

(別紙2)

## 論文審査の結果の要旨

論文提出者氏名 星野 綾子

社会の国際化の流れを受け語学学習の重要性が高まっている。語学学習においては、学習者の能力と学習達成度を測るための語学テストが重要であるが、語学教育の専門家の人手によるテスト作成と採点結果の評価は多大な労力を要するため、それらの自動化の推進が喫緊の課題である。この論文では、実用上重要であり、かつ現在の情報技術で処理可能な語学テストとして、多肢選択空所補充問題を対象にした自動問題生成において検討すべき評価基準を洗い出し、これに基づいて自動問題生成システムを設計、実装した。実装したシステムを用いて上記の評価基準に基づく実験と評価、分析を行った結果を記した論文である。なお、対象とする言語は実用上最も重要な英語に限定している。

本論文は「Automatic Question Generation for Language Testing and its Evaluation Criteria」と題し、英文で書かれており、以下の7章からなる。

第1章「Introduction」では、まず研究の動機を述べ、次に自動生成されたテストの評価基準として質、難度、診断力を目的として定めた。

第2章「Background」では、関連分野の理論的背景をまとめている。まず、テスト理論における種々の評価尺度を説明し、特に項目反応理論について詳述している。次に語学テストの自動生成などにおいて本研究で利用した自然言語処理技術の現状を説明している。

第3章「Previous and Related Work」では、語学テストの自動問題生成に関する先行研究の調査結果をまとめ、検討すべき課題を提示した。

第4章「Automatic Question Generation」では、本研究で用いる自動問題生成手法について説明している。このシステムでは、主にWWWからダウンロードしたニュース記事を問題文に用い、人手で作成した構文パターンを構文解析結果と一致させて適切な空所となる箇所を設定する。多くのタイプの文法問題では構文パターンの抽出性能としてF値で0.7以上を得た。生成した問題の難易度判定に関しては、英語学校におけるTOEFLの正解率から得た難易判定に対して、文長、品詞、選択肢と原文からの編集距離を素性として用いた予測を種々の機械学習アルゴリズムで比較した結果、SVMによって精度0.62、特に確信度における上位15位までの問題では精度0.70を得た。ちなみに人間の難易判定は精度0.7程度であり、提案したシステムは難易判定において高い性能を持つ。

第5章「Semi-automatically Generating Questions」では、4章で提案した手法を用いて開発した対話型の自動問題生成システムと、その評価実験の結果について述べている。具体的には10名の大学学部の英語教師が利用した結果、使い勝手、生成された問題の質の両面で高い主観評価を得た。

第6章「Automatically Generating and Administrating Questions」では、評価基準のひとつである診断力に関して自動生成された問題の回答結果を主成分分析し、to不定詞とSV0という異なる側面において各々の難易が判定できるという基礎データを得た。これによって、診断力を評価する手法として主成分分析が有効であることが示された。4章で述べた難易度予測を用いて、回答者の回答に適応して問題の難易度を自動的に変更するシステムについて実験的評価を行った結果、難易度の予測性能が従来の文長のみによる場合より実際の正答率を正確に予測できた。また適応的出題は、ランダムな出題に比べて変動の少ない正答率を得ることができた。

第7章「Summary and Conclusions」は、結論および将来の課題である。

以上を要するに、本論文は、語学テストの自動問題生成手法の提案と英語を対象言語とした自動問題生成システムの設計、実装および実験的評価を行い、冒頭で述べた語学テストの自動化を行うための道筋を明らかにした。よって、本審査委員会は、本論文が博士（学際情報学）の学位に相当するものと判断する。