

国立国語研究所学術情報リポジトリ

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

メタデータ	言語: Japanese 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: 玉岡, 賀津雄, 宮岡, 弥生, 林, 炫情, TAMAOKA, Katsuo, MIYAOKA, Yayoi, LIM, Hyun-jung メールアドレス: 所属:
URL	https://doi.org/10.15084/00002115

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——

玉岡 賀津雄

(広島大学留学生センター)

宮岡 弥生

(広島経済大学)

林 炫情

(広島大学, 日本学術振興会特別研究員)

キーワード

エントロピー, 冗長度, 尊敬と謙譲, 表現の多様性, 韓国語系日本語学習者

要 旨

本研究では、シャノンの通信の数学理論で知られる「エントロピー」と「冗長度」という指標を使って、韓国語系日本語学習者を対象にペーパーテストで測定した知識と、インタビューで測定した運用の差異について検討した。使用された尊敬表現と謙譲表現の種類と頻度を、それぞれ知識と運用とに分けて集計し、表現の種類と頻度のデータから、エントロピーと冗長度を算出した。まず、ペーパーテストとインタビューにおける表現の正答という観点からみると、尊敬と謙譲にも知識と運用にも違いがなかった。本研究では、あくまで基本的な10種類の動詞を扱っており、これらの動詞の知識を基にした運用では正答に大きな違いが無かったのであろう。本研究は、敬語表現の多様性をエントロピーと冗長度で考察することを目的としているので、正答における違いがないことが望ましい。さらに、エントロピーについて分析した結果、尊敬表現と比べて、とりわけ運用における謙譲表現の多様性と不規則性が明らかになった。また、冗長度の分析では、エントロピーが示すほど顕著ではないが、尊敬は知識でも運用でも、同じ表現が繰り返される傾向があるが、謙譲は知識よりも運用の方が表現のバリエーションが多かったことが示された。これらエントロピーと冗長度の二つの指標で描いたプロットは、知識と運用における尊敬と謙譲の違いを鮮明に示した。

1. はじめに

あることを表現する際に、一つしかその表現形式がないということは希である。例えば、誰かにその人がどこかへ行くかどうか質問する場合、「行きますか」、「行かれますか」、「行かれるでしょうか」、「いらっしゃいますか」、「いらっしゃるでしょうか」など多様な表現が可能である。しかし、すべての表現が同じ頻度で使われるわけではない。このような表現の多様性や規則性を一定の指標で測定して、全体的な傾向を考察する方法はないものであろうか。

クロード・シャノン (Claude Shannon) は、『通信の数学理論 (A Mathematical Theory of Communication)』(1948) で、「エントロピー (entropy)」と「冗長度 (redundancy)」という概念を発表した (エントロピーについては、有本, 1982; 堀, 1979; 海保, 1989; 甘利, 1970を参照)。エントロピーは情報量の尺度の一つであり、あいまいさや乱雑度の増減を示す指標である。エントロピーは、表現の種類と使用頻度から一つの値を算出して、不規則性を示すことができる。また、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一つの値を算出して、無駄の程度を表すことができる。つまり、エントロピーと冗長度の尺度を組み合わせることで、ある表現の多様性や規則性が簡単な数値で表すことができるのである。

エントロピーを日本語研究に応用した先行研究として、袋小路 (garden path) 文の研究がある (Den & Inoue, 1997; 井上, 2000; Inoue & Den, 1999)。これら一連の研究では、主語・目的語として動詞と共起する名詞の種類と頻度の一覧から、エントロピーを算出している。そして、エントロピーの高低と袋小路文の曖昧さへの陥りやすさ (これを、「ガーデンパス効果 (garden path effects)」と呼んでいる) とが関係していることを示した。袋小路文の研究 (Den & Inoue, 1997; 井上, 2000; Inoue & Den, 1999) では使用していないが、特定の表現の繰り返しの指標である冗長度も適切な指標となる可能性があるだろう。

本研究では、エントロピーと冗長度が応用できる比較的簡単な擬似実験的 (pseudo-experimental) 手法によるデータ収集を試みた。まず、韓国語母語話者で日本の大学で学んでいる留学生に対して敬語の知識¹を測定するためにペーパーテストを実施した。そして、その後1ヶ月余りの間をあけて同じ留学生に、同じ表現が表出するようにコンテキストを構成した教授と学生という設定でのインタビューを行い、敬語運用を測定した。本研究のインタビューは芝居に近い形式の擬似的な設定である。この種の方法は、2人が向かい合って地図の道順を説明する場面を設定して、特定の表現を収集し大規模な会話コーパスを作り上げたスコットランドのエジンバラ大学とグラスゴー大学のMap Task共同プロジェクト (Anderson, Bader, Bard, Boyle, Doherty, Garrod, Isard, Kowtko, Mcallister, Miller, Sotillo, Thompson, Weinert, 1991; Brown, Anderson, Yule & Shillcock, 1983) を、さらに条件を限定して応用したものである。ペーパーテストとインタビューで使用された尊敬表現と謙譲表現の種類と頻度を集計し、エントロピーと冗長度を算出して比較した。以上のプロセスをもとに本研究では、エントロピーと冗長度が、敬語表現の知識と運用の関係をどのくらい正確に表しうる指標であるかを検討した。

2. 実験の手順

表現の多様性を観察する一つの材料として、敬語表現を選んだ。そして、以下の手順で、知識を測定するためのペーパーテスト、運用を測定するためのインタビューを実施した。

2.1. 被験者の特性

本研究の被験者は、韓国語を母語とする日本語学習者24名である。被験者の性別は、男性が7名で、女性が17名であった。年齢は、最も若い日本語学習者が20歳6ヶ月で、最年長の日本語学

習者が38歳5ヶ月であった。平均年齢は29歳8ヶ月で、標準偏差が4歳5ヶ月であった。これらの被験者は、すでに日本の大学または大学院に入学して勉学および研究に従事している。日本語能力は、中級以上で、超上級も含まれており、日常的に日本語を使っている。

2. 2. 敬語知識および敬語運用テストの実施手続き

敬語テストは2種類ある。一つは、敬語知識を測定するためにペーパーテストの形式で行った。もう一つは、敬語運用を測定するためにインタビュー形式で行った。その手続きとして、まず、被験者24名全員に、ペーパーテスト形式の敬語テストを行い、同じ被験者に対して1ヵ月後に、インタビュー形式の敬語テストを行った。ペーパーテスト形式とインタビュー形式の敬語テストを1ヶ月の期間をおいて行ったのは、両テストでまったく同じコンテキストと敬語項目を用いたためである。ペーパーテストの内容を忘れた頃として、1ヶ月という期間をおいてインタビューを行った。

2. 3. 敬語テストの研究対象項目

本研究は、敬語のうち、尊敬表現と謙譲表現を対象とした。尊敬表現については、五段動詞として「行く」「言う」「帰る」「読む」「書く」の5つを、一段動詞および変格活用動詞として「見る」「食べる」「つける」「やる」「する」の5つの動詞を研究の対象項目として取り上げ、合計で10種類の動詞を研究対象とした。このうち、「行く」「言う」「見る」「食べる」「する」の5つの項目を謙譲表現としても採用した。残りの5つは、尊敬表現と難易度がほぼ同じであると考えられる「待つ」「返す」「届ける」「知らせる」「借りる」を選択した。したがって、謙譲表現も尊敬表現と同様に、ほぼ同じような10種類の動詞を使用した。これによって、尊敬表現と謙譲表現が直接比較できると考える。また、これらの動詞は、初級レベルで出てくる動詞である。本研究の被験者は中級以上の留学生であるため、理解できないということはないと思われる。また、用いる表現も、初級レベルのものに限っているので問題そのものが分からないということはないと考えられる。これらの敬語知識がどのように敬語運用において用いられるかを考察した。

2. 4. ペーパーテストによる敬語知識の測定

ペーパーテストで測定する敬語知識は、先行研究の方法を取り入れて (Miyaoaka & Tamaoka, 2001; 宮岡・玉岡, 2002), 一文ずつコンテキストを示したうえで、辞書形で提示した動詞を適切な表現に直すという形をとった。例えば、「初対面の大学の教授にたずねます」というコンテキストで、「秋の学会へ (行くか)」の下線部を正しい形に直すというものである。項目は、尊敬表現10問、謙譲表現10問、ダミー10問の合計30問である。これらをランダムに並べてペーパーテストの形式で被験者に課した。

2. 5. インタビューによる敬語運用の測定

ペーパーテスト形式の敬語テストを行った1ヵ月後に、同じ被験者に対して敬語運用を測定す

るためにインタビュー形式の実験を行った。この実験では、より自然な発話に近い形式を採集することを目的とした。しかし、完全な自然発話だと、研究の意図に沿った発話データがなかなか得られない。そのため、敬語を用いる状況と動詞をあらかじめ設定した上で、発話を促すという形式をとった。

まず、体系的な分析が可能な範囲内で、できるだけ自然な形で尊敬および謙譲の発話を促すためには、尊敬表現と謙譲表現の実験を別のスタイルで行う必要があった。つまり、尊敬表現は主語を話し相手に、謙譲表現は主語を自分自身にしないといけないことから、場面を2つに分けた。実験者は、被験者の話し相手である先生と、カードを提示する補佐役の計2人である。話し相手となる人物は、被験者が最も敬語を使う可能性が高いと思われる現役の大学の教授にお願いした。被験者の話し相手としては常に同一人物を設定し、話し相手の属性の微妙な違いによって被験者の反応が異なる可能性を排除した。補佐役は韓国語母語話者で、実験の教示はすべて韓国語で行った。従って、被験者は実験の方法を十分理解することができたはずである。インタビューは、すべてテープレコーダーで録音して、後で使われた敬語表現をすべて記述して、分析のデータとして使用した。なお、ターゲットとなる同じ敬語表現の繰り返しは起こらないように場面を設定しており、実際、そういう事例はみられなかった。

2. 5. 1 尊敬表現の運用をみるインタビュー設定

インタビュー形式のテストが始まる前に、教授に被験者と適当な雑談をしてもらい、自然な会話の状況を作った。まず、尊敬表現について実験を行った。はじめに被験者に、教授に質問をしてほしいと伝えてメモ用紙を渡し、教授の答えを書き留めるように指示した。その際、すべての質問項目が書かれた質問紙をあらかじめ被験者に渡しておく、回答を考える時間を与えてしまい、相手の発話に応じて瞬時にかつ臨機応変に答えを作り出さなくてはならない自然発話との乖離が大きくなる。そこで、質問項目を書いたカードを一枚ずつ被験者に手渡し、被験者はその都度それを見て教授である先生に質問するという形式とした。カードには、ペーパーテスト形式の敬語テストと同様に、用いる動詞の辞書形（終止形）を明示し、それを適切な形に変えて先生に質問するように指示した。例えば、「秋の学会へ（行くか）。」というカードを渡すと、被験者は、括弧で囲んだ「行くか」を先生に対する適切な表現に変えて実際に質問するように求められる。この場合だと、「秋の学会へいらっしゃいますか。」や「秋の学会へ行かれますか。」などと先生に聞くことになる。それに対して、質問に答える先生は、自然な会話が流れるように心がけた。この形式で、10種類のカードに10種類の動詞が示され、それを尊敬表現に変えて、先生に質問するというインタビュー形式の発話とした。

2. 5. 2 謙譲表現の運用をみるインタビュー設定

次に、謙譲表現について、尊敬表現とは異なる形式で実験を行った。謙譲表現は、主語が被験者自身でなくてはならないため、話し相手である教授と被験者とのやり取りの中で、謙譲表現を使うような状況となるようにした。教授には、あらかじめせりふとせりふの順序を指示し、すべ

での被験者に対して同じような応対をするように心がけてもらった。被験者には、教授の発話に対する反応を書いたA3の紙を被験者に良く見えるようにその都度提示した。用紙には、ペーパーテストと同様に、回答で用いる動詞を明示した。尊敬と謙譲とで用いる動詞をそろえ、同様に辞書形を紙によって提示している。例えば、用紙に大きく「先生の論文を借りてもいいか聞いてください。(使う動詞：借りる)」と書いたものを見せる。被験者は、これを見て、「先生の論文をお借りしてもいいですか。」や「先生の論文をお借りしてもよろしいでしょうか。」のように質問をする。教授からの働きかけがあるたびに、次の発話内容を用紙で見せ、その指示に従って即座に適切な表現を考えて応対するよう被験者に指示した。この形式で質問ごとに異なる動詞を指定し、10種類の発話を促した。

3. 指標の算出

ペーパーテストとインタビューから得られた動詞の敬語表現項目別のデータを基に、表現の得点、エントロピー、冗長さの指標を算出した²。

3.1. 表現の得点

表現の得点は、与えられたコンテキストの中で適切な表現が用いられているかどうかを基準として得点化した指標である。得点は、丁寧度が高いものを用いるほど高くなるように計算し、語形の誤りのような不正解は0点とした。しかし、尊敬と謙譲とでは丁寧度のバリエーションが異なるため、得点化も異なる。尊敬表現については、「です・ます」のみで答えているものは1点、「れる・られる」は2点、「お～になる」および交替形式の尊敬表現は3点とした。謙譲表現は、「です・ます」が1点、「お～する」および交替形式の謙譲表現は2点とした。したがって、各表現の最高得点は尊敬表現が72点（各動詞の最高得点が3点で24名の被験者）、謙譲表現が48点（各動詞の最高得点が2点で24名の被験者）である。

3.2. エントロピー

エントロピーはクロード・シャノンが情報量の乱雑度の増減を示す尺度として提唱した数式である。海保（1989）が比較的分かりやすく公式とその意味を説明している。シャノンによると、エントロピーは以下の式で計算される。

$$H = - \sum_{j=1}^j p_j \log_2 p_j$$

本研究では、答えとして得られた表現をもとにその種類と頻度を計算して、動詞ごとのエントロピーを計算した。インタビューの「見た」の疑問形の謙譲表現は、6種類が使われた。24名の被験者のうち1名だけは「ご覧になったですか」という表現を用いたので誤りとなり、頻度の集計から除外された。そのため、全体の延べ頻度は23である。まず、各表現の使用頻度を計算し、それぞれの使用頻度を全体の延べ頻度の23で割って比率を出す。各比率の底が2の対数を算出す

る。例えば、「見ましたか」であれば、使用頻度が10で、比率は10を23で割った0.435である。この対数は-1.202となる。そこで、比率の0.435と対数の-1.202を掛けて-0.522を出す。比率と対数の積から得られる値をすべて足して値を正に変えた（-1を掛ける）のがエントロピーである。本研究の被験者数は24名であり、各動詞の正しい表現の最大数は24となるので、表現の多様性を「尊敬と謙譲」と「知識と運用」の2要因の2条件（合計4条件）を同じエントロピーの指標で直接比較できる。

3.3. 冗長度

エントロピーと共にシャノンが提示した有名な概念は冗長度である。冗長度とは無駄の程度を表す指標である。ただし、本研究の場合は、無駄というより、表現の出現頻度に偏りがあることを意味し、同じような表現が繰り返し使われる度合いを示していると考えるべきであろう。シャノンによると、冗長度は、以下の公式で得られる。

$$R = (1 - H/H_{\max}) \times 100(\%)$$

Hはエントロピーであり、 H_{\max} はエントロピー最大値を意味する。エントロピー最大とは、すべてが等しい確率で生起する場合である。つまり、いずれが起こっても不思議ではない混沌とした無秩序の状態を意味する。重なり頻度（表現の数）をJとすると、以下の式でエントロピーの最大値が得られる。

$$H_{\max} = \log_2 J$$

例えば、本研究のインタビューでの謙譲表現「言う」であれば、24名の被験者が16種類の正しい表現をしたので、エントロピーの最大値は、 $H = \log_2 16$ で4.00となる。この数値は、16種類の表現がすべて等しく生起する場合のエントロピーである。さて、冗長度とは、尊敬または謙譲で表現された特定の表現について得られたエントロピーをそのエントロピー最大値で割り、その数値を1から引いて100倍しパーセントで示したものである。一見すると、エントロピーが高くなると冗長度は低くなるという両変数の逆相関を考えてしまいそうであるが、実際には相関係数が-1になることはなく、両変数は独立した指標である。また、エントロピーと同様に、冗長度も、尊敬と謙譲の知識と運用について、値を直接比較できる指標である。

4. エントロピーと冗長度からみる尊敬および謙譲表現

すべての動詞について、尊敬と謙譲の知識と運用の4条件についてエントロピーと冗長度を算出した。そこで、動詞「行く」を例に、エントロピーと冗長度がどれくらい敏感に表現の種類と頻度を反映しているかを検討する。

4. 1. 尊敬表現における知識と運用の違い

ペーパーテスト（知識）での「行く」の敬語表現は、図1に示したように、「行かれますか」（9回）、「行きますか」（7回）、「いらっしゃいますか」（6回）、「行かれるご予定はありますでしょうか」（1回）のわずかに4種類であった。そのため、エントロピーは低く1.75である。しかし、4種類の内3つの表現に頻度が分散しているので冗長度はあまり高くなく12.28%であった。一方、同じことをインタビュー（運用）で聞くと、「行きますか」（9回）、「行かれますか」（7回）、「いらっしゃいますか」（4回）、「出席しますか」（1回）、「いらっしゃるんですか」（1回）、「行かれるんですか」（1回）、「行かれますでしょうか」（1回）と7種類の表現が現れ、ペーパーテストに比べて多様な表現が使われている。その点で、エントロピーはやや高くなり2.24となる。冗長度は、「行きますか」と「行かれますか」の2つの表現が多用され、後の表現がいくつか使われる程度なので、ペーパーテストよりやや高くなり20.06%となる。以上のように、エントロピーと冗長度は、動詞「行く」の尊敬表現の多様性と規則性の様子を示していると言えよう。

4. 2. 謙譲表現における知識と運用の違い

動詞「行く」の謙譲については、尊敬に比べて表現の種類が多くなっている。この傾向は、ほぼすべての動詞について言えるようである。具体的な謙譲表現をペーパーテストとインタビューを別々に図2に示した。ペーパーテストの条件（知識）では、11種類の表現が使われており、延べ頻度は24名の被験者全員が正しい謙譲表現を使っているので、24である。この分布を基にして、エントロピーを計算すると、3.07となり、「行く」の尊敬表現の1.75と比べてより大きい値となっている。これは、ペーパーテストでの謙譲表現は、規則性が低く、多様であることを示している。ただ、冗長度は11.31%で、尊敬表現の12.28%とほとんど変わらない。これは、全体の分布パターンをみた場合、同じ表現が反復される傾向が類似しているからであろうと思われる。一方、イン

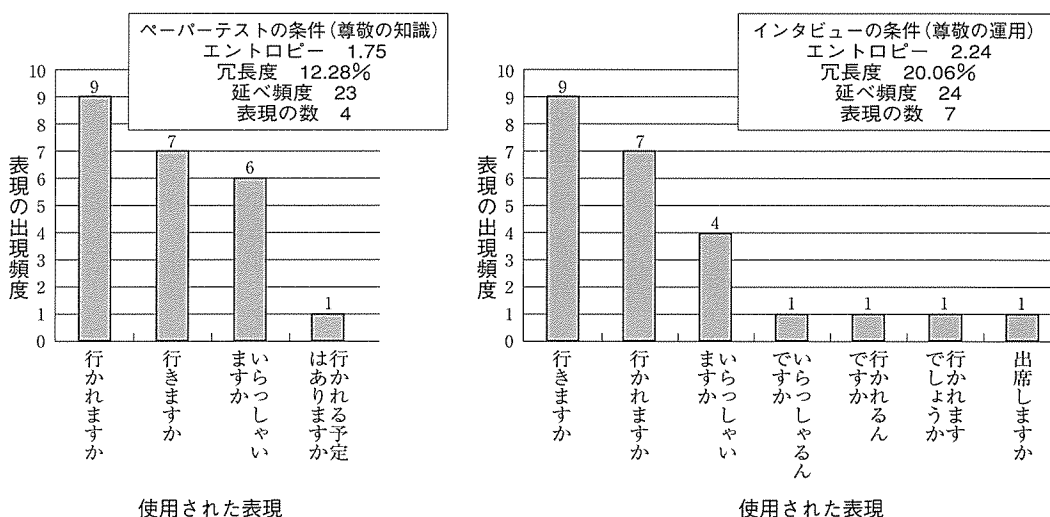


図1 ペーパーテストおよびインタビューでの「行く」の尊敬表現の種類と出現頻度

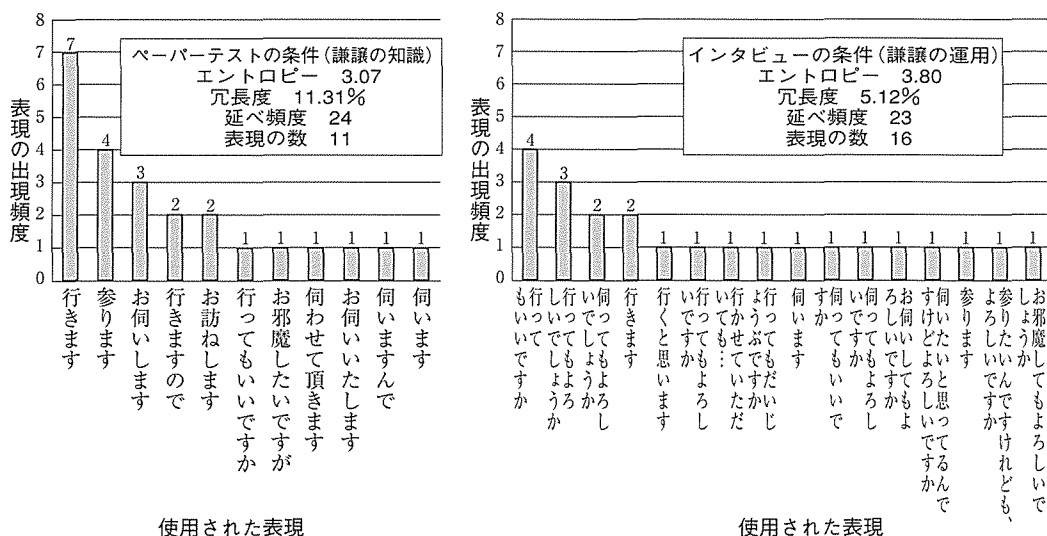


図2 ペーパーテストおよびインタビューでの「行く」の謙譲表現の種類と出現頻度

タビューでは、「行く」の謙譲表現が16種類で多く、表現の頻度は1名だけ誤った表現を使った被験者がいるため延べ頻度が23である。エントロピーはさらに大きくなり、3.80となっている。また、冗長度は、5.42%になり、極めて小さな値である。これは、インタビューでは謙譲表現が多様であるため、ほとんどの被験者がそれぞれ異なる表現を使い、繰り返しがないことを示している。以上のように、尊敬表現も加味して考えて、エントロピーと冗長度が謙譲表現の多様性あるいは規則性を微妙に表していると言えそうである。

5. 指標の分析

動詞「行く」を事例として、尊敬および謙譲表現についてペーパーテストとインタビューの違いをエントロピーと冗長度で考察したが、これらは表現の分布を具現する指標であると考えてよさそうである。そこで、10種類の動詞について、「尊敬と謙譲」および「知識と運用」の2つの変数による違いを、表現の得点、エントロピー、冗長度についての分散分析で検討した。それぞれの指標の平均と標準偏差は表1に示したとおりである。

5.1. 表現の得点についての分散分析

表現の得点については、尊敬が72点満点で謙譲が48点満点と配点が異なるので、これらを別々にして一元配置の分散分析（敬語知識、敬語運用）を行った。尊敬 [$F(1,9)=2.95, p=.120$] および謙譲 [$F(1,9)=2.76, p=.131$] とともに知識・運用の主効果は有意ではなかった。つまり、文法的に正しいかどうかを基準として見た場合、本実験の被験者は、尊敬と謙譲について、ペーパーテストで測定した知識とインタビューで測定した運用について同じような正確さで表現していたことを意味する。ここで、エントロピーと冗長度は、「尊敬と謙譲」の「知識と運用」について直接比

較できる指標であることはすでに述べたが、さらに、「尊敬と謙譲」の「知識と運用」の正答を示す「表現の得点」が、知識と運用で同じであったことは、もともとの正答数に違いがないことを示しているので、エントロピーと冗長度から「尊敬と謙譲」の「知識と運用」の2要因4条件を直接に比較することが保障されたことになる。

表1 敬語知識と敬語運用に関する平均と標準偏差

指標の種類		敬語知識		敬語運用	
		尊敬	謙譲	尊敬	謙譲
表現の得点	平均	47.00	34.50	43.90	31.70
	標準偏差	8.14	3.84	3.76	4.50
エントロピー	平均	1.94	2.19	1.99	3.53
	標準偏差	0.48	0.66	0.38	0.34
冗長度	平均	19.91	14.50	20.54	7.50
	標準偏差	10.71	6.97	6.80	4.25

注：反復測定の分散分析は、項目数20で行っている。被験者数は24名である。

5. 2. エントロピーについての分散分析

エントロピーの指標について、2（尊敬，謙譲）×2（知識，運用）の分散分析を行った。尊敬・謙譲の変数は項目間分析であり，知識・運用は反復測定による項目内分析である。まず，尊敬・謙譲 [$F(1,18)=27.79, p<.001$] の主効果が有意であった。これは，謙譲表現 ($M=2.86, SD=0.83$) の方が，尊敬表現 ($M=1.96, SD=0.42$) よりもエントロピーが高く，より多様で規則性が低い使われ方をしていることを意味している。次に，知識・運用 [$F(1,18)=24.28, p<.001$] も有意な主効果を示した。これは，運用 ($M=2.76, SD=0.85$) が知識 ($M=2.07, SD=0.56$) よりも有意にエントロピーが高いという結果である。知識に対して，それを運用するにあたり多様な使われ方がされていることを示している。さらに，両変数の交互作用も有意であった [$F(1,18)=26.96, p<.001$]。この結果は，謙譲の運用において，エントロピーが平均で3.53と他の3条件と比べて有意に高くなっていることを示していると思われる。たとえばペーパーテスト形式よりもインタビュー形式の方がバリエーションの出やすい状況にあるとはいえ，尊敬は知識と運用の差がわずかであることを考慮に入れると，謙譲のエントロピーの高さは特筆に値する。本研究のインタビューは擬似的な自然発話ではあるが，そのような限定的な条件であっても，24名の被験者が謙譲表現の運用において多様な表現を使い，規則性が低いことが証明されたと考えてもよいであろう。

5. 3. 冗長度についての分散分析

冗長度の指標についても，2（尊敬表現，謙譲表現）×2（敬語知識，敬語運用）の分散分析を行った。エントロピーと同様に，尊敬表現・謙譲表現の変数は項目間分析であり，敬語知識・敬語運用は反復測定による項目内分析である。まず，尊敬表現・謙譲表現 [$F(1,18)=10.63, p<.01$] の主効果が有意であった。これは，尊敬表現 ($M=20.23\%, SD=8.65\%$) の方が謙譲表現 ($M=11.00\%, SD=6.50\%$) よりも冗長度が高いことを示している。つまり，尊敬表現の方が，繰り返しが多くより一貫した表現が使われる傾向にあることを意味している。しかし，エントロピーの結果と異な

り、敬語知識・敬語運用には有意な主効果は見られなかった [$F(1,18)=2.81, p=.111$]。知識 ($M=17.21\%$, $SD=8.99\%$) と運用 ($M=14.02\%$, $SD=8.58\%$) には平均にわずかな違いがあるが、この 3.19%の違いは有意な差ではない。また、両変数の交互作用も有意ではなかった [$F(1,18)=4.01, p=.061$]。しかし、相互作用がかなり有意水準に近いので、反復測定の一元配置の分散分析を、尊敬表現と謙譲表現を別々にして、知識と運用の違いを検討した。その結果、尊敬の知識と運用の間には有意な差は見られなかったが [$F(1,9)=.801, p=.394$]、謙譲の知識と運用の間には有意な差が見られ [$F(1,9)=9.842, p<.05$]、知識よりも運用の方が冗長度が低かった。つまり、尊敬は知識も運用も、同じ表現が繰り返される傾向があるが、謙譲は知識よりも運用の方が表現のバリエーションが多かったことを示している。つまり、冗長度においては、尊敬と謙譲に違いがあり、その知識を運用するにあたっては、エントロピーが示すほど顕著ではないが、冗長度も、謙譲が知識よりも運用において表現の規則性が低いことを示している。

6. エントロピーと冗長度によるプロットからみた知識と運用における尊敬と謙譲の違い

これまで、知識と運用において尊敬と謙譲がどのような違いを呈するかを検討するために、エントロピーと冗長度をそれぞれの動詞について算出して分散分析を行った。そこで、今度は個々の動詞について全体的に考察するために、図3のように知識と運用について、冗長度を横軸、エントロピーを縦軸にして各動詞をプロットした。さらに各動詞について、灰色で塗ったのが謙譲表現であり、動詞の横に括弧で尊敬・謙譲の違いを示した。また、各動詞の平均のエントロピーと冗長度を表2に記した。これで、すべての動詞について、尊敬と謙譲が知識と運用でどのような使われ方をしているかが視覚的に捉えられるはずである。

まず、図3を見ると、知識では動詞の尊敬と謙譲が混在していることが分かる。つまり、ペー

表2 動詞項目ごとの尊敬と謙譲表現のエントロピーと冗長度

動詞の種類		表現の得点		表現の種類		エントロピー		冗長度		
		知識	運用	知識	運用	知識	運用	知識	運用	
尊敬	1	行く	45	43	4	7	1.75	2.24	12.28%	20.06%
	2	言う	62	46	4	3	1.05	1.19	47.33%	25.02%
	3	見	53	46	5	6	1.74	1.84	25.23%	28.86%
	4	食	56	49	6	4	1.97	1.42	23.95%	28.76%
	5	する	50	48	3	7	1.32	2.01	16.57%	28.33%
	6	帰	43	46	5	6	2.30	2.15	11.20%	17.01%
	7	読	43	42	7	7	2.38	2.20	15.21%	21.70%
	8	書	40	42	6	5	2.14	2.05	17.38%	11.77%
	9	つ	35	37	6	6	2.13	2.24	17.60%	13.46%
	10	や	43	40	8	7	2.63	2.52	12.39%	10.38%
平均		47.00	43.90	5.40	5.80	1.94	1.99	19.91%	20.54%	
標準偏差		7.72	3.56	1.43	1.33	0.46	0.38	10.16%	6.80%	
謙譲	1	行く	37	33	11	16	3.07	3.80	11.31%	5.12%
	2	言う	32	31	9	16	2.62	3.82	17.26%	4.48%
	3	見	34	23	5	12	1.89	3.24	18.43%	9.71%
	4	食	33	26	3	15	1.21	3.31	23.73%	15.20%
	5	する	40	31	5	13	2.08	3.39	10.23%	8.49%
	6	待	26	33	11	13	3.18	3.46	7.96%	6.38%
	7	返	38	38	3	16	1.54	3.86	2.87%	3.62%
	8	屈	35	31	5	16	1.93	3.89	16.96%	2.69%
	9	知	34	35	5	15	1.74	3.69	24.85%	5.56%
	10	借	36	36	8	10	2.66	2.86	11.43%	13.79%
平均		34.50	31.70	6.50	14.20	2.19	3.53	15.23%	8.18%	
標準偏差		3.64	4.27	2.87	1.99	0.63	0.32	6.58%	4.25%	

注：知識は、ペーパーテストでの敬語表現で、運用はインタビューでの敬語表現を示す。

パートテストで尊敬と謙譲の知識を測定すると、ほぼ同じような使われ方をしていると思われる。ただし、同じ動詞が尊敬と謙譲でほとんど重なっていないので、両敬語表現において、ある程度、表現の多様性に違いがあると言えそうである。実際、エントロピーも冗長度も分散分析で有意な違いを示しており、際立った違いは図では見えにくいですが、謙譲の方が尊敬よりも多様性に富み、一貫性はなさそうである。

運用については、尊敬ではエントロピーが比較的低く冗長度も低い下部に、謙譲ではエントロピーが高く冗長度が低い上部に偏在していることが一見してわかる。とりわけ、エントロピーが4に近い「届ける」、「返す」、「言う」、「行く」の4つの動詞は、多様な謙譲表現で使われているようである。エントロピーと冗長度の分散分析の結果でも、尊敬と謙譲に違いが見られ、やはり謙譲の方が尊敬よりも多様な表現が使われているのが分かる。本研究は擬似的なインタビューであるが、それでも敬語の運用において表現の多様性の違いが顕著に描きだされた。

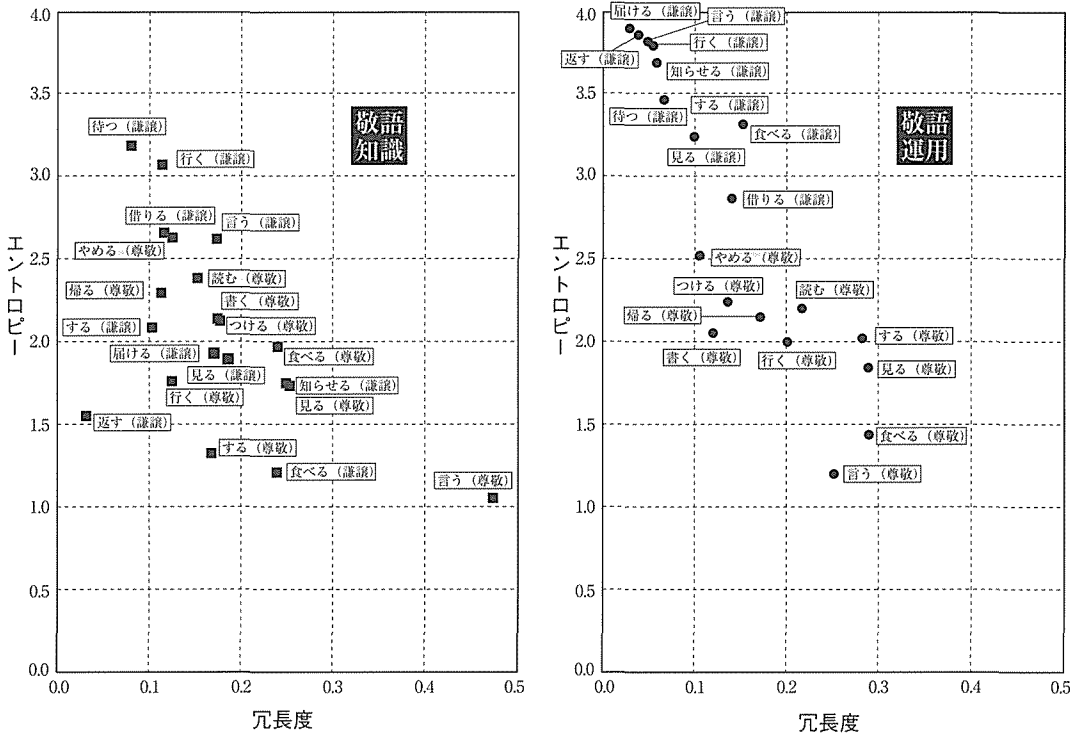


図3 敬語知識（ペーパーテスト）と敬語運用（インタビュー）における尊敬および謙譲表現のエントロピーと冗長度のプロット

さらに、敬語知識と敬語運用とを図3で比較すると、尊敬については、知識と運用にあまり鮮明な違いはなさそうに見える。しかし、知識については、謙譲でも、尊敬と表現の多様性はそれ程変わらないようである。一方、運用については、謙譲の知識と運用ではとりわけエントロピーに大きな違いが見られ、完全に尊敬と分離していることが分かる。このことは、エントロピーの

分散分析で、尊敬と謙譲および知識と運用の2変数が有意な交互作用を示したことによって裏付けられる。

以上のように、エントロピーと冗長度を動詞ごとにプロットした図を見ても、これら両変数が敬語の知識と運用における表現の多様性を適切に捉えていると言えるのではなかろうか。

7. 研究の要約と応用

本研究では、クロード・シャノンの通信の数学理論で知られるエントロピーと冗長度という指標を使って、韓国語系日本語学習者を対象にペーパーテストで測定した知識とインタビューで測定した運用の差異について検討した。まず、ペーパーテストとインタビューを表現の正誤という視点からみると、尊敬表現と謙譲表現にも知識と運用にも違いがなかった。本研究では、あくまで基本的な10種類の動詞を扱っており、これらの動詞の知識を基にした運用では正答に大きな違いが無かったのであろう。本研究は、敬語表現の多様性をエントロピーと冗長度で考察することを目的としているので、正答における違いがないことが望ましい。つまり、エントロピーと冗長度で尊敬表現と謙譲表現をその知識と運用について比較することの基準が同じであると仮定することができる。その上で、知識と運用の場面で使われる尊敬と謙譲の表現について、エントロピーと冗長度を算出した。エントロピーは、乱雑さを示す指標として知られているが、分散分析の結果、尊敬表現と比べて、とりわけ運用における謙譲表現の多様性と不規則性が明らかになった。これらエントロピーと冗長度の二つの指標で描いた図3のプロットは、知識と運用における尊敬表現と謙譲表現の違いを如実に示している。以上のように、この指標化の方法は、母語や第二言語の習得研究に容易に応用できることが分かった。本研究は、エントロピーと冗長度の二つの指標によって、これまで記述的な分析にとどまりがちであった言語表現の研究を、数量的かつ包括的に扱うことの可能性を示した。

さらに、エントロピーと冗長度が言語表現上の多様性と規則性を、表現の種類と頻度から示しえたことは、近年、コンピュータの普及とともに目覚ましく発展してきたコーパス研究への応用の可能性をも示唆している（コーパスおよびその研究の詳細については、前川・籠宮・小磯・菊池・小椋, 2001; 前川・菊池・籠宮・山口・小磯・小椋, 2001; 丸山, 1998; 松田, 2001; 滝沢, 2001; 横山・笹原, 2001などを参照）。実際、コーパス研究が「確率文法 (probabilistic grammar)」として発展する可能性をHalliday (1991) が早い時期に論じている。コーパス研究とは、電子化された大規模な言語資料から、瞬時に単語や文法構造を検索し、その種類と頻度を基に研究を進めていく方法である。例えば、「ぶる」の接尾辞は、通常、「学者ぶる」や「医者ぶる」などの肯定的な表現に多くとされている（森田, 1989, 1996）。しかし、実際には「不幸ぶる」のように否定的な表現に「ぶる」がつく場合があることを、『男はつらいよ』のシナリオのコーパスから『男はつらいよ・寅次郎夢枕』（山田洋次）で見つける（玉岡・宮岡・黄, 2002）など、一般的な文法規則を逸脱するような表現を見出すために手軽に使える手法である。しかし、このことは、コーパス研究がある表現をくまなく検索してその頻度を数えることに尽きるという極論に行き着いてしまう。ところが、種類と頻度からエントロピーや冗長度という新たな指標を算出し得たことは、これらが直接

比較可能な値であるため、松田（2001）が提言するように「コーパス研究への多変量解析の必要性」に応えることができるようになると思われる。「数を数える」コーパス研究から、多様な統計解析が使用できるコーパス研究へと発展しうる可能性を秘めている。

今後のエントロピーや冗長度などの指標は、コーパス研究の重要課題であるコロケーションへも応用できる。例えば、「食べ終わる」、「読み終わる」のように、「終わる」にはさまざまな動詞が結びついて複合動詞が作られる。こうした、2つの動詞の結びつきについてエントロピーと冗長度を算出することができれば、複合動詞全体の傾向が判るであろう。また、接辞「不」が作る「不親切」、「不道德」などの結合関係を指標化することもでき、接尾辞や接頭辞の違いを含んで接辞全体をエントロピーと冗長度で分析することも可能である。実際に、内元・関根・伊佐原（2001）は、形態素解析についてエントロピー最大値を応用した研究を展開している。さらにまた、滝沢（2002）のコーパス研究で、副詞と終助詞が共起する種類と頻度を調べているが、特定の終助詞に対する副詞のバリエーションをエントロピーと冗長度で表すことも可能である。本研究は、こうした可能性を模索する一つの試みであった。

注

- 1 厳密には「知識 (knowledge)」は「言語を実際に使用する場合、背後にあってその使用を可能にしている知識体系である」（中村・金子・菊池, 1989, p.9）とされる。この定義では、知識は脳の中にあり、書く行為によるペーパーテストであろうと話す行為によるインタビューであろうと、いったん産出されたものはすべて言語の「運用 (performance)」と考えられる。しかし、本研究では、ペーパーテスト条件は他者とのコミュニケーションを伴わないという意味で知識を測定していると仮定した。一方、インタビューは会話によるコミュニケーションであるという条件で、運用を測定していると仮定した。以下、本研究では、ペーパーテストは知識を測定し、インタビューは運用を測定するとして議論を進める。
- 2 本研究の敬語表現について、エントロピー最大値、慣用度（出現頻度のもっとも多い表現を延べ頻度で割った数値）、TTR (type-token ratio; 表現の種類を延べ頻度で割った数値) も算出したが、エントロピーとのピアソンの相関係数が、敬語知識（尊敬と謙譲を一緒にした数値）についてはエントロピー最大値で $r=.947$ ($p<.01$)、慣用度で $r=-.847$ ($p<.01$)、TTRで $r=.932$ ($p<.01$) と高く、また敬語運用についても、エントロピー最大値で $r=.947$ ($p<.01$)、慣用度で $r=-.934$ ($p<.01$)、TTRで $r=.973$ ($p<.01$) と高いので、これらの指標はエントロピーと類似した言語現象を表していると考え、本研究では議論しなかった。なお、本研究のエントロピーと冗長度のピアソンの相関係数は、知識が $r=-.562$ ($p<.01$) で、運用が $r=-.888$ ($p<.01$) である。

引用文献

- 甘利俊一（1970）『情報理論』ダイヤモンド社。
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., Mcallister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34, 351-366.

- 有本卓 (1982) 『確率・情報・エントロピー』 森北出版.
- Brown, G., Anderson, A., Yule, G., & Shillcock, R. (1983). *Teaching task*. Cambridge, U.K.: Cambridge University Press.
- Den, Y., & Inoue, M. (1997). Disambiguation with verb-predictability: Evidence from Japanese garden-path phenomena. M. G. Shafto & P. Langley (eds.). *Proceedings of the Nineteenth Annual Conference on the Cognitive Science Society* (pp. 179-184). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer & B. Altnberg (eds.). *English corpus linguistics* (pp. 30-43). New York: Longman.
- 堀淳一 (1979) 『エントロピーとは何か』 講談社ブルーバックス.
- 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』 大阪大学人間科学部博士論文.
- Inoue, M., & Den, Y. (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* (pp. 499-502). July 27-30, 1999, International Conference Center, Waseda University, Tokyo.
- 海保博之 (1989) 「第 1 講：情報をはかる—エントロピー・伝達情報量・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.
- 前川喜久雄・籠宮隆之・小磯花絵・菊池英明・小椋秀樹 (2001) 「日本語話し言葉コーパスの設計」 『音声研究』 4 (2), 51-61.
- 前川喜久雄・菊池英明・籠宮隆之・山口昌也・小磯花絵・小椋秀樹 (2001) 「『日本語話し言葉コーパス』の構築における計算機利用」 『日本語学』 20(13), 61-79.
- 丸山直子 (1998) 「数理的研究」 『国語学』 193, 115-122.
- 松田謙次郎 (2001) 「コーパス調査と計量的研究」 『日本語学』 20(5), 32-41.
- Miyaoka, Y., & Tamaoka, K. (2001). Use of Japanese honorific expressions by native Chinese speakers. *Psychologia*, 44, 209-222.
- 宮岡弥生・玉岡賀津雄 (2002) 「上級レベルの中国語系日本語学習者と韓国語系日本語学習者の敬語習得の比較」 『読書科学』 46, 63-71.
- 森田良行 (1989) 『基礎日本語辞典』 角川書店
- 森田良行 (1996) 『意味分析の方法—理論と実践』 ひつじ書房
- 中村捷・金子義明・菊池朗 (1989) 『生成文法の基礎—原理とパラメターのアプローチ』 研究社出版
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 (Part I) and 623-656 (Part II).
- 玉岡賀津雄・宮岡弥生・黄其正 (2002) 「『良妻ぶる』妻は『悪妻』か—接尾辞『ぶる』付加による価値評価の転換と収束」 『日本語教育』 114, 1-10.
- 滝沢直宏 (2001) 「コーパスを使った英語教育」 名古屋大学言語文化部・国際言語文化研究科公開講座委員会 (編) 『インターネットと英語教育』 開文社.
- 滝沢直宏 (2002) 「文末の接続助詞 (+終助詞) と不連続に共起し易い副詞—「から」の場合を中心に」 S. Makino (ed.), *The Tenth Princeton Japanese Pedagogy Workshop: Proceedings* (pp. 132-142). Princeton, N.J.: Princeton University.

内元清貴・関根聡・伊佐原均（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」『自然言語処理』8(1), 127-141.

横山詔一・笹原宏之（2001）「文字・表記分野での計量的研究概観」『日本語学』20(5), 157-165.

（投稿受理日：2002年10月28日）

（改稿受理日：2003年 7 月30日）

玉岡賀津雄（たまおか かつお）

広島大学留学生センター教授

739-8524 広島県東広島市鏡山 1 - 1 - 1 広島大学留学生センター

0824-24-6288

ktamaoka@hiroshima-u.ac.jp

宮岡弥生（みやおか やよい）

広島経済大学常勤講師

731-0192 広島県広島市安佐南区祇園 5 - 37 - 1 広島経済大学

082-871-1068

y.miya8411@hue.ac.jp

林炫情（イム ヒョンジョン）

日本学術振興会外国人特別研究員

739-8529 広島県東広島市鏡山 1 - 1 - 1 広島大学大学院国際協力研究科

0824-24-6905

lim@hiroshima-u.ac.jp