

同時的な日英対話翻訳のための日本語発話文の分割

笠 浩一朗[†] 松原 茂樹^{††} 稲垣 康善^{†††}

[†] 名古屋大学大学院情報科学研究科 〒464-8601 名古屋市千種区不老町

^{††} 名古屋大学情報連携基盤センター 〒464-8601 名古屋市千種区不老町

^{†††} 愛知県立大学情報科学部 〒480-1198 愛知県愛知郡長久手町大字熊張字茨ヶ廻間 1522-3

あらし 計算機による同時翻訳を実現するために、原言語の発話を文よりも短い単位に分割し、その単位ごとに翻訳する方法が考えられる。本論文では、同時的な日英対話翻訳を実現するための適切な翻訳単位について検討する。まず、日英同時通訳コーパスの日本語対話文を、独立かつ即時的に訳出可能な単位（同時翻訳単位）に人手で分割し、その単位境界の特徴を分析する。また、節境界、節境界の種類、発話単位境界、及び、境界前後の形態素情報の特徴に基づいて作成した同時翻訳単位境界を判定するルールについて述べる。さらに、日本語対話文を同時翻訳単位境界に分割する実験について報告する。

キーワード 音声翻訳, 翻訳単位, 同時通訳, 節境界, 発話単位境界

Japanese Sentence Segmentation for Simultaneous Japanese-English Spoken Dialogue Translation

Koichiro RYU[†], Shigeki MATSUBARA^{††}, and Yasuyoshi INAGAKI^{†††}

[†] Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

^{††} Information Technology Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

^{†††} Faculty of Information Science and Technology, Aichi Prefectural University, Nagakute-cho, Aichi-gun, Aichi-ken, 480-1198, Japan

Abstract To realize a simultaneous translation system, the system needs to segment a source utterance into the units which are shorter than the sentence and can be translated incrementally. In this paper, we discuss and propose a suitable translation unit for Japanese-English dialogue machine translation. At first, we segment Japanese utterances in a Japanese-English simultaneous interpretation corpus into the units by human and analyze the boundary of their units. Also, we propose the rules for segmenting translation units based on the analysis of translation unit and report our experiment using their rules.

Key words speech translation, translation unit, simultaneous translation, clause boundary, utterance unit boundary

1. はじめに

近年の音声翻訳技術の進展にともない、対話を対象とした音声翻訳システムの開発が盛んにおこなわれている [1], [2], [4], [6]. これらのシステムの多くは、ターンや文を処理の基本単位として用いているため、対話の効率や円滑さが低下するという問題がある。それに対して、同時通訳者のように話者の発話に追従して訳出することができれば、対話の効率や円滑さが改善されることが期待できる。

計算機で同時翻訳を実現するには、原言語の発話に対して文

よりも短い単語列を独立的に訳出可能な単位（以下、同時翻訳単位）として認識し、それを即座に訳出する必要がある。これまでに、独話の日英翻訳における同時翻訳単位として、「節」に着目した研究が報告されている [3]. しかし、対話における日英同時翻訳単位に関しては行われていない。

そこで本論文では、日英同時通訳コーパスの日本語対話文に、同時翻訳単位境界を人手で付与し、その同時翻訳単位の境界の特徴を分析する。分析では、節境界ならびに発話単位境界と、同時翻訳単位境界との関係について調査した。また、節境界の種類と同時翻訳単位境界との関係についても調査した。分析結

- 1 お店は道路沿いではないんですけど
- 2 林ビルの二階にあります。
- 3 林ビルはすぐ見つけていただけます。
- 4 テレビ塔という大きなタワーのすぐ横です。
- 5 多分日本語だと思いますので今からお書きします。
- 6 そうですね。せっかくお越しいただいているので
- 7 名古屋城を見られたらと思いますね。
- 8 名古屋にある代表的なお城です。

(a) 日本語対話文

- 1 The restaurant is not on the street but
- 2 It's on the second floor of Hayashi building.
- 3 You can find Hayashi building easily.
- 4 It's just next to a tall tower called TV tower.
- 5 Perhaps it is written in Japanese. So I write it for you.
- 6 I see. As you took a trouble to come here.
- 7 You should see Nagoya castle.
- 8 It is a famous castle in Nagoya.

(b) 逐語訳

図1 日本語対話文とその逐語訳

- 1 お店は
道路沿いではないんですけども
- 2 林ビルの二階にあります。
- 3 林ビルはすぐ見つけていただけます。
- 4 テレビ塔という大きなタワーのすぐ横です。
- 5 多分
日本語だと思いますので
今からお書きします。
- 6 そうですね。
せっかくお越しいただいているので
- 7 名古屋城を見られたらと思いますね。
- 8 名古屋にある代表的なお城です。

(a) 同時翻訳単位に分割された日本語対話文

- 1 The restaurant
is not on the street but
- 2 It's on the second floor of Hayashi building.
- 3 You can find Hayashi building easily.
- 4 It's just next to a tall tower called TV tower.
- 5 Perhaps
it is written in Japanese.
So I write it for you.
- 6 I see.
As you took a trouble to come here.
- 7 You should see Nagoya castle.
- 8 It is a famous castle in Nagoya.

(b) 同時翻訳単位に対応する逐語訳

図2 同時翻訳単位に分割された日本語対話文とその逐語訳

果に基づいて同時翻訳単位境界の判定ルールを作成し、日本語対話文を同時翻訳単位に分割する実験を行った。

本論文の構成は以下の通りである。2章では、日本語対話文への同時翻訳単位境界の付与について説明する。3章では、節境界及び発話単位境界との関係についてデータを分析する。4章では、同時翻訳単位境界を判定するルールを示し、5章で日本語対話文の分割実験について報告する。

2. 同時翻訳単位境界の付与

2.1 同時翻訳単位

同時翻訳単位とは、独立かつ即座に訳出可能な原言語の単位であるとする。例えば、日本語対話文

(J2.1) 今のところ予定通りですが出発が遅れる可能性がありますのでご了承くださいませ。

の英語訳を

(E2.1) For now, it is on time, but the departure might be delayed. Please understand it.

とする。このとき、構成要素間の対訳対応とその出現順序を考慮して、(J2.1)を同時翻訳単位の境界（以降、同時翻訳単位境界と呼び、“//”で記す）で分割すると、

(J2.2) 今のところ//予定通りですが//出発が遅れる可能性がありますので//ご了承くださいませ。

となる。

2.2 逐語訳データ

人手による同時翻訳単位の付与には名古屋大学同時通訳データベース [7] の対話データを利用した。このデータベースには、英日、日英の二人の同時通訳者を介した英語話者と日本語話者の対話の音声データとその書き起こしデータが収録されてい

る。同時通訳者による発話は、厳しい時間的な制約のために、意識や省略を多く含んでいる。このような発話は、独立性の観点からみて同時通訳単位境界を付与するには適さない。そのため、同時翻訳単位境界の付与には、逐語訳を利用した。逐語訳は、プロの翻訳者が以下の方針に従って作成した。

訳質: 原文の内容を聞き手が理解可能な訳文である

漸進性: 細かい単位で漸進的に訳出するために、原文の語順に近い訳文である

文脈依存度: 文脈への依存を避けるために、意識や省略を含まない訳文である

付与単位: 原則として発話単位ごとに対訳可能な訳文である

3. 同時翻訳単位境界の分析

3.1 基礎データ

1行が1発話単位に相当する日本語対話文とその逐語訳を図1に示す。日本語対話文と逐語訳の左側に付与された数字が日英間の対応関係を示している。図2(a)は、図1の日本語対話文を人手により同時翻訳単位に分割したものであり、図2(b)は、図2(a)の同時翻訳単位に対応するように逐語訳を分割したものである。

同時翻訳単位が付与された日本語対話文の基礎的な統計データを表1に示す。ここで、発話単位境界とは、原則として200ms以上のポーズごとに発話を分割した境界である。また、節境界とは、節境界解析プログラムCBAP [5]により分割した境界である。また、同時翻訳単位境界の数では、文末との一致を除い

ている。

3.2 言語的な境界との関係

形態素境界以外で同時翻訳単位境界になることはなかった。同時翻訳単位境界と節境界、あるいは発話単位境界が一致する割合を調査した結果を図3に示す。形態素境界が節境界であるかないか、あるいは発話単位境界であるかないかによって、同時翻訳単位境界のなりやすさに違いがみられた。そこでこれ以降では、同時翻訳単位境界を、A：節境界でかつ発話単位境界である場合、B：節境界でかつ発話単位境界でない場合、C：節境界でなく発話単位境界である場合、D：節境界でかつ発話単位境界でない場合に分けて分析する。

3.2.1 A：節境界でかつ発話単位境界である場合

節境界でかつ発話単位境界（図3のA）である場合の89.8%（53/59）が同時翻訳単位境界と一致した。次に、節境界の種類ごとに、それが同時翻訳単位境界になる割合を図4に示す。図4より、節境界の種類が「補足節」の場合、同時翻訳単位境界になることはない。節境界の種類が「主題ハ」、「並列節」の場合、同時翻訳単位境界になる場合とならない場合がある。

「並列節」

同時翻訳単位境界になる場合：

- 英語においても並列節の構造である

日本語： お店は道路沿いではないんですけども
/並列節ケレドモ/林ビルの二階にあります。

英語訳： The restaurant is not on the street
but it's on the second floor of Hayashi building.

同時翻訳単位境界にならない場合：

- 「けれども」が「前置き」の意味で生起する

日本語： 十二月二十日なんですけれども/並列節ケレドモ/
東京発ロサンゼルスラスベガス行きはすべて満席に
なっております。

英語訳： The flights from Tokyo to Los Angeles or to
Las Vegas on December twentieth are fully booked.

「主題ハ」

同時翻訳単位境界になる場合：

- 日本語の文頭で主語として生起して、英語でも文頭の
主語や前置詞句となる

日本語： チェックアウトは/主題ハ/
朝の十時となっております。

英語訳： The check-out time is ten a.m.

同時翻訳単位境界にならない場合：

- 疑問文である、または、述部が「存在」を表す動詞

日本語： お部屋のほうは/主題ハ/どうされますか。

英語訳： What type of room would you like?

日本語： コアラは/主題ハ/いません。

英語訳： You can not see koala bears.

3.2.2 B：節境界でかつ発話単位境界でない場合

節境界でかつ発話単位境界でない境界（図3のB）の50.0%（65/130）が同時翻訳単位境界と一致した。節境界の種類ごとに、それが同時翻訳単位境界になる割合を図5に示す。図5よ

表1 分析データの基礎統計データ

対話数	7 対話
文数	329 文
同時翻訳単位境界数 (文境界を除く)	164 個
発話単位境界数	167 個
節境界数 (文末除く)	189 個
形態素数	2677 形態素

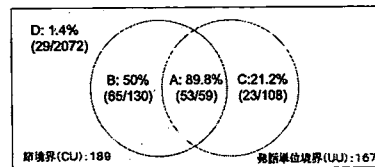


図3 同時翻訳単位境界と節境界及び発話単位境界との関係

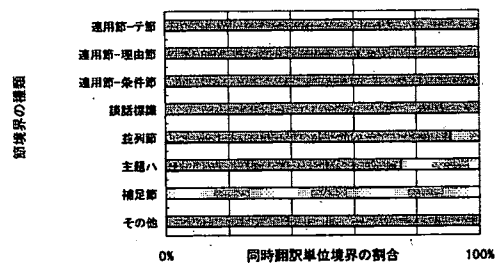


図4 節境界であり発話単位境界である場合 (A) における同時翻訳単位境界と節境界の種類との関係

り、節境界の種類が「談話標識」、「並列節」、「理由節ノデ」の場合のすべてが同時翻訳単位境界と一致した。また、「条件節タラ」、「主題ハ」、「テ節」は、同時翻訳単位境界になる場合とならない場合があった。「主題ハ」は、Aの場合の「主題ハ」と同様の特徴をもっていた。

「条件節タラ」

同時翻訳単位境界になる場合：

- 英語においても条件節の構造である

日本語： 何かございましたら/条件節タラ/こちらのほうに
ご連絡させていただきます

英語訳： If we have something tell you,
we will call the hotel.

同時翻訳単位境界にならない場合：

- 節の語順が日本語と英語で異なる場合

日本語： 名古屋の名物でしたら/条件節タラ/
別のお店が有まして

英語訳： I recommend you another restaurant,
if you want to eat a local specialty of Nagoya.

「テ節」

同時翻訳単位境界にならない場合：

- テ形の動詞「歩く」に接続助詞「て」が接続する場合
あるいは、テ形複合動詞になっている場合

日本語： 歩いて/テ節/十分ですわ。

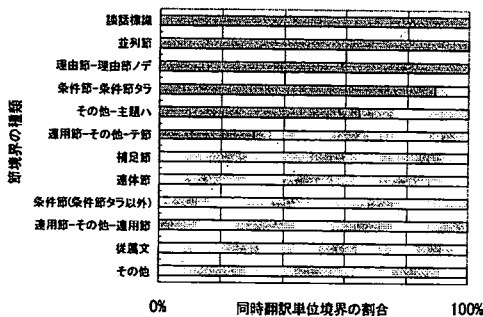


図5 節境界であり発話単位境界でない場合 (B) における同時翻訳単位境界と節境界の種類との関係

英語訳: It is ten minutes' walk.

日本語: 林ビルはすぐに見つけて/テ節/いただけると
思います。

英語訳: You can find Hayashi building easily.

3.2.3 C: 節境界でなくかつ発話単位境界である場合

節境界でなくかつ発話単位境界である境界 (図3のC) の21.2% (23/108) が同時翻訳単位境界と一致した。その境界の直前の単語ごとの同時翻訳単位境界になる頻度を表2に示す。表2の括弧内の数字は、同時翻訳単位境界にならない場合も含めた出現頻度を示している。表2より、境界の直前の形態素が「も」、「を」のときは同時翻訳単位境界になり、「で」、「が」、「に」のときに同時翻訳単位境界になりやすいことが分かる。

直前の単語: 「で」

同時翻訳単位境界になる場合:

- 境界が英語の文境界になるように訳出

日本語: 名駅から徒歩で/発話単位境界/
バス降り場から大きな郵便局が見えますので

英語訳: You can come here on foot from Nagoya station.

From the bus stop you will see a big post office.

同時翻訳単位境界にならない場合:

- 境界の直後に述部が続く場合

日本語: 二つまでですと無料で/発話単位境界/お預かり
できます。

英語訳: We can keep up to two bags for free.

直前の単語: 「が」、「に」

「で」における同時翻訳単位境界にならない場合と同様に、境界の直後に述部が続く場合には、同時翻訳単位境界になりにくい傾向がみられた。

直前の単語: その他 (「で」、「が」、「も」、「を」、「に」以外)

同時翻訳単位境界になる場合:

- 発話の途中で相手からの発話が入り込まれたために、
発話を途中で打ち切り、「はい」などの相槌を行った場合

日本語: 駅から/発話単位境界/はい。

英語訳: From the station. Yes.

表2 境界の前の単語と同時翻訳単位境界になる頻度との関係

項目	C: 節境界でなく 発話単位境界である	D: 節境界でなく 発話単位境界でない
で	5(9)	3(31)
が	3(6)	15(45)
も	2(2)	1(7)
を	2(2)	0(21)
に	2(5)	4(51)
その他	6(74)	5(1898)
合計	23(108)	29(2072)

3.2.4 D: 節境界でなくかつ発話単位境界でない場合

最後に、節境界でなくかつ発話単位境界でない形態素境界 (図3のD) の1.4% (29/2072) が同時翻訳単位境界と一致した。その境界の直前の単語ごとの同時翻訳単位境界になる頻度を表2に示す。表2より、直前の単語が「で」、「が」、「に」のときに、2回以上同時翻訳単位境界と一致した。

直前の単語: 「で」

Cと同様に境界の前の部分を英語で文として訳出できる場合に、同時翻訳単位境界になりやすかった。また、境界の直後に述部が続く場合には、同時翻訳単位境界になりやすかった。

直前の単語: 「が」

境界の直後に述部が続く場合には、同時翻訳単位境界になりにくい傾向がみられた。

直前の単語が「に」(副詞)

同時翻訳単位境界になる場合:

- 文頭で「最後に」、「次に」という副詞の場合

日本語: 最後に//クレジット番号と連絡先をお願いします。

英語訳: Lastly, could I have your credit number and
contact address, please?

境界の直前の単語: その他 (「で」、「が」、「に」以外)

同時翻訳単位境界になる場合:

- 文頭の副詞

日本語: 多分//日本語だと思いますので今からお書きます。

英語訳: Perhaps, it is written in Japanese.

So I write it for you.

4. 同時翻訳単位境界の判定ルール

4.1 ルール作成の方針

前節の分析結果より、形態素境界以外で同時翻訳単位境界になることはないので、形態素境界を同時翻訳単位境界の候補とする。実際の翻訳処理では、同時翻訳単位境界を漸進的に検出する必要があるため、漸進的に利用可能と考えられる統語情報のみを用いる。具体的には、節境界、発話単位境界、境界の前後の形態素情報を利用する。形態素境界を、節境界あるいは発話単位境界であるかないかによって、前節で示したA~Dの4種類に分類し、節境界である、すなわち、AとBの場合には、節境界の種類と前後の形態素情報で判定する。節境界でな

い、すなわち、CとDの場合は、前後の形態素情報のみで判定する。

4.2 同時翻訳単位境界の判定ルール

同時翻訳単位境界の判定ルールを図6に示す。また、規則のより詳しい条件をA~Dごとに以下に示す。

A：節境界でかつ発話単位境界である場合

- 節境界の種類：「主題ハ」、「補足節」以外
- 節境界の種類：「主題ハ」で、以下の条件を満たす
 - 境界の直後の単語が疑問詞ではない（疑問詞：「何」、「どこ」、「いつ」など）
 - 境界の直後の単語が「存在」を表す動詞ではない（「存在」の動詞：「いる」、「ある」、「おる」など）

B：節境界でかつ発話単位境界でない場合

- 節境界の種類：「談話標識」、「並列節」、「理由節ノデ」、「条件節タラ」
- 節境界の種類：「主題ハ」で、以下の条件を満たす
 - 境界の直後の単語が疑問詞でない（疑問詞：「何」、「どこ」、「いつ」など）
 - 境界の直後の単語が「存在」を表す動詞ではない（「存在」の動詞：「いる」、「ある」、「おる」など）
- 節境界の種類：「テ節」で、以下の条件を満たす
 - テ形の動詞でない：（テ形の動詞：「歩いて」）
 - テ形複合動詞でない（テ形複合動詞：「～ている」「～ていただく」など）

C：節境界でなくかつ発話単位境界である場合

- 直前の単語：「を」、「も」
- 直前の単語：「が」、「に」、「で」で、以下の条件を満たす
 - 直後の単語：述部でない
- 直後の単語：相槌（例：「はい」など）である

D：節境界でも発話単位境界でもない場合

- 直前の単語：「が」、「で」で、以下の条件を満たす
 - 直後の単語：述部でない
- 文頭の副詞（「最後に」、「次に」、「多分」など）である

5. 同時翻訳単位境界の分割実験

5.1 実験の概要

前節で示した同時翻訳単位境界の判定ルールを用いて日本語対話文を同時翻訳単位境界に分割する実験を行った。実験では、名古屋大学同時通訳データベース[7]のうち、3節の分析に用いなかった14対話分の対話データを用いた。実験データには、320個の同時翻訳単位境界が存在した。

5.2 実験の結果

前節で示したA、B、C、Dの場合のルールを組み合わせたときの実験結果を表3に示す。「A+B+C」の場合にF値が最大となった。また、A、B、C、Dの各々の実験結果を表4に示す。ただし、表4の再現率は、各場合の同時翻訳単位境界になるものうち、本ルールで同時翻訳単位境界と判定できた割合を示している。

5.3 考察

A：節境界でかつ発話単位境界である場合

同時翻訳単位境界になる割合が非常に高いため、精度、再現率ともに高い値を示した。また、同時翻訳単位境界であると誤って判定した10境界の節境界の種類は、主題ハ（3境界）、談話標識（2境界）、条件節バ（1境界）、テ節（1境界）、連体節（1境界）、引用節（1境界）、間投句（1境界）であった。それらのうち「主題ハ」は、疑問文であり文末まで見ないと同時翻訳単位境界でないとして判定できないものであった。「談話標識」は、「また」が文中で生じたものであった。それ以外では、「連体節」、「引用節」、「間投句」などは同時翻訳単位境界になりにくい節境界であるためであると推測される。「テ節」と「条件節バ」は、その境界の直後が述部であり、そのような場合には同時翻訳単位境界になりにくい。同時翻訳単位境界でないにも関わらず同時翻訳単位境界と判定された節境界の種類は、いずれも「主題ハ」であり。これは境界の直後に疑問詞が生じたため同時翻訳単位境界でないとして判定されたものの、同時翻訳単位境界になっている場合であった。

B：節境界でかつ発話単位境界でない場合

同時翻訳単位境界であると誤って判定した境界は60境界あった。その節境界の種類ごとの頻度は、「主題ハ」が36境界、「談話標識」が6境界、「テ節」が6境界、「並列節」が7境界、「条件節タラ」が4境界、「理由節ノデ」が1境界であった。「主題ハ」の誤りの主な原因としては、疑問文を認識できなかった、ルールに含めてなかった疑問詞が生じた、などが多数を占めた。「談話標識」の誤りは、Aと同様に「また」が文中に生じたり、談話標識「まず」が直後の副詞に係る場合であった。同時翻訳単位境界と判定できなかった24境界の節境界の種類は、「条件節ト」が5境界、「テカラ節」が3境界、「感動詞」が2境界、「間投句」が2境界、連用節が2境界、「従属文」が2境界、「理由節カラ」が2境界、それ以外が6境界であった。「条件節ト」、「理由節カラ」などは、分析データでは同時翻訳単位境界になるものがなかったが、実験データでは多くの場合、同時翻訳単位境界になっていた。これは、分析データでの出現頻度が1~3回という少ない頻度でしか出現していないため、十分にその特徴を分析できなかったことが原因である。また、「感動詞」や「間投詞」は分析データには出現していなかったため、同時翻訳単位境界として判定できなかった。

C：節境界でなくかつ発話単位境界である場合

同時翻訳単位境界であると誤って判定した境界が19境界あった。その境界の直前の単語は、「を」が8境界、「に」が6境界、それ以外が6境界であった。また、同時翻訳単位境界と判定できなかった境界が10境界であった。この主な原因は、文頭の「あと」、「今から」などの表現を翻訳単位として判定できなかったことによるものであった。

D：節境界でなくかつ発話単位境界でない場合

同時翻訳単位境界であると誤って判定した境界が49境界あった。その境界の直前の単語は、「で」が29境界、「が」が15境界、それ以外が5境界であった。直前の単語が「で」の場合、「～でよろしいですか」、「～でお願いします」などの場合に、う

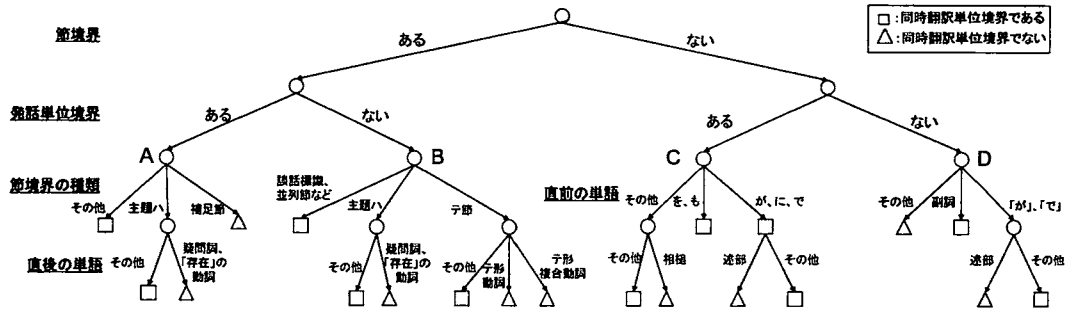


図6 同時翻訳単位境界の判定ルール

表3 実験結果

ルール	精度	再現率	F 値
A	87.2% (68/ 78)	21.3% (68/320)	34.2
A+B	74.5% (205/275)	64.1% (205/320)	68.9
A+B+C	71.3% (221/310)	69.1% (221/320)	70.2
A+B+C+D	61.2% (256/418)	80.0% (256/320)	69.4

表4 各 A~D の場合ごとの実験結果

場合	精度	再現率	F 値
A	87.2% (68/ 78)	97.1% (68/ 70)	91.9
B	69.5% (137/197)	85.1% (137/161)	76.5
C	45.7% (16/ 35)	61.5% (16/ 26)	52.5
D	41.0% (34/ 83)	54.0% (34/ 63)	46.6

まく述部の判定ができないため、誤って同時翻訳単位境界として判定された。また、同時翻訳単位境界と判定できなかった境界が 29 境界であった。この主な原因は、文頭の副詞「明日」、「今から」などの表現を翻訳単位として判定できなかったものであった。また、別の原因として文頭の談話標識を CBAP が認識できなかったものが 6 境界あった。

6. おわりに

本論文では、同時的な日英対話翻訳のための翻訳単位について検討するために、日英同時通訳コーパスの日本語対話文を独立かつ即時的に訳出可能な単位に人手で分割し、その境界の特徴を分析した。また、節境界、節境界の種類、発話単位境界、及び、境界の前後の形態素情報の特徴に基づいて作成した同時翻訳単位境界を判定するルールについて述べた。さらに、それらのルールを用いて日本語対話文を同時翻訳単位境界に分割する実験を行った。その結果、「A+B+C」の場合のルールを用いたときに F 値が 70.2 で最大となった。今後は、実験により得られた知見を利用して、統計的な方法により同時翻訳単位境界を判定する方法について検討する予定である。

謝辞 日頃御指導下さる名古屋大学大学院教授の坂部俊樹先生に深く感謝致します。また、同時翻訳単位境界の付与に多大な貢献をして頂いた名古屋大学の工学部の Zhe Ding さんに感謝致します。

文 献

[1] R. Frederking, A. Blackk, R. Brow, J. Moody, and E. Stein-brecher, Field Testing the Tongues Speech-to-

Speech Machin Translation System, *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)* pp. 160-164, 2002.

[2] R. Isotani, K. Yamada, S. Ando, K. Hanazawa, S. Ishikawa and K. Iso.

Speech-to-Speech Translation Software PDAs for Travel Conversation, *NEC Research and Development*, 44, No.2 pp. 197-202, 2003.

[3] H. Kashioka, T. Maruyama, H. Tanaka, "Building a Parallel Corpus for Monologues with Clause Alignment," *MT Summit IX*, pp. 216-223, 2004.

[4] F. Liu, Y. Gao, L. Gu and M. Picheny. Noise Robustness in Speech to Speech Translation, *IBM Tech Report RC22874*, 2003.

[5] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英脚, "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.

[6] T. Takezawa, T. Morimoto, Y. Sagisaka, N. Campbell, H. Iida, F. Sugaya, A. Yokoo and S. Yamamoto. A Japanese-to-English Speech Translation System:ATR-MATRIX, *Proceedings of 5th International Conference on Spoken Language Processing*, pp. 957-960, 1998.

[7] 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*, pp. 1585-1588, 2005.