

漸進的日本語生成に基づく英日同時通訳手法

笠 浩一朗†

水野 敦‡

松原 茂樹§

稲垣 康善‡

†名古屋大学大学院工学研究科 ‡名古屋大学大学院情報科学研究科

§名古屋大学情報連携基盤センター ‡愛知県立大学情報科学部

k_ryu@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

自然で円滑な多言語間コミュニケーション支援環境の実現のため、同時通訳機の開発が望まれている。実際、同時通訳技術に関する研究がいくつか試みられており、今後の進展が待たれている [1, 5, 4]。そのような背景のもと、著者は、構文トランスファー方式に基づく同時通訳手法を提案しており、これまでに英日同時通訳の実験システム *LINAS* を開発している [8]。*LINAS* は、入力された英語文を漸進的に依存解析し [2]、それをもとに、日本語対訳文の語順決定、及び、語彙変換を実行する。

LINAS では、日本語係り受け制約に基づく訳文生成をベースとしており、入力英語文の語の生起順序にできる限り準拠した、一定の品質を備えた訳文を生成することができるが、その一方で、その制約が強力に作用するために翻訳処理の同時性が損なわれることがある。すなわち、*LINAS* におけるこれまでの日本語生成手法では、生成する訳文の文法として係り受け文法を採用したために、生起順序に関する英語文への準拠の程度が小さくなり、結果的に入力に追従した訳出が難しいといった問題があった。

そこで本稿では、日本語係り受け制約の一つである後方修飾性を逸脱することを認め、英日同時通訳における倒置を用いた日本語生成手法を提案する。本研究では、倒置を右から左の方向性をもつ文節間係り受け関係とし、そのような現象を訳文の生成に利用する。倒置現象を含んだ訳文を認めることにより、より英語文の語順に近い日本語文の生成が可能となる。また、実際の倒置現象の発生傾向について調査した結果を考慮することにより、容認可能な程度の品質を備えた訳文を生成できる。本手法を英日同時通訳の実験システム *LINAS* に導入し、ATR 音声言語データベースを用いた翻訳実験を実施した。実験の結果、入力文に追従可能な訳文出力の実現に対する本手法の有効性を確認した。

2 漸進的な日本語生成手法

2.1 倒置を用いた日本語生成

倒置や言い直しを活用した日本語文生成について説明するために、例として以下の英語文 (2.1) について考える。

(2.1) I'll go to the park with your brother.

この標準的な日本語訳は、例えば (2.2) である。

(2.2) あなたの弟と公園に行きます。

文 (2.1) を漸進的に訳すとき、原理的に “with your brother” が入力されるまで訳文を生成することはできない。しかし、日本語訳として (2.3) を生成するのであれば、“to the park” の入力終了した時点で “公園に行きます” を生成できる可能性がある。

(2.3) 公園に行きます、あなたの弟と。

この日本語訳には、右から左への係り受け、すなわち、倒置現象が含まれており、必ずしも理想的な訳文であるとはいえない。しかし、実際の話し言葉には倒置現象がしばしば現れることを考慮すると、適切に出現するのであれば、倒置を含む訳文であっても容認可能であると考えられる。そのため本研究では、訳文生成に倒置を積極的に活用し、より遅れの少ない同時通訳処理の実現を目指す。

2.2 話し言葉における倒置現象

倒置現象は、任意の文節間に生じるわけではなく、その出現にはある傾向が存在する。そこで、話し言葉データを用いてその出現傾向について調査した。調査では、名古屋大学 CIAIR で構築されている音声対話コーパス [3] に収録されている係り受け構造データ [7] を使用し、右から左に係る文節間関係を対象とした 7781 発話に存在する全ての倒置現象 (計 230 箇所) を観察した。その結果、倒置となる係り受け関係について、以下の特徴が明らかになった [6]。

1. 受け文節のほとんどが述部である。
2. 受け文節に係る文節数は高々 2 つである。
3. 係り文節の最終形態素の品詞の多くは、助詞、名詞、形容詞のいずれかであり、助詞の種類は、“は”、“で”、“に”などに限定される。
4. 一発話における倒置の数は高々一つである。

2.3 言い直しを用いた日本語生成

本生成手法では、上述の性質を満たしたときのみ倒置を使用可能であるとする。ただし、そのような性質を満たすかどうかは、一般には、入力の最後まで決定できな

表 1: 通訳実験結果

手法	遅延	翻訳文評価		
		正解	理解可能	不正解
従来手法	2.65	18	8	46
本手法	2.03	17	7	48

い。そこで、倒置を生成した後になって上述の性質に反することが明らかになったときには、倒置を構成する受け文節を再び生成することにより、倒置を含まない訳文を生成する。例えば、入力文が (2.1) ではなく実は (2.4) であったとする。

(2.4) I'll go to the park with your brother by his car next Sunday.

この場合、“with your brother” までの入力に対して (2.3) が生成されていることになるので、引き続き “by his car next Sunday” を訳出した後、倒置の受け文節「行きます」を再度出力する。すなわち、(2.5) を訳出する。

(2.5) 公園に行きます、あなたの弟と、彼の車で次の日曜日に行きます。

この場合、「あなたの弟と」は文末の「行きます」に係るとみなせるため、倒置を含まない訳文となるとともに、結果的に「行きます」の内容が言い直されたことになっている。以上により、処理の同時性と訳文の品質を兼ね備えた同時翻訳が可能となる。

3 実験

同時通訳システム LINAS[8] の構成図を図 1 に示す。本手法を LINAS に導入し、英日同時通訳実験を行った。実験対象として ATIS コーパスの 578 文を用いた。

倒置が使用された対訳文に対して、従来のシステムで日本語生成した場合と、本手法を用いて日本語生成した場合について、システムの同時性および翻訳品質を評価した。同時性の評価尺度として、入力単語に対する訳語出力の平均遅延時間を計測した。以下の式で、平均遅延時間を求めた。

$$D = \sum_k \frac{d_k}{n} \quad (1)$$

ここで d_k は、 k 番目に入力された単語の対訳語が出力されるまでの経過時間である（ただし、言い直された場合は、言い直されるまでの経過時間とする）。経過時間は、1 単語が入力されるたびに 1 単位時間経過とした。また、 n は英語話者発話の単語数を表す。翻訳の精度については、人手により 3 段階で評価した。実験結果を表 1 に示す。実験結果より、578 文中の 72 文に対して倒置が使用され、うち 12 文で言い直しが使用された。また、本手法によりシステムの遅延が約 23% 減少することが確認された。一方、本手法を用いることにより訳文が理解しづらくなった文は、2.8%にとどまった。

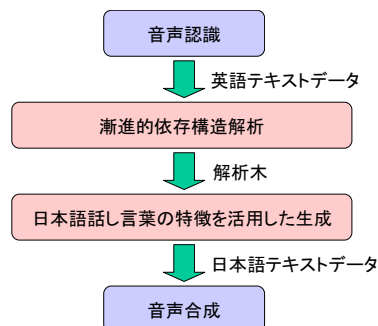


図 1: 同時通訳システム LINAS

4 おわりに

本稿では、英日同時通訳における英語入力に対する日本語出力の同時性の向上を目的として日本語生成手法を提案した。提案手法を実験システム LINAS に導入し翻訳実験を行った結果、本手法により、システムの翻訳品質を保ちつつ、同時性が向上することを確認した。

参考文献

- [1] J. W. Amtrup, “Incremental Speech Translation”: Lecture Notes in Artificial Intelligence, 1735 (1999).
- [2] 加藤 芳秀, 松原 茂樹, 外山 勝彦, 稲垣 康善: 主辞情報付き文脈自由文法に基づく漸進的な依存構造解析, 電子情報通信学会論文誌, Vol.86-D-II, No. 1, pp. 86-97 (2003).
- [3] 河口信夫, 松原茂樹, 山口由紀子, 武田一哉, 稲垣康善: CIAIR 実走行車内音声データベース, 情報処理学会研究報告, pp139-144 (2003).
- [4] S. Matsubara and Y. Inagaki: “Incremental Transfer in English-Japanese Machine Translation”, IE-ICE Transactions on Information and System, Vol. E80-D, No.11, pp. 1222-1129 (1997).
- [5] H. Mima, H. Iida, O. Furuse: Simultaneous Interpretation Utilizing Example-based Incremental Transfer, *Proceedings of 15th International Conference on Computational Linguistics*, pp855-861 (1998).
- [6] 水野敦, 松原茂樹, 稲垣康善: 話し言葉コーパスを用いた倒置現象の分析, 平成 14 年度東海支部連合大会論文集, p.251 (2002).
- [7] T. Ohno, S. Matsubara, N. Kawaguchi and Y. Inagaki: Spiral Construction of Syntactically Annotated Spoken Language Corpus, *Proceedings of 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, pp. 477-483 (2003).
- [8] 渡邊善之, 松原茂樹, 外山勝彦, 稲垣康善: 英日同時翻訳のための漸進的日本語生成, 言語処理学会第 6 回年次大会発表論文集, pp.272-275 (2000).