

# 平成15年度 情報工学コース卒業研究報告要旨

石井(健)研究室	氏名	小澤 猛 志
卒業研究題目	話し言葉によるテキスト対話のための同義語処理	
<p>本研究の目的は、人とコンピュータとの日本語による対話の実現である。対話を通じて、利用者が癒しを感じ、楽しむことを目指す。本目的を達成するには、格式ばった書き言葉ではなく、自然で親しみやすい話し言葉での対話実現が必要になる。また将来、音声による対話を実現するには、話し言葉の理解は不可欠である。</p> <p>1966年、MITのJoseph Weizenbaumによって自然言語での対話プログラムELIZAが開発・発表されている。ELIZAとの対話では、ELIZAを人間だと錯覚する利用者がいることや対話において癒しを感じた利用者がいることが確認されている。しかし、ELIZAが扱っている言語は英語である。そこで、本研究ではELIZAを参考にし、日本語による対話プログラムKELDIC (Ken's Laboratory Dialog Computer) の設計を行った。日本語には英語にはない特徴があり、また話し言葉にも独自の特徴がある。それらの特徴が原因で、ELIZAをそのまま真似ただけでは対話が成り立たない。そこで、形態素解析プログラムやシソーラス辞書、同義語辞書、和英辞書を用いた処理を加えた。日本語の話し言葉の特徴に対処するため、まず本研究では同義語の処理を進める。</p> <p>日本語の話し言葉には様々な表現のバリエーションがある。元来使われている大和言葉に加え、中国から影響を受けた漢語や欧米諸国の影響を受けたカタカナ語といった外来語が数多くある。英語で一人称はIであるが、日本語の場合、「私」や「僕」のほかにも「我」や「ミー」というものまで含まれる。話し言葉では肯定を表す言葉として「はい」だけではなく、「うん」等もよく使われる。また、通称と正式名称が違うもの、意味や読みは同じだが違う漢字で表記するものもある。このような日本語を扱う上で同義語の処理は欠かせない。しかし、市販されている同義語辞書をそのままKELDICに用いただけでは、期待した結果は得られない。そこで、市販の辞書に収録されている同義語の中からKELDICにおける出力文生成処理で許容される範囲の同義語を選び出す方法を考案した。KELDICは与えられた入力文の中から、スクリプトとして登録されている単語(キーワード)を探索する。この探索では語彙不足によりキーワードを発見できないことも起こりうる。しかし、入力文内のある単語を同義語とみなし、許容範囲外のキーワードと一致させる誤りのほうがより致命的である。すなわち、入力文を認識できないならば話題を変えるなどの対応で対話を続けることが可能であるが、誤認識してしまうと見当違いの返答文ばかりか、意味の通じない返答文を返し、対話の破綻を引き起こしかねない。そこで、本研究では誤認識を防ぐということを中心とした同義語辞書編成をめざした。実験の結果、同義語に対する誤認識の比率を大幅に下げることが成功した。</p>		