

平成 17 年度 情報工学コース卒業研究報告要旨

吉川 研究室	氏 名	ディン ジョウ
卒業研究題目	日本語対話文の節分割に基づく 同時通訳単位の獲得	
<p>近年，音声・言語処理技術の発展，及び，コンピュータを介した異言語間対話における自然で円滑なコミュニケーションの要求を背景に，音声翻訳技術に関する研究が活発に行われている．現在開発されている対話を対象とした音声翻訳システムの多くは，話者の発話の終了を待ってから訳出を開始する逐次通訳の形態を採用している．しかし，対話に出現する発話には，重文や複文などの長文の出現も少なくないため，発話の途中で訳出を開始する同時通訳が望ましい．同時通訳は，話者の発話を同時通訳に適した単位（同時通訳単位）で分割し，その単位ごとに訳を出力することによって実現される．同時通訳に適した単位とは，同時的にかつ独立に翻訳可能な単位である．しかしながら，対話における同時通訳単位を形成する適切な言語単位については，これまでほとんど明らかにされていない．</p> <p>同時通訳単位を明らかにするために，実データの定量的な観察が必要である．一つの方法として，大量の同時通訳単位に分割する境界（同時通訳単位境界）が付与されたコーパスの利用が考えられる．しかしながら，対話における同時通訳単位境界が付与されたコーパスは未だに存在しておらず，また，人手で大量の同時通訳単位境界を付与することは大きなコストを要するため難しい．</p> <p>そこで本論文では，日英対訳コーパスの日本語対話文に同時通訳単位境界を自動付与する手法を提案する．同時通訳単位として文単位が適用される場合，一文が長くなる傾向があり，同時性が低下する．また，単語や文節は前後の文脈に影響されやすく，独立に訳出することが困難である．従って，本研究では，一文よりも短く，単語や文節よりも独立に訳出可能な言語単位として，述語を中心としたまとまりである「節」に着目した．日本語の節境界と同時通訳単位境界との関係を調べるために，まず事前調査を行った．その結果，節境界はその種類によって同時通訳単位境界となる傾向が異なること，ならびに，原言語の同時通訳単位の出現順序とそれに対応する目的言語の出現順序が異なる場合があることが判明した．そこで，本手法では，まず，同時通訳単位境界となる割合の高い節境界の種類を限定し，対話における同時通訳単位境界の候補を絞る．次に，同時通訳単位の出現順序が交差する場合を排除するために，単語対応の情報を使用して候補を更に絞る．ここで単語対応の情報とは，対訳コーパスに出現する日英両単語の対応を示すキーワードの組とその出現順序を意味する．</p> <p>名古屋大学 CIAIR 同時通訳コーパスに収録された日英対訳コーパスを用いて実験を行った．実験の結果，本手法の有効性を確認した．また，本手法により獲得できなかった同時通訳単位境界，及び，誤って獲得された境界について分析を行った．</p> <p>【発表実績】</p> <ul style="list-style-type: none"> ● 丁，笠，松原，吉川，“日本語対話文における同時通訳単位 - 音声対訳コーパスを用いた分析 -，”言語処理学会第 12 回年次大会発表論文集 (2006.3) (発表予定) ● Z. Ding, K. Ryu, S. Matsubara, and M. Yoshikawa, “ Interpreting Unit Segmentation of Conversational Speech in Simultaneous Interpretation Corpus, ”Oriental-COCOSDA, pp.148-152 (2005.12) ● 丁，笠，松原，吉川，“日本語対話文の節分割に基づく同時的な翻訳単位に関する考察，”平成 17 年度電気関係学会東海支部連合大会講演論文集 (2005.9) 		