わせだけを教えたほうがよいのか，ということを考える基盤となるため，教育上有用である．

2.2　生産性指数とは

生産性とはバラエティの豊富さの尺度であり，Type 頻度に注目すればよいが，Type 頻度は Token 頻度に大きな影響を受けるため，補正が必要である．中俣（2015a）では下式を生産性指数として提案した．これは，Guiraud Index または Root TTR（Type Token Ratio）と呼ばれる語彙多様性（Lexical Diversity）の式を流用したものである．

$$\text{生産性指数} = \frac{\text{共起項目の Type 粗頻度}}{\sqrt{\text{対象項目の Token 頻度}}}$$

2.3　生産性指数の問題点とその解決策

生産性指数は計算も簡便な指数であるが，１つ問題がある．大きな母集団からサンプルを抽出した時の計算結果が，母集団全体に対しての計算結果と異なるという点である．これは生産性指数の計算式が，サンプルに存在するすべてのデータを計算に利用することから発生する問題である．BCCWJ に出現する「ている」の生産性指数がサンプル数によって変化する様子を下の表１に示す[1]．

表 1 :「ている」のサンプル数の変化に伴う生産性指数（中俣 2015a）の変化

サンプル数	985,113 （母集団）	50,000	10,000	5,000	1,000	500
生産性指数	21.06	23.74	23.41	22.34	16.63	13.28

サンプルを抽出した時に計算結果が異なるということが具体的にどのような問題を引き起こすかというと，多義語における各意味ごとの生産性を計算する場合にネックとなる．仮に，コーパス調査で得られたすべての用例を意味分類した上で，各意味の生産性を計算する場合には問題は生じない．しかし，例えば BCCWJ における「てい

[1] データは中俣（2014）を作成するために 2011 年 10 月に中納言 Ver1.0RC2 を用いて収集したものを利用した．キーの品詞を動詞にし，後方共起の１語目の語彙素を「ている」にして検索した．長単位検索モードを使い，動詞の直後に「ている」が接続するパターンと，助動詞を１つ挟んで「ている」が接続するパターンの双方を調べた．ただし，現在この方法で検索を行うと，1,004,880 例となり，19,767 例のズレがある．このズレは当時のバージョンの中納言で結果を 10 万例以下にしてダウンロードするために，条件を非常に細かく分割して複数回検索した結果，若干の漏れがあったことによるものと判明した．最も大きな漏れは出現書字形が「でい」となるもの（連用形と未然形）で 19,673 例が該当する．

る」の出現数は 100 万例に近く，このすべてを意味分類することは非現実的である．このような場合は，ランダムにサンプルを抜き出し，そのサンプルに対して人手で意味分類を行い，各意味の「ている」について記述していくという手法が穏当であろう．ランダムに抽出されたサンプルの特性はその数が十分であるならば母集団の特性と一致することが保証されるからである．しかし，上述の生産性指数の計算式では，サンプル抽出した対象に対しては使用することができない．

そこで，本稿では，サンプル抽出を行った場合には，別の計算式を用い，その結果を本来の生産性指数に変換するという手法を提案する．

中俣（2015a）では生産性指数の計算式を選定する際，7 つの指標を候補にあげ，その中から Guiraud Index が選ばれた．この時，他の指標候補の中で Guiraud Index と最も高い相関を示した指標に標準化 TTR（Standardized TTR）がある．

標準化 TTR とは母集団を適宜切り分け，その各サンプルごとの TTR を平均するというものである．中俣（2015a）では 1,000 サンプルごとの TTR を平均した．計算手法にサンプリングを含んでいるため，サンプル抽出されたものに対して計算を行っても，母集団全体に対して計算した結果に相違しないという利点がある[2]．なお，Guiraud Index で同様に 1,000 サンプルごとの値を求めたとしても，Guiraud Index サンプル抽出をした段階で元の値から変化するため，それを平均したとしても全体からかけ離れた値になってしまう．あくまでも TTR を使う必要がある．

先ほどと同様にサンプル数の変化に伴う標準化 TTR の値を示すと，表 2 のようになり，ほぼ安定していることがわかる．

表 2：「ている」のサンプル数の変化に伴う標準化 TTR（1,000 サンプル）の変化

サンプル数	985,113（母集団）	100,000	50,000	10,000	5,000	1,000
標準化 TTR	0.5102	0.5103	0.5094	0.5193	0.5232	0.5260

この標準化 TTR と中俣（2015a）の生産性指数（Guiraud Index の式を使ったもの）の相関係数は 0.815 であり，散布図にすると図 1 のごとく，ほぼ直線的な関係にあることがわかる．Excel で回帰式を計算すると，一次関数 y=39.151x+1.6202 が得られる．

[2] このような利点にも関わらず中俣（2015a）において標準化 TTR を採用しなかった理由は，まず計算の難易の度合いが挙げられる．昨今，一般の日本語教師でも利用しやすい日本語文章難易度判定システムや，Ninjal LWP for BCCWJ といった web アプリケーションで，延べ語数と異なり語数は標準的に表示されるようになってきており，これらを代入すれば Guiraud Index は容易に求められる．一方，標準化 TTR は母集団をサンプルに切り分けると言葉で書けば簡単だが，実際に検索結果などに対してそれを行うにはある程度プログラミングまたは Excel に習熟する必要があり，手間もかかる．また別の理由としては計算結果のレンジの違いが挙げられる．103 項目に対して計算を行った結果，標準化 TTR では 0.03 から 0.55 と狭い範囲に収まるが，Guiraud Index ではこれが 0.53 から 27.10 と幅広く分布する．様々な議論をする上で，後者の方が利用しやすいと判断した．

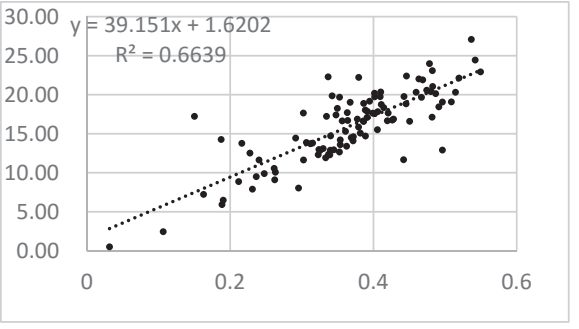

図 1：生産性指数（縦軸）と標準化 TTR（横軸）の関係

よって，サンプル抽出した結果に対して，まず標準化 TTR を計算し，得られた値を前述の回帰式に代入すれば中俣（2015a）の生産性指数の推測値が得られることになる．また，表 2 によれば，標準化 TTR と 1000 サンプルの TTR はほぼ等しいため，ランダムに抽出された 1,000 例さえあれば，生産性指数の推測値を求めることも可能である．ただし，サンプリングにはバラつきも存在するため，可能であればそれよりも多いサンプルから標準化 TTR を算出して使うほうが望ましい．

次節以降では，標準化 TTR を用いて「ている」の意味ごとの生産性指数を計算する．

3．手法

生産性指数の新しい計算方法を「ている」について試行した．以下，3.1 でデータの収集方法について述べ，3.2 で採用した意味分類について述べる．3.3 では「ていた」と「ていない」の処理について述べる．

3.1　データについて

コーパス検索アプリケーション「中納言」を利用し，『現代日本語書き言葉均衡コーパス』（BCCWJ）から長単位検索を用い，「ている」に前接する動詞 817,688 例を抽出した．中俣（2014，2015a）ではこれに加え，動詞の後に助動詞 1 つを挟んで「ている」が接続する 167,425 例も対象にしたが，今回は含めなかった．これは，「ている」データのタグ付けを通して，それぞれの動詞がどの意味分類になりやすいかというデータをまとめることも目的としているからであり，あくまでも動詞の意味分類を考えるために，動詞が直前にあるもののみを選んだ．助動詞を含んだ場合の生産性指数は 21.06，含まない場合の生産性指数は 20.58 である．他方，この方法では「ている」のみならず「ていた」「ていない」なども含まれるが，それらはデータに含め，同様に意味分類を行った．また，「ている」だけでも十分なデータが得られたため，縮約形の「てる」は対象としなかった．

3.2　「ている」の意味分類

次に得られた 817,688 例からランダムに 20,000 例を抽出し，「動詞＋ている」を【継

続】【結果存続】【経験】【状態】に分類した.

分類のカテゴリー数であるが，例えば庵（2001）は「進行中」「結果残存」「繰り返し」「効力持続」「記録」「完了」「反事実」「単なる状態」の 8 つに分類しているが，ここまで細かい分類はアスペクト研究としてはともかく，本研究が目的とする教育への応用には馴染まない．極論すれば「継続」の用法と「結果存続」の用法のどちらを先に導入すればよいのかに答えるデータを作るのであれば，動詞の継続性の有無に着目し，この 2 つに分類してしまえばよい．しかしながら，実際のデータに目を通していると，2 分類に振り分けるのが困難な例に出会うのも事実であり，ある程度一貫して分類するために 4 つのカテゴリーを用意した.

以下に説明と例を示す．例文の後の記号は BCCWJ のサンプル ID である.

①【継続】　継続性を持った事態が，完了せずに継続していることを表す.

(1)　一日中，寝られるわけなんてなくて，薬を飲んで寝ているのをわかっているから起こそうとかしない．（OY05_00235）

(2)　僕は，今博多に行くかどうか迷っている．（PB49_00131）

(1)(2)とも継続性のある事態がまだ完了せず，発話時点においても継続していることを意味する．(2)のように心理的な内容を表すものも含める.

②【結果存続】　動詞で表される事態は参照時にすでに完了しており，その結果状態が発話時においても継続していることを表す．きっかけとなった事態がある.

(3)　シールは濡れていると正常に機能しません．（PB3n_00085）

(4)　もちろん，大学教育研究センターは，そういう努力をこれまでもしてきておられることを承知しています．（PB43_00094）

(3)は濡れる＝水滴がかかるという事態があり，その後水滴が付着している状態を表している．(4)も承知する＝情報を得るという事態は発話時以前に存在し，その情報を得た後の知識の状態を述べている.

③【経験】　過去に起こった事態のうち，その直接の結果が参照時まで残っているとは言えないもの.

(5)　一年前，ノリ不作の犯人は諫早干拓だとさんざん国民を煽ったが，このノリ大豊作に対してはほう（ママ）かむりを通した．さすがに地元紙や地元テレビ局は逃げることはできず，このノリ豊作を報道している．（PB33_00098）

(6)　グリム兄弟は，ハーナウで兄ヤーコブが 1785 年に，弟ウィルヘルムが 1786 年に生まれている．（LBh9_00235）

これらはどちらも過去，時間軸の中の 1 点で起こった事態である．この用法は工藤（1995）のパーフェクトに含まれ，現在まで「効力」が残っているか否かについて議論がある（江田 2013）．しかし，本研究では効力の有無は捨象し，過去の事態でその直接の結果について述べているのではないものを【経験】とした.

④【状態】　過去にきっかけとなる事態はなく，現時点での状態のみを示すもの.

(7)　八十八年実績でみると，米国は二国間 ODA の八十四％を占めている．（OW3X_00307）

(8)　決して無分別な好戦者ではなく，待つべきは待ちつづける我慢強さを備え持っています．（LBq6_00021）

(7)は占めるという具体的な事態があったわけではなく，一定の割合があることを「占めている」と表現しているだけである．(8)もある特定の時点で「備え持つ」という変化が起こったということはできないため，【状態】とした．

しかし，現実には【結果存続】と【状態】で迷う例もあった．

(9)　大阪南港は九州・四国・南西諸島とを結ぶ，大阪の海の玄関．市内からニュートラムや高速道路が伸びていて足の便が良い．（PM21_00718）

この「伸びていて」はあくまでも交通手段が存在していることを表しており，「伸びた」結果を表しているとは考えにくい．しかし，この「伸びた」を「建設した」と解釈すれば，建設するという事態の結果とも考えられる．このような場合も，必ずどちらか 1 つのカテゴリーに振り分けた（(9)は【状態】）．【状態】は「似ている」のように明らかに変化の結果とは考えられない例などのために設けたカテゴリーである．

また，先行研究でよく設けられる意味分類に「習慣」（吉川 1973）「反復」（工藤 1995）「繰り返し」（許 2000，庵 2001，江田 2013）と呼ばれるものがある．本研究では時間的な多回生起は「ている」文に限られたことではないこと[3]，また動詞と「ている」形の関係を見たいことから，日本語記述文法研究会（2007）の立場に従い，(10)のような時間的な繰り返しは【継続】に分類した．

(10)　毎日生ニンニク食べているので，部屋に来られると臭いが残ってたまらん！！（OC09_09619））

3.3　「ていた」「ていない」の意味分類

「ていた」「ていない」に関しても，同様に分類した．なお，全体の 21.8%が「ていた」，4.8%が「ていない」（「ていなかった」を含む）であった．(11)から(14)は「ていた」の例であり，(11)が【継続】，(12)が【結果存続】，(13)が【経験】，(14)が【状態】である．(15)から(18)は「ていない」の例であり，(15)が【継続】，(16)が【結果存続】，(17)が【経験】，(18)が【状態】である．

(11)　たったいま声もたてずに泣いていた彼女からは想像もつかないほどの感情のほとばしりだった．（LBk9_00150）【継続】

(12)　正月休みに，雪絵は晃児の実家に戻っていた．（PM51_00697）【結果存続】

(13)　家なき子に出ていたんですが，安達ゆみをいじめてる役の女の子なんですが，目が結構きつい感じの子，今はどうしてるのですか？（OC01_09627）【経験】

(14)　東の方に，山と谷間のすばらしい風景が広がっていた．（OB4X_00105）【状態】

(15)　それにエーブは，ゆりかごのなかにおとなしくしていないで，元気よく自分を主張しました．（LBhn_00026）【継続】

(16)　六十年代には大衆はものごとがわかっていない，だからわれわれが教えてあげなければならない，と思ってたの．（LBi3_00078）【結果存続】

(17)　というより，現在残る記録によると，りんは帰国後七年にわたってイコンを描いていないのだ．（PB47_00066）【経験】

[3] ル形の文であっても，「今日の昼はうどんを食べる」は 1 回生起であるのに対し，「昼はいつもうどんを食べる」は多回生起で習慣を表す．

(18)　同一世代に属していない.（LBn3_00031）【状態】

4．結果と考察

4.1　意味分類ごとの生産性

意味分類ごとの生産性を表 3 に示す.「ている」全体の生産性（中俣 2015a）と比較するため，標準化 TTR を 2.3 で述べた回帰式に代入した結果を示す.

表 3：「ている」の意味分類と生産性

	継続	結果存続	経験	状態
Token 頻度	8,818	8,031	701	2,450
Type 頻度	1,635	1,454	223	451
標準化 TTR	**0.430**	**0.399**	**0.318**	**0.223**
生産性指数	**18.5**	**17.2**	**14.1**	**10.4**

「ている」全体の生産性は【継続】が最も生産性が高く，【結果存続】がそれに次ぐ.【経験】は生産性がやや下がり，【状態】は生産性が少なく，前に接続する語が限られていることを意味する．つまり，「似ている」「占めている」といった用法は，日本語教育の文脈では文法項目よりも語彙項目に近いといえる.

参考までに，中俣（2015a）では「ている」全体の生産性指数は 21.05，103 項目の生産性指数の平均値は 15.95 であった.

4.2　意味分類の割合

得られた「ている」に前接する動詞は 3,168 種類あった．個々の動詞について，用例を元に，各意味分類の割合を付与した．Token 頻度の上位 10 語の例を表 4 に示す.

表 4：Token 頻度上位 10 動詞の意味分類の割合

動詞	継続	結果存続	経験	状態
する	64.2%	23.2%	4.5%	8.1%
成る	1.6%	74.8%	1.8%	21.8%
思う	94.5%	5.5%	0.0%	0.0%
持つ	0.7%	76.4%	0.0%	22.9%
知る	0.0%	100.0%	0.0%	0.0%
言う	39.3%	35.2%	21.0%	4.5%
見る	90.9%	7.0%	2.1%	0.0%
考える	98.2%	1.8%	0.0%	0.0%
遣る	94.0%	3.5%	2.5%	0.0%
入る	1.0%	91.1%	0.0%	7.8%

このようなデータを 3,168 の動詞にわたって作成した．このデータは自然言語処理の分野においても重要な役割を果たすと考えられるので，本論文の公開に合わせて著者のウェブサイトで公開する[4]．

上記の表では，上から 5 つ目の「知る」が 100%【結果存続】となっているが，頻度が 50 以上の動詞で継続が 100%を占めるものには「待っている」「感じている」「働いている」「住んでいる」「困っている」があり，【経験】が 100%を占めるものに「述べている」があり[5]，【状態】が 100%を占めるものに「似ている」がある．また，上記の表では，上から 6 つ目の「言う」はどの意味分類が優勢か決定しかねるが，全体としてはこのような動詞は稀で，最大の割合が 50%以下のもの（つまり，過半数を超える意味割合がないもの）は全部で 116 動詞，全体の 3.7%にすぎない．他方，90%以上が 1 つの意味分類になる動詞は 2,761 動詞，全体の 87.2%を占める．つまり，ほとんどの動詞はどれか 1 つの意味分類に偏り，これを利用した動詞の分類が可能となる．

4.3　習得難易度との比較

本研究の結果は，習得研究の成果とも合致する．中俣（2015b）は生産性が高い形式（19 以上）＞生産性が低い形式（13 以下）＞生産性が中程度（13～19）の形式の順に習得されると主張している．ただし，この境界の数値は便宜的に定められたもので，実際にはクリアカットには分断できない．表 3 を見ると，19 に近い【継続】は生産性が高いため習得が早く，続いて，生産性が 13 以下の【状態】が習得され，生産性が中程度の【結果存続】と【経験】が遅いと予測される．

「ている」の大規模な習得研究である許（2000）によれば「ている」の習得順序は下記の通りである．（【】は本研究の分類名）

運動の持続（±長期）【継続】→性状（＋可変性）【状態】→性状（－可変性）【状態】→繰り返し【継続】→結果の状態【結果存続】→状態の変化【結果存続】→経歴・経験【経験】

先に立てた予測と実際の習得研究のデータは一致しており，生産性指数の計算方法の妥当性や難易度との関係を支持するデータと言えよう．

5．おわりに

本稿では Guiraud Index の代わりに標準化 TTR を使うことで，用法別に生産性を計算する方法を提案し，実際に「ている」の用法別の生産性を計算した．【継続】の生産性が高く，【結果存続】，【経験】と続き，【状態】は低いという結果が得られた．難易度や日本語教育における導入順を考える上で参考になるデータである．また，研究の過程で 20,000 例の「ている」の意味分類を行ったが，その結果，多くの動詞が 1 つの意味分類に偏っていることがわかった．この「どのような動詞がどのような意味になりやすいか」という割合も汎用性のあるデータであると言える．

---

[4] なお，公開データは頻度を表示したものにする．

[5] 江田（2013）は効力持続（本稿の「経験」）の「ている」は新書において，ほとんどが思考・言語の動詞に偏ることを指摘しており，本研究のデータもこれと同様であった．このような「ている」は他者の文章を引用する際に欠かせないものであり，教育上重要である．

謝辞

本研究の遂行にあたっては平成 27～30 年度科研費基盤研究（C）「言語使用実態に基づく日本語記述文法の計量的評価法と応用方法の開発」（15K02583，研究代表者森篤嗣）の助成を受けた．また，執筆にあたっては同科研のメンバー，特に森篤嗣氏と茂木俊伸氏の助言を得た．記して感謝申し上げる．

文献

庵功雄（2001）「テイル形，テイタ形の意味の捉え方に関する一試案」『一橋大学留学生センター紀要』4: 75-94.

許夏珮（2000）「自然発話における日本語学習者による「ている」の習得研究」『日本語教育』104:20-29.

工藤真由美（1995）『アスペクト・テンス体系とテクスト』ひつじ書房.

江田すみれ（2013）『日本女子大学叢書 14「ている」「ていた」「ていない」のアスペクト―異なるジャンルのテクストにおける使用状況とその用法―』くろしお出版.

中俣尚己（2014）『日本語教育のための文法コロケーションハンドブック』くろしお出版.

中俣尚己（2015a）「初級文法項目の生産性の可視化―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（2015b）「生産性から見た文法シラバス」庵功雄・山内博之（編）『現場に役立つ日本語教育研究 1 データに基づく文法シラバス』109-128.くろしお出版.

日本語記述文法研究会（編）（2007）『現代日本語文法 3　第 5 部アスペクト第 6 部テンス第 7 部肯否』くろしお出版.

吉川武時（1973）「現代日本語動詞のアスペクトの研究」（金田一春彦（編）1976『日本語動詞のアスペクト』155-327.むぎ書房 に所収）

関連 Web サイト

国立国語研究所（2011）『現代日本語書き言葉均衡コーパス（通常版）』
http://pj.ninjal.ac.jp/corpus_center/bccwj/index.html
（中納言 Ver1.0RC2 による利用）

（2016 年 3 月 18 日受付，2016 年 8 月 1 日再受付）

Mathematical Linguistics, Vol.30 No.7 (December 2016) pp.417-426.

Note

Meaning Classification and the Productivity of *-teiru*

NAKAMATA Naoki (Kyoto University of Education)

Abstract:
This article provides an improved method of calculating productivity index in order to calculate the productivity of each usage of polysemous or multifunctional words. The new method utilizes the standardized TTR, which strongly correlates with the index previously proposed. This case study classifies 20,000 examples of the aspect marker *-teiru* into four meaning groups: continuation, result, experience, and state. The productivity of each group is then calculated, resulting in a value decrease in the same sequence. This result also matches previous studies of the acquisition of *-teiru.* Furthermore, as a byproduct of this study, a list showing which meaning a verb tends to have was created. Most verbs tend to have one of four meanings.

Keywords: productivity index, standardized TTR, BCCWJ,
list of verbs and the meaning of their *-teiru* forms, sequence of acquisition