

## 日本語学習者支援のための敬語変換タスクの提案

## Proposal of Honorific Style Transfer Task for Supporting Japanese Language Learners

松本 悠太<sup>\*1†</sup>  
Yuta Matsumoto林崎 由<sup>\*1†</sup>  
Yu Hayashizaki北山 晃太郎<sup>\*1</sup>  
Kotaro Kitayama舟山 弘晃<sup>\*1\*2</sup>  
Hiroaki Funayama三田 雅人<sup>\*2</sup>  
Masato Mita乾 健太郎<sup>\*1\*2</sup>  
Kentaro Inui<sup>\*1</sup>東北大学  
Tohoku University<sup>\*2</sup>理化学研究所  
RIKEN

Honorific Language is considered important for smooth communication in social life in Japan, but it is difficult for non-native learners of Japanese to use honorific language properly. However, there has been little discussion about the development of a system that automatically converts ordinary language style, and there is no data to evaluate such a system. In this study, we propose a new task called honorific conversion, which automatically converts arbitrary ordinary sentences into appropriate honorific sentences while preserving the meaning. We also provided a data annotation scheme and an evaluation dataset to establish this task. Furthermore, we show the validity and feasibility of solving the salutation transformation task as a type of style transformation task by investigating the use of sentence decomposition methods.

## 1. はじめに

敬語とは同じ意味のことを述べるのに話題中の人物・聞き手・場面などに応じて複数の表現を使い分ける待遇表現である。日本において敬語は社会生活上でのコミュニケーションを円滑に行う上で重要だとされている。一方で、敬語を正しく使うためには高度な日本語に対する知識が必要となる。例えば、食べる → 召し上がるのような動詞ごとの不規則変化に関する知識や、同じ動詞であっても話者と聞き手の関係、動作主の違いによって尊敬語（言う → おっしゃる）や謙讓語（言う → 申し上げる）といった敬語表現を使い分けることが必要となる。このことから、敬語を正確に使い分けることは日本語を母語としない日本語学習者にとって非常に困難である。

このような背景から、敬語表現の誤用に関する分析 [丸元 01] や会話文中の敬語の誤用を自動的に指摘するシステムの開発 [白土 04, 白土 05, 白土 06] などがこれまで研究されてきた。一方で常体文を敬体文に変換するようなシステムの開発についてはほとんど議論されていない。数少ない先行研究として、野口ら [野口 07] や飛鳥井 [飛鳥井 15] はルールによって常体の動詞を敬語へと自動変換するシステムの研究を行っているが、その対象となる表現や変換の評価は限定的なものである。実用的な日本語学習者支援のためには、話者と聞き手の関係などの文脈を考慮しつつ常体文を敬体文へと変換するようなシステムや、そのシステムを評価する尺度が必要となる。そこで本研究ではスタイル変換タスクの一種として、任意の常体文に対して、意味を保持しつつ適切な敬体文へと自動変換を行う**敬語変換**という新たなタスクを提案する（図 1）。本研究では、敬語変換タスクの確立に向けた最初の一步として、敬語変換のための評価用データセットを構築する。さらに、文表現の分解手法を用いた調査によって敬語変換をスタイル変換タスクの一種として見なして解くことの妥当性および実現可能性を示す。

本研究の貢献は以下の 3 点である：（1）日本語学習者支援のための敬語変換タスクを提案した。；（2）敬語変換タスクの実現に向けて、敬語変換タスクのための評価データセットを提供した。；（3）文表現の分解手法を用いた調査によって敬語変換

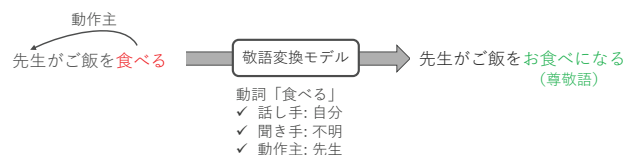


図 1: 敬語変換タスクの概要。

タスクをスタイル変換タスクの一種として見なして解くことの妥当性および実現可能性を示した。

## 2. 敬語変換タスク

本節では、敬語変換のタスク定義を行う。敬語変換とは、与えられた任意の常体文に対し、文の意味を保持しながら適切な敬語表現を含む文（敬体文）へと自動変換するタスクである。敬語変換タスクは、常体や敬体といった日本語固有のスタイルを変換することと同義のため、既存のスタイル変換（Style Transfer）タスク [McDonald 85, Hovy 88] の一種として見なして定義できる。スタイル変換を形式的に定義すると、変換先の文を  $x'$ 、変換先のスタイル属性を  $a'$ 、変換元のスタイル属性  $a$  をもつ変換元の入力文を  $x$  と定義したとき、スタイル変換の目的は  $p(x'|a, x)$  をモデル化することである。ここで敬語変換では、 $a'$  および  $a$  はそれぞれ敬体および常体となる：

入力文  $x$ : 先生がご飯を食べる。 入力属性  $a$ : 常体  
変換文  $x'$ : 先生がご飯をお食べになる。 変換属性  $a'$ : 敬体

最も敬語変換タスクに関連するスタイル変換として、主に英語を対象にした、くだけた文体（Informal）から形式ばった文体（Formal）へのスタイル変換（Formality に関するスタイル変換）が挙げられる [Rao 08, Briakou 21]。しかし、敬語変換は Formality に関するスタイル変換を含む従来のスタイル変換と大きく次の 2 点で異なる：（1）変換先のスタイルが複数あること；（2）話し手・聞き手・動作主など文脈を考慮しながら複数のスタイル候補の中から妥当なスタイルを同定する必要があること。（1）に関しては、具体的に敬体というスタイルには尊敬語や謙讓語といったさらに粒度の細かい複数のス

連絡先: yuta.matsumoto.q8@dc.tohoku.ac.jp(松本悠太)

† equal contribution

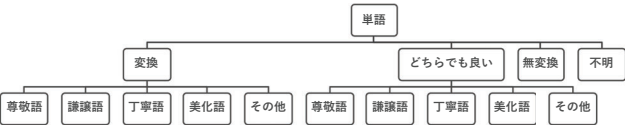


図 2: 敬語変換のアノテーションラベル構造。

イルが存在する。(2)に関しては、例えば、図 1 に示す例においては、着目する単語「食べる」を適切な敬語に変換するためには、話し手と「食べる」の動作主である「先生」との関係を考慮したうえで妥当なスタイルとして「尊敬語」を同定する必要がある。このように、本研究が提案する敬語変換タスクにおいては、従来のスタイル変換よりも文脈に応じた高度なスタイル変換技術が要求される。

3. データセット構築

本節では敬語変換のための評価データセットの構築手順およびそのアノテーションスキームについて述べる。

3.1 データの選定

常体から敬体への変換アノテーションを実施するデータの選定にあたり、本研究では次の 3 つの基本要件を定めた: (1) 常体文が多く含まれていること; (2) 日本語学習者による作文であること; (3) 文法誤りは修正済みであること。要件 1 は、敬語変換タスクを成立させるための必要条件であるといえる。要件 2 に関して、1. 節でも述べた通り、敬語変換タスクの主な目的は日本語学習者支援であるため、入力データの書き手も日本語学習者であることが望ましいと考える。要件 3 は、日本語を母語としない日本語学習者の多くは、敬語の正確な使い分けと同様に文法的に正しい文を書くことが困難であるため学習者文には文法誤りも含まれることが考えられる。しかし、自然言語処理分野にはすでに文中に含まれる文法誤りを自動訂正する文法誤り訂正 (Grammatical Error Correction) タスクが確立されており、日本語を対象とした研究も近年活発に行われていること [小川 20, Koyama 20], また本研究はあくまで敬語変換に焦点を当てることを目的としているため事前に文法誤りが修正済みである学習者文が望ましいと考える。上記の 3 つの要件を満たすため、本研究では、TEC-JL [Koyama 20] の“訂正文”を元データとして選定した。TEC-JL は学習者コーパスの一つである Lang-8 コーパス [水本 13] 中の日本語学習者文に対して、人手で訂正アノテーションを行った日本語文法誤り訂正の評価用データセットである。Lang-8 コーパスは SNS 型語学学習サービスに基づいて作られていることから、データの性質上作文のトピックは日記調のものが多いため、常体文が自然に含まれていると考えられる。

3.2 アノテーション

本研究では、敬語表現への知識と十分な日本語読解能力を有する 2 名の日本人作業員によりアノテーションを実施した。なお、本研究ではアノテーション作業にあたり作業員には以下のような指示を与えた。

- 1. 敬体に変換すべき、ないしは変換可能な文に対して、元文の意味を保持して変換を実施する。
- 2. 敬体への変換だけではなく、より丁寧な表現がある口語はより丁寧な表現に言い換える。
- 3. 敬体に変換可能かどうかは、文章中の話し手・聞き手・動作主・行為の受け手の関係を総合的に考慮して実施する。

表 1: 変換必要性ラベルに関する統計量。

	作業員 1		作業員 2	
	#	%	#	%
変換	31	15.5	20	10.0
どちらでも良い	47	23.5	70	35.0
無変換	119	59.5	107	53.5
不明	3	1.5	3	1.5

表 2: 変換した敬語の種類に関する統計量。

	作業員 1		作業員 2	
	#	%	#	%
変換: 尊敬語	6	7.14	4	3.88
変換: 謙譲語	3	3.57	1	0.97
変換: 丁寧語	24	28.57	5	4.85
変換: 美化語	0	0.00	1	0.97
変換: その他	2	2.38	13	12.62
どちらでも良い: 尊敬語	2	2.38	10	9.71
どちらでも良い: 謙譲語	1	1.19	5	4.85
どちらでも良い: 丁寧語	35	41.67	49	47.57
どちらでも良い: 美化語	0	0.00	2	1.94
どちらでも良い: その他	11	13.10	13	12.62

- 4. 敬体へ変換する際にどういった変換を表すかに関する変換ラベルを付与する。

ここで項目 4 に関して、アノテーションする変換ラベルの構造を図 2 に示す。作業員はまず大分類として、敬体に変換すべきであれば「変換」、聞き手や行為の受け手の不明瞭なことから変換してもしなくても良い場合は「どちらでも良い」、敬体に変換すべきでなかったりすでに正しい敬体文である場合は「無変換」、元文の意図が汲み取れない場合は「不明」の全 4 種類のラベルを付与する (図 2 の 2 段目参照)。次に、小分類として、「変換」「どちらでも良い」のラベルを付与する場合、単語単位で「尊敬語」、「謙譲語」、「丁寧語」、「美化語」、「その他」のさらに粒度の細かいラベルを付与する (図 2 の 3 段目参照)。例えば、尊敬語に変換すべき箇所では「変換: 尊敬語」のようにラベルを付与する。このように、本データセットでは、単にどのタイプの敬語に変換するだけでなく、必ず変換すべきなのか、どちらでも良いものなのかといった変換必要性の強弱が表現可能なアノテーションスキームとなっている。上記の手順でアノテーションを行った結果、2 つのリファレンスを持つ 390 文対の評価データセットを構築した。

3.3 分析

まずはじめに、構築した評価データにどのような種類の敬語変換がどの程度含まれているか分析する。表 1 は、構築された評価データからランダムに 200 文抽出してきたサブセットにおける大分類ラベルの分布、つまり変換必要性に関する分布を示している。この結果から、どちらの作業員も半分以上のデータについては変換する必要がないとしていることがわかる。これは、日本語学習者が書く文においては、入力文体および文脈によって敬体への変換の必要が無い文が一定量含まれることを示唆している。そのため、敬語変換モデルは、「変換しない」という選択肢も含めて変換すべきかどうかを正確に判別でき

表 3: 作業者間で変換必要性の強弱が一致しなかった例.

変換前	変換後	付与ラベル	
		作業者 1	作業者 2
外国人が日本に来て数年で花粉症になったという話もよく聞く。。。。	外国人が日本に来て数年で花粉症になったという話もよく聞きます。。。。	変換: 丁寧語	どちらでも良い: 丁寧語
漫画とアニメはヤンキーが人気になって、たくさんの日本人はヤンキーになりたいと思う。	漫画とアニメはヤンキーが人気になって、たくさんの日本人はヤンキーになりたいと思います。	変換: 丁寧語	どちらでも良い: 丁寧語

表 4: 作業者間で変換後の文が一致しなかった例.

変換前	変換後	
	作業者 1	作業者 2
「いいアイデアが一つあったら、金持ちになれますよ」と言った人がいる。	「いいアイデアが一つあったら、金持ちになれますよ」と言った人が <b>いらっ</b> しゃいます。	「いいアイデアが一つあったら、金持ちになれますよ」と言った人が <b>います</b> 。
まず、言っておきたいのは Mikkom さん、コメントしてくれてありがとうございます。	まず、 <b>申し上げたい</b> のは Mikkom さん、コメントしてくださってありがとうございます。	まず、 <b>申し上げておきたい</b> のは Mikkom さん、コメントしてくださってありがとうございます。
虎の王（キング）は八百匹の虎を率いて開封の城の下へやってきた。	虎の王（キング）は八百匹の虎を率いて開封の城の下へやって <b>きました</b> 。	虎の王（キング）は八百匹の虎を率いて開封の城の下へやって <b>来られました</b> 。

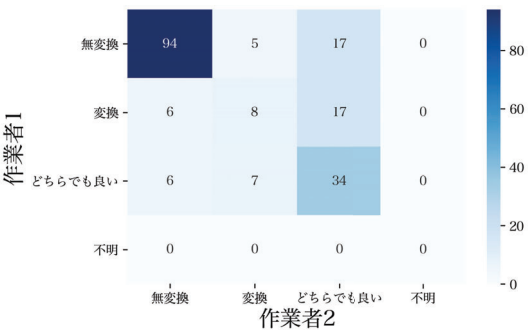


図 3: 作業者間における変換ラベルの混同行列.

ることが期待される。次に、変換が行われた場合、つまり「変換」あるいは「どちらでも良い」が選ばれた際の小分類ラベルの分布を表 2 に示す。この結果から、変換の種類としては丁寧語が最も多いこと、尊敬語の訂正が謙譲語の訂正の 2~4 倍含まれることがわかる。

次に、作業者間でどの程度変換が一致したか分析するために、同じサブセットに対して変換必要性に関する一致度を算出したところ 68%であった。ここで、作業者間でどのように判断が揺れているのかをより詳細に分析するために、変換必要性ラベル付与に関する混同行列を図 3 に示す。変換を行わない文に対しては高い一致度となっているが、「変換」と「どちらでも良い」の変換必要性の強弱に関しては作業者間で大きく判断が揺れていることがわかる。表 3 にその具体例を示す。表 3 の例にあるように、変換後の文が同じであっても、その変換必要性の強弱には不一致が見られる。敬体に変換すべきかどちらでも良いかは話し手と聞き手の関係に依存するものである一方、本データセットでは文単位でのアノテーションを行ったため、一文から話し手と聞き手の関係を推測することが難しかったことがこのような不一致につながったものだと考えられる。また、同様のサブセットにおいて変換後の文自体の一致率を算出したところ、無変換を含む全サブセットにおける一致率は 65%であり、両作業者ともに編集した文に限定して一致率を測ったところ 44%という結果となった。ここで、作業者間の変換が一致しなかった場合について定性的に分析する。表 4 は作業者間で変換後の文が一致しなかった例である。表 4 の 1 段目で見られた「います」と「いらっしゃいます」の例に示

されるように、敬語変換における妥当な変換は必ずしも一意に決まらないことがわかる。

4. 調査

本節では、敬語変換を常体文から敬体文へのスタイル変換と見なして解くことの妥当性およびその実現可能性について調査することを目的とする。具体的には、文表現の分解手法を用いて、常体および敬体がそれぞれ分解可能なスタイル属性なのかについて調査する。

4.1 文表現の分解手法

本研究では文表現の分解手法として ADNet [Romanov 19] を用いた。ADNet は系列変換モデルに基づく敵対的文章分解モデルであり、文を複数の構成要素に分解する。Romanov ら [Romanov 19] は、ADNet をスタイル変換タスクに適用することで、文章を意味ベクトルとスタイルベクトルに符号化できることを示した。本研究では、常体文集合および敬体文集合に対して ADNet を用いてそれぞれ意味ベクトルとスタイルベクトルに符号化し、常体文におけるスタイルベクトルと敬体文におけるスタイルベクトルが分離可能かについて調査する。

4.2 実験設定

ADNet を訓練するにあたり、次に示す方法で訓練データの構築およびモデルの訓練を行った。まず訓練データの構築にあたり、大規模 Web クロールデータである Common Crawl\*1 から日本語テキスト [Buck 14] を抽出し、整形スクリプト\*2 を用いて整形した。具体的には、同一行内で句点に挟まれている部分を文として抽出し、長さが 200 文字を超える文、バイナリ文字を含む文、および空白を削除した。次に、整形した日本語テキストに対し、MeCab[工藤 13] により形態素単位で分かち書きを行い、蒲谷 [蒲谷 09] の敬語使い方辞典に基づき人手作成した辞書およびルールを用いたパターンマッチングにより敬体文と常体文をそれぞれ 50 万文抽出した。具体的には、敬語表現には、「言う」→「おっしゃる」のような語幹自体が変わる**不規則変化の敬語**と、「話す」→「お話になる」、「話される」のような語幹は変わらず一定の規則により変化する**規則変化の敬語**の 2 種類が存在する [蒲谷 09]。本研究では、不規則変化の敬語表現に関しては人手で辞書を作成、規則変化の敬語

\*1 <https://commoncrawl.org/>  
\*2 <https://github.com/cacaho/ja-sentence-distiller>



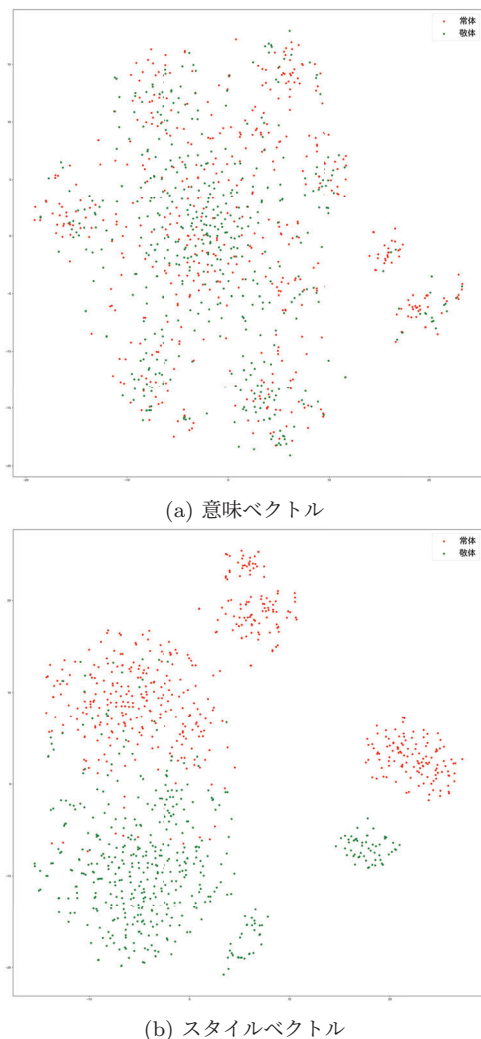


図 4: 意味ベクトルとスタイルベクトルの可視化結果。

表現に関しては「お/ご～になる」「お/ご～する/いたす」のような規則を正規表現で表し、それぞれパターンマッチを用いて抽出した。なお、一文中に 2 回以上当該敬語パターンが含まれていた場合はその文は敬体、文中に一度も当該敬語パターンを含まず、かつ 11 字以上の場合には常体と判定して抽出した<sup>\*3</sup>。

単語埋め込みには学習済みの日本語 fast-Text [Bojanowski 17] を使用した。学習はバッチサイズ 128 文で 1000 エポックまで更新し、検証データにおける損失関数の値が最も低いモデルを採用した。語彙数は 50000 語とし、その他のハイパーパラメータは Romanov らの設定に従った。

### 4.3 結果

図 4 は、検証用データの一部 (1000 文) における常体文集および敬体文集に対し、訓練した ADNet を適用して意味ベクトルとスタイルベクトルに分解しそれぞれ符号化したものを t-SNE アルゴリズム [Maaten 14] により 2 次元に次元削減して可視化したものを示している。

図 4 (a) から、スタイルの違いによる意味表現の偏りは見られず、ほぼ均質な分布となっていることが確認できる。また、

<sup>\*3</sup> パターンが 1 回のみ含まれている文は、「これらのものは個別に判定される」など、常体であるにも関わらず敬語のパターンに当てはまる誤分類が多かったため訓練に使用しなかった。

図 4 (b) から、スタイルベクトルは敬体文と常体文で 2 つのクラスターに明らかに分かれていることがわかる。このことから、常体および敬体がそれぞれ分解可能なスタイル属性であることが示されたため、敬語変換をスタイル変換の枠組みとして解ける可能性があることを示唆された

## 5. おわりに

本研究では任意の常体文から適切な敬体文へ変換する敬語変換タスクを提案し、その確立に向けて評価データセットを構築した。さらに、文表現の分解手法を用いて、敬語変換をスタイル変換の一種とみなして解くことの妥当性および実現可能性について明らかにした。今後は実際に任意の常体文を敬体文に変換するシステムの開発、そして今回構築したデータセットによる評価を目指したい。

## 参考文献

- [Bojanowski 17] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T.: Enriching Word Vectors with Subword Information, *TACL*, Vol. 5, pp. 135–146 (2017)
- [Briakou 21] Briakou, E., Lu, D., Zhang, K., and Tetreault, J.: Olá, Bonjour, Salve! XFORMAL: A Benchmark for Multilingual Formality Style Transfer, in *NAACL*, pp. 3199–3216, Online (2021)
- [Buck 14] Buck, C., Heafield, K., and Ooyen, van B.: N-gram Counts and Language Models from the Common Crawl, in *LREC* (2014)
- [Hovy 88] Hovy, E. H.: Generating Natural Language Under Pragmatic Constraints (1988)
- [Koyama 20] Koyama, A., Kiyuna, T., Kobayashi, K., Arai, M., and Komachi, M.: Construction of an Evaluation Corpus for Grammatical Error Correction for Learners of Japanese as a Second Language, in *LREC*, pp. 204–211 (2020)
- [Maaten 14] Maaten, van der L.: Accelerating t-SNE using Tree-Based Algorithms, *JMLR*, Vol. 15, No. 93, pp. 3221–3245 (2014)
- [McDonald 85] McDonald, D. D. and Pustejovsky, J. D.: A Computational Theory of Prose Style for Natural Language Generation, in *EACL* (1985)
- [Rao 08] Rao, S. and Tetreault, J.: Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer, in *NAACL*, pp. 129–140 (2008)
- [Romanov 19] Romanov, A., Rumshisky, A., Rogers, A., and Donahue, D.: Adversarial Decomposition of Text Representation (2019), <https://arxiv.org/abs/1808.09042>
- [蒲谷 09] 蒲谷 宏: 敬語使い方辞典, 新日本法規 (2009)
- [丸元 01] 丸元聡子, 白土保, 井佐原均: 敬語表現の誤用に関する統計的分析, 言語処理学会第 7 回年次大会 (2001)
- [工藤 13] 工藤 拓: MeCab: Yet Another Part-of-Speech and Morphological Analyzer (2013), <https://taku910.github.io/mecab/>
- [小川 20] 小川 耀一郎, 和英: 日本語文法誤り訂正における誤り傾向を考慮した擬似誤り生成, 言語処理学会 第 26 回年次大会 (2020)
- [水本 13] 水本智也, 小町守, 永田昌明, 松本裕治: 日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得, 人工知能学会論文誌, Vol. 28, No. 5, pp. 420–432 (2013)
- [白土 04] 白土保, 丸元聡子, 村田真樹, 井佐原均: 会話文中の敬語の誤用を自動的に指摘するシステムの試作 –教科書的知識に基づくルール構築–, 言語処理学会 第 10 回年次大会 (2004)
- [白土 05] 白土保, 丸元聡子, 村田真樹, 井佐原均: 会話文中の敬語の誤用を自動的に指摘するシステムの試作 –プロトタイプの開発–, 言語処理学会 第 11 回年次大会 (2005)
- [白土 06] 白土保, 丸元聡子, 村田真樹, 井佐原均: 会話文中の敬語の誤用を自動的に指摘するシステムの試作 –ウチソトの考慮, 及び GUI の開発–, 言語処理学会 第 12 回年次大会 (2006)
- [飛鳥井 15] 飛鳥井 元晴, 岸 義樹: 敬語文章変換システムの作成, 情報処理学会第 77 回全国大会 (2015)
- [野口 07] 野口 聡, 南條 浩輝, 吉見 毅彦: 動詞の通常表現から敬語表現への換言, 言語処理学会 第 13 回年次大会 (2007)