社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－SLP－74（22）
2008 / 12 / 10

音声翻訳単位の推定における句読点情報の効果

清水 徹[†,††,‡]　中村 哲[†,††]　河原 達也[†,‡]

†独立行政法人 情報通信研究機構 知識創成コミュニケーション研究センター
††ATR 音声言語コミュニケーション研究所
‡京都大学情報学研究科

†,††〒619-0288 京都府「けいはんな学研都市」光台 2-2-2
‡〒606-8501 京都府京都市左京区吉田本町

E-mail: {tohru.shimizu, satoshi.nakamura}@nict.go.jp, kawahara@i.kyoto-u.ac.jp

あらまし 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自然な話し言葉を，適切な単位に区分化する処理が求められている．筆者らは，区分化の単位として従来用いられている文や節に代わる，プロの同時通訳者が原言語からターゲット言語に変換する自然なタイミングである音声翻訳単位を提案し，同単位の特徴と，言語情報ならびに韻律情報と SVM チャンカ用いた音声翻訳単位境界の推定手法について検討を行ってきた．一方，書き言葉では，区分化の単位として，句読点が広く用いられている．本稿では，音声翻訳単位境界と句読点との関連性と，プロの通訳者が付与した音声翻訳単位境界情報と句読点情報の双方を用いた場合の音声翻訳単位境界推定への効果について述べる．日本語話し言葉コーパス(CSJ)を用いた実験おいて，句読点既知データの場合 F 値 0.88，句読点未知データの場合 F 値 0.86 と，プロの通訳者による F 値 0.84 に相当する性能を達成した．

キーワード 音声翻訳単位，句読点，チャンキング，SVM

Effect of punctuation marks for speech translation unit boundary detection

Tohru SHIMIZU[†,††,‡] Satoshi NAKAMURA[†,††] and Tatsuya KAWAHARA[†,‡]

† Knowledge Creating Communication Research Center, National Institute of Information and Communication Technology
††ATR Spoken Language Communication Research Labs.
2-2-2 Hikaridai, Keihannna Science City, 619-0288, Japan
‡ School of Informatics, Kyoto University
‡〒606-8501 Sakyo-ku, Kyoto, 606-8501, Japan

E-mail: † ‡ {tohru.shimizu, satoshi.nakamura}@nict.go.jp, kawahara@i.kyoto-u.ac.jp

Abstract As automatic speech recognition and translation of long and complicated utterance cause more errors, there is increasing requirement for utterance segmentation techniques. We proposed speech translation unit (STU), which is a segment of an utterance which the human interpreter treats as a single cognitive unit, and also proposed STU boundary detection method using a SVM based chunker which combines lexical features and prosodic features. It is well known that comma and period are the most widely used punctuation marks in written text. In this paper, characteristics of STU and punctuation marks are investigated, and a STU boundary detection method which combines both STU boundary information and punctuation marks is proposed. An experimental evaluation using CSJ corpus shows STU boundary detection achieved a F-measure of 0.88 for input text with punctuation marks and 0.86 for input text without punctuation marks, which is better than or equal to the STU boundary detection accuracy of human interpreters (F-measure of 0.84).

Keyword Speech translation unit boundary(STU), punctuation marks, chunking, SVM

1. まえがき

良く知られているように，自然な話し言葉(特に独話)では，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観察される．しかも，これらの現象は人により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における基本的な処理単位として，文やポーズを採用しにくいという課題があった．一方，音声認識や翻訳の処理において，長い入力は誤りが生じやすいことが経験的に知られており，また，発声内容をより少ない遅延時間で処理する必要性から，入力を何らかのトリガーにより短い単位に分割する区分化処理が求められている．

これまで，話し言葉を文より短い単位に区分化する試みとしては，主に音声から生成したテキストの自動整形を目的として，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]，同処理の精度向上を目的として，ポーズ・形態素情報に加えて人手で与えた韻律情報（X-JtoBI のトーン層・BI 層のラベル）を素性として用いる手法[2]，句・文境界や句読点を挿入する手法[3]が提案されているほか，話し言葉の翻訳処理を目的として，節境界やポーズの前後の形態素情報に基づいてチャンキングする手法[4]，プロの同時通訳者が原言語からターゲット言語（例えば，日本語から英語）に変換する自然なタイミングである音声翻訳単位(Speech Translation Unit (STU))を，形態素情報，ポーズ情報，入力音声の基本周波数(F_0)情報に基づき自動的に算出した各形態素の F_0 の平均的傾きに基づいてチャンキングする手法[5] などが提案され，その有効性が示されている．しかし,文献[5]で指摘されているように，区分化の精度は入力にポーズがある個所では高く,ポーズのない個所では低い傾向がある．

一方，書き言葉を区分化する単位として句読点が広く用いられている．特に，読点は文中で用いられることから，区分化の処理において読点情報が利用可能な場合においては，読点情報が区分化の精度向上に寄与する可能性がある．また，読点情報が利用可能でない場合においても，読点情報の統計的な出現傾向を用いることにより確度の高い境界を得ることが期待できると考えられる．

そこで本稿では，文献[5]で区分化の単位として用いられている音声翻訳単位境界と句読点との関連性について述べるとともに，句読点が既知の場合と未知の場合のそれぞれについて，句読点情報が音声翻訳単位境界推定精度に与える影響について述べる．

2. 音声翻訳単位コーパス，句読点コーパスとその特徴

2.1. 音声翻訳単位コーパス

日本語の話し言葉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の入力を適宜分割して英語に変換することが適当な場合，その日本語入力に対する分割位置を音声翻訳単位境界と定義する[5]．音声翻訳単位境界付与作業は，3 名のプロの同時通訳者により講演の書き起こしテキストについて行った．作業者には可能な範囲で短く区切るように指示を与えている．作業者間のばらつきを考慮して，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音声翻訳単位境界とした．(本作業は，書き起こしテキストを用いて行っており，作業者は音声を聞いていない．書き起こしテキストは，句読点を付与されていない．)

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．図 1 に音声翻訳単位の一例を示す．

飛行機も好きですが T
車も好きということで T
ちょっといいと思っていました TP
一年間飛んでいて海外には行けるのです TP
あんまり P
もっと長く飛んでいれば T
ベテランになってくれば T
色々行けるところもあるのです TP
いわゆる航空会社ですから T
ただみたいなもので海外に行けるのです TP
ただただでは行けるけれど TP
休みを取って行くまでに至らないです TP
一年ぐらいだと P
まだ飛んでいることが精いっぱいです TP
<中略>
しかし何だそれは P
確かめなければと思っていました TP

“T”は音声翻訳単位の境界位置，”P”は 200 ミリ秒を越えるポーズ位置を示す

図 1 音声翻訳単位コーパスの例

2.2. 句読点コーパス

句読点のない書き起こしテキストに句読点を挿入した．作業は，3 名の作業者が行い，作業者間のばらつきを考慮し，3 名中 2 名以上の作業者が共通に境界

と認定した箇所を句読点とした．コーパス作成には，音声翻訳単位コーパスと同様に日本語話し言葉コーパス(CSJ)を用い，音声翻訳単位コーパスの作成に用いた46講演を含む247講演を用いた．図2に句読点の一例を示し，表1に句読点コーパスの規模と特徴を示す．

飛行機も好きですが M
車も好きということで M
ちょっといいと思っていました MP
一年間飛んでいて海外には行けるのです MP
あんまり M
もっと長く飛んでいれば M
ベテランになってくれば色々行けるところもあるのです MP
いわゆる航空会社ですから M
ただみたいなもので海外に行けるのです MP
ただ M
ただでは行けるけれど MP
休みを取って行くまでに至らないです MP
一年ぐらいだと M
まだ飛んでいることが精いっぱいです MP
<中略>
しかし M
何だ M
それは MP
確かめなければと思っていました MP

“M”は句読点の位置(多数決の時点で句点と読点の区別をしていない)，”P”は200ミリ秒を越えるポーズ位置を示す

図2 句読点コーパスの例

2.3. 音声翻訳単位コーパスと句読点コーパスの特徴

表1に音声翻訳単位コーパス(CSJ-TU)と句読点コーパス(CSJ-PM)の規模と特徴を示す．表1から分かるように，ポーズがある個所では，音声翻訳単位境界と句読点ともにかなりの割合で置かれており，ポーズのない個所では,句読点(多くは読点)が音声翻訳単位境界に比較して多く置かれていることが分かる．その結果境界間の平均長は句読点間の方が短い．

なお,音声翻訳単位あたりの平均形態素数を，先行研究における区分化単位の平均形態素数で比較すると，文献[1]における節境界間の平均形態素数9.4と同程度，文献[3](文献[6]のコーパスを使用)における同時翻訳単位境界あたりの平均形態素数5.4の約2倍であることが分かる．

表1 音声翻訳単位コーパス(CSJ-TU)と句読点コーパス(CSJ-PM)の特徴

	CSJ-TU	CSJ-PM
形態素数	82,680	366,408
ポーズ(200msec 以上)がある箇所	3,804	17,784
ポーズ単位平均長(形態素数)	21.7	20.6
境界数	7,663	41,258
内訳：末尾に		
ポーズあり	3,786 (49.4%)	17,738 (43.0%)
ポーズなし	3,877 (50.6%)	23,520 (57.0%)
境界間平均長(形態素数)	10.8	8.9

音声翻訳単位あたりの平均形態素数10.8，句読点間の平均形態素数8.9と，文献[1]における節境界間の平均形態素数9.4と同程度であったことから，これまで良く区分化の単位として用いられてきた節単位との比較を行う．CSJにおいては，日本語節境界検出プログラムCBAP[7]による節境界を求め，これを人手により修正することにより節単位としていることから，節単位境界のもとになった節境界と音声翻訳単位境界，句読点との比較結果を表2に示す．CBAPが出力する節境界は0～3の4種で，0が最も強い境界である(CBAPでは句点に関するルールがあることから，表2の算出時にのみ句点の挿入を行っている.その他の実験では，句点あるいは文末情報は用いていない)．

表2に示すように，音声翻訳単位境界や句読点の多くは節境界と重なっているものの，境界が弱くなるに従って節境界であるが音声翻訳単位境界や句読点でないものも増えている．また，節境界でないものも一定数存在する．これらの結果から，節境界と音声翻訳単位，句読点は異なる特徴を有することが分かる．

次に，音声翻訳単位境界と句読点がどの程度一致しているかを調べた．音声翻訳単位コーパスと句読点コーパスの双方に共通な46講演について調べた結果を表3に示す．表3より,音声翻訳単位境界の多く(89.5%=6,862/7,663)は句読点に含まれるものの,句読点のうち音声翻訳単位境界と共通なのは72.8%(=6,862/9,429)であり，残りの27.2%は音声翻訳単位境界ではない．特に，音声翻訳単位境界と共通ではない句読点はポーズがない個所に存在している．

2.4. プロの通訳者の音声翻訳単位境界推定精度の評価

表1の音声翻訳単位コーパス作成にあたった3名のとは別のプロの通訳者1名により音声翻訳単位境界の付与作業を行い計4名の作業結果を得た．この4名のうち任意の3名の多数決結果と残り1名の作業結果から，4名の作業者の平均のF値を求めることができる．表4にF値を示す．

表2 音声翻訳単位境界,句読点と節境界との関係

a) 音声翻訳単位コーパス(CSJ-TU)

節境界レベル	音声翻訳単位境界あり	音声翻訳単位境界なし
境界あり	6,709(87.6%)	7,268
内訳：0	3,796(49.5%)	76
1	470(6.1%)	253
2	1,809(23.6%)	1,500
3	634(8.3%)	5,439
境界なし	954(12.4%)	67,749
計	7,663 (100%)	75,017

b) 句読点コーパス(CSJ-PM)

節境界レベル	句読点あり	句読点なし
境界あり	31,866(77.2%)	30,949
内訳：0	7,895(19.1%)	417
1	2,073(5.0%)	1,005
2	6,841(16.6%)	5,279
3	15,057(36.5%)	24,248
境界なし	9,392(22.8%)	294,201
計	41,258 (100%)	325,150

表3 音声翻訳単位境界と句読点との比較

a) 音声翻訳単位コーパス(CSJ-TU)

	音声翻訳単位境界あり	音声翻訳単位境界なし	計
句読点あり	6,862 (3,781 / 3,081)	2,567 (13 / 2,554)	9,429
句読点なし	801 (5 / 796)	72,450 (5 / 72,445)	73,251
計	7,663	75,017	8,2680

括弧内は，(ポーズあり/ポーズなし)の内訳

表4 プロの通訳者の音声翻訳単位境界推定精度
(作業者4名の平均)

	再現率(%)	適合率(%)	F値
全体	86.4	82.2	0.843
ポーズあり	99.5	99.8	0.996
ポーズなし	75.2	68.7	0.718

プロの通訳者のF値は0.843で，単位末にポーズを伴う場合と伴わない場合を区別した場合，ポーズを伴う音声翻訳単位境界のF値はほとんど1に近く，ポーズを伴わない音声翻訳単位境界のF値は0.7を若干上回る値であった．このプロの通訳者によるF値を音声翻訳単位境界の自動推定時の上限値と考え，3節以下の自動推定結果との比較に用いることとした．

3. SVMチャンカを用いた音声翻訳単位境界，句読点の自動推定

3.1. 音声翻訳単位境界と句読点の推定精度

文献[5]と同様に形態素情報(表層，品詞，活用形)，ポーズの有無，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[8]を用いて音声翻訳単位の学習・評価を試みた．

YamChaは以下の設定とした．

- 参照範囲：形態素：連続する7(前3,後3)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

評価データは,音声翻訳単位コーパスと句読点コーパスの双方に共通の46講演とし,音声翻訳単位については,データ量が少ないことから10分割交叉検定とした．表5に音声翻訳単位と句読点の推定結果を示す．

表5 SVMチャンカを用いた境界推定精度

a) 音声翻訳単位

	再現率(%)	適合率(%)	F値
全体	82.2	87.1	0.846
ポーズあり	99.8	99.6	0.997
ポーズなし	64.9	73.3	0.688

b) 句読点

	再現率(%)	適合率(%)	F値
全体	82.6	85.0	0.838
ポーズあり	99.9	99.8	0.999
ポーズなし	71.0	74.6	0.728

表5 a)の結果を表4の結果と比較するとSVMチャンカによる推定結果は，プロの通訳者による性能に匹敵していることが分かる．また，表5の a) b)より,音声翻訳単位境界の推定精度と読点の推定精度に大きな差はなかった．

3.2. 句読点情報を用いた音声翻訳単位境界の自動推定

表3に示したように，音声翻訳単位境界の多くは句読点に含まれることから，入力テキストとして句読点が利用可能な場合には，句読点情報を積極的に利用することにより，音声翻訳単位の推定精度が向上することが期待できる．また，入力テキストとして句読点が利用不可能な場合であっても，音声翻訳単位境界と句読点双方の自動推定結果を統合することで，自動推定された音声翻訳単位境界の中でどの境界がより確度の高い境界であるかを知ることができると期待できる．

a) 句読点が既知の場合

SVMチャンカの学習素性に句読点情報を加える．
自動推定時にも句読点情報を与える．

b) 句読点が未知の場合

音声翻訳単位境界の推定結果と句読点の推定結果のアンドを確度の高い音声翻訳単位とする．

表6に音声翻訳単位境界の推定結果を示す．a)句読点が既知の場合にF値0.880，b)句読点未知の場合F値0.856と,プロの通訳者人間によるF値0.843を超える性能が得られた．また，b）句読点未知の場合の適合率は87.5%と高いことから，表5で得られた境界の中での確度の高い音声翻訳単位境界を知ることができる．

表6 SVMチャンカを用いた音声翻訳単位境界推定精度(句読点情報を加味した場合)

a) 句読点が既知の場合

	再現率(%)	適合率(%)	F値
全体	86.2	89.8	0.880
ポーズあり	99.8	99.8	0.998
ポーズなし	73.0	79.2	0.760

b) 句読点が未知の場合

	再現率(%)	適合率(%)	F値
全体	78.0	94.9	0.856
ポーズあり	99.7	99.7	0.997
ポーズなし	56.7	87.5	0.688

4. むすび

本稿では，話し言葉の区分化を目的とした音声翻訳単位コーパスと句読点コーパスの構築，ならびにSVMチャンカを用いた音声翻訳単位境界，句読点の推定精度について述べた．音声翻訳単位と句読点の推定精度はF値およそ0.84であり，音声翻訳単位の推定精度はプロの通訳者のそれにほぼ匹敵する性能が得られた．また，音声翻訳単位境界情報に加えて句読点情報の双方を用いることにより，句読点が既知の場合でF値0.88，句読点が未知の場合であってもF値0.86を得た．音声翻訳単位コーパスの構築にはプロの通訳者が必要なことから大規模なコーパスを構築するにはコストが大きいが，句読点コーパスは比較的コストが低く済むことから，今後，さらに大規模な句読点コーパスを用いた場合の自動推定精度，人間の作業者との性能比較について検討を進めて行きたい．

本研究の一部は、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。

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] 尾嶋憲治，秋田祐哉，河原達也，“局所的な係り受けと韻律の素性を用いた話し言葉の節・文境界推定”，情報処理学会研究報告，2007-SLP-67, pp.13-18, 2007.

[3] 秋田祐哉，河原達也，“会議録作成のための話し言葉音声認識結果の自動整形”，日本音響学会秋季研究発表会講演論文集，pp.103-104, 2008.

[4] 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.

[5] 清水徹，中村哲，河原達也，“同時通訳者の知識と韻律情報を用いた講演文章のチャンキング”，情報処理学会研究報告，2008-SLP-72, pp.81-86, 2008.

[6] 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.

[7] 丸山岳彦，柏岡秀紀，熊野正，田中英輝“日本語節境界検出プログラム CBAP の開発と評価”，自然言語処理，11，3，pp. 39-68，2004.

[8] T. kudo, Y. Matsumoto, “Chunking with support vector machines”, Proc. of the 2nd meeting North American Chapter of the Association for Computational Linguistics, 2001.