自然言語処理 128−15
(1998. 11. 6)

類推によるコーパス中の対応関係の推定

タンテリ アンドリアマナカシナ　荒木 健治　栃内 香次
北海道大学大学院工学研究科

本論文では原文と翻訳文間の対応関係の検索手法について述べる．実例に基づく機械翻訳では翻訳例に対応関係が含まれていれば翻訳パターンを容易に抽出することができる．統計的な対応関係の検索手法から大量なコーパスがないと良質な結果が得られない．本手法では，対応関係が含まれている既存の翻訳例コーパスを用いて，新しい翻訳例の対応関係を推定し，対応関係付きの翻訳例を増加させる．大量なコーパスがまだ存在していない言語には適切な手法と考える．さらに，翻訳例中の対応関係を利用することによって，一対多や多対多の対応関係をうまく推定することができる．本手法を用いた実験により仏日会話の文を利用し，1,000 翻訳例から 80%以上の正対応関係が得られることが確認された．

Sub-Sentential Alignment by Analogy

Tantely Andriamanankasina　Kenji Araki　Koji Tochinai
Graduate School of Engineering, Hokkaido University

This paper describes a method for searching the word correspondences between a sentence and its translation sentence. In the Example-Based Machine Translation, translation patterns can be extracted easily if word correspondences between pair of translation sentences are determined. Statistical approaches are not able to produce reliable result unless a huge bilingual corpus is available. We propose a method for incrementing a word correspondences-included initial corpus automatically. It is appropriate for new languages whose huge corpus are still not available. In addition, links involving multiple tokens can be predicted by having the word correspondences included within the corpus, as reference. The method was evaluated with French-Japanese spoken language texts. As the number of translation examples goes beyond 1,000, more than 80% of exact correspondence rates were earned.

1　はじめに

実例に基づく機械翻訳では，バイリンガルコーパスの間の対応関係を正確に見つけ出すことは極めて重要である．翻訳例間の原文と訳文の対応関係が分かれば，翻訳パターンを簡単に取り出すことができる．例えば，フランス語文 **"je suis malade"**[1] とその日本語の翻訳文 **"私 は 病気 です"** では，もし **"malade"** と **"病気"** の対応関係が決定されれば，翻訳パターン **"je suis X ： 私 は Y です"** を取り出すことができ，X と Y はさまざまな単語対と置き換えることができる．

このような対応関係を見出すことは，辞書を利用して容易に行うことが可能に思えるが，登録されていない単語の出現，単語の活用形，辞書の見出し語と形態素解析結果との相違という問題がある．また，辞書の見出し語は一語単位であるので様々な複

[1] フランス語の文や日本語の例文は太字で表示される

合語には対応できない．さらに，何よりも同じ単語が同じ文に2回以上現れればそれぞれの対応関係を正確に決定するのは辞書では不可能である．そこで，辞書によらず，コーパスを用いて対応関係を求める手法が提案されている [4, 5, 6].

コーパス中の対応関係の決定では統計的な手法による研究が多く行われて来た [4, 5, 6]．統計的手法の問題は，コーパスのサイズが限定されると，良質な結果は得られないことである．それゆえ，あまり研究されていない言語では，コーパス及び辞書が整備されていないため，良好な結果が得られないことになる．さらに，バイリンガルコーパス中の一対多や多対多の対応関係を全て正確に決定することはできないという問題がある．文献 [4] では，一対多あるいは多対多を決定できないことが失敗の原因であり，文献 [5] では一対一の対応関係しか対象としていない．翻訳パターンの取り出しや機械翻訳という応用から見ると対応関係の誤りがシステム全体の性能に大きな悪影響を与える．一方，付属語の対応関係を決定しないでそれらを固定の言葉として扱う手法もある [6] が，その結果抽出される翻訳パターンは比較的質が良いが量は少ない．つまり，この場合も実用的な翻訳システムを作成するのに十分な翻訳パターンを抽出するためには大量のコーパスが必要になる．これらの点がこの手法の問題点である．

我々は対応関係を有する翻訳例を用いた仏日機械翻訳について研究を進めている [1, 2]．仏日の翻訳は比較的新しい分野なので，大量なコーパスや辞書はまだ存在していない．上述のように，実例に基づく機械翻訳手法では翻訳例の数が大量にあればあるほど結果に対する信頼度が高まる．しかし，翻訳例の人手による作成には時間と労力がかかる．したがって，小規模な対応関係を有するコーパスを利用して自動的に大量の対応付けをする必要がある．

本論文では，原文と訳文の組の対応関係の推定に対応関係決定済み翻訳例から類似翻訳例を抽出し，その対応関係を基にその組の対応関係を推定する手法を提案する．既存の翻訳例コーパスの中に全く同じ部分が現れなければ，品詞情報を利用して，対応関係を推定する．本手法では，これによって，量的に限られているコーパスでは利用できないという統計的な手法の問題点を回避できる．

また，一対多や多対多の対応関係抽出問題も，予め翻訳例に対応関係が含まれていることで解決できる．例えば，**“voulez - vous”** と **“いただけ ますか”** が現れたら，**“voulez”** だけが **“いただけ”** と対応するか，**“voulez - vous”** にするか，あるいは **“いただけ ます か”** と対応するかいずれも考えられる．統計的な結果は文の意味による結果といつも同じ結果ではない．それに，決定手法が一対一にするか否かによって結果も異なる．しかし，翻訳例の中に例えば **“voulez - vous”** と **“いただけ ますか”** かそれに近い **“pouvez - vous”** と **“もらえ ます か”** があれば，それらの間に対応関係がどうなっているかを参考にして簡単に問題が解決できる．

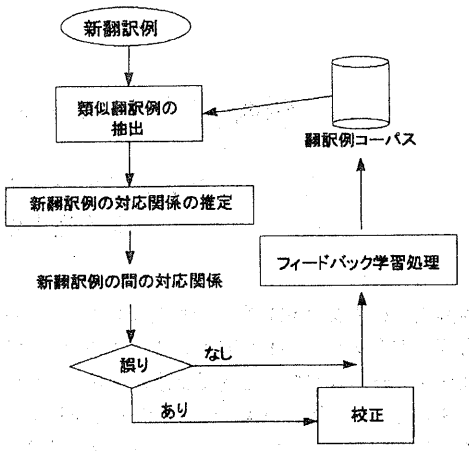

図 1: 対応関係の推定手法の概要

次章で，本手法の概要を紹介そして説明する．次に，実験方法と結果を記述し，考察する．最後に全体をまとめる．

2 対応関係の推定手法の概要

対応関係の推定の流れを，**図 1**に示す．原文とその翻訳文を入力し，翻訳例コーパスから類似翻訳例を抽出し，それらの対応関係を基に入力された原文と翻訳文間の対応関係を推定し，決定する．その結果に誤りが含まれていれば，人手で校正し，新しい翻訳例を翻訳実例コーパスに追加する．一方，対応関係の推定結果から何が正解か何が誤ったかによりフィードバック学習処理を行い，後の入力翻訳例の対応関係の推定の精度の向上させる．各文に，同手順を繰り返せば，対応関係付き大量なコーパスを少しずつ作ることができる．　翻訳例コーパスの中

表 1: 翻訳例の構造

フランス語文	"vous/PRV avez/ACJ un/DTN cendrier/SBC ?/?"
日本語文	"灰皿/6 は/9 あり/2 ます/14 か/9 。/1"
対応関係	2/3 4/1 5/6

PRV: pronoun DTN: determinant SBC: common noun ACJ: verb "avoir"
1: 特殊 2: 動詞 6: 名詞 9: 助詞 14: 接尾辞

の実例の構造は，表1に記述される．一つの形態素は「単語/品詞」形式で表示される．フランス語の形態素解析では，INALF[2]の EBTI プログラムを，日本語では CHASEN1.51[7] を利用した．INALF が提供したフランス語の品詞の数は 40 である．しかし，類似文の抽出にマイナスの影響を引き起こすことから，男性単語と女性単語，そして単数と複数は区別しないことにした．日本語では，品詞が木構造で分類されているが，我々は一番上のレベルに相当する 14 品詞を利用した．なお，本手法では構文解析，意味解析を行わず，形態素鵜解析結果を用いている．構文解析や意味解析の結果を利用して，より正確な推定ができると思われるが，これらのツールの精度は依然として不十分である．一方，最近の形態素解析ツールは非常に精度が高い．そこで，品詞情報のみを利用することとした．

対応関係は wpf1,.../wpj1,... 形式で表示される．wpfi はフランス語文の中の単語の位置で，wpjj は日本語文の中の単語の位置である．表 1の例では，2/3 は **"avez"** が **"あり"** と対応することを意味する．同じように **"cendrier"** は **"灰皿"** と，そして **"?"** は **"。"** と対応する．対応関係を人手で校正するとき，文の意味を理解した上で，対応関係を決定した．従って，さまざまな場合が存在している．日本語の "は" のような冠詞やゼロ代名詞のように対応がない形態素も考えられるし，**"s' il vous plaît"** と **"ください"** の関係のように複数の形態素と対応する場合も考えられる．また，**"voulez - vous"** と **"いただけ ます か"** の関係のように複数の形態素と対応する複数の形態素もあり，隣接しない部分は一つの形態素と対応する場合もある．例えば，**" n' ai pas"** の **"n' pas"** と **"あり ません"** の **"ません"** を対応させる．

[2]Institut National de la Langue Française

3 類似翻訳例の抽出

対応関係を推定するために，原文か翻訳文の類似文だけを探すのではなく，原文とも翻訳文ともに類似している翻訳例を抽出する．二つの文とも類似していなかったら，その文は抽出しない．入力文とその類似文のそれぞれの訳文が類似しているか確認する必要がある．

類似文の抽出方法を次に述べる．翻訳例コーパス中より入力翻訳例の原文と類似する原文を持つ翻訳例を検索する．選択された翻訳例の訳文と入力文の訳文の類似を調査する．ここで，訳文同士が類似している場合に限りその翻訳例を類似翻訳例として抽出する．次に，入力翻訳例の訳文から同じ手順を行い，抽出された翻訳例を合わせて対応関係の推定に利用する．

3.1 文の部分間の類似

ある言語で二つの文が類似していても他の言語ではそれらの翻訳文が類似するとは限らないことから，原文と翻訳文が類似している類似翻訳例を見つける可能性は低いと考えられる．そこで，類似文ではなく類似文の部分という考え方を導入する．文の一部だけを観察すれば，原文とも翻訳文とも類似している文の部分が見つかる可能性が高まる．最初に類似文の部分の検索を各言語で行う．入力原文の各形態素にその形態素が入っている部分と類似する n 文の部分を検索する．アルゴリズムは下記の通りである．

> 入力文の各形態素に字面が同じ形態素を例文の中で検索する．見つかれば，その形態素の位置から，前方と後方の比較を行う．字面が一致しない位置から品詞を調べて比較を続ける．一致しない形態素とぶつかったら比較は終了にする．

上述のアルゴリズムにより，字面で一致する部分のうちその両側の品詞が一致するものの隣接する

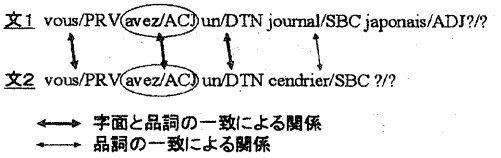

図 2: 原文の部分間のマッチング手法

対応関係を決定する．原文と例文を比較するときも，訳文同士の類似を確認するときも，同じアルゴリズムを利用する．**図 2**に例を示す．下記の類似度を利用して n 文の部分の最も類似するものを選択する．

$$SM = \alpha * NE + NP \quad (1)$$

SM[3]は類似度，NE[4]は字面が一致する形態素の数，NP[5]は品詞が一致する形態素の数である．

図2の文を見ると，形態素 **"avez/ACJ"** に関し，文の二番目の形態素が字面で一致する．そこから，前方で字面の一致 **"vous/PRV"** と **"vous/PRV"** を発見する．後方では字面の一致 **"un/DTN"** と **"un/DTN"** と品詞一致 **"journal/SBC"** と **"cendrier/SBC"** を見つけ出す．合計では三つの字面一致と一つの品詞一致が存在する．従って，類似度は式 (1) の NEに 3 を，NPに 1 を代入して，$SM = \alpha * 3 + 1$.

一文中では，同じ形態素が複数回現れる場合もある．この場合，各形態素は異なる部分に属するものと考えて扱う．

3.2 抽出する類似部分の数

節 3.1では，各形態素に n の類似部分が抽出されるので，原文と訳文に対し，それぞれ n 倍のその文の形態素の数の類似部分を抽出する．例えば，形態素個数が 6 の文だと，各形態素に n の類似部分を検索し，その文に対し全部で $6n$ の類似部分を抽出する．但し，同じ文の部分が取り出される可能性もあるし，字面一致がいつもあるとは限らないため類似部分が見つからない場合もある．例えば，図 2 の文 1 を入力文とする．形態素 **"avez"** に対し，文 2 の **"vous avez un cendrier"** という文の部分が抽出されるとする．しかし，その文の部分も形態素 **"vous"** に対し，隣接する字面一致 **"vous avez un"** のおかげで類似度も高くなって，抽出される．一方，翻訳例コーパスにはまだ入っていない形態素が現れたら，それに対し字面で一致する形態素がないので，類似部分は一つも抽出されない．つまり，実際に見つけた異なる類似部分の数は多くても n ということである．6 は文の長さすれば，文全体では多くても $6n$ の類似部分が抽出される．

[3]Similarity Metric
[4]Number of Exact matches
[5]Number of Part-of-speech tag matches

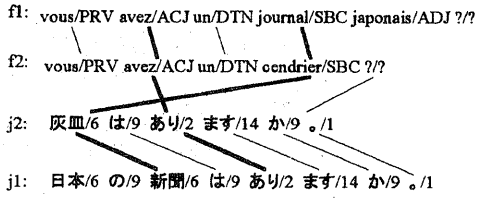

図 3: 類推による対応関係の検索

4 新翻訳の対応関係の推定

この章では，類似部分に存在している対応関係を利用して，入力原文と翻訳文の対応関係を推定する．過程は図 3に示される．(f1,j1) は新翻訳例で，(f2,j2) は取り出された類似例である．f2 と j2 の間の対応関係はあらかじめ翻訳例と共に与えられる．f1 と f2，そして j1 と j2 の間の対応関係は類似翻訳例の抽出の時に決定される．ここでは，一つの f1 の形態素から一つの j1 の形態素までの経路を全て検索する．図 3の例では，**"avez/ACJ"** から **"あり/2"** まで，そして **"journal/SBC"** から **"新聞/6"** までの二つの経路がある．それらが二つの文間の対応関係である．

原文と類似文の類似度を訳文同士の類似度と合計する．対応関係の推定は合計の高いものから利用して，開始する．一対多や多対多の対応関係は同じ類似部分を利用したときのみ認める．つまり，形態素が以前推定された対応関係に含まれていたら，その形態素を巻き込む後の対応関係は，拒否される．

5 フィードバック学習処理

様々なユーザがシステムを使用するので，異なる分野の文や誤りのある文が入力される．また，対応関係の校正の時に様々な方法や間違いもある．システムがそれらの問題に耐えることができるように

表 2: 利用したコーパス

翻訳例合計数	2,600
日本語文の平均長さ	7.74 形態素
フランス語文の平均長さ	7.84 形態素
一文に対する対応関係の数	7.27

ここでフィードバック学習を行う．文献 [3] で提案されたフィードバック学習処理の考えを本手法に導入した．各文にあるパラメータ FP[6] を利用する．校正結果と対応関係の推定結果を比較して，それらの対応関係の正誤を決定をする．推定された対応関係が誤ったら，使用された文の FP の値を一つ減らす．または，対応関係が正しければ，使用された文の FP を一つ増やす．FP は後の入力翻訳例に対しどの例を取り出せばよいかを決定する時に類似度に加えて使う．ここでは，FP に関し二つの条件を導入する．

1. 誤った対応関係があるとき，FP の減少はいつも行うが，正しい時には FP が初期値に戻っていない文のみ FP の増加を行う．これは，仮に正しいときに FP を増加させると特定の翻訳例のみが用いられるようになることを防ぐためである．

2. FP はマイナスの値であるが，それの絶対値を使用する．それが高ければ，文が選択される可能性に悪影響を与える．その値を 1 と初期化し，類似文を抽出するときに類似度に加えて，以下の尺度を使用する．

$$NM = SM/abs(FP) \qquad (2)$$

SM[7] は類似度で，NM[8] は類似文を選択するときの新しい尺度である．NM は文と文の類似度ではなく，入力翻訳例の対応関係の推定にどの文を使用すればよいかを決定するための尺度である．

6 実験と結果

実験に利用したコーパスを**表 2**に示す．文章は仏日会話の本 [8, 9] から取った．会話には短い文が多く存在することから文の平均長がやや短くなっている．翻訳例コーパスは最初空である．翻訳例を一つ一つシステムに入力し，対応関係を推定し，もし誤りが含まれていれば人手により校正する．最後に，翻訳例コーパスに追加する．対応関係推定の適合率と再現率を以下の式で定義する．

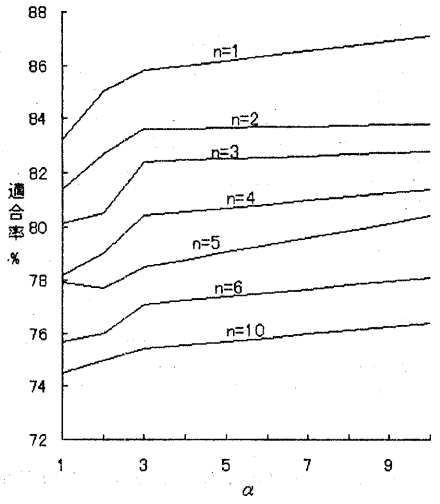

図 4: パラメータの変化による適合率の変化

$$\text{適合率} = 100 \times \frac{\text{正対応関係の数}}{\text{推定された対応関係の数}} \qquad (3)$$

$$\text{再現率} = 100 \times \frac{\text{推定された対応関係の数}}{\text{校正後の対応関係の数}} \qquad (4)$$

式 (1) のαと**節 3.1**の n は次に述べる予備実験により決定した．800 の対応関係付き翻訳例コーパスを使用し，新しい対応関係のない 200 翻訳例を入力した．αを 1, 2, 3, 10 まで，そして n は 1, 2, 3, 4, 5, 6, 10 に変化させて，結果の適合率と再現率を観察した．適合率を**図 4**に示す．αが 1 になると高い適合率を得ることができないがαが大きくなると適合率が高くなる．n による変化を見ると，n が大きくなると高い適合率が得られない．再現率を**図 5**に示す．αによる変化が n が大きければ大きいほど再現率が高くなる．高い適合率と再現率ともに考えるとαが 10，そして n が 5 となる．つまり，各形態素に五つの類似部分を抽出した．αは類似度に利用するパラメータで結局類似度は $SM = 10*NE+NP$になる．αが大きければ類似度は字面一致の数の影響が大きくなる．字面一致の数がほぼ同じ場合に品詞一致の数の影響が出る．次に，2,600 翻訳例を一つずつランダムにシステムに入力した．推定された対応関

[6] Feedback Parameter
[7] Similarity Metric
[8] New Metric

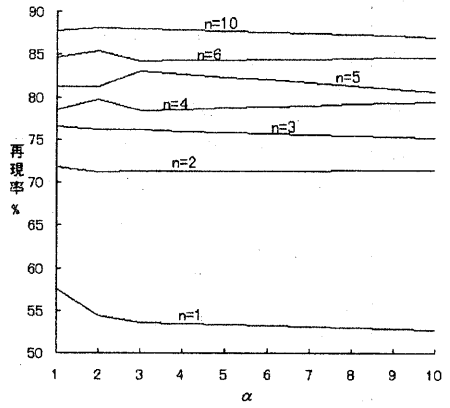

図 5: パラメータの変化による再現率の変化

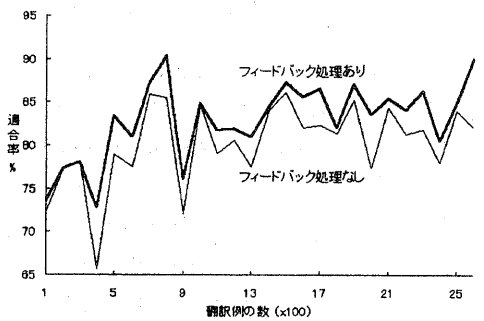

図 6: 適合率の変化

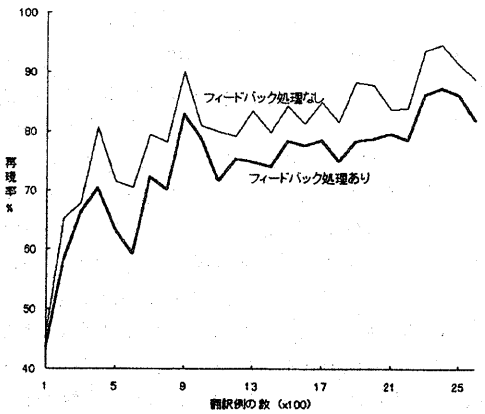

図 7: 再現率の変動

表 3: 品詞情報の必要性

	全ての対応関係	正対応関係	適合率
品詞情報から	33.1%	26.0%	65.2%
字面から	66.9%	74.0%	91.9%
合計/平均	100.0%	100.0%	83.1%

係が誤ったら，人手により校正してから翻訳例コーパスに追加する．結果を 100 例ずつにわけて適合率と再現率を観察した．フィードバック学習を行わない場合も実験し，フィードバックの効果も考察した．適合率を図 6に，再現率を図 7に示す．次に品詞情報が本手法においてどう役に立ったかを考察する．品詞情報によって推定された対応関係と字面一致で抽出された対応関係を数えた．その結果を表 3に示す．品詞情報によって推定された対応関係は全ての結果の 33.1%であるが，正対応関係中では 26.0%に下がった．品詞情報の利用の適合率は 65.2%である．

7 考察

図 6より，翻訳例が増加すればするほど適合率が上昇して行くことが分かる．傾きが小さいが下がることはない．1,000 の翻訳例からつねに 80%以上の適合率が得られた．また，2,600 までの翻訳例では適合率は最大 90.0%である．フィードバック処理がある場合の方がない場合よりも上回っていることより，フィードバック学習処理の有効性を示すことができた．一方，図 7より，翻訳例が増えれば増えるほど再現率が増加して行くことがはっきり確認できた．この値はフィードバック処理がある場合の方が下回っているが，再現率より適合率のほうが重要である．その理由としては，精度の高い翻訳パターンを少数でも取り出すことが翻訳システムにとってより重要であると考えられるからである．2,400 の翻訳例で再現率は最大 87.2%である．これは本手法が有効であることを示している．翻訳例の数の少なさ，そして非文が多い会話の文ということを考えると非常に良い結果が得られたと考えられる．二つのグラフを同時に見ると，適合率，再現率ともにしだいに上昇し，しかも高い値を示している．このことにより，本手法の有効性を確認することができた．

実験結果の例を図 8に示す．**"quelle heure"** と **"何時"** の対応関係のように一対多や多対多の対応関係がうまく決定されている．論理的に **"何"** は

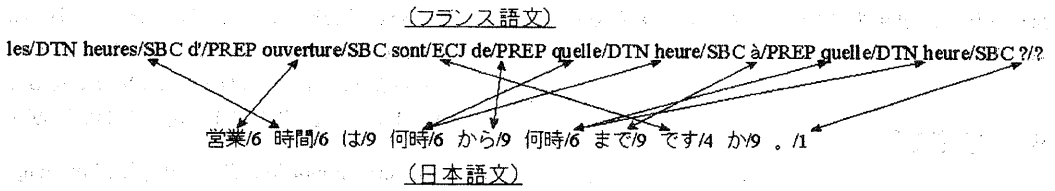

図 8: 結果の例

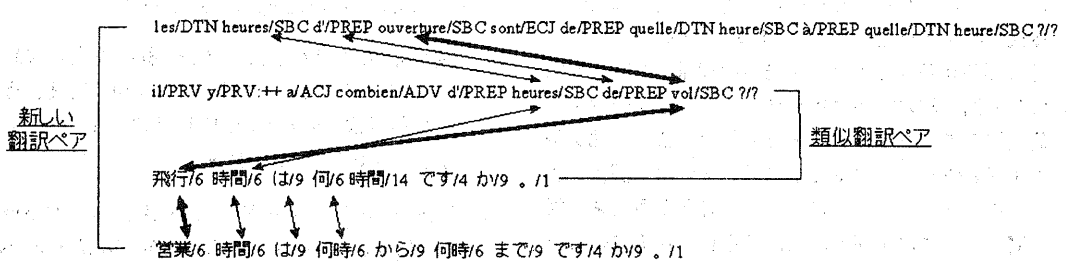

図 9: 品詞情報によって推定できた対応関係

"quelle" と，そして **"時"** は **"heure"** と一致する．しかし，**"何時"** が一つの形態素になったことで **"quelle"** と **"何時"** そして **"heure"** と **"何時"** という対応関係になった．類似文の部分が既に翻訳例の中に存在することから，このような一対多や多対多の対応関係を推定することが可能である．

さらに，本手法は文の中に複数現れている同じ形態素のそれぞれの対応関係を決定することができた．図 8では，**"quelle heure"** と **"何時"** はともに二度出現しているが，それぞれの対応関係を正しく決定できた．それが可能になったのは単語ではなく部分を見ることにより隣の単語などを見ているからである．単語単位で考えている統計的な手法ではこのような処理は行えない．

決定されたすべての対応関係の 33.1%は品詞情報によって決定された．これは品詞情報の必要性を示している．翻訳例コーパスにはまだ入っていなかった形態素なので，品詞情報と周りの単語だけで対応関係を決定することができた．品詞情報を利用しなければ再現率が 75–80%から 50–55%に減少する．図 **9**に **"ouverture"** と **"営業"** の関係から品詞情報の有効性を見ることができる．選択された部分の中に出ているそれらの対応している部分は字面で一致しないにも関わらず品詞情報を観察することによって正しい対応関係に決定できた．つまり，本手法では知らない単語が新翻訳例に存在しても対応関係を決定することができる．適合率は字面で一致する対応関係の推定より低いが，品詞情報は未登録語の対応関係の推定のために必要な情報と考えられる．

類似部分の抽出の結果は両言語の品詞の数に影響を受ける．品詞の数が少なければ長い類似部分を見つけることができる．しかし，少なすぎると高い回数で出現する品詞が誤抽出を起こす．一方，品詞の数が多いと品詞で一致する部分が短く，未登録語の対応関係を探すのは困難になる．それは推定できなかった対応関係の原因の一つであった．フランス語の 40 の品詞に対し日本語の 14 の品詞は少ないと考えられ，バランスを考慮する必要がある．

もう一つの誤りの原因は類似部分の検索のときに対応関係のないものが出現することである．類似文の中に対応関係のない形態素があると再現率が低下する．一方，一対多や多対多の対応関係では，同じ翻訳例を利用して推定されたら場合，その対応関係に決定するが，異なる翻訳例を利用していた場合には決定しない．それらの問題は翻訳例の増加につれて少しずつ解決されていくが，手法の改良のために，類似文で対応関係がない形態素の対応関係の推定や異なる類似翻訳例を利用した一対多や多対多の

対応関係の決定手法について研究する必要があると考える．

8 むすび

本論文では類推を用いた対応関係の推定手法について述べた．統計に基づく手法では大量なコーパスがないと良質な結果は得られないことから大量なコーパスがまだ存在しない言語間では実例に基づく機械翻訳手法を適用するのが困難である．ここでは，対応関係が決定されている既存の翻訳例を用いて，新しい翻訳例の対応関係を推定し，対応関係付きの翻訳例を増加させる手法を提案した．新しい翻訳例の対応関係を推定するときに類似文の対応関係を利用する．そうすることによって一対多や多対多の対応関係も一文に数回現れた同じ単語の対応関係もうまく抽出できる．さらに，フィードバック学習処理の導入によって，失敗を繰り返すのを避けることができる．それは，後の翻訳例の対応関係の推定に利用される．

実験結果では抽出された対応関係の数が翻訳例の増加に伴って増加して行くことが確認できた．適合率も少しずつ増加している．2,600 の翻訳例では対応関係の 80.0%以上を 90.0%の適合率で抽出できた．使用した会話文に非文の多いことを考えるとこれはよい結果と考えられる．一方，フィードバック学習処理が行った場合が行わない場合より上回ったことからフィードバック学習処理の有効性が確認できた．さらに，一対多や多対多の対応関係や一文に数回出現する同じ単語の対応関係がうまく抽出できることや品詞情報の必要性が確認できた．

今後の課題としては，両言語の品詞の数のバランスや類似文で対応関係がない形態素の対応関係の推定や異なる類似翻訳例を利用した一対多や多対多の対応関係の決定手法を考えられる．

謝辞 本研究の一部は文部省科学研究費補助金 (第 10680367 号及び第 09878070 号) によって行われている．フランス語の形態素解析を提供した "Institut National de la Langue Française" に，謹んで感謝の意を表する．

参考文献

[1] Andriamanankasina, T., Araki, K., Miyanaga, Y., and Tochinai, K.: *Method for Searching the Best-Matching Sentence in Example-Based Machine Translation*, Technical Report of IEICE, Vol. NLC97-10, pp. 15–20, (1997).

[2] Andriamanankasina, T., Araki, K., Miyanaga, Y., and Tochinai, K.: *Machine Translation Based on the Relations between Words*, Proc. of NL Symposium Towards Useful Natural Language Procesing Japan (1997).

[3] 荒木健治, 高橋祐治, 桃内佳雄, 栃内香次：帰納的学習を用いたべた書き文のかな漢字変換, 電子情報通信学会論文誌, Vol. J97-D-II, No. 4, pp. 391–402 (March 1996).

[4] Brown, P. F., Della Pietra, S. A., Della Pietra V. J., and Mercer R. L.: *The Mathematics of Statistical Machine Translation: Parameter Estimation*, Computational Linguistics, Vol. 19, No. 2, pp. 263–309 (1993).

[5] Melamed, D.: *A Word-to-Word Model of Translational Equivalence*, 35th Conference of the Association for Computational Linguistics (ACL'97), Spain (1997).

[6] 北村美穂子, 松本裕治: 対話コーパスを利用した対話表現の自動抽出, 情報処理学会論文誌, Vol. 38, No. 4, pp. 727–736 (1997).

[7] Yamashita, T.: *ChaSen Technical Report*, Nara Advanced Institute of Science and Technology (1996).

[8] Meguro, S.: *Manuel de Conversation Française*, Hakusuisha, Tokyo (1987)

[9] Sato, F.: *Locutions de base*, Hakusuisha, Tokyo (1990).