

平成 19 年度 情報工学コース卒業研究報告要旨

石井 研究室	氏 名	河田純奈
卒業研究題目	コーパスを用いた話し言葉・書き言葉変換ルールの自動生成	
<p>近年，人間と対話のできるコンピュータ（対話システム）は様々な分野で利用されている．チケット予約のような，タスクを達成するために対話をするのではなく，対話すること自体が目的である対話システムは非タスク指向型対話システムと呼ばれている．対話システムの実現のためには性能の定量的評価が不可欠であるが，タスク指向型システムとは異なり，非タスク指向型システムの定量的評価は困難である．非タスク指向型システムの定量的評価法としては，磯村らの HMM を用いた評価法がある．この評価法では，発言の種類を表す「タグ」を各発言に付与する必要がある．しかし，手動によるタグ付与はコストがかかるため，磯村らはこのタグ付与を自動で行う手法を提案している．</p> <p>この自動タグ付与には発言中の単語の出現確率を用いるため，同じ意味の発言でも，用いる表現の種類が多いと自動タグ付与の精度が低くなるという問題がある．そのため，表現を統一する手法が求められている．</p> <p>本研究では統一した表現として書き言葉に着目し，自動タグ付与の精度向上のために話し言葉から書き言葉へ変換する手法を提案する．本研究では，話し言葉から書き言葉への変換ルールをコーパスから自動生成する．本手法では話し言葉の対話と，それを人間が書き言葉に変換した対話をコーパスとし，それらの差分をもとに文節単位で変換ルールを生成する．差分に含まれる単語をそのまま使うのではなく，動詞や形容詞などを一般化して汎用的なルールを生成する．また，誤変換への対策として，変換ルール生成時に前後の単語を付与する処理を行う．</p> <p>使用するルール数を変えて変換の精度を調べる実験を行った結果，適合率が 94%，再現率が 60%であった．</p> <p>本手法を適用した対話に自動タグ付与を行った結果，意味的に近いものを許容した正解率は 88.35%となった．本手法を適用しない場合は 87.26%であった．これにより，本手法が自動タグ付与の精度向上に有効なことが確認された．</p>		