中間言語を中心とした多言語医療音声認識翻訳システムの構築

中尾雪絵 [1] Manny Rayner[2] Pierrette Bouillon[2]
Beth Ann Hockey[3] 神崎享子 [4] 井佐原均 [4]

1) ナント大学　2) University of Geneva　3) University of Santa Cruz　4) NICT

yukie.nakao@univ-nantes.fr, Emmanuel.Rayner@issco.unige.ch,
Pierrette.Bouillon@issco.unige.ch, bahockey@ucsc.edu,
kanzaki@nict.go.jp, isahara@nict.go.jp

1. はじめに

本稿の目的は、中間言語と音声認識向けの文法ベース言語モデルの接点を求め、その新しい有益性を探ろうというものである。

多言語翻訳システムにおける中間言語は、異なる言語間の翻訳において中心的な機能を果たす中立的表現である。中間言語の長所については過去にも述べられているが(Arnold *et al.*, 1994)、最大の利点は N^2 問題を回避できることであろう。原言語から目標言語へ単純に移し変える翻訳では、移し変え規則を N^2 個記述しなければならない。しかし、中間言語を使えば、記述する規則は 2N 個で済む。

音声認識向け文法ベース言語モデルに関しては、多くの音声認識システムで音響モデルと言語モデルの 2 つが知識ソースとして使われている。音響モデルが音素と波形の連関を規定するのに対し、言語モデルは使用可能な語の配列の分布を規定する。言語モデルは、文法ベースシステムにおいて文法として定義づけられる。ここでいう「文法」の構築には、最低限の統語制約が必要だが、これだけでは充分とは言えない。モデル強化さらに音声認識の向上のためには、意味またはドメインによる制約を加える必要がある。単純な事例における標準的な解決策としては、文法に意味属性の制約を加えると、認識器の性能が上昇することがわかっている(Rayner *et al*., 2006, §11.4)。

意味属性の制約は、文法構造を大きく変えずに加えられるのが利点だが、複雑な制約の場合はやや困難である。一般に、音声認識システムでは、先立つ談話の文脈に依存した翻訳が有用だが、文脈から作られる発話に対する制約を文法ベースの言語モデルに含めることは容易ではない。言語処理技術の点から望ましいのは、言語モデルを 2 つまたはそれ以上に分割し、それぞれが独自の制約を持つことであろう。

本稿は上記 2 点の、多言語音声認識システムにおける融合を目標とする。一般的な中間言語の定義は既に多くの意味文法の特性を備えているが、我々はさらに一歩進め、中間言語＝意味文法ととらえたい。ここでいう文法は、可能な中間言語表現の範囲を定義づけるだけでなく、原言語認識器で使用される言語モデルに含めるのが厄介な「ドメインによる制約」のエンコードにも使われるものである。

我々は、これらの考案を医療音声翻訳プロジェクト Medslt へ適用した。Medslt(Bouillon *et al*, 2005)は、多言語オープンソースのドメイン限定医療音声認識翻訳システムで、2003 年よりジュネーヴ大学で開発が進められている。システムは、医師と患者が共通言語を有さない場合の医療会話の補助を目的とし、日本語を含む 6 ヶ国語に対応する。医療会話は安全性の面から、再現率より適合率が遥かに重要である。このため、我々は制御言語法を取り、基本設計は規則ベースとした。よく言われる脆弱さの問題には、ヘルプモジュールで対応した(Chatzichrisafis *et al.*, 2006)。

音声認識には NUANCE を使用している。ここに装備されているのは、オープンソース Regulus プラットフォームを用い(Rayner *et al*., 2006)、言語学的知見に基いた文法をコンパイルした、文法ベース言語モデルである。Regulus 設計の総合的な目標は、密接な関連のある大量の文法を記述・維持する手間を、文法間の内的一貫性を保って簡素化することである。中間言語ベースの翻訳設計を選んだのは、文法ベースの文法を採用

した理由と同様、システムを開発・維持しつつ、どの言語ペアでも良い性能を出すという課題に対応できるからである。

次節では、意味文法としての中間言語の定義がRegulus概念でどのように構造されるかを説明する。

2. 意味文法としての中間言語の定義

中間言語の定義に実行可能な適格さを求める理由は、1)中間言語を認識フィルターとして用い、2)翻訳規則の構築プロセスを簡略化するためである。2)に関しては、中間言語の定義がなくても、原言語→中間言語の翻訳規則(to-interlingua)と中間言語→目標言語の翻訳規則(from-interlingua)を組み合わせ、原言語から目標言語に翻訳することは可能である。しかし、この方法では 2 つの規則が共存し、エラーの際に原因がどちらの規則にあるのか特定が難しくなる。

From-interlingua 規則が効率的かつ完全であるためにはどうすればよいか、という点も考慮する必要がある。原言語から生成される中間言語文は類似した内容が多いが、同一ではない。このため、from-interlingua 規則をただ 1 つの原言語から構築しても、この規則が想定する文を、関連する他の原言語に使うことはできない。逆に、全ての原言語のテストセットを構築すると、中間言語は各言語で共有されているにも関わらず、テスト量が 2 倍になる。また、扱う全言語に対して言語処理技術者と目標言語話者(informant)を揃えなければならないという実際面での問題も起こる。システムの扱う言語が増えれば、問題はなお深刻になる。ソフトウェア工学の面から望ましいのは、翻訳規則の構築タスクをモジュール化し、to/from-interlingua 規則をそれぞれ別に構築することである。これを実現するには、中間言語用の読解が容易な注釈(gloss)が必要である。作業者は原言語→中間言語、中間言語→目標言語双方向の翻訳に取り組むことになるからである。Medslt は以下のような注釈を取り入れている。

```
s:[] --> vp:[] ka
vp:[] --> p:[] wo v:[subcat=transitive]
```

ここまで述べてきたような中間言語設計のための必要条件をまとめる際、我々が望むのは中間言語を標準的なフォーマリズムに基づき適格に定義すること、それにより構造の複雑な文を簡潔に表現したり、文構造が簡単に解読できる表層形式を構成することである。これらの必要条件は文法定義に近く、中間言語の設計＝文法とも言えよう。「文法」とは何か。チョムスキーに遡る伝統的な考え方では、多くの文法はもともと表層的な言語の定義づけに使われ、表層の言語は付随的に意味と結びつく。順序を逆にすると、我々の関心は文法を用いて意味構造を定義づけ、付随的にこれらの構造を表層の言語へと関連づけることである。即ち、中間言語の設計＝意味文法とも言える。またほとんどの文法の関心は「構文解析」にある。表面的な記号列は文中で正しく形成されているか、もし正しければ統語構造や付随する意味構造はどうか定義づける必要がある。特に「生成」について、中間言語の意味形式（原言語からの翻訳）は中間言語レベルで適切な意味構造を持っているか、もし持つとすれば付随する意味構造や表層的な形式はどうか、定義づける必要がある。

意味文法の使用に関わる主な問題は、意味内容を記述せずに、表面的な統語制約をコード化してしまうことである。統語の自立性は、制約がドメイン特定の意味概念とは独立的に取り込まれるのが最適である。筆者らの扱う中間言語設計文法は、それとは逆に、簡潔な人工統語と組み合わせてできる真の意味での意味文法と言えるだろう。

```
np:[sem= @np_d_nbar_sem(Det, N), agr=3, agr=Agr, wh=Wh,
      sem_n_type=Type, conj=n, gapsin=GIn, gapsout=GIn,
pronoun=n,
      @takes_pps(PPs)] -->
   d:[sem=Det, agr=Agr, wh=Wh],
   noun:[sem=N, agr=Agr, sem_n_type=Type, @takes_pps(PPs)].

location:[sem=concat(BodyPart, Part, Side)] -->
   body_part:[sem=BodyPart],
   ?part:[sem=Part],
   ?side:[sem=Side].
```

図 1　英語ならびに中間言語文法の規則例

なお、意味文法の規則と、原・目標言語を定義づける文法規則との間には重要な違いがある。原言語の表現は、対象言語の文法から作られる。これは言語学的なベースから成る複雑な特性文法であり、非終端記号・特性には S, NP, agreement, gapping 等、標準的な言

語学概念を用いている。構造は簡素化され、非終端記号にはSYMPTOM, LOCATION, BODY-PART 等の概念をあてはめる。図 1 の例では、英文法の規則で NP 構造を定義づけている。LOCATION 表現は、例えば BODY-PART(head) と任意の選択肢 PART(front)、SIDE(left)と連結して成立する。

3. Medslt の中間言語

文法ベースの中間言語が有益なのは to-interlingua 規則と from-interlingua 規則を別々に構築することで、翻訳規則の構築をモジュール化できることである。Medslt の from-interlingua 規則は、全ての中間言語表現を含む 1 つのコーパスから構築されている。コーパスに含まれる表現は、取り扱う全ての原言語コーパス→中間言語の翻訳から作られる。例えば中間言語表現 'YN-QUESTION persistent pain be PRESENT ACTIVE'に対し、原言語コーパスは *is it a persistent pain, is the pain persistent* (英), *la douleur est elle persistente* (仏), *itami wa shitsukoi desu ka, shitsukoi itami desu ka* (日), *el dolor es persistente* (西), 目標言語コーパスは *is the pain persistent* (英), *la douleur est-elle persistente* (仏), *shitsukoi itami desu ka* (日), *el dolor es persistente* (西)などの文を持つ。

中間言語は、形式的にはシステムが扱う自然言語と同じ要領で規定される言語であり、中間言語を扱うツールの多くが標準文法仕様のユーティリティまたは直接的な使用を目指している。中でも中間言語表現の妥当性を確かめるツールは、Regulus 生成器・コンパイラを中間言語文法に応用して作られている。中間言語は自然言語文法に比べれば遥かに単純な文法であり、中間言語表現からの生成が可能である。これを利用して、設計された単純かつ強力なデバッグツールは、不適格な中間言語表現 T に対し、新たな情報を追加または消去した、的確なヴァリアント表現 T^1を抽出できる。To/from 中間言語翻訳の回帰テストは標準的な Regulus の回帰テストツールで行う(Rayner *et al*., 2006, 6 章)。

中間言語のもう 1 つの機能は、言語モデルコンポーネントの役割を果たしていることである。Medslt は N-best リスコアリングを用い、原言語の音声認識において異なる N 個の認識仮説を展開する。N-best リストから仮説を展開する際には、認識器と中間言語から展開された知識を信頼度ソースに使い、信頼度を元にランクづけされた認識仮説の中から、適切な中間言語表現を作る仮説を選ぶ。これだけで音声認識の性能は向上する。1-best 処理であれば、不適切な中間言語により翻訳文を得ない場合でも、N-best 処理は N 個の仮説文から適切なものを選ぶことで結果を出すことができる。図 2 は N-best 処理が性能を向上する典型例である。

実際の発話:
"ni jikan ijou itami wa tsuzuki masu ka"
N-best 仮説文:
1: "nijikai zutsu de tsuzuki masu ka"
2: "ni jikan ijou sex de tsuzuki masu ka"
3: "ni jikan ijou iki de tsuzuki masu ka"
4: "ni jikan ijou itsu tsuzuki masu ka"
5: "ni jikan ijou totsuzen tsuzuki masu ka"
6: "ni jikan ijou itami ga tsuzuki masu ka"

図 2　N-best 認識と中間言語フィルタリングの組み合わせが、どのように音声認識性能を向上させるかを示した日本語例。6 つの仮説文はいずれも文法の想定文だが、6 つ目の文のみが適切な中間言語を作る。

着想は単純だが、利点は多い。特に言語モデルのドメイン限定による意味制約は、特定言語から独立した中間言語を可能にするため、システムが扱う全言語の認識への適用が容易となる。次節で紹介する評価結果の 1 つは、中間言語の知識ソースを加えた音声認識の性能が大きく向上したことを証明している。

4. 評価と結び

1 つ目の評価として、文法ベースの中間言語の定義により、構築処理の効率が上昇したかどうかを調べた。評価に先立ち我々は、仮に訳文を原言語→中間言語、中間言語→目標言語の二部に分けても、原言語→目標言語を直接評価したものと同じ結果が出るだろうかという点に着目した。

評価の内容は次のようなものである。まずはシステム内で構築があまり進んでいない言語ペア、日本語→フランス語の翻訳の「直接評価」を行った。この結果を「内在的評価」、すなわちフランス語→中間言語の組み合わせ（本来は仏語→英語、仏語→アラビア語の文脈で構築）並びに中間言語→日本語（本来は英語→日

本語の文脈で構築）の結果と比較することにした。まず、フランス語→日本語をオフラインモードで実行し、フランス語 507 文に対する日本語訳を得た。その後、システム作業者が、関連する仏語→日本語、仏語→中間言語、中間言語→中間言語、仏語翻訳それぞれに対し評価を行った。過去に行っている評価と同様、評価は good/ok (完全ではないが許容できるもの)/bad の 3 段階である。評価の後、直接評価と内在的評価を組み合わせて比較したところ、496 文のうち、81 文が 2 種類の評価間で異なる結果を得た。この 81 文のうち、79 文の異評価の原因は good と ok の違いによるものであった。概して good と ok は評価者の主観によって評価が分かれやすい部分である。いっぽう、直接評価が good であるのに、内在的評価が bad となっている例もあった。似た例文を検討したところ、これらの文に対する適切な評価は bad と ok の中間あたりに位置すると考えられる。全体として、内在的評価は、ほぼ直接評価と似た結果を得たと考えられる。

2 つ目の評価は、知識ソースとして中間言語を使用することで、音声認識が向上するかを調べたものである。我々はオフラインモードでフランス語(Chatzichrisafis *et al*., 2006)、英語(Rayner *et al*., 2005a)、日本語(Rayner *et al*., 2005b)の音声データを集めて書きおこし、適切な中間言語表現であるかどうかを人手で評価した。

言語	Subset	#発話	#不適切な中間言語		向上	
			N-best-	N-best+	Abs.	Rel.
仏	全て	2130	30.5%	27.6%	2.9%	**9.5%**
英	全て	870	42.0%	38.0%	4.0%	**9.4%**
日	全て	544	39.2%	37.7%	3.5%	**8.9%**
仏	想定内	1583	12.7%	9.7%	3.0%	**23.6%**
英	想定内	515	11.2%	9.1%	2.0%	**19.0%**
日	想定内	331	10.6%	7.8%	2.8%	**26.5%**

表 1　仏英日 3 言語を対象に、中間言語を知識ソースに用いた N-best リスコアリングを加えた音声認識性能の向上を示す。Subset は「全て」または「想定内」の文、「#発話」は処理された発話数、「#不適切な中間言語」は正しくない中間言語の例分数と N-best の使用・不使用別の結果、「向上」は正しい訳文を得た発話の割合の絶対的(Abs.)・比較的(Rel.)な減少率を表す。

結果は表 1 の通りである。文法ベースの音声アプリケーションには 1)発話すべて、2)文法適用内の発話のみの 2 つの場合を設けた。なお、Medslt にはユーザに文法想定内の発話を促すヘルプシステムがあり、Subset の性能はユーザの直感的な評価を示したものと考えられる。興味深いのは、各言語の結果の類似性であろう。どの言語とも、中間言語を用いた N-best リスコアリングを加えると意味エラー率が約 9%、想定内データでは約 20%も下がっている。

全体として、これら 2 つの評価結果は満足のいくものであった。少なくとも現段階で、我々の考える中間言語構造が構築と評価プロセスを大幅に簡略化しつつも、音声認識の性能の水準を良いものに保っていることが裏付けられたかと思う。今後、本稿に述べた方針に従い、システムの開発を進めていきたいと考えている。中でも、ユーザビリティ向上を目指した中間言語の注釈機能を明確化することが当面の目標である。

参考文献

Arnold D. *et al*. (1994). *Machine Translation: An Introductory Guide*, Oxford : Blackwell.

Bouillon P. *et al*. (2005). A Generic Multi-Lingual Open Source Platform for Limited-Domain Medical Speech Translation. In *Proceedings of the 10th Conference of the European Association for Machine Translation (EAMT)*, p-50-58, Budapest, Hungary.

Chatzichrisafis N. *et al.* (2006). Evaluating Task Performance for a Unidirectional Controlled Language Medical Speech Translation System. In *Proceedings of the HLT-NAACL International Workshop on Medical Speech Translation*, p. 9-16, New York.

Rayner M. *et al*. (2005a). A Methodology for Comparing Grammar-Based and Robust Approaches to Speech Understanding. In *Proceedings of the 9th International Conference on Spoken Language Processing (ICSLP)*, p. 1103-1107, Lisboa, Portugal.

Rayner M. *et al*. (2005b), Japanese Speech Understanding Using Grammar Specialization. In *HLT-NAACL 2005: Demo Session*, Vancouver, British Columbia, Canada: Association for Computational Linguistics.

Rayner M. *et al*. (2006). *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler.* Chicago: CSLI Press.