

日本語発話文に含まれる敬語の誤用を自動的に指摘するシステム ーウチソトの考慮, 及び GUI の開発ー

白土保* 丸元聡子** 村田真樹* 内元清貴* 井佐原均*

* 独立行政法人 情報通信研究機構, ** 計量計画研究所

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. このような特徴を持つ言語は日本語以外では世界中に少数しか見られない.

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. 日本社会における敬語の誤用は, 言語によるコミュニケーションを通じた社会的人間関係の構築を妨げる場合がある. 特にビジネスの場面における敬語の誤用は, 時として円滑なビジネスを進める上での障害にもなり得る. このため, 一般的には敬語の誤用はできるだけ避けることが望ましい. 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語形上の誤用, 及び運用上の誤用を指摘するシステムの構築を行っている. 敬語の誤用は大きくこの2種類に分類できる. 語形上の誤用を指摘する機能は, 基本的には語形上の誤用に関する辞書の構築, 及び拡張により実現可能である. 一方, 運用上の誤用を指摘するためには, 発話に関わる人物間の社会的関係と発話文との整合性のチェックが必要である.

一般に人物間の社会的関係を表す主要な要因としては, 社会的な上下関係 (以下“上・下”と記す) ウチ・ソト, 親・疎の3つがある. このうち, 上・下, 及びウチ・ソトは, 敬語が規範的か否かを判断する際の重要な要因であることが, 日本語の敬語に関する多くの文献において明記あるいは暗示的に述べられている. しかし, 3つ目の要

因である親・疎は, 敬語として規範的か否かという判断基準にはなじまないと考えられる. なぜなら, 互いに“親”であると認識している人の間では規範的な敬語を使わなくても問題になることはそれほど多くはないと考えられるため, そもそも敬語として規範的か否かという観点から表現の妥当性を判断する必要性が小さいからである. 即ち, 敬語としての規範性が問題になるのは, 基本的には人物間の関係が“疎”の場合であると考えられる.

以上から本システムでは, 発話に関係する全ての人物間の関係を基本的に“疎”と仮定した上で, 発話に関係する人物間の社会的関係を人物間の上・下の関係とウチ・ソトの関係の組み合わせによって表すこととした. 尚, 前回我々が提案したシステム[1]では, 運用上の誤用に関し上・下の要因しか考慮していなかったが, 本システムではウチ・ソトの要因も考慮していることが新しい点である. 更に本システムでは, ユーザの便宜を向上させるための GUI (Graphical User Interface) も実装した.

敬語 (正確には, 敬語を含む言語一般) は時代の経過と共に変化するため, 敬語としての規範をどこに置くかは重要な問題である. この問題に対し本研究では, 敬語に関し様々な国語学研究者の間で共通して述べられていることがらを規範と考えこの規範にできるだけ厳密に準拠する, という立場を取る. このため, 現代の日本社会において敬語として概して許容されている表現であっても, 本システムではそのような表現を規範的な敬語として見なさない可能性がある. しかしこのことは, 少しでも誤用の可能性のある表現をできるだけ漏らさずピックアップできる, という利点として考えることもできる.

2. 敬語の誤用

前述のように、敬語の誤用は語形上の誤用、及び運用上の誤用に分類できる。

2. 1 語形上の誤用

その語形が敬語として規範的ではない表現の使用のこと。本研究では、敬語のいろいろな文献を参考に語形上の誤用を定義した。本システムで用いている語形誤り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。尚、一般的に日本語として誤っているか否かという観点からは語形上の誤用を無数に想定することができるが、現時点では、特に敬語に関わる表現の中で最も典型的な語形上の誤用が定義されている。

表1: 語形誤り表現リスト (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	尊敬語と謙讓語の混用

2. 2 運用上の誤用

語形は正しいが、発話に関わる人物間の社会的関係と整合しない表現の使用のこと。ここで、発話に関わる人物間の社会的関係に基づいた運用上の規範をどのように定義するかは、本システムが対象とする発話文の定義にも依存するため、次章で述べる。

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

本システムでは、以下3つの制約を満たす発話文を対象とする。

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

制約2: 発話に関わる人数は2名～4名。2名の時は話者(名前は“山田”に固定)、聞き手(名前は“佐藤”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“高橋”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“木村”に固定)。

制約3: 話者(山田)／聞き手(佐藤)／高橋／木村が述語の主語あるいは補語の場合はそ

の人物の名前を明記する(ただし、話者／聞き手に関しては、“私”／“あなた”等の代名詞も使用可とする)。

以上の制約は、現行の構文解析システムや意味解析システムを用いた際の文構造の解析(特に、述語の主語、補語の同定等)における解析エラーの生じる可能性があるような複雑な文を排除するために設けた。従って、高い精度の文解析手法が将来開発されれば、これらの制約はより緩やかにできると考えられる。

3. 2 敬語特徴パターン

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に対して、表2に示す敬語タイプ辞書を用いて、4つの要素 s , o , e , p からなるパターンを作り、このパターンで文の敬語的特徴を表すことにする(これを以下、“敬語特徴パターン”と呼ぶ)。要素 s は、述語の主語の敬称の有／無、に応じて1/0の値を取る。要素 o は、述語の補語の敬称の有／無、に応じて1/0の値を取る。要素 e は、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、述語の敬語タイプが、常体／尊敬語／謙讓語 a ／謙讓語 b ／二方面敬語、に応じて0/1/2/3/4の値を取る。ここで、“謙讓語 a ”とは、補語を高める謙讓語、“謙讓語 b ”とは、補語を高めない謙讓語、“二方面敬語”とは、尊敬語かつ謙讓語 a の語を指す。

表2: 敬語タイプ辞書 (一部) (“～” は動詞)

形態素の部分的並び	敬語タイプ
佐藤+さん	敬称
高橋+氏	敬称
お+～+する	謙讓語 a
ご+～+する	謙讓語 a
頂く	謙讓語 a
申す	謙讓語 b
お+～+なさる	尊敬語
ご+～+に+なる	尊敬語
です。	丁寧語

3. 3 入出力

システムの入出力例を図1に示す。この例では、高橋と山田(話者)がウチの関係、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(聞き手)はソトの関係にある。

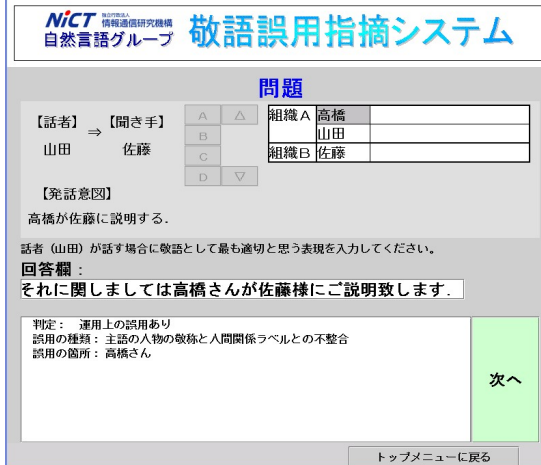


図1：入出力例

システムは入力文（回答欄に自由記述）の語形上の誤用、及び入力文と人間関係情報（上・下×ウチ・ソトの情報、図1右上の枠の形式で自由に指定可）の整合性をチェックし、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”）と人間関係情報（話者と高橋は、聞き手とはソトの関係）が整合しないため、“運用上の誤用”と判定されている。

3. 4 処理の流れ

本システムにおける処理フローを図2に示す。入力文はまず形態素解析され、形態素の並びに変換される。次に、語形誤り表現リスト（表1）を用いて、語形上の誤用がチェックされる。語形上の誤用が見つからなかった場合には、文構造解析によって述語の主語と補語の同定を行う。文構造解析の結果、及び敬語タイプ辞書を用いて敬語特徴パターンを抽出する。最後に整合表を用いて、敬語特徴パターンと人間関係情報（上・下×ウチ・ソト）の整合性がチェックされる。整合表は、敬語特徴パターンの各変数の値と、与えられた発話状況の整合性を定義した表である。“発話状況”とは、システムに与えられた人間関係情報（上・下×ウチ・ソト）と主語／補語の情報を合わせたものである。整合表から、登場人物3名、 s に関する部分を抜粋したものを表3に示す。表3の上半分では、 $s=1$ と整合する発話状況が、論理式： $(L>A>S)(p1:A)$ or \dots or $(S)(A)(p2:S)$ として定義されている。ここで、記号“S”，“L”，“A”はそれぞれ“山田”，“佐藤”，“高橋”を指す。同

じ“()”内に現れた記号が指す人物はウチ、そうでない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。“ $(p1:X)$ ”，“ $(p2:Y)$ ”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。

図1の例では、 $s=1$ （高橋の敬称あり）と（高橋>山田）（佐藤）との整合性がチェックされたが、 $s=1$ に対応する発話状況を表す論理式群の中で $(A>S)(L)(p1:A)$ を満たす論理式が見つからなかったため、“主語の敬称有り”が運用上の誤用として判定されたことになる。整合表は、学習データに基づいて機械学習（決定リスト）で構築した。学習データは、全ての社会的関係（発話に関わる人物数 $2 \cdot 3 \cdot 4$ 名×上・下×ウチ・ソト、の全ての組み合わせ=273通り）のそれぞれの状況に対応する文として規範的に正しいと考えられる文であり、国語学に関する様々な文献において共通して明示的あるいは暗示的に述べられていると解釈できる規範に基づいて我々が作成した。

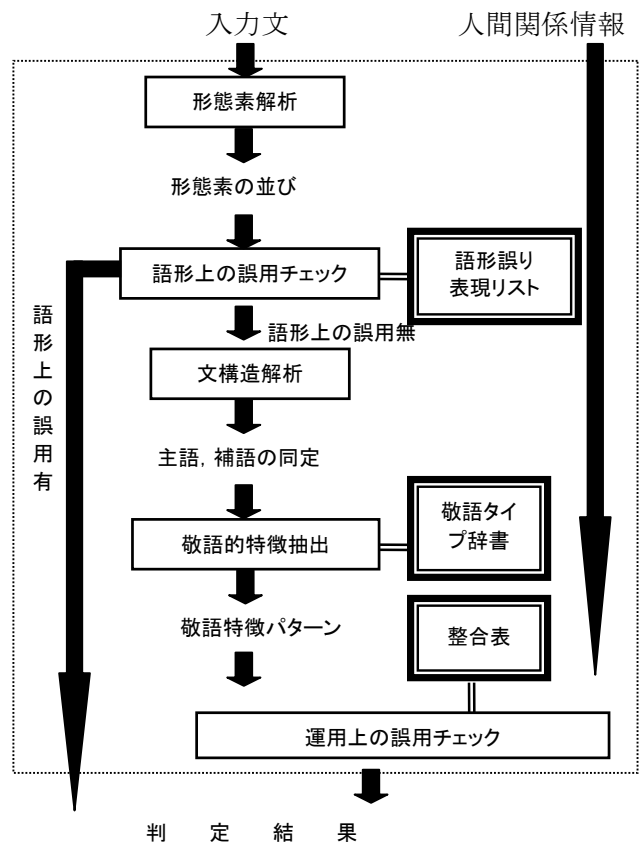


図2：処理フロー

表 3: 整合表 (発話に関わる人物 3 名, s 抜粋)

敬語特徴	発話状況
$s=1$	(L>A>S)(p1:A) (L=A>S)(p1:A) (A>S=L)(p1:A) (A>L>S)(p1:A) (A>S>L)(p1:A) (L>A>S)(p2:S) (A>S=L)(p2:S) (A>L>S)(p2:S) (A>S>L)(p2:S) (L=A>S)(p2:S) (L>S)(p1:L) (L=A>S)(p2:L) (L>A>S)(p2:L) (A>S>L)(p2:L) (A>S=L)(p2:L) (A>L>S)(p2:L) (S)(A)(p1:A) (S)(L)(p1:L) (S)(A)(p2:L) (S)(A)(p2:S)
$s=0$	(A>S)(p1:S) (L>S)(p1:S) (S=A)(p1:S) (S=L)(p1:S) (S>L)(p1:S) (L>A)(p1:S) (S>A)(p1:S) (L=A)(p1:S) (A>L)(p1:S) (A)(L)(p1:S) (A)(S)(p1:S) (S)(L)(p1:S) (S>L)(p1:L) (S=L)(p1:L) (S>A)(p1:A) (S=A)(p1:A) (A>S)(L)(p1:A) (S>A)(p2:S) (S=A)(p2:S) (S=A)(p2:L) (S>A)(p2:L) (S>L)(p2:A) (S=L)(p2:A) (L)(A>S)(p2:L) (A>S)(L)(p2:S) (S>L=A) (S=A>L) (S=L>A) (S>A>L) (S>L>A) (S=L=A)

4. システムの妥当性の検証

本システムの妥当性を検証するため、テストセット 1 : 著者らが作成した正例+負例 (“正/負例”) とは規範的/非規範的な例文を指す。ここで用いた正例は機械学習に用いたものと同じ、負例は機械学習には未使用), 及びテストセット 2 : 第三者 (著者らの研究グループとは独立した国語学研究グループ) が作成した正例+負例を用いた実験を行った。テストデータ 1, テストデータ 2 に含まれる文の総数はそれぞれ, 約 1,600 文, 約 8,200 文 (いずれも, 正/負例概ね半分ずつ) で

ある。これら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, 全ての正例に対し “文は正しい” と判定し, 全ての負例に対し “誤用” と判定した。一方, テストデータ 2 に関しては, 正例の約 99.4% に対し “文は正しい” と判定したが, 残り約 0.6% については “誤用” と判定した。“誤用” と判定された主な文は, 作成者のミスと思われるものであった。また, システムは負例の約 98% に対し “誤用” と判定したが, 残り約 2% に対しては “文は正しい” と判定した。

“正しい” と判断した文のうち典型的なものは, 「あげる」や「くれる」を含む文であり, 本研究ではこれらの表現を常体と見なしていることに起因する。この取り扱いは今後の課題である。

以上の結果は, 本研究の枠組み (敬語の規範に関する考え方, 発話状況の取り扱い, 及び対象とする文に関する制約) の下で, 敬語として規範的な表現と非規範的な表現を, 本システムが概ね正しく判断できることを示唆する。

5. むすび

日本語発話文, 及び発話に関わる人物間の社会的関係の情報を入力とし, 入力文における誤用の有無, 及び誤用が含まれる場合にはその種類と箇所を出力するシステムを開発した。本システムは, 敬語に関する語形上の誤用, 及び運用上の誤用の両方を指摘することができる。発話に関わる人数は 2 名~4 名を取り扱うことができ, 人物間の社会的関係としては上・下, 及びウチ・ソトが考慮されている。規範的な例文と非規範的な例文を用いてシステムの妥当性を検証したところ, 本システムが概ね妥当な出力を行うことが確認できた。

実用的なシステム構築のためには, 言語の専門家ではない一般の人々が実際にどのような敬語を運用しているかの調査を行い, その結果に基づいて誤用の深刻さの重み付けを行うことが重要であると考えられる。更に, 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できることが望ましい。今後は, 以上の課題に関する機能拡張を行う。

参考文献

[1] 白土保, 丸元聡子, 村田真樹, 井佐原均, “日本語発話文に含まれる敬語の誤用を自動的に指摘するシステム —プロトタイプの開発—”, 言語処理学会 2005 年年次大会 (2005 年 3 月)。