

同時通訳者の知識を用いた講演文章のチャンキング*

清水徹

ATR 音声言語コミュニケーション研究所

1. はじめに

自然な話し言葉(特に独話)では、良く知られているように、文の区切りが明確でない、一文が長くなる、文の途中に間(ポーズ)が空くなどの現象が見られる。これら文区切りやポーズ位置は、話者等により大きなばらつきがあることから、不特定話者を対象とする音声アプリケーションでの基本的な処理単位に文やポーズを用いるのは問題がある。

これまで、話し言葉を文より短い単位に分割して処理する試みとして、ポーズと係り受け情報に基づいて文や節にチャンキングする手法[1]、節境界やポーズの前後の形態素情報に基づいて同時的な翻訳単位にチャンキングする手法[2]などが提案され、その有効性が示されているが、文や節という単位の妥当性や、推定精度の目標値の設定法についての議論は必ずしも十分ではないと考えられる。

本稿では、日本語から英語への変換を目的としたタスクにおいて、同時通訳者が原言語からターゲット言語に変換する単位を翻訳単位(TU)と定義し、翻訳単位の特徴ならびに翻訳単位を推定する方法について評価を行ったので報告する。

2. 翻訳単位コーパスの作成とその特徴

日本語を漸次的に英語に通訳することを想定し、日本語文の長さが長く、日本語文を適宜分割して英語に変換することが適当な場合、その日本語文の分割位置を翻訳単位境界と定義した。作業は、3名のプロの同時通訳者が講演の書き起こしテキストについて行い、作業者間のばらつきを考慮し、3名中2名以上の通訳者が共通に境界と認定した箇所を翻訳単位境界とした。

翻訳単位の例) “/” は翻訳単位の境界位置

目的地に到着するとまず昼食を取って/その後午後は日によって各学年ごとに集まって色々楽しく過ごしたり/各パーティーごとに色々な出し物を考えてそれを各パーティーごとに発表し合ったり/といった遊ぶ時間もたくさん取られています/

コーパス作成には、日本語話し言葉コーパス(CSJ)の46講演を用いた。表1にコーパスの規模と特徴を示す。

表1 翻訳単位コーパスの規模

形態素数	82,680
ポーズ(200msec 以上)がある箇所	3,758
翻訳単位境界数	4,449
うち末尾にポーズあり	3,627 (81.5%)
うち末尾にポーズなし	822 (18.5%)
翻訳単位平均形態素数	18.6

本コーパスでは、翻訳単位あたりの平均形態素数は、18.6であり、文献[2]における同時翻訳単位境界あたりの平均形態素数5.4や、文献[1]における節境界間の平均形態素数9.4よりかなり長い単位であること、翻訳単位境界の多くは末尾にポーズを伴っているが、ポーズを伴わない翻訳単位境界も約2割弱あることがわかる。

また、3名の作業者間のばらつきから、ここで認定した翻訳単位に対するプロの同時通訳者のF値を求めることができる。表2にF値を示す。

表2 プロの通訳者の翻訳単位境界推定精度

翻訳単位	再現率	適合率	F値
全体	93.3%	89.3%	0.912
末尾にポーズあり	97.1%	99.4%	0.982
末尾にポーズなし	76.1%	56.9%	0.651

プロの通訳者のF値は、0.912でかなり高い。但し、末尾にポーズを伴わない翻訳単位に限ればF値は0.651にとどまり、作業者によってばらつきが大きいことが分かった。

3. 翻訳単位の推定

形態素情報(表層、品詞、活用形)、ポーズの有

* Segmentation of spoken monologue using human interpreter's knowledge.

無、当該形態素が音声翻訳単位末か否かを素性とし、SVM チャンカである YamCha[3]を用いて音声翻訳単位の学習・評価を行った。

YamCha は以下の設定とした。

参照範囲：

形態素：連続する 9(前 7,後 2)形態素

境界情報：境界の前 3 形態素

多項式のカーネルの次数：2 次

多クラスの識別：pairwise 法

データ量が少ないためデータを各講演毎に 46 分割し交叉検定を行った。表 3 に翻訳単位の推定結果を示す。

表 3 YamCha を用いた翻訳単位推定精度

素性	再現率	適合率	F 値
ポーズの有無あり	84.1%	94.3%	0.889
末尾にポーズあり	98.9%	98.3%	0.986
末尾にポーズなし	18.9%	48.4%	0.271
ポーズの有無なし	83.0%	93.6%	0.880
末尾にポーズあり	96.9%	99.2%	0.980
末尾にポーズなし	21.8%	44.5%	0.293

素性としてポーズの有無は、末尾にポーズがある箇所の若干の F 値の向上に貢献している。表 2 のプロの通訳者の推定精度と比較すると、末尾にポーズがある箇所では、プロの通訳者に匹敵する性能が得られているものの、末尾にポーズがない箇所の F 値は低い。特に再現率が著しく低いことが分かる。

4 翻訳単位の推定の頑健性

3 節の実験は、学習データと評価データが同一コーパスであることから、学習データが異なる場合の性能検証を行った。学習には、名古屋大学同時通訳データベース[4] (独話)の 16 講演を用い、評価には、表 3 と同じデータを用いた。表 4 に学習データの規模と特徴を示す。表 1 のデータと比較して、ポーズの出現頻度が高く、翻訳単位あたりの平均形態素数も少ない。

表 4 比較用翻訳単位コーパスの規模と特徴

形態素数	61, 882
ポーズ(200msec 以上)がある箇所	8, 826
翻訳単位境界数	5, 891
うち末尾にポーズあり	5, 227 (88. 7%)
うち末尾にポーズなし	664 (11. 3%)
翻訳単位平均形態素数	10. 5

表 5 に翻訳単位の推定結果を示す。表 3 と比較すると、素性としてポーズを用いた場合には、末尾にポーズがある箇所では、CSJ を学習データとした場合と同程度の性能が得られ、末尾にポーズがない箇所では、適合率が大きく低下している。素性としてポーズがない場合には、さらに性能が低下する。

表 5 別の学習データを用いた場合の YamCha を用いた翻訳単位推定精度

素性	再現率	適合率	F 値
ポーズの有無あり	85.8%	74.2%	0.796
末尾にポーズあり	98.6%	97.4%	0.980
末尾にポーズなし	29.6%	16.4%	0.211
ポーズの有無なし	87.6%	58.1%	0.699
末尾にポーズあり	95.1%	97.9%	0.965
末尾にポーズなし	54.5%	14.1%	0.223

5 むすび

本稿では、日本語の話し言葉から英語への変換を目的としたタスクにおいて、同時通訳者の翻訳単位をコーパスとして整備し、同コーパスを用いて YamCha による推定実験を行った。この結果、翻訳単位末にポーズがある箇所では、同時通訳者に匹敵する性能が得られたが、翻訳単位末にポーズがない箇所の推定精度は低かった。今後、ポーズがない箇所の推定精度の向上に必要な素性の検討を進める予定である。

本研究は、総務省戦略的情報通信研究開発推進制度(SCOPE)(071707004)の支援により実施したものである。

参考文献

- [1] 西光雅弘, 高梨克也, 河原達也, “係り受けとポーズ・フィラーの情報を用いた話し言葉の段階的チャンキング”, 電子情報通信学会技術研究報告, SP2005-137, NLC2005-104, 2005.
- [2] 笠浩一郎, 松原茂樹, 稲垣康善, “同時的な日英対話翻訳のための日本語発話文の分割”, 電子情報通信学会技術研究報告, NLC2006-56, SP2006-112, 2006.
- [3] T. kudo, Y. Matsumoto, “Chunking with support vector machines”, Proc. of the 2nd meeting North American Chapter of the Association for Computational Linguistics, 2001.
- [4] H. Tohyama, S. Matsubara, N. Kawaguchi, Y. Inagaki, “Construction and utilization of Bilingual Speech Corpus for Simultaneous Machine Interpretation Research”, Proc of 9th European Conf. on Speech Communication and Technology, 2005.