

平成14年度 情報工学コース卒業研究報告要旨

稲垣 研究室	氏 名	松本 和之
卒業研究題目	語形変化を含むチャット文の日本語形態素解析手法	
<p>インターネットの普及により、ネットワークを用いて行われるチャットが盛んになった。それに伴い日本語チャットの履歴を利用する研究が増えてきている。チャット文には、“(̄▽̄)”などの顔文字、“がっこう”に対する“がっこー”のような語形変化、“うわああん”などのバリエーション豊かな感情表現といった、チャット特有の表現がよく見られる。日本語を処理する上で、形態素解析が必要になる場合が多いが、こういった表現を含むチャット文を、従来の形態素解析システムで正しく解析することは難しい。そこで本研究では、チャット文に現れる特徴の一つである語形変化という現象に着目し、従来の形態素解析システムの辞書に語形変化形を追加することによって、チャット文に対する形態素解析の正解率を向上させる。</p> <p>まず、チャットの履歴から2356発言を調査し、どのような品詞においてどのような語形変化が起こるかという統計をとり、収集した語形変化から、“ou→o-”のようなローマ字表記の語形変化規則を作成した。次に、既存の日本語形態素解析システム MAJO 付属の辞書に登録されている単語を、この語形変化規則に従って語形変化させることで、新たな単語を得た。例えば先に挙げた規則を“gakkou(がっこう)”という単語にこの規則を適用することで、新たに“gakko-(がっこー)”という単語が得られる。このようにして得た単語250,313語を、MAJOの辞書に追加登録した。最後に実験として、辞書登録の前後でチャット文の形態素解析結果にどのような変化があるのか調べたところ、正解率は54.8%から59.9%に向上した。</p>		