

報告番号	※甲	第	号
------	----	---	---

## 主 論 文 の 要 旨

論文題目 同時通訳データの分析に基づく  
漸進的な英日翻訳に関する研究  
氏 名 笠 浩一朗

### 論 文 内 容 の 要 旨

近年、グローバル化の進展により、国際会議や国際シンポジウムが盛んに開催されている。また、海外に拠点を置く会社も増えており、国際的なビジネスも盛んに行われている。このような異言語間コミュニケーションを必要とする場面を支援する情報技術として、音声翻訳システムが注目されており、その開発が進められている。

欧米では、TC-STAR や GALE と呼ばれる音声翻訳システムの開発を目的とした大規模なプロジェクトが進められている。国内でも、成田空港や関西国際空港で音声翻訳システムの被験者実験が実施されている。さらに、音声翻訳に関する国際ワークショップが毎年開催されており、音声翻訳システムの翻訳性能を競うコンテストが実施されるなど、システムの実用化に向けた動きが加速しつつある。

実際に市販された音声翻訳システムとしては、音声翻訳機能を搭載したモバイルマルチメディアプレーヤ「VoToL」、中国語と英語との会話を自動翻訳できる携帯電話サービス「JAJAH Babel」、英語とスペイン語の双方向の音声翻訳に対応した iPhone アプリ「Jibbigo」などが挙げられる。

しかし、現在の音声翻訳システムの多くは、話者の発話が終了してから訳出を開始する逐次通訳である。逐次通訳は、講演など一人の話者が話し続ける独話を対象とした場合、話者が発話を休止することがないため利用することができない。また、二人以上の話者が交代で話す対話を対象とした場合、話し手が話してから聞き手に伝わる間に時間差が生まれ、コミュニケーションの円滑さを妨げることになる。これらの問題を解消するために、話者の発話に追従して訳出する同時通訳システムの開発が進められている。

同時通訳は人間の言語活動において最も高度であるとされており、システムを実現するための有効な方法については、これまでほとんど明らかにされていない。こ

のような現状を打破するために、プロの同時通訳者の振る舞いを観察し、得られた知見を活用することは一つの方法である。

そこで本研究では、大規模な同時通訳データを用いて実際の同時通訳者の訳出方略に関する知見を獲得し、得られた知見をもとに対話を対象とする英日同時通訳システムを開発した。さらに、独話を対象とした同時通訳システムの開発を目的に、同時通訳データベースの講演データを用いて、同時通訳者の話速に関する定量的な分析を与えた。

本論文は、全5章から構成される。第1章は本論文の序論であり、同時通訳システムの開発に関する研究の動向を示すとともに、本論文の位置づけとアプローチを述べたものである。

第2章では、同時通訳者の訳出方略を獲得するための基盤データとなる同時通訳データベースの設計、構築、及び、利用について詳述する。本データベースの構築では、多様な側面から同時通訳者の発話を分析するために、対話と独話の英日双方向の通訳音声収録する。また、研究資源としての利便性を考慮し、音声データを文字化し、フィラーや言い直しなどの非文法的な言語現象に談話タグを付与する。さらに、発話単位での時間情報や対訳対応関係などの言語情報を人手によりアノテーションする。

第3章では、英日同時翻訳のための訳文生成手法を提案する。英語と日本語のように、大きく異なる性質を持つ言語間の同時翻訳では、対応する語の生起順序が異なるため、標準的な訳文を生成したのでは同時性の高い翻訳を実現することは困難である。例えば、英語はSVO構文であるのに対し、日本語はSOV構文であるため、英語話者発話のOの部分の発話が完了するまでVの訳出を待たなければならない。他にも、英語では時間を表す副詞が一般に文末に生起するのに対して、日本語では文頭に生起するため、発話が完了するまで訳出を開始できない。一方、同時通訳データベースを観察すると、英日同時通訳者は、標準的な訳文ではなく、原文の語順に近い語順の日本語文を訳文とすることにより、早い段階での訳出を可能にしていることがわかる。提案手法では、そのような訳出方法を参考に、日本語として容認可能な範囲で、原文の語順に近い語順の日本語訳を入力と同時的に生成する。本論文では、日本語として容認可能な文が満たすべき性質を、依存構造の観点から検討する。また、そのような性質を満たす訳文を生成する手法について述べる。提案手法では、日本語の語順に関する制約は比較的緩やかであるという特徴を利用し、日本語文として容認可能な範囲内で、原文の語順に準拠した訳文を生成する。訳文としての容認可能性は、日本語依存文法に基づいて検査する。すなわち、依存文法に基づく制約を満たし、かつ、英文の語順に最も近い日本語文を同時翻訳における訳文とする。ただし、日本語文として不自然でない場合に限り、倒置、すなわち、後方修飾性を満たさない依存関係を含む訳文についても生成を許容する。

提案手法を実装したシステムを用いて翻訳実験を行い、翻訳品質、及び、同時性の観点から本手法を評価した。その結果、容認可能な翻訳品質をもち、かつ、入力に対して同時性が高い英日同時翻訳を実現するために、提案手法が有効であることを確認している。

第4章では、英日同時通訳における通訳者の話速に関する定量的分析について述べる。本分析では、特に、同時通訳者の話速に影響を与えると考えられるいくつかの要因に焦点をあて、その影響の程度について分析を与える。具体的には、大局的な要因として、「講演の話速」及び「通訳者の個性」、また、局所的な要因として、講演中の「話し手の発話状態」及び「訳出の進行状況」に着目し、話速データを比較することによりそれらの要因が及ぼす影響の程度を明らかにする。定量的分析のために、本研究では、同時通訳データベースの英日通訳データを用いた。分析では、このデータベースの中から22の英語講演データに対する複数の通訳者による88例の英日同時通訳データを抽出し、使用した。また、音声認識ツールを用いて発声時刻情報を自動推定することにより、単語レベルでの話速データを作成し、分析に使用した。分析の結果、講演者の平均話速と通訳者の平均話速の間には相関関係はなく、また、講演の違いに比べ通訳者による話速の違いの方が大きいことを確認した。さらに、原発話が確定前より確定後の方が通訳者の話速が速く、講演者の発話が休止中の方が発話中に比べて通訳者の話速が速いことを確認した。

第5章では、本研究についてまとめる。最後に今後の課題として、同時通訳システムの開発のための研究課題と将来の展望について述べる。