自然言語処理 116－17
（1996. 11. 19）

待遇表現の計算モデル
－ 語尾の付加による待遇値変化について －

白土 保, 井佐原 均
郵政省通信総合研究所

待遇表現の計算モデルが提案されている．このモデルでは，それぞれの待遇表現及び語尾に対し，各表現が持つ話し手と聞き手の待遇関係に応じた丁寧さを表す待遇値が一定の確率分布をとり，その確率分布は一次元の正規分布である，と仮定されている．そしてこの仮定に基づき，待遇表現に語尾を付加した際の待遇値の変化量が，付加の際得られる情報量に基づいて定義されている．いくつかの待遇表現，及びそれぞれの待遇表現に語尾を付加した表現の待遇値を心理実験によって求めたところ，語尾の付加による待遇値の変化は提案されたモデルによって予測された傾向に従い，モデルの妥当性が支持された．

A computational model for polite expression.
- modification of politeness magnitude by ending word adding -

Tamotsu SHIRADO, Hitoshi ISAHARA
Communications Research Laboratory, M.P.T.

A computational model for polite expression is presented. In the proposed model, politeness magnitudes corresponding to expressions and ending words are assumed to be distributed in normal distribution. Modification of politeness magnitude on an ending word adding to a polite expression is defined by the amount of information. The result of the psychological experiment support the validity of the model.

1 はじめに

我々は，感性を取り扱うことのできる自然言語処理システム構築に向けた基礎研究のひとつとして，待遇表現の計算モデルに関する研究を行っている．

待遇表現とは，話し手が，聞き手及び話題に含まれる人物と自分との間に，尊卑，優劣，利害，疎遠等どのような関係があるかを認識し，その認識を言語形式の上に表したものである[1]．本研究ではこれらの関係を総称して，“待遇関係”と呼ぶことにする．ただし，今回は言語表現が用いられる状況として，第三者に関する話題が含まれないような状況に限定した．

いま，“待遇表現が持つ情報”＝“意図を伝えるのに最小限必要な情報”＋“話し手と聞き手の待遇関係に応じた丁寧さ（あるいはぞんざいさ）を伝える情報”と考えると，同じ意図に対する複数の待遇表現の間で表現の丁寧さの程度に関する比較が可能である，と期待される．本研究では，待遇表現が持つ情報のうち話し手と聞き手の待遇関係に応じた丁寧さを伝える情報を計量化したものを，“待遇値”と呼ぶ．

待遇表現の計算モデルに関する研究としては，数学的な形式化に重点を置いた水谷の研究[2]や，心理実験による待遇表現の計量化に重点を置いた

荻野の研究[3] がある．しかし，待遇表現の構成要素から待遇表現全体の待遇値を計算するモデルの提案，及び心理実験に基づくモデルの検証は，これまでほとんど行われて来ていない．

本研究では，短い待遇表現と語尾によって構成される待遇表現について，待遇表現全体の待遇値を構成要素から計算するモデルを提案している．モデルでは，待遇表現及び語尾それぞれに対する待遇値が一定の確率分布をとり，その分布は一次元の正規分布である，と仮定した．この仮定に基づき，待遇表現に語尾を付加した際の待遇値の変化量を，付加の際得られる情報量に比例する量として定義した．

この計算モデルの妥当性を検証するため，ある事項について“**知っている**”という意図を聞き手に伝えるための表現として，“知ってる”,“知っている”,“知ってます”,“知っております”,“分かる”,“分かってる”，“分かってます”，“存じてます”，“存じ上げてます”の９種類の待遇表現を用い，またそれぞれの待遇表現に付加する語尾として終助詞“よ”を用いた心理実験を行った．実験により得られた待遇値に基づき，“よ”の付加による待遇値の変化を調べたところ，その変化はモデルによって予測される傾向に従い，モデルの妥当性が支持された．

2 待遇表現の計算モデル

2.1 待遇表現が用いられる状況についての制約

今回は待遇表現が用いられる状況として，第三者に関する話題が含まれないような状況に限定したモデルを提案している．ただし第三者に関する話題が含まれる状況に対しても，待遇表現を,(1) 話し手と聞き手の待遇関係，(2) 話し手と第三者の待遇関係，及び (3) 聞き手と第三者の待遇関係，の３つの要因（あるいはそのうち主要な２つの要因）に対応した構成要素に分けることができる範囲内で，今回提案するモデルの組み合わせによってある程度対処が可能であると考えられる．

2.2 待遇値

荻野はクロス集計表に基づく待遇表現の計量化の研究において，ほとんどすべての待遇表現の待遇値は一次元の値として表現できることを示した[4]．本研究ではこの結果をふまえ，すべての待遇表現の待遇値はスカラーで表すことができるものとする．以下，計量化によって得られる待遇表現 $\mathbf{P}$ の待遇値を $V(\mathbf{P})$ と記す．

2.3 待遇表現及び語尾に対する待遇値の確率分布

前述の通り，待遇値は待遇表現が持つ情報のうち話し手と聞き手の待遇関係に応じた丁寧さを伝える情報を計量化したものであり，一次元の値として表すことができると考えられる．

ところがひとつの待遇表現は，いろいろな待遇関係の場面において用いることが可能である．従って，それぞれの待遇表現に対する待遇値は本来ある一定の確率的な広がりを持っており，待遇表現 $\mathbf{P}$ の計量化においては，この確率分布の平均値が $V(\mathbf{P})$ として得られている，と考える方が自然である（ただし，心理実験などによって実際に得られる待遇値は，その実験に用いた計量化の手法に依存した尺度空間上で定義された値となる）．本研究では，この確率分布として正規分布を仮定する．以上の考察をまとめ，仮定１及び仮定２を提案する．

仮定１： 任意の待遇表現 $\mathbf{P}$（待遇値 $V(\mathbf{P})$）に対し，確率変数が一次元の正規分布 $N(V(\mathbf{P}),\sigma_P^2)$ に従う (分散 σ_P^2 は $\mathbf{P}$ に依存した値) ような待遇値の分布が存在する．

図１に待遇表現 $\mathbf{P}_1$: “知ってます”, 及び $\mathbf{P}_2$: “存じてます” に対する確率分布 $N(V(\mathbf{P}_1),\sigma_{P_1}^2)$, 及び $N(V(\mathbf{P}_2),\sigma_{P_2}^2)$ の例を示す．

仮定１を語尾について拡張したものが，仮定２である．

仮定２： 任意の語尾 $\mathbf{E}$ に対し，確率変数が一次元の正規分布 $N(\mu_E,\sigma_E^2)$ に従う (平均 μ_E，分散 σ_E^2 は $\mathbf{E}$ に依存した値) ような待遇値の分布が存在する．

仮定2において，語尾単独の待遇値を直接評価することは実際の状況ではありえないが，ここでは，$N(\mu_E,\sigma_E^2)$ は語尾Eを含むあらゆる待遇表現に対する待遇値の分布の平均的な分布を表すものと考える．図2に語尾$\mathbf{E}_1$: “よ”，及び$\mathbf{E}_2$: “です”に対する確率分布$N(\mu_{E_1},\sigma_{E_1}^2)$，及び$N(\mu_{E_2},\sigma_{E_2}^2)$の例を示す．

2.4 語尾の付加による待遇値変化

いま待遇表現Pに語尾Eを付加した際の待遇値の変化を，話し手と聞き手の待遇関係に関する情報の獲得の観点から考察する．ただし，ここで想定するPとEの組み合わせは，すべて文法的に接続可能なものとする．

仮定1により，語尾が付加される前の待遇表現P（待遇値$V(\mathbf{P})$）に対しては，確率変数が一次元の正規分布$N(V(\mathbf{P}),\sigma_P^2)$ に従うような待遇値の分布が存在する．

この状態において，Pに語尾Eが付加されることを知ったとき，もし$N(V(\mathbf{P}),\sigma_P^2)$ と$N(\mu_E,\sigma_E^2)$がほぼ同じ分布ならば，待遇表現Pによって期待された話し手と聞き手の待遇関係に適合するような語尾Eが付加されたに過ぎず，Eの付加は話し手と聞き手の待遇関係について何ら新しい情報をもたらさない．一方$N(V(\mathbf{P}),\sigma_P^2)$ と$N(\mu_E,\sigma_E^2)$ が異なる時，Eの付加は，Pによってそれまで期待されていた話し手と聞き手の待遇関係に修正を促す情報をもたらすことになる．

この修正量の大きさは，$N(V(\mathbf{P}),\sigma_P^2)$,$N(\mu_E,\sigma_E^2)$ それぞれに対する確率密度関数が囲む領域の共通部分の面積が小さい程大きい，と考えられる．ここでは，この共通部分の面積を適合度と呼び，Cと記すことにする．図3にP: “知ってます”（分布$N(V(\mathbf{P}),\sigma_P^2)$）及びE: “よ”（分布$N(\mu_E,\sigma_E^2)$）に対する適合度$C$の例を示す．

適合度Cは，Pによって期待されるすべての待遇関係の空間の中で，Eによって期待される待遇関係が生じる確率を表す．従ってシャノンの情報量[5]の定義によれば，Eの付加によって$\log_e(1/C)$ の情報量（ここでは対数の底をeとした単位で情報量を定義する）を得たことになる．

語尾Eの付加による話し手と聞き手の待遇関係についての修正量は待遇値の変化量として評価されるが，ここで待遇値の変化量と情報量$\log_e(1/C)$との間で以下の関係を仮定する．

仮定3： 待遇表現Pに語尾Eを付加した際の待遇値の変化量の絶対値は，Eの付加によって得られる情報量に比例する．

以上の考察に基づき，待遇表現Pに語尾Eを付加した際の待遇値の変化量$\Delta(\mathbf{P},\mathbf{E})$ は次式で定義される．ここでは，$\mu_E < V(\mathbf{P})$ を仮定したが，$V(\mathbf{P}) < \mu_E$ の場合も同様の式で定義できる．

$$\Delta(\mathbf{P},\mathbf{E}) \triangleq K\log_e(1/C) \qquad (1)$$

$$C \triangleq \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{X} e^{-\frac{(x-V(\mathrm{P}))^2}{2\sigma_{\mathrm{P}}^2}}dx + \frac{1}{\sigma\sqrt{2\pi}}\int_{X}^{+\infty} e^{-\frac{(x-\mu_{\mathrm{E}})^2}{2\sigma_{\mathrm{E}}^2}}dx \qquad (2)$$

式(1)のKは比例定数であり，$\mu_E < V(\mathbf{P})$ の時負の値，$V(\mathbf{P}) < \mu_E$の時正の値をとる．式(2)のXは，$N(V(\mathbf{P}),\sigma_P^2)$，$N(\mu_E,\sigma_E^2)$ それぞれに対する確率密度関数の交点のx座標を表す．

3 計算モデルから予測される，語尾の付加による待遇値変化

前節で提案した待遇値変化に関する計算モデルから予測される性質として，待遇値変化$\Delta(\mathbf{P},\mathbf{E})$の待遇値$V(\mathbf{P})$ に関する特性に注目する．

ここでは，おおまかな特性を調べることを目的とするが，この目的からは数学的な取扱いの簡便さのため$\sigma_P=\sigma_E=\sigma$と仮定して差し支えないと考えられる．このとき式(2)で定義された適合度Cは次式で表すことができる．

$$C = \frac{2}{\sigma\sqrt{2\pi}}\int_{-\infty}^{X} e^{-\frac{(x-V(\mathrm{P}))^2}{2\sigma^2}}dx \qquad (3)$$

いま，$(x\text{-}V(\mathbf{P}))/\sigma=t$ と置くと，$dx=\sigma\, dt$ より式(3)は以下の式に書き直すことができる．

$$\frac{2}{\sqrt{2\pi}}\int_{X'}^{+\infty} e^{-\frac{t^2}{2}}dt \qquad (4)$$

ただし，$X'=(V(\mathbf{P})-X)/\sigma=(V(\mathbf{P})-\mu_E)/2\sigma$.

ここでX'を変数と見なし，式(4)がX'に関しどのような関数になっているかを調べることにより，$\Delta(\mathbf{P},\mathbf{E})$を$V(\mathbf{P})$の関数として表すことを考える．

式(4)に示す関数($C(X')$と記す)は解析的には解けないので，正規分布表[6]を用いて$e^{-poly(X')}$で近似する（ただし$poly(X')$は，X'の多項式）．回帰関数を$e^{-poly(X')}$に選んだ理由は，$C(X')$が$X'\rightarrow\infty$では$e^{-\frac{X'^2}{2}}\cdot f(X')$（ただし$f(X')$は$X'$に関する無限級数）の形の関数で良く近似できる[7]ためである．

いま，どのような$\mathbf{P}$と$\mathbf{E}$の組み合わせに対しても，$|V(\mathbf{P})-\mu_E|\leq 6\sigma$が満たされると仮定する．この仮定は，$N(V(\mathbf{P}),\sigma^2)$に対する確率密度関数が囲む面積の中で，$N(\mu_E,\sigma^2)$に対する確率密度関数が囲む部分との共通部分の面積が占める割合が少なくとも0.6％以上存在する場合にあたり，この条件はほぼ一般に満たされると考えて良い．この条件のもとで，$C(X')$を$\mathrm{poly}(X')=a+bX'$(a,bは回帰係数，$b>0$)とした関数$e^{-(a+bX')}$により回帰する．このとき回帰曲線のデータへの説明力を表すR^2（ただしRは決定係数[8]）は約0.97となり，この関数が$C(X')$の良い近似となっていることが分かる．

以上から待遇表現$\mathbf{P}$に語尾$\mathbf{E}$を付加した際の待遇値の変化量$\Delta(\mathbf{P},\mathbf{E})$は，$\Delta(\mathbf{P},\mathbf{E})=K\log_e(1/C)\simeq K\log_e(e^{a+bX'})=K(a+bX')=K(a+b(V(\mathbf{P})-\mu_E)/2\sigma)=K(a-b\mu_E/2\sigma)+Kb/2\sigma\cdot V(\mathbf{P})$となる．よって待遇値の変化量$\Delta(\mathbf{P},\mathbf{E})$は待遇表現$\mathbf{P}$の待遇値$V(\mathbf{P})$の一次式となり，$b>0$より直線の傾きの符号は比例定数$K$の符号と同じとなる．

4　モデルの妥当性の検証実験

4.1　実験に用いる表現

実験に用いた待遇表現は，ある事柄について**“知っている”**という意図を伝える際に用いる表現P：　“知ってる”,“知っている”,“知ってます”,“知っております”,“分かる”,“分かってる”，“分かってます”，“存じてます”，“存じ上げてます”の９種類とした．待遇表現$\mathbf{P}$に付加される語尾$\mathbf{E}$は，終助詞“よ”とした．語尾として“よ”を用いたことは，その語尾の付加によって待遇値が減少するような状況を調べたことになる[9]．

ここですべての$\mathbf{P}$及び$\mathbf{E}$に関し$\mu_E<V(\mathbf{P})$と仮定すると，式(1)の比例定数Kは負となるため，$\mathbf{E}$を付加したことによる$\mathbf{P}$の待遇値の変化量は，$V(\mathbf{P})$に関して傾きが負の一次式で近似できると期待される．

4.2　待遇値を求める実験

すべての待遇表現，及びそれぞれの待遇表現に語尾を付加した待遇表現の待遇値は，一対比較法に基づくサーストンの比較判断の法則による計量化手続きにより求められた．

[実験刺激に用いた待遇表現]

“知ってる”,“知っている”,“知ってます”,“知っております”,“分かる”，“分かってる”，“分かってます”，“存じてます”，“存じ上げてます”，“知ってるよ”,“知っているよ”,“知ってますよ”,“知っておりますよ”,“分かるよ”，“分かってるよ”，“分かってますよ”，“存じてますよ”，“存じ上げてますよ”の１８種類の待遇表現．

[被験者]

２０代～３０代の日本人男性４名．

[一対比較法による実験手続き]

１８種類の待遇表現の中から任意のふたつの表現の対全ての組み合わせ($_{18}C_2=153$通り)を作り，被験者に表現の対をひとつづつ呈示した．被験者は各対のうちより丁寧だと思う方の表現に○印を，両方とも丁寧さが同じだと思う場合には△印を記入するよう求められた．図４に回答例を示す．

[サーストンの比較判断の法則に基づく計量化]

実験刺激に用いた（“よ”を付加した表現も含む）１８種類の待遇表現を$\mathbf{P}^{\dagger}_i$, $i=1,2,..,18$と表す．待遇表現$\mathbf{P}^{\dagger}_j$と待遇表現$\mathbf{P}^{\dagger}_k$の対において，$\mathbf{P}^{\dagger}_j$が$\mathbf{P}^{\dagger}_k$より丁寧だと判断した割合の被験者全員の平均をp_{jk}とする．また，p_{jk}をZ得点[10]で表した値をZ_{jk}とする．

一対比較法では，ふたつの刺激$\mathbf{P}^{\dagger}_{j}$及び$\mathbf{P}^{\dagger}_{k}$に対する（丁寧さの大小に関する）弁別過程を調べることになるが，ここではそれぞれの刺激に対する弁別過程が正規分布に従うと仮定し，更にそれぞれの平均，標準偏差 μ_j, σ_j, μ_k, σ_k，及び両弁別過程の相関係数r_{jk}に関し，$\sigma_j=\sigma_k=\sigma$，$r_{jk}=0$ と仮定したとき｛サーストンの比較判断の法則のケースV｝次式が成立する(11)．

$$\mu_j - \mu_k = Z_{jk}\sqrt{2}\sigma \quad (5)$$

従って最も小さいμの値を0と置くと，式(5)によって他の全てのμ_iがμからの相対値として順次決められる．ただし，μ_iの値は$\sqrt{2}\sigma$を単位とした値である．以上の方法によって得られたμ_iが待遇表現$\mathbf{P}^{\dagger}_{i}$の待遇値とされる．表1に，実験で得られた待遇値を示す．

表1 実験により得られた待遇値

待遇表現P	待遇値	P+“よ”	待遇値
“知ってる”	2.07	“知ってるよ”	1.15
“分かってる”	2.36	“分かってるよ”	1.49
“分かる”	3.09	“分かるよ”	1.87
“知っている”	4.10	“知っているよ”	2.73
“分かってます”	5.41	“分かってますよ”	4.97
“知ってます”	6.63	“知ってますよ”	4.68
“存じてます”	10.55	“存じてますよ”	8.68
“知っております”	13.41	“知っておりますよ”	11.33
“存じ上げてます”	15.19	“存じ上げてますよ”	12.95

4.3 モデルの予測との比較

表1を基に，終助詞“よ”をつける前の待遇表現Pの待遇値$V(\mathbf{P})$を横軸に，Pに“よ”を付加した際の待遇値変化$\Delta(\mathbf{P},\text{“よ”}) = V(\mathbf{P}+\text{“よ”}) - V(\mathbf{P})$を縦軸にプロットしたものを図5に示す．

図5に示したデータ点に対し，直線$a+bx$を用いて回帰したところ，回帰係数の推定値は$a=-0.72$, $b=-0.1$となり，R^2は0.65となった．これは，直線回帰によって説明できる待遇値変化の分散の割合が65％であることを意味する．

更に回帰直線の傾きbが有意に負であることを示すため，帰無仮説H_0: $b=0$，に関する検定を行った．傾きbの標本標準偏差をs_bとすると，この検定は検定量$t=b/s_b$，自由度＝サンプル数－2によるt検定により行うことができる(12)．$b=-0.1$, $s_b=0.029$，サンプル数＝9を代入し，検定量$t\simeq-3.4$，自由度＝7となった．このとき，帰無仮説H_0は危険率1％で棄却される（$-3.4 < t_7(0.01) \simeq -3.0$）．従って回帰直線の傾きが有意に負であることが示された．

以上の結果は，心理実験に基づく待遇値変化が，モデルから予測される傾向にほぼ従っていることを示唆する．

5 まとめ

待遇表現の計算モデルを提案し，心理実験によるモデルの妥当性の検証を行った．今回実験に用いた待遇表現及び語尾については，実験によって得られた待遇値がモデルから予測される傾向にほぼ従うことが示され，モデルの妥当性が支持された．

今回は，語尾として終助詞“よ”のみを用いたため，待遇値が減少するような状況のみを調べたことになる．またモデルの検証に用いた待遇値は僅か4名の被験者による実験で得られた値であるため，統計的信頼性の問題が残った．今後はより多数の被験者を用い，“です”などの語尾の付加や，“おり”などの語の挿入による待遇値が増加する状況での待遇値の変化を調べることにより，モデルの追検証を行う．

参考文献

(1) 鈴木一彦他編,研究資料日本文法9．敬語法編(明治書院,1989), pp. 2-7.
(2) 水谷静夫,待遇表現概要(計量計画研究所,1995).
(3) 荻野綱男,“聞き手に対する敬語行動の理論”，国語学,158集(1989), pp. 27-35.
(4) 荻野綱男,“待遇表現の社会言語学的研究”,日本語学 第5巻12号, pp. 55-63.
(5) N. Abramson(宮川洋訳),情報理論入門(好学社,1983), pp. 12-13.
(6) 鈴木七緒他, 確率と統計演習 (共立出版, 1991), pp.246-248.
(7) 森口繁一他， 数学公式II（岩波書店,1987), p. 167.
(8) 竹内啓他, 統計学事典(東洋経済,1989), p. 155.
(9) 中川裕志他,“日本語の終助詞の機能 －「よ」「ね」「な」を中心として－”,言語処理学会誌, Vol. 3, No.2 (1996), pp. 3-18.
(10) 田中良久， 心理学的測定法(東大出版会,1977), pp.56-57.
(11) 同上， pp. 159-164
(12) G.W. Snedecor(畑村又好他訳)，統計的方法(岩波書店,1991), pp. 131-134.

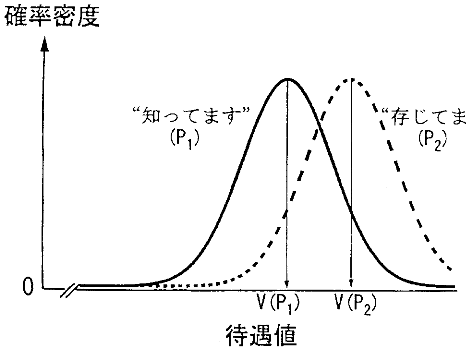

図１　確率密度分布の例.
(“知ってます” 及び “存じてます”)

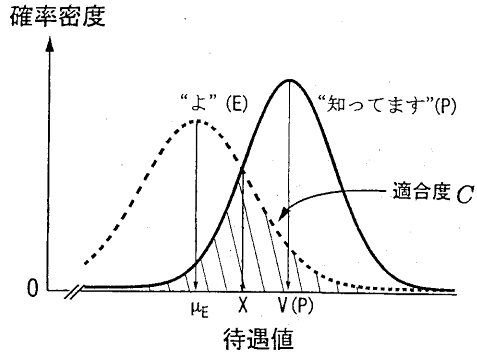

図３　適合度の例（“よ” 及び “知ってます”）.

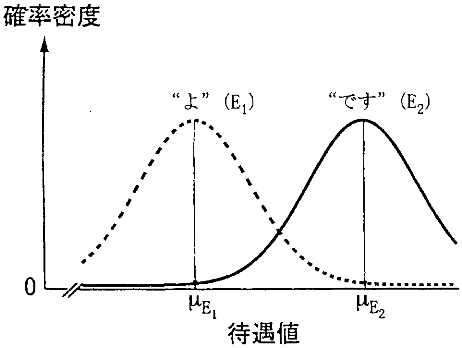

図２　確率密度分布の例.
(“よ” 及び “です”)

知ってるよ　　○ 知ってます
知っております ○　　知っているよ
分かってる　　△　　知ってる
. . . .

図４　回答例.

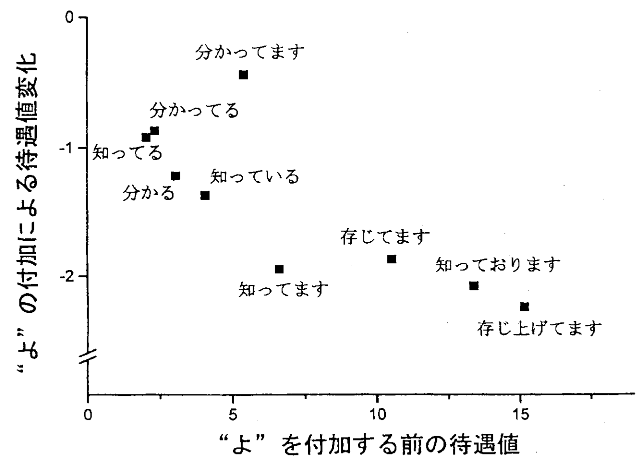

図５　“よ” の付加による待遇値変化の値.