

2-5 敬語誤用指摘システムの開発

2-5 Development of a System to Point Out Misuse of Japanese Honorific

白土 保

SHIRADO Tamotsu

要旨

近年、日本語の敬語に関する様々な誤用が指摘されてきている。敬語の誤用は人物間の適切な社会的関係の認識・構築に悪影響を及ぼす場合があるため、可能な限り避けることが好ましい。敬語の誤用を避けるには、敬語の規範に関する正しい知識の習得が不可欠である。このような知識習得を効率的に行うため、敬語学習を支援する計算機システムの実現が期待される。以上の背景の下、我々は日本語発話文に含まれる敬語の誤用を指摘するシステムを開発した。本システムは、日本語発話文及び発話にかかわる人物間の社会的関係を入力とし、入力された発話文における誤用の有無及び誤用が含まれる場合には誤用の箇所と種類を出力する。例文を用いた実験によってシステムの妥当性を検証したところ、ごく一部のケースを除き、本システムがおおむね妥当な出力を行うことが確認された。本システムは、特に敬語の初学者に対する学習支援システムとして有用と考えられるが、その他の人々にとっても、日本語文書の入力におけるケアレスミスをチェックする等の用途として幅広く活用できると考えられる。

We developed a computational system to indicate the misuse of honorifics in Japanese speech sentences. The misuse was checked by constructing a consistency table. The consistency table defined the consistency between the honorific features of sentences and the social relationship among the people involved in the sentence. The social relationship was represented by combinations of [relative social position among the people] × [in-group/out-group relationship among the people]. The proposed system was verified using test data prepared by the authors and also by third-party linguistic researchers. The results showed that the system was able to discriminate between the correct and the incorrect honorific sentences in all but a few cases.

【キーワード】

敬語, 誤用, 日本語, 社会的関係, 規範

Honorific expression, Misuse, Japanese, Social relationship, Norm

1 まえがき

敬語は日本語の重要な特徴の一つとされており、日本語の敬語は単に依頼、要求あるいは人を示す代名詞において見られるだけでなく、言語体系及び言語行動のほぼ全般にわたって発達している。ところが現代の日本社会において、日本語の敬語に関する様々な誤用が指摘されてきている[1][2]。日本社会における敬語の誤用は、人物間の適切な社会的関係の認識・構築に悪影響を及ぼす場合が

ある。敬語の誤用を避けるには、敬語の規範に関する正しい知識の習得が不可欠である。このような知識の習得を効率的に行うため、敬語学習を支援する計算機システムの実現が期待される。

以上の背景の下、我々は日本語発話文に含まれる敬語の誤用を指摘するシステムを開発した。本システムは、日本語発話文及び発話にかかわる人物間の社会的関係を入力とし、入力された発話文における誤用の有無及び誤用が含まれる場合には誤用の箇所と種類を出力する。

本システムの妥当性を検証するため、筆者らが用意したテストデータ及び筆者とは別の研究グループに属する日本語学研究者が作成したテストデータを用意し、これらのテストデータをシステムに判定させる実験を行ったところ、ごく一部のケースを除き、本システムがおおむね妥当な出力を行うことが確認できた。

2 敬語の誤用

敬語の誤用は語形上の誤用及び運用上の誤用に大きく分類できる。本システムにおいては、語形上の誤用を指摘する機能は語形上の誤用に関する表現リストの構築により実現している。また、運用上の誤用を指摘する機能は、(1) 発話文の敬語的特徴、(2) 発話文の述語の主語・補語及び(3) 発話にかかわる人物間の社会的関係、の間の整合性をチェックする規範ルール(整合表)の構築により実現している。

2.1 語形上の誤用

語形上の誤用とは、語形が敬語として誤っている表現の使用を意味する。本システムで用いている語形上の誤用の表現リストの一部を表1に示す。表中“～”は任意の動詞を表す。

2.2 運用上の誤用

運用上の誤用とは、語形は正しいものの、発話にかかわる人物間の社会的関係と整合しない表現の使用を意味する。ここで、発話内容に関係する人物間の社会的関係は、社会的な上下関係(以下では単に“上・下”と呼ぶ。)と社会的なウチ・ソト(同じ社会的グループ、例えば、同じ会社、同じ家族、に属するか否か)の組合せによって表した。なお、本システムでは、上・下が明確な場合、すなわち人物間の社会的上下関係が明確に認識できる場合だけでなく、人物間の社会的地位が等しい

場合も取り扱うことができる。

3 敬語誤用指摘システム

3.1 本システムが対象とする発話文

システムには日本語発話文及び発話にかかわる人物間の社会的関係が入力される。システムは、人物間の社会的関係が常に正しいものとして処理を行う。発話文としては以下の制約を満たす文を対象とする。

制約1: 発話文には述語が一つだけ含まれ、その述語の主語、補語はそれぞれ一つ。

制約2: 発話にかかわる人数は2名～4名。2名の時は話者(名前は“S”に固定)、聞き手(名前は“L”に固定)、3名の時は話者、聞き手及び発話文中で参照される人物1名(名前は“A”に固定)、4名の時は話者、聞き手及び発話文中で参照される人物2名(4人目の名前は“B”に固定)。

制約3: 話者(S)／聞き手(L)／A／Bが述語の主語あるいは補語の場合は人物を示す記号を明記する。

3.2 敬語特徴パターン

敬語は尊敬語、謙譲語及び丁寧語に大きく分類できる。尊敬語の中で最も典型的な語は、尊敬語を表すための形式を持つ述語及び人物に対する敬称(例えば、“様”)である。謙譲語の中で最も典型的な語は、謙譲語を表すための形式を持つ述語である。更に謙譲語は、補語を高める謙譲語(以下、“謙譲語a”と記す)及び補語を高めない謙譲語(以下、“謙譲語b”と記す)に分類できる。丁寧語は主に文末における丁寧な語(例えば、“～です。”)である。

制約1～3に従う文の敬語的な特徴は、述語の主語の敬称の有／無、述語の補語の敬称の有／無、文末が丁寧／丁寧でない及び述語の敬語的な特徴、すなわち尊敬語／謙譲語a／謙譲語b／謙譲語a+尊敬語／常体で表すことができると考えられる。本研究では以上を総称して“敬語タイプ”と呼ぶ。このため本システムでは、発話文を形態素解析して得た形態素の並び及び発話文を文構造解析して得られた主語・補語の情報に基づき、表2

表1 語形上の誤用の表現リストの一部

表現(“～”は動詞)	誤用のタイプ
お／ご～になられる	二重敬語
お／ご～なされる	二重敬語
お／ご～をなされる	二重敬語
お／ご～される	尊敬語+謙譲語の混用

表2 敬語タイプ辞書の例(“～”は動詞)

形態素の部分的並び	敬語タイプ
“L” + “様”	敬称
“A” + “さん”	敬称
“B” + “氏”	敬称
“お” + “～” + “する”	謙譲語 a
“ご” + “～” + “する”	謙譲語 a
“ご” + “～” + “致す”	謙譲語 a
“頂く”	謙譲語 a
“申す”	謙譲語 a
“お” + “～” + “なさる”	尊敬語
“ご” + “～” + “に” + “なる”	尊敬語
“おっしゃる”	尊敬語
“いらっしゃる”	尊敬語
“頂く” + “～られる”	謙譲語 a+尊敬語
“です” + “.”	丁寧語
“ます” + “.”	丁寧語

表3 敬語特徴パターンの定義

敬語特徴パターンの各要素の値	文の特徴
$s = 0$	subj の敬称なし
$s = 1$	subj の敬称あり
$o = 0$	obj の敬称なし
$o = 1$	obj の敬称あり
$e = 0$	文末が丁寧でない
$e = 1$	文末が丁寧
$p = 0$	述語が常体
$p = 1$	述語が尊敬語
$p = 2$	述語が謙譲語 a
$p = 3$	述語が謙譲語 a + 尊敬語
$p = 4$	述語が謙譲語 b

に示す敬語タイプ辞書を用いて表3に示すパターンを作り、このパターンで発話文の敬語的特徴を表すこととする(以下、このパターンを“敬語特徴パターン”と呼ぶ)。

敬語特徴パターン(表3)は四つの要素 s 、 o 、 e 、 p からなる。要素 s は、述語の主語(“*subj*”と標記)の敬称の有/無、に応じて 1/0 の値を取る。要素 o は、述語の補語(“*obj*”と標記)の敬称の有/無、に応じて 1/0 の値を取る。要素 e は、発話文の文末が丁寧/丁寧でない、に応じて 1/0 の値を取る。要素 p は、述語が常体/尊敬語/謙譲語 a / 謙譲語 a + 尊敬語/謙譲語 b、に応じて 0/1/2/3/4 の値を取る。

3.3 入出力

システムの入出力例を図1に示す。発話にかかわる人物名は固定されている。具体的には、話者は“山田”、聞き手は“佐藤”、3人目の人物は“高

図1 システムの入出力

橋”である。また、図1には登場していないが、4人目の人物は“木村”である(“山田”、“佐藤”、“高橋”、“木村”は後の処理でそれぞれ、記号:“S”、“L”、“A”、“B”に置き換えられる)。図右上には発話にかかわる人物間の社会的関係が指定されている。この例では、高橋と山田がいずれも組織Aに所属(すなわち、ウチの関係)、かつ高橋の社会的地位は山田より上である。そして、この2人と佐藤(組織Bに所属)はソトの関係にある。システムでは組織名:“A”、“B”、“C”、“D”と表示されたボタンを用いて、任意の人物を任意の組織に割り当てることができ、ボタン“↑”、“↓”を用いて同じ組織に属する人物間の社会的地位(上/下/等しい)を指定することができる。

ユーザはGUIに示された人物間の社会的関係の下で、正しい敬語であり、かつ上記の制約1~3に従う文を自由に記述することを求められる。

システムは入力された文の語形上の誤用及び運用上の誤用をチェックし、誤用の箇所が見つかった場合には、誤用の種類及び箇所を出力する。図1の例では、語形上の誤用は見つからなかったが、主語の敬称(“さん”)と社会的関係が整合しない(詳細は3.5で説明する)ため、“運用上の誤用あり”と判定されている。

3.4 処理のフロー

本システムにおける処理のフローを図2に示す。入力、すなわち発話文及び発話にかかわる人物間の社会的関係は、図1に示すGUIにおける入力を変換したものである。具体的には、図1の回答欄に記された文の“山田”、“佐藤”、“高橋”(“木村”)をそれぞれ、記号:“S”、“L”、“A”、

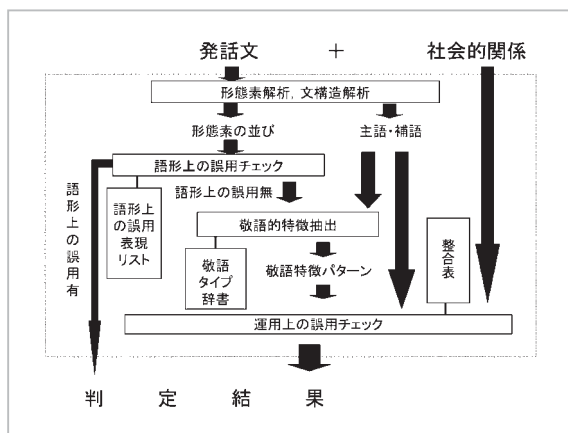


図2 システム処理フロー

“B”に置き換えたものが図2に記された“発話文”である。また、図1のGUI上で指定された人物間の社会的関係に応じて生成された記号列（この際、人物名は上記と同じ記号を用いて表される）が図2に記す“社会的関係”である。例えば、図1から生成される記号列は、(A>S)(L)である。ここで、同じ“()”内に現れた記号が指す人物はウチ、それ以外の人物はソトの関係にあることを意味する。そして、同じ“()”内の“>”の左側の記号が指す人物の社会的地位は右側の記号が指す人物より上、“=”の両側の記号が指す2人の人物間の社会的地位は同じ、であることを意味する。

発話文はまず形態素解析され、その結果得られた形態素の並びに対して、語形上の誤用の表現リスト（表1）を用いて語形上の誤用がチェックされる。例えば、形態素の並びの中に、“お”+動詞+“に”+“なる”+“れる”が現れた場合には、語形上の誤用“お～になられる”であると判断し、その旨を出力して処理を終了する。また、文構造解析により、述語の主語と補語が同定される。例えば、図1に示した発話文に対しては、主語がA（“高橋”）、補語がL（“佐藤”）と同定される。

語形上の誤用がなく、かつ文構造解析が成功した場合、形態素の並び及び主語・補語の情報に基づき、発話文に対する敬語特徴パターンが求められる。この際、敬語タイプ辞書（表2）及び敬語特徴パターンの定義（表3）が用いられる。例えば、図1に示した発話文に対する形態素の並びは、“A”“さん”“が”“L”“様”“に”“ご”“説明”“する”“ます”“。”、となる。これは、主語（A）の敬称があり

（“A”“さん”）、補語（L）の敬称があり（“L”“様”）、文末が丁寧であり（“ます”“。”）、述語が謙譲語a（“ご～する”）であるので、 $s=1$ 、 $o=1$ 、 $e=1$ 、 $p=2$ となる。

最後に、整合表を用いて、敬語特徴パターンと、発話にかかわる人物間の社会的関係の間の整合性がチェックされ、運用上の誤用の有無に関する判定結果が出力される。

3.5 整合表

整合表（整合表の一部を表4に示す）は、[発話にかかわる人物間の社会的関係]×[発話文の述語の主語・補語]と敬語特徴パターンの各要素の値との間の対応に関する敬語上の規範を記したルールである。ここで[発話にかかわる人物間の社会的関係]×[発話文の述語の主語・補語]は $\wedge$ 、 $\vee$ を含む論理式で表される。表記法は前述と同様である。すなわち、記号“S”、“L”、“A”、“B”は人物を示す記号である。同じ“()”内に現れた記号が指す人物はウチ、それ以外の人物はソトの関係にあることを意味する。“>”の左側の記号が指す人物の社会的地位は右側の記号が指す人物より上、“=”の両側の記号が指す2人の人物間の社会的地位は同じであることを意味する。また、“{subj:X}”、“{obj:Y}”はそれぞれ、述語の主語がX、補語がYであることを意味する。例えば、図1に示した入力に対しては、(A>S)(L) $\wedge$ {subj:A} $\wedge$ {obj:L}となる。図1の例では、整合表の $s=1$ （主語の敬称有り）に対応する論理式群の中でこれ

表4 整合表（発話にかかわる人物が2名（すなわち、話者と聞き手のみ））

敬語特徴パターンの各要素	[発話にかかわる人物間の社会的関係]×[述語の主語・補語]
$s = 0$	$(S=L) \vee (S>L) \vee [(L>S) \wedge \{subj:S\}] \vee [(L>S) \wedge \{obj:L\}] \vee [(S(L) \wedge \{subj:S\}) \vee \{(S(L) \wedge \{obj:L\})]$
$s = 1$	$[(L>S) \wedge \{subj:L\}] \vee [(L>S) \wedge \{obj:S\}] \vee [(S(L) \wedge \{subj:L\})] \vee [(S(L) \wedge \{obj:S\})]$
$o = 0$	$(S=L) \vee (S>L) \vee [(L>S) \wedge \{subj:L\}] \vee [(L>S) \wedge \{obj:S\}] \vee [(S(L) \wedge \{subj:L\})] \vee [(S(L) \wedge \{obj:S\})]$
$o = 1$	$[(L>S) \wedge \{subj:S\}] \vee [(L>S) \wedge \{obj:L\}] \vee [(S(L) \wedge \{subj:S\})] \vee [(S(L) \wedge \{obj:L\})]$
$e = 0$	$(S=L) \vee (S>L)$
$e = 1$	$(L>S) \vee (S(L))$
$p = 0$	$(S=L) \vee (S>L)$
$p = 1$	$[(L>S) \wedge \{subj:L\}] \vee [(L>S) \wedge \{obj:S\}]$
$p = 2$	$[(L>S) \wedge \{subj:S\}] \vee [(L>S) \wedge \{obj:L\}]$
$p = 0 \vee 1$	$[(S(L) \wedge \{subj:L\})] \vee [(S(L) \wedge \{obj:S\})]$
$p = 0 \vee 2 \vee 4$	$[(S(L) \wedge \{subj:S\})] \vee [(S(L) \wedge \{obj:L\})]$

を満たす論理式が見つからなかったため、“主語の敬称が有る”、ことが運用上の誤用として判定されたことになる。

4 システムの妥当性の検証

システムの妥当性を検証するため、テストデータを用いてシステムに判定させる実験を行った。テストデータとしては、テストデータ 1(著者らが作成したテストデータ)及びテストデータ 2(著者らの研究グループとは別の日本語学研究グループが作成したテストデータ)を用意した。テストデータは、正例データと負例データから成る。ここで“正例データ”/“負例データ”とは、人物間の社会的関係とペアとなる発話文が敬語として、正しい/(運用上)正しくない、データを指す。

テストデータ 1 は、約 800 の正例データと、同数の負例データから成る。テストデータ 2 は、約 4,000 の正例データと、同数の負例データから成る。

以上のテストデータをシステムに入力したところ、テストデータ 1 に関しては、すべての正例データに対し“文は正しい”と判定され、すべての負例データに対し“誤用”と判定された。

一方、テストデータ 2 に関しては、正例データのうち約 99.4 % は“文は正しい”と判定されたが、残り約 0.6 % については“誤用”と判定された。“誤用”と判定された典型的な例を図 3 上半分に示す。この例では、人物 S、L、A がすべてウチの関係で上下関係は $A > S > L$ 、かつ主語は L である。このとき整合表では、 $(A > S > L) \wedge \{subj: L\}$ に整合する p の値は $p = 2$ (すなわち、述語は謙譲語 a) である(表 4 では省略)。ところがこの文の述語は常体 ($p = 0$) であるため、“誤用”と判定されたことになる。テストデータ 2 の作成者は、このケースでは述語は必ずしも敬語でなくても非規範とは言えないと認識していたことが示唆される。この解釈を取る日本語学研究者は他にも少なからずいると考えられる。“誤用”と判定された他の例文の多くも、敬語規範に関する認識の違いに起因したケースであった。

また、テストデータ 2 における負例データのうち約 97.3 % は“誤用”と判定されたが、残り約 2.7 % は“文は正しい”と判定された。テストデータ 2 の

【“誤用”と判定された正例の例】

君が A さんに説明した。 ($A > S > L$)

【“正しい”と判定された負例の例】

私が A さんに電話をしてあげます。 ($A > L$)(S)

図 3 システムの予測とは一致しないテストデータ 2 の例

負例のうち“正しい”と判断された典型的な例を図 3 下半分に示す。この例では、登場人物 L と A がウチの関係で上下関係は $A > L$ 、かつこの 2 人と S がソトの関係、かつ主語は S である。このとき整合表では、 $(A > L)(S) \wedge \{subj: S\}$ に整合する p の値は、 $p = 0 \vee 2 \vee 4$ である(表 4 では省略)。このとき文の述語(“あげます”)は本システムでは単に常体 ($p = 0$) と見なされ、 $p = 0$ は $p = 0 \vee 2 \vee 4$ に含まれるため“正しい”と判定される。しかし日本語学研究者の間では、“あげます”は常体ではあるがソトの人物や目上に使うべきでない表現であるとされることが多い。テストデータ 2 の作成者もこの解釈に基づいて、“あげます”を負例として用いたと考えられる。このような表現を適切に扱えるようにするためには、システムで取り扱うことのできる述語のタイプを増やす等の拡張が必要と考えられる。

5 むすび

日本語発話文に含まれる語形上の誤用及び運用上の誤用を指摘するシステムを開発した。本システムは、日本語発話文及び発話にかかわる人物間の社会的関係を入力とし、入力された発話文における誤用の有無及び誤用が含まれる場合には誤用の箇所と種類を出力する。例文を用いた実験によってシステムの妥当性を検証したところ、ごく一部のケースを除き、本システムがおおむね妥当な出力を行うことが確認された。本システムは、特に敬語の初学者に対する学習支援システムとして有用と考えられるが、その他の人々にとっても、日本語文書の入力におけるケアレスミスをチェックする等の用途として幅広く活用できると考えられる。敬語学習の便宜のためには、誤用の指摘だ

けでなく正しい表現の候補を例示することが望ましい。今後は、このような課題に関する機能の拡

張を行うことによって、より実用的なシステムの構築を目指す。

参考文献

- 1 菊地康人, “敬語”, 講談社, 1997.
- 2 石野博史, “敬語の乱れー誤用の観点からー”, 文化庁「ことば」シリーズ 24 続敬語, 1986.

白土 保

知識創成コミュニケーション研究センター推進室研究マネージャー(旧情報通信部門けいはんな情報通信融合研究センター自然言語グループ主任研究員) 博士(工学)
感性情報処理