統計的機械翻訳

永田　昌明

「統計的機械翻訳」(statistical machine translation) は，すでに翻訳されたある言語と別の言語の文のデータから翻訳規則や対訳辞書などに相当する数学的なモデルを学習し，ある言語の任意の文を別の言語に翻訳する技術であり，近年，自然言語処理の分野において最も注目を集めている技術といって過言ではない．本稿では，IBM 翻訳モデル，句に基づく翻訳，階層的句に基づく翻訳など統計的機械翻訳の代表的な手法を概観する．

キーワード：機械翻訳，対訳コーパス，言語モデル，翻訳モデル，デコーダ

1．はじめに

コンピュータを利用してある言語を別の言語に翻訳する技術を「機械翻訳」(machine translation) と呼ぶ．機械翻訳の研究はコンピュータの誕生とほぼ同時に 1950 年代から始まり，今日では様々な言語の間での機械翻訳ソフトウェアが製品化され，インターネット上で無料の自動翻訳サービスが提供されている．

従来の機械翻訳へのアプローチは，言語学の専門家がコンピュータ処理向けの文法規則と辞書を作成する「知識に基づく機械翻訳」(knowledge-based machine translation) が主流であった．近年，すでに翻訳されたある言語と別の言語の文のデータから翻訳規則や対訳辞書などに相当する数学的なモデルを学習し，ある言語の任意の文を別の言語に翻訳する「統計的機械翻訳」(statistical machine translation) というアプローチが提案され，賛否両論はあるが少なくとも学会レベルでは主流になっている．

本稿では，ほぼ歴史的な経緯に沿って統計的機械翻訳の基本的な考え方と代表的な手法を概観する．

2．言語翻訳の生成モデル

統計的機械翻訳の研究は，1980 年代後半に IBM のワトソン研究所の音声認識グループで始まった[1]．初期の研究がフランス語から英語への翻訳を対象としていたため，統計的機械翻訳の分野では，原言語 (source language，翻訳元言語) をフランス語 $\mathbf{f}$，目的言語 (target language，翻訳先言語) を英語 $\mathbf{e}$ と表記する習慣があり，本稿もこれに従う．

一般に，あるフランス語の文に対して様々な英語の文への翻訳が考えられる．統計的機械翻訳では，あるフランス語の文 $\mathbf{f}$ に対してすべての英語の文 $\mathbf{e}$ が翻訳になりうると考え，すべての文の対 $(\mathbf{e}, \mathbf{f})$ に対して「翻訳者が $\mathbf{f}$ を $\mathbf{e}$ に翻訳する可能性」に相当する確率 $P(\mathbf{e}|\mathbf{f})$ を割り当てる．このとき，与えられた $\mathbf{f}$ に対して確率 $P(\mathbf{e}|\mathbf{f})$ が最大にする $\hat{\mathbf{e}}$ を選べば，フランス語を英語に翻訳する際の誤りを最小にできる．ベイズの法則により結局 $P(\mathbf{e})P(\mathbf{f}|\mathbf{e})$ を最大にする文を探せばよい．

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) = \arg\max_{\mathbf{e}} P(\mathbf{e})P(\mathbf{f}|\mathbf{e}) \qquad (1)$$

式(1)は，「雑音のある通信路モデル」(noisy channel model) を言語翻訳に適用したことを意味する．翻訳すべきフランス語の文は，非常に雑音の多い通信路において英語がフランス語に変形したとみなし，これを元の英語の文へ復元する処理が言語翻訳であると考える．

一般に，英語の文の事前確率 $P(\mathbf{e})$ を計算するためのモデルを言語モデル (language model)，英語の文が与えられたときのフランス語の文の条件付き確率 $P(\mathbf{f}|\mathbf{e})$ を計算するためのモデルを翻訳モデル (translation model) と呼ぶ[1]．また言語翻訳は雑音のある通信路による符合化 (encode) の逆過程という解釈から，$P(\mathbf{e})P(\mathbf{f}|\mathbf{e})$ を最大化する英語の文を探索する処理をデコード (decode，復号)，復号を実行する処理系をデコーダ (decoder，復号器) と呼ぶ．

ながた　まさあき
NTT コミュニケーション科学基礎研究所
〒 619-0237 京都府相楽郡精華町光台 2-4

[1] ベイズの法則を用いた式(1)のせいで，翻訳モデルの原言語 (英語)/目的言語 (フランス語) は翻訳システムの原言語 (フランス語)/目的言語 (英語) と逆になる．この混乱を避けるために原言語と目的言語ではなくフランスと英語を使う習慣になった (と思われる)．

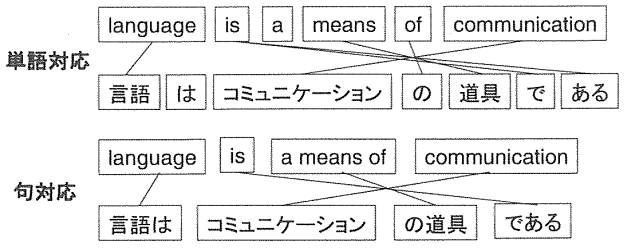

図1　単語対応と句対応

同じ内容を二つの言語で記述したテキストの集合を対訳コーパス（parallel corpus）または二言語コーパス（bilingual corpus）と呼ぶ．翻訳モデルや言語モデルは対訳コーパスから学習する．対訳コーパスとしては，カナダの国会議事録（Hansards），EUの議会議事録（Europarl），香港の議会議事録，国連の刊行物，多国籍企業の製品マニュアル，通信社のニュース記事などが統計的機械翻訳の研究に利用されている．

3．単語に基づく翻訳

1990年前後にIBMは順番に少しずつ複雑になるモデル1からモデル5までの5つの翻訳モデルを提案した[1]．この「IBM翻訳モデル」は，対訳文において互いに翻訳になっている単語を結び付ける単語対応（word alignment）という概念を基本としている．図1の上段に単語対応の例を示す．一般に単語対応は多対多対応であり，英語の冠詞や日本語の助詞のように相手言語に対応する単語がない場合もある．

IBM翻訳モデルでは，フランス語の文$\mathbf{f}$と英語の文$\mathbf{e}$が互いに翻訳になっており，その単語対応が$\mathbf{a}$であるような同時確率分布$P(\mathbf{f},\mathbf{a},\mathbf{e})$を考え，$P(\mathbf{f}|\mathbf{e})$をすべての単語対応に関する条件付き確率$P(\mathbf{f},\mathbf{a}|\mathbf{e})$の和として表す．

$$P(\mathbf{f}|\mathbf{e})=\sum_{\mathbf{a}} P(\mathbf{f},\mathbf{a}|\mathbf{e}) \tag{2}$$

IBM翻訳モデルでは，英語からフランス語への翻訳においてフランス語の単語は最大1つの英語の単語に対応する，すなわち，英語の1つの単語に対応するかまたは対応する単語がない（位置0の空の単語に対応する）という制限を加える．これにより，長さlの英語の文$\mathbf{e}=e_1^l=e_1\cdots e_l$と長さ$m$のフランス語の文$\mathbf{f}=f_1^m=f_1\cdots f_m$に対して，単語対応$\mathbf{a}$は$a_1^m=a_1\cdots a_m(a_j=i, 0\leq i\leq l)$と表現できる．何も制約を加えなければ，単語対応は文全体で2^{lm}通りある．上記の制約を加えると，言語対に対して翻訳モデルが非対称になるという問題はあるが，単語対応の総数が大幅に減って$(l+1)^m$通りとなり，パラメタの推定が容易になる．

各翻訳モデルのパラメタはEMアルゴリズムを用いて推定する．局所的な最適値に陥ることを避けるために，同じ訓練データに対して，より簡単なモデルのパラメタの推定値を次のモデルのパラメタの初期値とするという手順によりモデルを推定する．

最近では，IBM翻訳モデルは単語対応を求める目的で使用され，翻訳モデルとして使用されることは少ないため，本稿では説明を省略する．詳しくは文献[9]等を参照してほしい．

4．句に基づく翻訳

2000年前後から翻訳の基本的な単位を単語から句に拡張する研究が盛んになり成功を収めた．ここでいう句（phrase）は，名詞句や動詞句といった言語学的な文の構成要素ではなく，単に連続した単語列を指す．句を基本単位にすることにより，局所的な並び替え，複数単語から構成される表現，局所的な文脈に依存する挿入と削除などを翻訳モデルの中で表現することができる．

図1に句対応の例を示す．例えば，“a means of”と「の道具」を句単位で対応させているように，名詞と助詞の語順や冠詞の有無の違いは句を適切に選択することで解決できる．また句を単位として並び替えを行うことで語順の操作が簡単になる．

本稿では句に基づく統計翻訳（phrase-based SMT）の代表例として，文献[4]で提案された翻訳モデルとデコーダを紹介する．この翻訳モデルでは，まず原言語の文$\mathbf{f}$をI個の句の列$\bar{f}_1^I=\bar{f}_1\cdots\bar{f}_I$に分割し，原言語の各句$\bar{f}_i$を目的言語の句$\bar{e}_i$に翻訳し，句を並び替える．翻訳確率$P(\mathbf{f}|\mathbf{e})$は句翻訳確率（phrase translation probability）$\phi(\bar{f}_i|\bar{e}_i)$と相対的な句歪み確率（phrase distortion probability）$d(a_i-b_{i-1})$の積で近似する．

$$p(\bar{f}_1^I|\bar{e}_1^I)=\prod_{i=1}^{I}\phi(\bar{f}_i|\bar{e}_i)d(a_i-b_{i-1}) \tag{3}$$

ここでa_iは，i番目の目的言語句に翻訳された原言語句の開始位置であり，b_{i-1}は，$(i-1)$番目の目的言語句に翻訳された原言語句の終了位置である．

句翻訳確率は抽出された句の相対確率から求める．

$$\phi(\bar{f}|\bar{e})=\frac{count(\bar{f},\bar{e})}{\sum_{\bar{f}'} count(\bar{f}',\bar{e})} \tag{4}$$

ここで$count(\bar{f},\bar{e})$は，原言語句$\bar{f}$と目的言語句$\bar{e}$の対応付けの頻度である．句歪み確率は，適当に決め

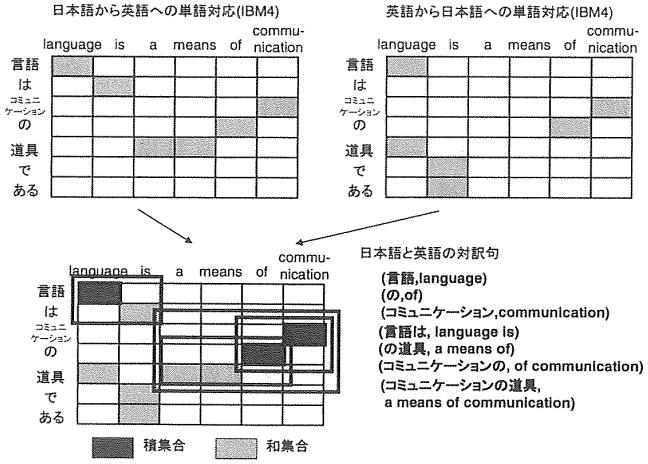

図2　単語対応付けからの対訳句の抽出

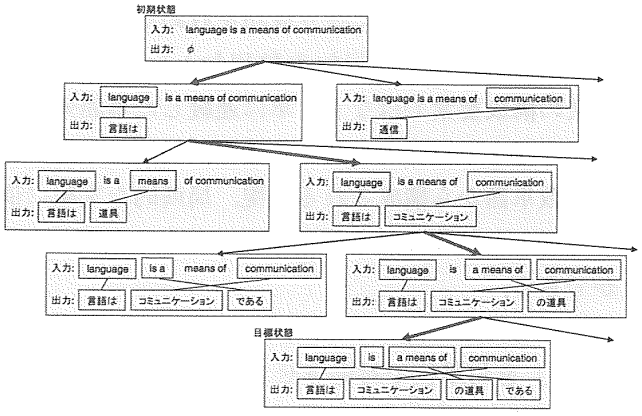

図3　ビーム探索によるデコーディング

たパラメタ α を用いて，句の移動距離に対して指数的に大きくなるペナルティを与える．

$$d(a_i-b_{i-1})=\alpha^{|a_i-b_{i-1}-1|} \tag{5}$$

上記の句歪み確率は，距離や移動方向，原言語句や目的言語句への依存性など，大局的な句の並び替えの傾向を表現していない．文献[5]では，日本語と英語のような語順が大きく異なる言語向けに，より詳細な句の並び替えモデルを提案している．

互いに翻訳となる句は，単語対応付けされた対訳コーパスからヒューリスティクスを用いて抽出する．図2に例を示す．まずIBM翻訳モデルを用いて原言語から目的言語および目的言語から原言語の双方向の単語対応を求め，両者の積集合（intersection）と和集合（union）を求める．積集合の対応点（alignment point）は信頼度が高いと考えられるので，積集合の対応点を起点に新しい対応点を加えて句を作る．新しい対応点は和集合の中から既存の対応点に隣接し句に新しい単語を加えるものを選ぶ．

原言語の入力文に対する目的言語の翻訳出力は，ビーム探索により文頭から文末方向に部分的な翻訳を生成する．図3に例を示す．まず空（empty）の初期仮説（initial hypothesis）から出発し，ある仮説において1つの句を翻訳して新しい仮説を作るステップを繰り返す．1つの句の翻訳では，原言語の文で未翻訳の単語列から原言語の句を1つ選び，対応する目的言語の句を目的言語の部分文の文末側に付加する．仮説は優先順位付きキュー（priority queue）で管理し，原言語の文のすべての単語を翻訳したら探索は終了する．

既に翻訳した単語数が多いほど仮説の確率は小さくなるので，翻訳した単語数ごとに優先順位付きキューを用意する．言語モデルとして単語ngramモデルを使っている場合，既に翻訳した原言語の単語集合が同じ，かつ，目的言語の最後の $n-1$ 単語が同じである2つの仮説があれば，確率が大きい仮説だけを残す．これを仮説の再構成（hypothesis recombination）という．さらに各優先順位付きキューごとに，最大仮説数による枝刈り（histgram pruning）および最大確率の仮説からの相対値による枝刈り（threshold pruning）を行う．

5.　言語翻訳の識別モデル

5.1　対数線形モデル

機械学習の分野における生成モデル（generative model）から識別モデル（discriminative model）への流れに呼応して，2000年代前半から雑音のある通信路モデルに代わって対数線形モデル（log linear model）を用いて事後確率 $P(\mathbf{e}|\mathbf{f})$ を直接モデル化する方法が主流になった．対数線形モデルでは，M 個の素性関数 $h_m(\mathbf{e},\mathbf{f})$ と，各素性に対する重み λ_m を考え，翻訳の事後確率 $P(\mathbf{e}|\mathbf{f})$ を次式により求める．

$$p_{\lambda_1^M}(\mathbf{e}|\mathbf{f})=\frac{\exp\sum_{m=1}^{M}\lambda_m h_m(\mathbf{e},\mathbf{f})}{\sum_{\mathbf{e}'}\exp\sum_{m=1}^{M}\lambda_m h_m(\mathbf{e}',\mathbf{f})} \tag{6}$$

入力文に対する翻訳を求める際には式(6)の分母を計算する必要はなく，素性と重みの線形和を最大とする候補を探索すればよい．

$$\hat{\mathbf{e}}=\arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f})=\arg\max_{\mathbf{e}}\sum_{m=1}^{M}\lambda_m h_m(\mathbf{e},\mathbf{f}) \tag{7}$$

この式は，$h_1(\mathbf{e}|\mathbf{f})=p(\mathbf{e})$, $h_2(\mathbf{e}|\mathbf{f})=p(\mathbf{f}|\mathbf{e})$, $\lambda_1=\lambda_2=1$ とすれば式(1)と同じであり，このモデルは雑音のある通信路モデルを包含している．パラメタ λ_1 と λ_2 の最適化は，最適なモデルの重み（model scaling factor）を求めることに相当する．通常，素性 $h_m(\mathbf{e},\mathbf{f})$ としては，翻訳モデル，言語モデル，歪みモデル，単語の長

さなどが用いられる．

訓練データとしてS個の文の対からなる対訳コーパス $\{(\mathbf{e}_s, \mathbf{f}_s)|s=1, \cdots, S\}$ が与えられたとき，モデルパラメタ λ_1^M は最尤推定，すなわちコーパスの事後確率を最大にするように求める．対数線形モデルの尤度は凸（convex）関数であり，一般化反復スケーリング（Generalized Iterative Scaling）や勾配（gradient）に基づく最適化法により大局的な最適値を求められる．

$$\hat{\lambda}_1^M = \arg\max_{\lambda_1^M} \sum_1^S \log p_{\lambda_1^M}(\mathbf{e}_s|\mathbf{f}_s) \tag{8}$$

学習の際に式(6)の分母，すなわち，入力文のすべての翻訳候補に関する和を求める必要がある．通常は，確率が大きい順に上位N個の翻訳候補を求め，このN-best候補の確率の和で分母を近似する．

5.2 翻訳品質の評価尺度

翻訳の精度をどうやって評価するかは難しい．文の正しい翻訳は何通りも考えられるので，人手による翻訳との完全一致のような単純な方法は不適切である．人間による主観的な評価は，流暢さ（fluency）や適切さ（adequancy）などの様々な要素を総合的に判断するので最も信頼できるが時間もお金もかかる．近年，BLEU，NIST，METEORなど，低コストかつ高速に計算でき，人間による評価との相関が高い評価尺度が幾つか提案された．これらの自動評価尺度は，システムの開発と評価を短いサイクルで繰り返すことを可能にし，機械翻訳の研究開発に革新をもたらした．ここでは最も標準的な評価尺度であるBLEU（Bilingual Evaluation Understudy）[7]を紹介する．

BLEUは，機械による翻訳はプロの翻訳者による翻訳（参照訳，reference）に類似しているほど良いと考え，類似度を0から1の間の数値で表す．具体的には，システムが出力した1つの翻訳候補と正解集合（複数の参照訳）の間の異なる長さの単語ngramの適合率（precision）p_n の幾何平均に，短い文へのペナルティである BP を掛けたものである．

$$BLEU = BP \times \exp\left(\frac{1}{N}\sum_{n=1}^N \log p_n\right) \tag{9}$$

ここで単語ngramとは連続する n 個の単語列であり，単語ngramの適合率 p_n とは候補に含まれるすべての単語ngramのうち正解集合の含まれる単語ngramと一致したものの割合である．通常は $N=4$ を用いる．一般に単語unigramの適合率は適切さ（訳語の精度）に関連し，長い単語ngramの適合率は流暢さ（語順の精度）に関連する．また同義語や言い替えに対応するために参照訳は4つ以上が望ましいとされる．

単語ngramの適合率は，不確かな訳語をあえて出力するより省略した方が高い数値になるという問題点がある．そこで翻訳候補が参照訳に比べて短い場合に次式のような簡略化ペナルティ（brevity penalty）BP を与える．

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{if } c \le r \end{cases} \tag{10}$$

ここで，c は候補の単語長，r は正解の単語長である．なお，p_n の計算は文単位の単語ngramの一致数をテストデータ全体で集計するので，BLEUの値はテストデータ全体に対して与えられる．

5.3 最小誤り率学習

式(8)に基づくモデルパラメタの最尤推定の問題点は，尤度最大のパラメタが翻訳精度を最大にする保証がないことである．BLEUのような自動評価尺度が誕生したことにより，翻訳精度の評価尺度を直接最大化するパラメタ推定法である「最小誤り率学習」（Minimum Error Rate Training）[6]が考案され普及した．

参照訳 $\mathbf{r}$ に対する翻訳候補 $\mathbf{e}$ の誤りを評価する関数を $E(\mathbf{r}, \mathbf{e})$ とする．例えばBLEUを最適化する場合は $E=1-BLEU$ とすればよい．訓練データとして対訳コーパス $\{(\mathbf{e}_s, \mathbf{f}_s)|s=1, \cdots, S\}$ が与えられたとき，誤り最小学習では，訓練データにおける最適候補と参照訳の誤りの総和が最小になるようにモデルパラメタ λ_1^M を求める．

$$\hat{\lambda}_1^M = \arg\min_{\lambda_1^M} \sum_{s=1}^S E(\mathbf{e}_s, \arg\max_{\mathbf{e}} p_{\lambda_1^M}(\mathbf{e}|\mathbf{f}_s)) \tag{11}$$

ここで $\arg\max_{\mathbf{e}} p_{\lambda_1^M}(\mathbf{e}|\mathbf{f}_s)$ は，パラメタ λ_1^M のモデルで $\mathbf{f}_s$ を翻訳したときの確率最大の候補である．

式(11)の誤り関数は凸関数ではないので，勾配に基づく最適化法は使えない．そこで，まずランダムに選んだ λ_1^M から出発し，他のパラメタを固定して1つのパラメタ λ_m について最小化することを繰り返す．

複数の翻訳候補から確率最大の候補を選ぶ問題において，他のパラメタを固定して λ_m だけを変数とすると，ある文の対 $\mathbf{e}$ と $\mathbf{f}$ に対する $\log P(\mathbf{e}|\mathbf{f})$ は λ_m の1次式で表現できるので，ある文 $\mathbf{f}$ に対する複数の翻訳候補に対する確率の最大値は区分的に線形な λ_m の関数になる．確率最大の候補が変化する λ_m とその誤りを記憶し，これを訓練データ全体でマージすると，誤りの総和の最小値を求めることができる．

6. 構文に基づく翻訳

大局的な語句の並び替えをうまく扱うために，構文理論（syntactic theory），特に自然言語の階層構造を翻訳モデルの中で利用する「構文に基づく統計的機械翻訳」（syntax-based SMT）が2000年代前半から現在まで盛んに研究されている．これまでに言語学的な構文解析に基づくものから形式言語論的な木変換（tree transduction）に基づくものまで様々な翻訳モデルが提案されているが，ここでは現在最も精度が良いとされる「階層的句に基づく翻訳」（Hierarchical Phrase-Based Translation）[2]を紹介する．

階層的句に基づく翻訳は「同期文脈自由文法」（Synchronous Context Free Grammar, SCFG）に基づいている．同期文脈自由文法の規則は一般に次式で表される．

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle$$

ここでXは非終端記号，γとαは終端記号と非終端記号の列，$\sim$はγとαに含まれる非終端記号の間の1対1対応を表す．

階層的句に基づく翻訳規則は，ある句が他の句を含むことを許す．これにより非連続な句（変数を含む翻訳規則）や句の並び替え規則を表現できる．以下に例を示す．ここで枠付きの添字は$\sim$でリンクされていることを表す．

$$X \rightarrow X_{\boxed{1}} \text{である}, \text{ is } X_{\boxed{1}} \quad (12)$$

$$X \rightarrow X_{\boxed{1}} \text{の } X_{\boxed{2}}, X_{\boxed{2}} \text{ of } X_{\boxed{1}} \quad (13)$$

$$X \rightarrow \text{言語は}, \text{ language} \quad (14)$$

$$X \rightarrow \text{コミュニケーション}, \text{ communication} \quad (15)$$

$$X \rightarrow \text{道具}, \text{ a means} \quad (16)$$

最初の規則は変数を含む翻訳規則の例であり，二番目の規則は句の並び替えの例である．三番目以降の規則は前節で説明した句に基づく翻訳の対訳句と同じであり，階層的句に基づく翻訳は句に基づく翻訳の拡張になっていることが分かる．

同期文脈自由文法の導出（derivation）は，対応付けられた開始記号の対から始まり，各ステップにおいて1つの規則の右辺にある2つの要素を使って2つのリンクされた非終端記号を書き換える．一般にはXを開始記号としてもよいが，文献[2]では，以下の2つの接着規則（glue rule）を導入し，入力文を並び替えのないチャンク（chunk）の列に分割することを許し，頑健性を高めている．

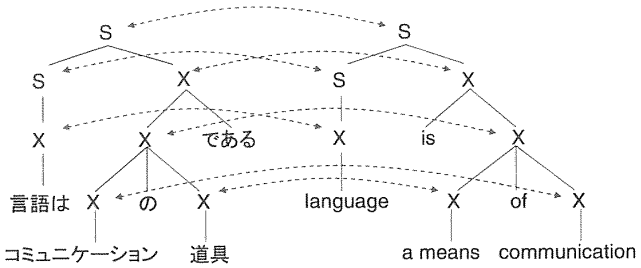

図4　階層的句に基づく対訳文の導出

$$S \rightarrow S_{\boxed{1}} X_{\boxed{2}}, S_{\boxed{1}} X_{\boxed{2}} \quad (17)$$

$$S \rightarrow X_{\boxed{1}}, X_{\boxed{1}} \quad (18)$$

図4にこれらの規則を用いた対訳文の導出の例を示す．

階層的な句は単語対応付けされた対訳テキストからの句の抽出を一般化することにより抽出する．まず句に基づく翻訳の場合と同様に，単語対応付けと矛盾しない句を抽出し，これを初期句（initial phrase）とする．次に句の内部に他の句を含む場合，含まれる句を非終端記号に置換する．例えば図2において（コミュニケーションの，of communication）という句は（コミュニケーション，communication）という句を含むことから，$X \rightarrow X_{\boxed{1}}$の，of $X_{\boxed{1}}$という規則を抽出できる．

上記の方法では非常に多くの規則が生成されるので，実際には，初期句の長さは最大10単語，非終端記号は最大2個，少なくとも1つの対応付けされた単語を含むなどの制約を加える．翻訳モデルの重みは，対数線形モデルの最小誤り学習により推定する．

デコーディングは，同期文脈自由文法の原言語側の規則を使って原言語の文を解析して原言語の構文木を作成し，これを目的言語の構文木に写像して終端記号を読み出すことにより目的言語の文を得る．各規則に含まれる非終端記号を最大2個までに制限しているので，解析にはCKYアルゴリズムが使える．

構文に基づく翻訳のデコーダの実装において最も悩ましいのは，言語モデルの重みを統合する方法である．句に基づく翻訳のデコーダでは，目的言語を文頭から文末方向へ連続する単語列として生成するため，単語ngramによる言語モデルを簡単に組み込むことができ，早期の枝刈り（ビーム探索）が可能となる．構文に基づく翻訳のデコーダでは，解析の途中段階において，目的言語を必ずしも連続した単語列として生成できないので言語モデルを組み込めず，効率的な探索が難しい．

文献[8]は，翻訳規則を抽出する際に，目的言語側

に右辺の先頭要素は必ず終端記号であるというGreibach標準形と同じ制約を加え，目的言語が文頭から文末方向へ連続する単語列として生成されるように原言語のearleyアルゴリズムによるトップダウン解析を制御することにより，言語モデルを簡単に適用できるようにする方法を提案した．文献[2]は，目的言語の言語モデルを組み込んだCKYアルゴリズムによる原言語の部分解析のk-best候補を効率良く計算するcube pruningを提案した．文献[3]は，cube pruningに遅延評価（lazy evaluation）を導入することによりさらに計算量を削減したcube growingを提案した．

7. おわりに

アラビア語と英語などの一部の言語対では統計的機械翻訳は既に実用化されており，従来手法より精度が高いといわれている．またIBM翻訳モデルを作成するGIZA++や句に基づく翻訳デコーダmoses[2]などオープンソースの統計的機械翻訳ツールが公開されており，対訳コーパスさえあれば，誰でも簡単に統計的機械翻訳という技術を体験できる．

統計的機械翻訳は，定跡よりも力任せの探索を重視することで世界チャンピオンと互角に戦うレベルに達したコンピュータチェスに似ているといわれる．この分野の研究者の一人として，コンピュータがプロの翻訳者を超える日を楽しみにしている．

参考文献

[1] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra and R. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2): 263-311, 1993.

[2] D. Chiang. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2): 201-228, 2007.

[3] L. Huang and D. Chiang. Forest rescoring: Faster decoding with integrated language models. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL-07)*, pp. 144-151, 2007.

[4] P. Koehn, F. J. Och and D. Marcu. Statistical phrase-based translation. In *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL-03)*, pp. 127-133, 2003.

[5] M. Nagata, K. Saito, K. Yamamoto and K. Ohashi. A clustered global phrase reordering model for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-06)*, pp. 713-720, 2006.

[6] F. J. Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 160-167, 2003.

[7] K. Papineni, S. Roukos, T. Ward and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Lnguistics (ACL-02)*, pp. 311-318, 2002.

[8] T. Watanabe, H. Tsukada and H. Isozaki. Left-to-right target generation for hierarchical phrase-based translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-06)*, pp. 777-784, 2006.

[9] 永田. 第II部確率モデルによる自然言語処理, 言語と心理の統計, pp. 59-128. 統計科学のフロンティア10. 岩波書店, 2003.

[2] http://www.statmt.org/moses/