英日同時翻訳のための漸進的日本語生成

渡邊善之† 松原茂樹†§ 外山勝彦†§ 稲垣康善†

†名古屋大学大学院工学研究科計算理工学専攻 ‡名古屋大学言語文化部

§名古屋大学統合音響情報研究拠点 (CIAIR)

{ynabe,matu,toyama,inagaki}@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

機械翻訳を介した自然な多言語間対話を行うためには, 実時間音声翻訳システムの実現が不可欠である. これまでに多くの機械翻訳手法が提案されてきたが, それらは書き言葉を対象とした文単位での翻訳処理に基づくものがほとんどである. そのようなシステムでは, 1文全体が入力された後でないと翻訳結果の出力を開始できないため, 対話の結束性が大きく損なわれることになる. 同時通訳者のように, 相手の発話途中で翻訳結果の出力が可能になれば, ユーザもまた, 早い段階で相手の発話内容を理解し, 応答することができる. その結果, 会話の待ち時間が減少するため, スムーズな多言語間会話の実現が期待できる.

同時通訳システムの実現にあたっては, 原言語と目標言語との間の同時進行性がポイントとなる [1, 3, 6, 7]. 英語と日本語では語の生起順序が異なるため, 英語入力に対してできる限り同時進行的に日本語を生成することは一般には困難であるものの, 話し言葉に特有な言語現象を含んだ日本語文の生成を容認することにより, 英日同時通訳が実現可能となることが確認されている [6]. しかし, そのための具体的な日本語生成手法は明らかではなかった.

そこで本稿では, 英日同時翻訳のための漸進的日本語生成手法を提案する. 本研究では, 話し言葉に特有の言語現象のうち, 語順の入れ換えに着目する. また, 本手法では, 日本語を生成するための文法として依存文法を用いる. 依存文法には, 非交差性, 後方修飾性, 係り先の唯一性という日本語の語順に関する制約があり, この制約を満たすことにより, 正しい意味内容をもつ日本語を作り上げることができる. また, 依存文法では, 自然な語の並びに関する制約が比較的緩いため, 日本語文が英語の語順に従う度合を高めることができる. すなわち, これらの制約や性質を用いて, 音声翻訳システムの翻訳結果として容認可能な日本語文を漸進的に生成する.

本稿で提案する生成手法をもとに英日同時翻訳の実験システムを作成した. システムは, チャートをベースとする漸進的な解析, 変換, 生成から構成され, それらが入力に対して同時進行的に振る舞う. ATR 音声言語データベースの1255 文をテストデータとして翻訳実験を行った結果、本手法の利用可能性を確認した.

本稿の構成は以下の通りである. 次の2節では, 話し言葉に特有の表現の活用について説明する. 3節では, 依存文法を用いた漸進的日本語生成手法について述べ, 4節では, 英日同時翻訳システムについて説明する. 5節では, 翻訳実験の結果を報告する.

2 話し言葉に特有の表現の活用

英語と日本語の場合, 語の生起順序が大きく異なるため, 漸進的に日本語を生成することは一般には困難である. 例えば, 英語文

入力	出力
I	
prepare	私は
the	
room	
with	
a	
bath	
for	
you	あなたに浴室のある部屋を用意します

図 1: 日本語翻訳文 (2.2) の出力タイミング

(2.1) I prepare the room with a bath for you.

に対する標準的な日本語翻訳文は

(2.2) 私は, あなたに浴室のある部屋を用意します.

である[1]. (2.1) の入力に対して, できる限り早い段階で (2.2) を生成した場合のタイミングを図1に示す. 図1における上から下への順序は, 入力及び出力の時間的順序を表す. 図1が示すように, 日本語の "あなたに" という訳が日本語文中の比較的早い段階で出現するのに対して, 英語では, "for you" が英語文の一番最後に出現するため, 英語文全体が入力されるまで, "私は" 以降の翻訳結果を生成することはできない.

しかし, 日本語の語順に関する自由度は比較的高いため, 同じ依存先をもつ構成要素の間であれば, 互いに語順を入れ換えても日本語として容認できる場合がある. ただし, 全ての構成要素の語順が入れ換えられるわけではなく, 少なくとも日本語係り受け制約を満たす必要がある. 例えば, 日本語翻訳文 (2.2) の "私は", "部屋を", "あなたに" については, いずれも "用意します" に依存しており, それらの間で語順を入れ換えても意味内容は通じる. しかし, "浴室のある" はその依存先が "部屋を" であるために, 後方修飾性により, それを "部屋を" の前に置かなければならない. また, 非交差性から "部屋を" に依存する文節以外の文節の間で語順を入れ換えることはできない. これらの条件を考慮し, できる限り英語の語順に従って例文 (2.1) を翻訳すると, 次の (2.3) のような翻訳結果を生成できる.

(2.3) 私は, 浴室のある部屋をあなたに用意します.

[1]対話では, "私は" や "あなたに" といった 人称及び 人称が省略されることが多いが, 本稿では省略については考慮しない

入力	出力
I	
prepare	私は
the	
room	部屋を
with	
a	
bath	あのー，浴室のある部屋を
for	
you	あなたに用意します

図 2: 日本語翻訳文 (2.4) の出力タイミング

さらに，同時翻訳システムでは 1 文全体が入力される前に翻訳処理を実行するため，入力途中の段階で誤った翻訳結果を生成する可能性がある．例えば，“I prepare the room” までが入力された段階で，“部屋を” が生成された場合，その後 “with a bath” が入力され，それが “部屋を” に依存することが判明すれば，語順の制約を満たした日本語文に翻訳することはできない．それに対しては，言い淀み “あのー” を生成した後で，言い直すことにより，翻訳処理を続行することができる．例えば，例文 (2.1) に対して，日本語翻訳文

(2.4) 私は，部屋を，あのー，浴室のある部屋をあなたに用意します．

を生成できる．図 2に日本語翻訳文 (2.4) の出力タイミングを示す．語順が入れ換わっており，かつ，言い淀み “あのー” や言い直しを含んでいるものの，日本語話し言葉としてその意味は正しく通じる．

3 漸進的日本語生成手法

前節では，話し言葉に特有な表現を活用することにより，できる限り英語の語順に従い，かつ，日本語話し言葉として容認可能な日本語文を生成できることを述べた．本節では，それを依存文法を用いて実現する手法について説明する．依存文法は構成要素間の関係に基づく文法であり，英語及び日本語の双方ともに依存関係を見出すことができる．英語には，構成要素間に主辞 (head) と補語 (complement) という関係があり，それらは補語が主辞に依存するという関係として定めることができる．一方，日本語には，文節間に係り受けと呼ばれる依存関係が存在し，“係り” の文節が “受け” の文節に依存するという関係として定めることができる．そして，日本語の係り受け関係には，係り先の唯一性，非交差性，後方修飾性という制約があり，この制約を用いれば，容認可能な日本語の語順を決定できる．しかし，同じ “受け” 文節をもつ複数の “係り” 文節の間における語順の制約は定められておらず，“係り” 文節間では，日本語係り受け制約を満たす限り，自由に語順を入れ換えることが可能である．

本手法では，依存文法を用いて，生成する日本語の語順を決定する．また，できる限り英語の生起順序に従うために，同じ依存先をもつ構成要素に対しては，英語文において早く出現した構成要素を先に生成する．

3.1 日本語生成順序の決定

日本語依存関係から，日本語生成順序を決定する．すなわち，日本語の生成順序は，日本語係り受け制約をもとに，

```
1. <subj,(私,1),(用意します,2)>
2. <     ,( ,3),(部屋,4)>
3. <obj ,(部屋,4),(用意します,2)>
4. <with ,([with],5),(部屋,4)>
5. <     ,( ,6),(浴室,7)>
6. <     ,(浴室,7),([with],5)>
7. <for ,([for],8),(用意します,2)>
8. <     ,(あなた,9),([for],8)>
```

図 3: 例文 (2.1) に対する日本語依存関係

```
1) <subj,(私,1),(用意します,2)>
   <obj ,(部屋,4),(用意します,2)>
   <for ,([for],8),(用意します,2)>
                    ↓
   [(私,1)(部屋,4)([for],8)(用意します,2)]
2) <     ,(あなた,9),([for],8)>
                    ↓
   [(私,1)(部屋,4)(あなた,9)([for],8)(用意します,2)]
3) <     ,( ,3),(部屋,4)>
   <with ,([with],5),(部屋,4)>
                    ↓
   [(私,1)( ,3)([with],5)(部屋,4)(あなた,9)
   ([for],8)(用意します,2)]
4) <     ,(浴室,7),([with],5)>
                    ↓
   [(私,1)( ,3)(浴室,7)([with],5)(部屋,4)
   (あなた,9)([for],8)(用意します,2)]
5) <     ,( ,6),(浴室,7)>
                    ↓
   [(私,1)( ,3)( ,6)(浴室,7)([with],5)
   (部屋,4)(あなた,9)([for],8)(用意します,2)]
```

図 4: 日本語生成順の決定

“係り” となる構成要素を “受け” となる構成要素の前方に置き，かつ，依存関係同士が交差しないようにする．また，同じ係り先を持つ構成要素同士は，英語において先に出現した構成要素の順に生成する．

以下では，依存関係は 3 項組 $< f, \alpha, \beta >$ で表し，構成要素 α が β に f という関係で依存することを意味することとする．また，構成要素には，英語文における出現の順番に関する情報を付与する．

3.2 生成処理の例

例として，例文 (2.1) に対する依存関係を用いたときの生成処理の流れを説明する．ただし，本節では，変換処理の段階で既に図 3のような日本語依存関係が作成されているとする．まず，日本語依存関係を用いて日本語の生成順を決定する．他に依存先をもたない構成要素 “用意します” に関する依存関係 1, 3, 7 を用いて生成順を決定する．係り受けの後方修飾性により，“用意します” の生成順は一番最後となる．“私”，“部屋”，“[for]”[2]は，同じ依存先をもつため，英語の出現位置の情報を用いて英語の出現順になるように日本語生成順を決定する．この場合は，図 4の 1) に示す順になる．

[2] “for” は単独ではその訳を決定できないため，この時点では “[for]” と表す

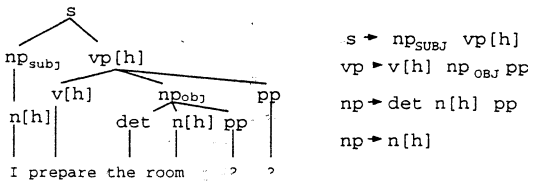

図 5: 主辞情報付き英語句構造と文法規則

表 1: 翻訳実験の結果 (1255 文)

	文数	割合 (%)
(1) 翻訳正解 (言い直しなし)	601 文	47.9
(2) 翻訳正解 (言い直しあり)	212 文	16.9
(3) 翻訳失敗	314 文	25.0
(4) 解析失敗	128 文	10.2

次に, 1) で決定した生成順の中の "用意します" の 1 つ前にある構成要素 "[for]" を依存先にもつ依存関係に対して, 先程と同じ操作を繰り返す. この場合, 2) のように "あなた" を "[for]" の前に生成することになる. 他の構成要素についても同様の処理を行う (図 4の 3) ~ 5)).

この結果, 日本語生成順は

(私), (NIL), (NIL), (浴室), ([with]), (部屋), (あなた), ([for]),(用意します)

となる.

この結果に助詞を補うことにより, 日本語文 "私 は 浴室 のある 部屋を あなた に 用意します" を最終的に生成することができる.

4 英日同時翻訳システム

4.1 システム構成

前節で述べた漸進的日本語生成手法をもとに, 英日同時翻訳システムを実現した. システムは, 漸進的な解析, 変換, 生成の 3 つのモジュールから構成される. 解析モジュールでは, 漸進的チャート解析に基づき主辞情報付き英語句構造を作成する. 変換モジュールでは, 英語句構造から依存関係を定める構造変換, 及び, その各構成要素に対する語彙変換を行う. 生成モジュールでは, 3 節で述べた手法により日本語生成順を決定し, 出力可能な部分までの生成を行う.

4.2 漸進的チャート解析

解析モジュールでは, 漸進的チャート法 [5] を用いて, 英語単語が入力されるごとにそれまでの入力に対する構文構造を作成する. 漸進的チャート解析は, 従来の上昇型チャート解析 [2] に, 活性弧に文法規則を適用する操作, 及び活性弧の項の最左未決定項を別の活性弧の項で置き換える操作を導入した枠組である. 漸進的チャート解析では, i 番目の語 w_i が入力されたとき, 以下の a) ~ c) の手続きを順に実行する. なお, 弧のラベルが項 σ であるとき, σ をその弧の項と呼ぶ.

a) **辞書引き** 語 w_i の範疇が X ならば, 項 $[w_i]_X$ をラベルとしてもつ不活性弧をチャートの節点 $i-1$ と節点 i の間に追加する.

b) **文法規則の適用** チャートの節点 $i-1$ と節点 i を結び, 項 $[\cdots]_X$ をラベルとしてもつ弧に対して, 文法規則 $A \rightarrow XY\cdots Z$ が存在するならば, 項 $[[\cdots]_X[?]_Y\cdots[?]_Z]_A$ をラベルとしてもつ弧をチャートの節点 $i-1$ と節点 i の間に追加する.

c) **項の置き換え** チャートの節点0と節点 $i-1$ を結ぶ活性弧の項 σ の最左未決定項を $[?]_X$ とする. このとき, チャートの節点 $i-1$ と節点 i を結ぶ弧の項 τ の範疇が X ならば, σ の最左未決定項を τ で置き換えた項をラベルとしてもつ弧をチャートの節点0と節点 i の間に追加する.

また, 英語の構成要素間の依存関係を求めるために, すべての文法規則に対して, その右辺の要素の中で, 主辞及び補語をあらかじめ定めておく [8]. 主辞は右辺に必ず 1 つ存在し, 残りの範疇はいずれも補語となる. 例えば, 図 5の文法規則 vp → v[h] np_{obj} pp は v が主辞であり, 残りの範疇 np, pp が補語であることを示す. この文法を用いることにより, 主辞情報付き英語句構造を作成できる. "I prepare the room" が入力された段階で作成される英語句構造を図 5に示す.

4.3 漸進的な変換処理

構造変換は, 句構造から依存構造への変換を行う. 解析モジュールで作成された主辞情報付き英語句構造に対して, 各節点ごとに依存関係を作成する. 例えば, 図 5の英語句構造の場合, まず, 範疇 s を根とし, その子節点として左から範疇 np, vp をもつ構造に対して, 構造変換を行う. このとき, 対応する文法規則 s → np_{subj} vp[h] より, vp が主辞, np が補語であり, また, np には必須格情報 *subj* があることがわかる. よって, それを依存関係 < *subj*, np, vp> に変換できる. 次に, 主辞要素, 補語要素の英単語を特定する. これは, 範疇 np, vp 以下の部分構造について, それぞれ主辞となる子節点の範疇を求めて, 最終的に得られる英単語または "?" とその英文における出現位置を依存関係の構成要素とする. その結果, 依存関係 < *subj*, (I,1), (prepare,2)> を作成できる. これらの操作を英語句構造の各節点に対してトップダウンに行う. ただし, 子節点が英単語及び "?" である場合には操作を行わない.

語彙変換では, 英語単語を対応する日本語に変換する.

4.4 漸進的な生成処理

3 節で述べた日本語の生成順序を決定する手法に基づき, 出力可能な部分までを生成する. そのとき, 対応する依存関係上の必須格情報または任意格情報を用いて助詞決定処理を行う.

4.5 翻訳処理の例

例文 (2.1) に対して, 同時翻訳システムの処理過程を表 2に示す. 表 2における上から下への順序は, 英語の入力の時間的順序を示す. 左から右への順序は, 各入力に対する処理の流れを示す.

5 翻訳実験

ATR 音声言語データベース [9] に収録されている 63 対話 (ホテルのフロント係による英語発話 1255 文) を用いて, UltraSparcII ワークステーション (512MB, 248MHz) 上で翻

表 2: "I prepare the room with a bath." に対する翻訳処理の流れ

入力	英語句構造	英語依存関係	日本語依存関係	出力
I	[[I]np[?]vp]s	< *subj*(I,1)(?,2)>	< *subj*(私, 1)(?, 2) >	
prepare	[[I]np[[prepare]v[?]np]vp]s	< *subj*(I,1)(prepare,2)> < *obj*(?, 3)(prepare,2)>	< *subj*(私, 1)(用意します, 2) > < *obj*(?, 3)(用意します, 2) >	私は
the	1.[[I]np[[prepare]v[[the]$_{det}$[?]n]np]vp]s 2.[[I]np[[prepare]v[[the]$_{det}$[?]n[?]pp]np]vp]s	< *obj*(?, 4)(prepare, 2) > < *nil*(the, 3)(?, 4) >	< *obj*(?, 4)(用意します, 2) > < *nil*(NIL, 3)(?, 4) >	
room	1.[[I]np[[prepare]v[[the]$_{det}$[room]n]np]vp]s 2.[[I]np[[prepare]v[[the]$_{det}$[room]n[?]pp]np]vp]s	< *obj*(room, 4)(prepare, 2) >	< *obj*(部屋, 4)(用意します, 2) >	部屋を
with	2.[[I]np[[[prepare]v[[the]$_{det}$[room]n[[with]p [?]np]pp]np]vp]s	< *obj*(room, 4)(prepare, 2) > < *with*(with, 5)(room, 4) > < *nil*(?, 6)(with, 5) >	< *obj*(部屋, 4)(用意します, 2) > < *with*([with], 5)(部屋, 4) > < *nil*(?, 6)([with], 5) >	あの一.
a	2.[[I]np[[prepare]v[[the]$_{det}$[room]n[[with]p[[a]$_{d}$ [?]n]np]pp]np]vp]s	< *obj*(room, 4)(prepare, 2) > < *with*(with, 5)(room, 4) > < *nil*(?, 7)(with, 5) > < *nil*(a, 6)(?, 7) >	< *obj*(部屋, 4)(用意します, 2) > < *with*([with], 5)(部屋, 4) > < *nil*(?, 7)([with], 5) > < *nil*(NIL, 6)(?, 7) >	
bath	2.[[I]np[[prepare]v[[the]$_{det}$[room]n[[with]p[[a]$_{d}$ [bath]n]np]pp]np]vp]s	< *obj*(room, 4)(prepare, 2) > < *with*(with, 5)(room, 4) > < *nil*(bath, 7)(with, 5) >	< *obj*(部屋, 4)(用意します, 2) > < *with*([with], 5)(部屋, 4) > < *nil*(浴室, 7)([with], 5) >	浴室 のある 部屋を 用意します

訳実験を行った. 入力文の平均単語数は 8.4 単語である. 文法規則数は 254 規則, 生成規則数は 834 規則, 辞書登録単語数は 1524 語である.

翻訳に用いた英語文をその了解性に従って 4 つに分類した. 表 1に示すように, (1) または (2) に分類された 64.8% が翻訳正解であり, これにより, 本手法の利用可能性を確認した.

本稿では, 日本語依存文法を導入することにより, 日本語の生成順序を決定する手法を提案した. 同じ依存先をもつ構成要素は, 英語入力と同時進行的に翻訳するために, 英語において先に入力された構成要素を優先して翻訳処理を実行する. しかし, "standard breakfast for children" をこの手法で翻訳すると, "標準的な子供たちのための朝食" となり, "標準的な" が "子供たちの" に係ると捉えられ得る日本語文を作成することになる. より自然な日本語文を生成するために, 日本語係り受け制約以外の日本語生成順序に関する制約が必要となる.

6 まとめ

英日同時通訳のための漸進的日本語生成手法を提案した. 日本語生成に依存文法を導入することにより, 日本語係り受け制約を満たす日本語を生成することができる. また, 同じ依存先をもつ構成要素同士では, 英語において先に出現した構成要素を優先することにより, できる限り英語の生起順序に従った日本語を生成できる. プロトタイプシステムを作成し, ATR 音声言語データベースに収録された英語発話 1255 文について翻訳実験を行い, 本手法の利用可能性を示した.

本手法では, 英語文の入力途中で随時, 翻訳処理を実行するために, 解析モジュールにおいて多くの英語構造が作成される. 変換モジュールでは, 作成された英語句構造の中から 1 つを選択し, 処理を実行するため, 選択結果によっては誤った依存関係が作成される場合がある. 今後は, 解析モジュールで作成される構造数を減らす手法, 及び適切な英語構造を選択する手法の検討が必要である.

参考文献

[1] Furuse, H. and Iida, H: Incremental Translation Utilizing Constituent Boundary Patterns, *Proc. of 16th Int. Conf. on Computational Linguistics*, pp.412-417 (1996).

[2] Kay, M : Algorithm Schemata and Data Structures in Syntactic Processing, *Technical Report CSL-80-12*, Xerox PARC (1980).

[3] Kitano, H.: Incremental Sentence Production with a Parallel Marker Passing Algorithm, *Proc. of 13th Int. Conf. on Computational Linguistics*, pp.217-222 (1990).

[4] 児玉 徳美: 依存文法の研究, 研究社出版(1987).

[5] Matsubara, S., Asai, S., Toyama, K. and Inagaki, Y.: Chart-based Parsing and Transfer in Incremental Spoken Language Translation, *Proc. of 4th Natural Language Processing Pacific Rim Symposium*, pp.521-524 (1997).

[6] 松原 茂樹, 浅井 悟, 外山 勝彦, 稲垣 康善 : 不適格表現を活用した漸進的な英日話し言葉翻訳手法, 電気学会論文誌, Vol.118-C, No.1, pp.71-78, (1998).

[7] Mima, H., Iida, H. and Furuse, O.: Simultaneous Interpretation Utilizing Example-based Incremental Transfer, *Proc. of 36th Annual Meeting of the Association for Computational Linguistics and 17th Int. Conf. on Computational Linguistics*. pp.855-861 (1998).

[8] 村瀬 隆久, 松原 茂樹, 外山 勝彦, 稲垣 康善 : 依存関係を用いた漸進的構文解析の効率化, 言語処理学会第 6 回年次大会発表論文集 (2000).

[9] 浦谷 則好, 竹沢 寿幸, 松尾 秀彦, 森田 千帆: 音声言語データベースの構成, テクニカルレポート TR-IT-0056, ATR 音声翻訳通信研究所 (1994).