

平成 30 年 6 月 29 日現在

機関番号：44101

研究種目：若手研究(B)

研究期間：2013～2017

課題番号：25770146

研究課題名(和文) 多角的な観点に基づく同時通訳者の通訳プロセスの定量的な分析

研究課題名(英文) Quantitative Analysis for Simultaneous Interpreters using Large Scale Corpus

研究代表者

笠 浩一郎 (RYU, Koichiro)

三重短期大学・その他部局等・准教授

研究者番号：40397451

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では、世界最大規模(100万語、182時間、200名の通訳者)の音声対訳データベースであり、かつ、基礎的なアノテーション(単語レベルの言語情報・時間情報、発話単位の対応情報)が付与された、名古屋大学同時通訳データベースに対して、高度なアノテーション情報(構文レベルの言語情報、モーラ単位の時間情報、単語単位の対応情報)を付与し、かつ、高機能な視覚化ツールを開発・利用することで、多角的な視点に基づいた定量的な分析を実施し、これまで明らかにされていない同時通訳者の通訳プロセスを解明することを目的に実施した。

本研究により、同時通訳者のタイプ分類、及び、通訳の欠落現象に関して成果があった。

研究成果の概要(英文)：In one study, we have focused on the features of utterances that were spoken by simultaneous interpreters and we have classified the type of interpreters based on the features. In this classification, we have adopted a method that consisted of two phases, the principal component analysis (PCA) phase and the k-means clustering phase. In the result, we have acquired two principal components by PCA. Moreover, we have classified the interpreters into six types. The other study is statistical analyses of missing translations in simultaneous interpretations. We analyzed the relationship between missing translations and various factors. The analyses revealed that the proportion of missing translations is high when the speech rate is high and delay is large. We also found that a high proportion of adverbs were missed in the translations, and that words at deeper positions in the syntactic structure were more likely to be missed.

研究分野：自然言語処理

キーワード：コーパス分析

1. 研究開始当初の背景

これまでに、同時通訳データベースを構築し、分析に必要な基礎的なアノテーション情報を付与し、同時通訳者の通訳プロセスを定量的に分析することで、同時通訳者の通訳プロセスの解明において、一定の成果を納めてきた。しかし、その一方で、まだその解明が十分に進んでいるとは言えない。今後、同時通訳者の通訳プロセスの解明を進めるには、より高度なアノテーション情報を付与し、付与したアノテーション情報を効率的に利用するための高度な機能を備えた視覚化ツールを利用し、多角的な観点から分析を進める必要がある。

2. 研究の目的

本研究の目的は、世界最大規模(100万語、182時間、200名の通訳者)の音声対訳データベースであり、かつ、基礎的なアノテーション(単語レベルの言語情報・時間情報、発話単位の対応情報)が付与された、名古屋大学同時通訳データベースに対して、高度なアノテーション情報(構文レベルの言語情報、モーラ単位の時間情報、単語単位の対応情報)を付与し、かつ、高機能な視覚化ツールを開発・利用することで、多角的な視点に基づいた定量的な分析を実施し、これまで明らかにされていない同時通訳者の通訳プロセスを解明することである。

3. 研究の方法

本研究では、平成25年度に英日通訳の独話データ、平成26年度に日英通訳の独話データ、平成27年度に英日通訳の対話データ、平成28年度に日英通訳の対話データに対して高度なアノテーション情報を付与する。また、分析ツールに関しては、25年度に音声データと言語データとか同期して観察できる機能を備えた分析ツールを作成し、26年度にアノテーションの検索機能を追加し、27年度以降は、分析ツールの機能を適宜修正する。さらに、多角的な分析に関しては、アノテーションの付与状況に合わせて実施する。

4. 研究成果

本研究により、主に以下の2点について研究成果を得ることができた。

- ・同時通訳者のタイプ分類
- ・訳出が欠落する状況の分析

「同時通訳のタイプ分類」に関する成果について説明する。同時通訳は、非常に高度な能力が求められる言語処理であり、その手法については明確なものは存在していない。実際、同じ原発話に対する複数の同時通訳者の訳出データを観察しても、通訳者ごとに異なる訳出内容、異なる訳出タイミングで訳出していることが分かる。しかしその一方で、共通する同時通訳の方略も観察することができる。そこには、いわゆる通訳のタイプみた

いなものを発見することができる。例えば、同時通訳には基本スタイルとして、最小訳出単位ごとに次々と訳出していく方式である「即時処理」と可能な限り長く聞いてから訳出する方式である「遅延処理」の2つがあると言われている。上記で述べた方式分類は、基本的な分け方であり、訳出する単位、及び、訳出されるまでの時間に着目したものである。同時通訳者の訳出にみられる特徴量には、それら以外にも、発話速度、訳出内容、ポーズの生起状況などがあり、これらに着目することで、通訳者のスタイルをより詳細に分類することができる。

そこで本研究では、同時通訳の訳出に表出する多数の特徴量に着目し、それらの特徴量に対して主成分分析とクラスタリング手法を適用することで、同時通訳者の通訳のタイプを分類する。本分類は、第1段階として主成分分析により、多数の特徴量を少数の指標(主成分)に圧縮する。次に、主成分分析により得られた主成分により同時通訳者をプロットしなおし、k-means法によるクラスタリングにより、同時通訳者を分類する。分類結果を評価するため、多くの同時通訳者と接する機会をもつ研究協力者にも、同じ同時通訳データをタイプに分類してもらい、本手法による自動分類結果と人手の分類結果を比較し、考察した。

10種類の特徴量をもとに、主成分分析を実施した。分析に利用する特徴量の単位およびスケールが異なるため、相関行列による主成分分析を実施した。主成分分析では、主成分の採用基準として、累積寄与率が7割以上になるように第1主成分から順次取り上げるという方法がある。本分析では、第3成分まで主成分として採用する累積寄与率が7割を超えるが、2次元の散布図でプロットした方が視覚的にわかりやすいため、今回は第2成分まで採用した。因子負荷量が絶対値で0.5を超えるものを主成分に寄与するものと解し、0.7を超えるものは主成分に大きく寄与するものと解する。

表1 因子負荷量

記号	主成分1	主成分2
X1 発話速度の平均	-0.051	-0.347
X2 発話速度の分散	.670	.026
X3 発話率	.809	.066
X4 平均ポーズ長	-.839	-.223
X5 フィラー率	.718	-.295
X6 訳出率	.441	.635
X7 訳出遅延時間の平均	-.271	-.912
X8 予測率	.061	.531
X9 文数	-.592	.379
X10 文長	.699	-.520

表1より、第1主成分では、「発話率」と「フィラー率」が強くプラスの効果を持っており、「ポーズ長」が強くマイナスの効果を持って

いる。さらに、「発話速度(分散)」、「文長」がプラスの効果を持っており、「文数」がマイナスの効果を持っている。これらのことから、第1主成分は、正の値が大きいと通訳者の『連続度』・『編集度』の傾向が強いことを、負の値が大きいと通訳者の『離散度』・『忠実度』の傾向が強いことを示す指標と言える。さらに、第2主成分では、「訳出遅延時間(平均)」が強くマイナスの効果を持っており、「訳出率」、「予測率」がプラスの効果、「文長」がマイナスの効果を持っている。そこで、第2主成分は、正の値が大きければ『即時度』が高いことを、負の値が大きければ『遅延度』が高いことを示す指標と言える。

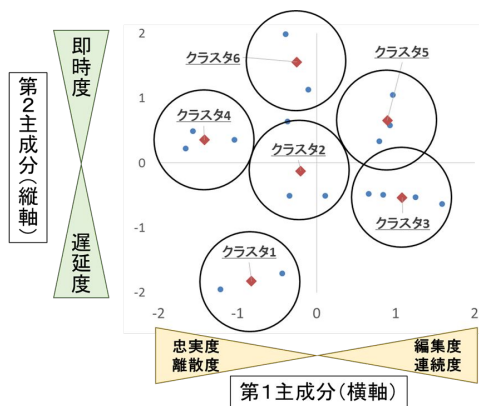


図1 タイプ分類結果

第1主成分と第2主成分に対して、各通訳者の主成分スコアを算出しプロットしたものを図1に示す。図1の横軸が第1主成分のスコア、縦軸が第2主成分のスコアを示している。丸のプロットが各通訳者を、ダイヤモンドのプロットが6つのクラスタに K-means 法でクラスターリングした場合の各クラスタの中心を示している。K-means 法では、クラスタ数を指定する必要があるが、クラスタに所属する通訳者の数が最低でも2個以上になるように、クラスタ数を調整した。得られた6個のクラスタについて、考察する。クラスタ1は、第2主成分は負に大きくなっている。このことから、遅延度が高い通訳者であり、『遅延タイプ』と言える。クラスタ2は、第1主成分と第2主成分の値が、比較的原点に近いので、『バランスタイプ』と言える。クラスタ3は、第1主成分が正に大きい値を示し、第2成分が少し負の値を示しているので、『編集・遅延タイプ』と言える。クラスタ4は、『忠実タイプ』であると言える。クラスタ5は、「即時性」が高く、「文整形力」は平均的であるので、『編集・即時タイプ』といえる。クラスタ6は、第2主成分が負に大きく、第1主成分は平均的であるので、「即時タイプ」であると言える。

次に、「訳出が欠落する状況の分析」について報告する。

講演者の話す速度と欠落の関係では、図2のような結果が出ており、話す速度が速いほど欠落率が上昇することが確認できた。

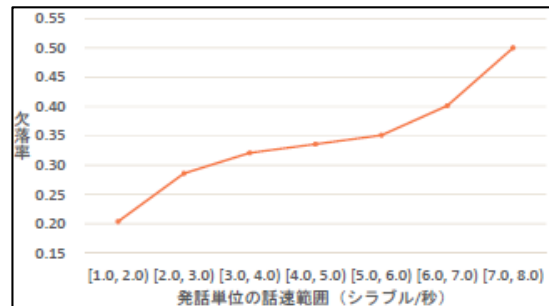


図2 講演者の発話速度と欠落率

次に、訳出遅延時間と語の欠落の関係では、図3のような結果が出ており、訳出遅延が大きくなるほど、欠落率が上昇することがわかる。

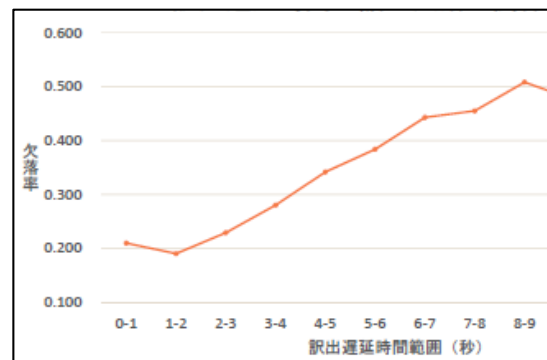


図3 同時通訳の訳出遅延時間と欠落率

さらに、語の構文構造の深さと欠落率の関係では、図4のような結果が出ており、構文構造が深くなるほど、欠落率が上昇することが分かった。

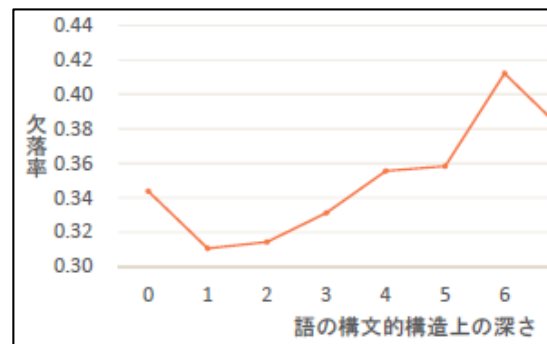


図4 語の構文構造の深さと欠落率

最後に、語の品詞と欠落率の関係では、図5のような結果が出ており、品詞によって欠落率に大きな違いがあることがわかった。

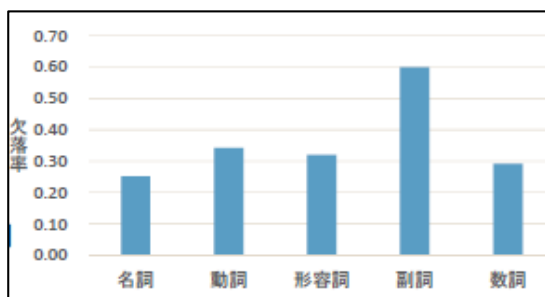


図5 品詞と欠落率

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計3件)

Zhongxi Cai, Koichiro Ryu and Shigeki Matsubara, Statistical Analysis of Missing Translation in Simultaneous Interpretation Using A Large-scale Bilingual Speech Corpus, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018

蔡仲熙、笠浩一郎、松原茂樹、同時通訳における語の欠落に関する定量的分析、日本通訳翻訳学会 18 回年次大会、2017 年

笠浩一郎、松原茂樹、主成分分析を用いた同時通訳者のタイプ分類、日本通訳翻訳学会 16 回年次大会、2015 年

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年：
 国内外の別：

取得状況(計0件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 取得年：
 国内外の別：

〔その他〕
 ホームページ等

6. 研究組織

(1) 研究代表者

笠浩一郎 (RYU, Koichiro)
 三重短期大学・生活科学科・准教授
 研究者番号：40397451