

4 同時通訳の工学と科学

一次世代自動通訳技術の実現に向けて—

松原 茂樹
(名古屋大学情報連携基盤センター)

自動通訳技術が進化し、同時通訳機の登場も遠い未来の話ではなくなりつつある。本稿では、同時通訳技術に関する研究開発の一端を紹介し、実用化に向けた技術的課題について論じる。まず、同時通訳の要素技術として、音声言語の同時的な解析ならびに生成技術について解説し、それらを基盤としたシステム開発の試みを紹介する。続いて、プロの同時通訳者の音声を大規模に収録したコーパスとして、同時通訳データベースについて概説する。また、通訳データの観察により明らかになった、訳出方略上の知見をいくつか紹介する。

同時通訳と逐次通訳

同時通訳機は、人が話し始めるとそれに追従しながら通訳するという、まさに夢の機械である。同時通訳は、長年にわたって情報処理技術の究極的な存在であり続けてきた。新聞や雑誌で「技術予測」なるものが特集されれば、予測対象として取り上げられ、実用化時期の見通しが提示されてきた。同時通訳機の登場は、英語を母語としない日本人の切実な願いである。そしてその思いは、世界のグローバル化が進むにつれ、ますます膨らんでいる。ただし、これまで示されてきた予測の当り外れはともかく、2008年6月現在、同時通訳機が実用化されるまでには至っていない。本稿では、同時通訳機の実現に向けた研究開発の一端を紹介し、今後の技術的課題について論じたい。

通訳はその様式により、**逐次通訳** (consecutive interpretation) と **同時通訳** (simultaneous interpretation) に大別できる。逐次通訳は、話し手が一区切り発言し終わった後で訳出を開始するものであり、同時通訳は、話し手が発言し終わる前に訳出を開始し、発言と並行して訳出を進めるものである。このうち同時通訳は、人間にとって最も高度な言語活動であるとされており、能力の養成には長期間にわたる訓練を必要とする。同時通訳の難しさは、「聴く」「話す」という2つの行為を並行して行わなければならないこと、また、続々と入ってくる話し手の発言内容を記憶し保持しなければならないこと、

にある。しかし、人間にとって大きな困難が伴う並行処理や短期記憶は、計算機にとってはむしろ困難性の低い部分であり、同時通訳機の技術的な難しさの一端は、「聴きながら理解する」「生成しながら話す」という、人間にとって当然の振舞いを実現することにある。

さて、本特集の記事でも論じられているように、自動通訳技術の研究開発は近年急速に発展してきた。しかし、これまでに開発された通訳機は、いずれも逐次通訳の形態を採用しており、同時通訳機能を備えたシステムが開発された例は世界中を見渡してもほとんどない。これは、音声処理や言語処理、機械翻訳など、自動通訳の要素技術の多くが、1文の処理を前提に開発されてきたという経緯による。旅行会話などで逐行される短い文を対象とするのであれば、あえて同時通訳機能を搭載してもそのメリットは小さい。開発コストを考えれば、同時通訳を採用するという選択肢はこれまでは存在しなかった。

しかし、講演の通訳では同時通訳が不可欠である。また、インタビューや会議などの会話では長い文が頻出するため、同時通訳を導入するメリットは大きくなる。自動通訳の技術開発が実用化の段階を迎え、通訳機の新たな進化の方向性を検討する時期にある。今後は、同時通訳機の実現に向けた研究開発の動きも活性化していくものと予想される。

本稿の以下では、まず、同時通訳への工学的アプローチを概観し、続いて、実データの収集と分析による科学的アプローチを紹介する。

This flight leaves Tokyo

図-1 文(1)の断片 "This flight leaves Tokyo" の構造

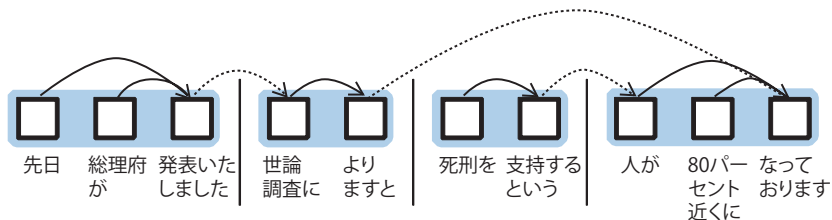


図-2 節と文構造の関係

同時通訳機の要素技術

自動通訳技術を評価するとき、話し手が発した言語(以下、**起点言語** (source language))での表現が、聞き手が受け取る言語(以下、**目標言語** (target language))においてどう表現されるかという、いわゆる how-to-say が問題となる。一方、同時通訳においては、how-to-say に加え、目標言語がいつ訳出されるかという、いわば when-to-say も重要となる。これは、訳文の正しさと訳出の同時性の双方を備えた通訳方式が必要となることを意味している。

自動通訳機は、通常、音声・言語処理におけるいくつかの要素技術の組合せにより実現され、そのパフォーマンスは要素技術の性能に強く依存する。この節では、特に、言語の解析技術と生成技術を取り上げ、同時通訳に適した言語処理方式について論じる。

* 同時的な解析技術

自動通訳における最初のステップは、起点言語の内容を正確に把握することであり、解析処理により言語的な構造を明らかにすることができる。この技術は、すでに音声・言語処理の分野で実現されているが、いずれも文を解析単位としており、同時通訳には適さない。同時通訳のためには、起点言語文の入力途中の段階でそれまでの入力に対する解析の結果を生成する必要がある、入力が進むにつれて順次結果を提示する技術、すなわち、同時的な言語解析技術が検討されてきた。

同時性を重視した解析

同時的な言語解析として考えられる方法の1つは、文の先頭から単語 (word) が入力されるごとに処理を実行するというものである。このような方法で、

(1) This flight leaves Tokyo for Bangkok at nine.

を同時的に解析するとは、たとえば "Tokyo" が入力された段階で、名詞 "flight" が動詞 "leaves" の主語であるとか、"Tokyo" は目的語であるなどといったことを認識するということである。これにより、それまでの部分に対して日本語訳「このフライトは東京を出発します」を(この時点で音声出力するかどうかはともかくとして)作り出すことができる。このような単語と単語の関係^{☆1}は**依存関係**(または、修飾・被修飾関係)と呼ばれる。単語 *a* が単語 *b* に依存する(単語 *b* を修飾する)ことは *a* から *b* への矢印で表現され、文の構造は依存関係にある単語対の集合で表現される。"This flight leaves Tokyo" という断片の構造は、**図-1**のように示される。

依存関係が付与された文脈自由文法を用いて、そのような構造を同時進行的に生成するアルゴリズムが提案されている¹⁾。単語が入力されるたびに解析処理を実行する方式であり、高い同時性を達成している。

精度を重視した解析

一般に、解析の同時性と精度はトレードオフ関係にある。同時性の高い解析では参照できる情報が十分になく、それが解析誤りの原因となる。一方で、文よりも小さく、単語よりも大きな言語単位を解析処理の単位として採用することが考えられる。節 (clause) は、述語を中心とする文法的にまとまった単位であり、上述のような単位として有力である。たとえば、

(2) 先日総理府が発表いたしました世論調査によりますと死刑を支持するという人が80パーセント近くになっております。

は4つの節「先日総理府が発表いたしました」「世論調査によりますと」「死刑を支持するという人が」「80パーセント近くになっております」から構成されている。この文の構造を**図-2**に示す。節を用いることの利点は、

^{☆1} あるいは、句 (phrase) や文節などの文構成素間の関係。



図-3 日本語文(4)の構造

文の構造が節の中で閉じやすいことにある。ここで、構造が閉じているとは、節の最終文節を除くすべての文節がその節の中の文節に依存することを意味する。たとえば、「先日総理府が発表いたしました」は、最終の文節「発表いたしました」以外の文節はその節内の文節に依存している。このような性質が存在すれば、節の解析をそれ以降の入力とは独立に実行することができる。

日本語で話された講演音声、節単位で解析する手法が提案されており³⁾、高い精度を備えた同時的な言語解析が可能となっている。このような解析手法は、節が入力されるごとにその訳を生成し出力するような同時通訳機において利用できる。

* 同時的な生成技術

通訳機の同時性は、訳出タイミングによって評価できる。入力に対して同時に訳を生成し、それをできる限り早く出力することがポイントとなる。しかし、英語と日本語のように構造的な違いが大きい言語間の場合、同時進行的に生成することは難しい。このような問題に対して、個々の言語が持つ性質を積極的に利用することにより、通訳の同時性を高める生成方法が開発されている⁵⁾。

この方法の特徴は、英語会話文に対して生成する日本語訳を工夫している点にある。たとえば、

(3) You can pick your ticket up at the counter today.
の標準的な訳文は、

(4) 今日カウンターでチケットをお受け取りいただけます。

である。しかし、文頭の「今日」に対応する英語表現 "today" が文末に生起しているため、英語文の入力が完了するまで日本語訳文の出力を開始することは原理的にできない。これは、この例文に固有の現象ではなく、英語と日本語の語順の違いにより生じる、いわば構造的な問題である。それに対して、出力する訳文を

(5) チケットをカウンターで今日お受け取りいただけます。

とすれば、"your ticket" まで入力された段階で「チケットを」を訳出できる可能性があり、通訳の同時性が高まる。このアイディアは、語順の自由度が高いという日本語の性質を利用するものである。このような技法は、同

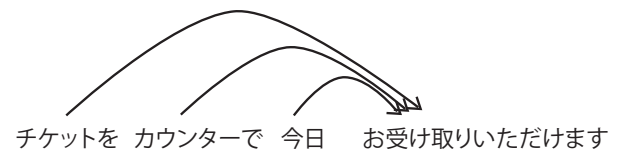


図-4 日本語文(5)の構造

時通訳者も活用している。事実、上述の訳文(5)は実際のプロの同時通訳者による実例である。通訳機においても、このような技術を駆使することにより、入力音声に追従した訳出が可能となる。

この技術は以下の方法で実現することができる。標準的な訳文(4)の依存構造を図-3に示す。日本語の場合、依存関係を表す矢印の方向は左から右であり、その制約に逸脱すると不自然で意味が通じない文となる。逆にいえば、その方向を遵守すれば、たとえ語順を変更したとしても、多くの場合、容認可能な訳文となる。実際、訳文(5)の構造は、図-4に示す通り上記の制約を満たしている。

訳出タイミングの制御についても同様の考え方により実現できる。すなわち、訳文の構造において、ある文節に依存する文節が存在しなければ、その文節は出力可能であるといえる。"your ticket" が入力された段階で、「チケットを」に依存する文節が存在しないことが英語文の解析結果から判明すれば、その時点で「チケットを」を出力する。

同時通訳機を実現する試み

前章で紹介した要素技術を背景に、同時通訳機能を備えた機械翻訳システムを開発する試みが筆者らによって進められている。ここでは、開発中の英日翻訳システム LINAS について簡単に紹介する⁵⁾。

LINAS は、英語の音声言語文を日本語文に同時に変換する。規則に基づく解析・変換・生成の3つのフェーズから構成されており、それぞれ入力に対して同時に処理を実行する。このうち、解析には前節で紹介した同時性を重視した解析方式¹⁾を、また、生成には同時的な生成方式を、それぞれ用いている。変換では、英語文の断片的な構造をもとに日本語の構造を漸次的に作り上げる。英語文における単語の語順と依存関係を考慮し、訳出可能な日本語表現が検出されれば随時出力する。入力が十分でない段階で出力を確定するため、入力が進むにつれてすでに出力した訳文に誤りがあることが判明することがあるが、システムは倒置や言い直しを生成する機能を備えており、聴き手が正しく理解できる訳文へと修

談話タグ	発話ID	発話開始時間	発話終了時間
0001 - 00:05:264-00:09:399 N: The theme for this speech is going to be the American	0001 - 00:06:440-00:08:207 I: (F え)次のテーマですが		
0002 - 00:09:840-00:11:032 N: Presidential debate	0002 - 00:08:944-00:09:783 I: アメリカの		
0003 - 00:11:424-00:13:391 N: and who would be the	0003 - 00:10:296-00:12:775 I: (F え)大統領に関するディベート		
0004 - 00:13:640-00:15:215 N: better president for America<SB>	0004 - 00:13:096-00:14:424 I: そして誰が		
0005 - 00:16:272-00:18:327 N: (F um) Let's see, today is	0005 - 00:14:648-00:18:255 I: より良い大統領とアメリカのために		
0006 - 00:18:640-00:20:400 N: December fifteenth	0006 - 00:18:728-00:19:263 I: なり得るかということですが		
0007 - 00:20:696-00:24:407 N: and it's been about a month	0007 - 00:19:528-00:21:887 I: 今日が		
0008 - 00:22:472-00:24:711 I: and a half since	0008 - 00:22:472-00:24:711 I: 十二月の十五日ですので		
	そして(F まあ)一ヶ月半ほど		

図-5 文字化データ(左が英語話者発話, 右が英日通訳者発話)

復することができる。

LINAS は、現状では、規則、語彙ともに実用規模には満たないものの、これまでの通訳機・翻訳機にはみられなかった、同時通訳を指向した言語技術をいくつか搭載している。

同時通訳データベース

同時通訳は人間にとっても究極的な言語活動であり、実際のプロの同時通訳者の振舞いやノウハウを詳細に分析し、得られた知見を通訳機の開発に活用することは効果的である。しかし、同時通訳のデータ収集には多大なコストを要するという事情もあり、研究に利用可能なデータは、これまで少量のサンプルが存在するにすぎなかった。

このような認識のもと名古屋大学では、同時通訳データベースの構築を進めてきた^{☆2}。このデータベースの特徴として、

- 第一線で活躍するプロの同時通訳者による実音声を収集している。
 - 約 60 万語相当の通訳者音声を収録しており、同時通訳データとしては世界最大規模である。
 - 同一の講演に対して複数の通訳音声を収録しており、通訳者による訳出方略の違いを観察できる。
 - すべての発話に開始・終了時刻を付与しており、起点言語と目標言語の間での時間的関係を見ることができる。
- などを挙げることができる²⁾。以下では、データベースの内容について紹介する。

☆2 名古屋大学統合音響情報研究拠点 (CIAIR)(1999～2003)における音声データベース整備の一環として実施された。

* 通訳音声の収録

講演通訳、会話通訳の双方を収録の対象とした。これは、社会の通訳需要の多くは講演通訳にあること、現状の通訳機の主要ターゲットが会話通訳であることによる。ただし、対象言語は、英語および日本語のみとしている。

講演音声は、政治や経済、環境など社会性のあるテーマを選定し、模擬講演の通訳環境を設けた。同時通訳者は通訳用ブースに入り、ヘッドホンから流れる話者音声に対して講演者の振舞いを見ながら通訳作業を遂行する。なお、同時通訳は豊富な訓練に基づく職人的技術であり、その方法論は通訳者によって異なる。そこで、同一講演に対して経験年数が異なる複数の同時通訳者(最大 4 名)を設定している。

一方、会話音声としては海外旅行をドメインとし、空港やホテルでの異言語間会話を模擬的に収録している。収録の事前設定は話者役割と会話タスクのみであり、できる限り自由な会話音声の収集に努めた。また、通訳の品質を確保するため、英日、日英それぞれに通訳者を用意した。

すべて同一のスタンドマイクを使用し、サンプリング周波数 16kHz、16 ビットでデジタル化し、複数チャネル環境で収録している。

* 通訳音声の文字化と視覚化

収集した音声データの文字化は、すべて人手で実施した。講演における文字化データのサンプルを図-5 に示す。データの言語学的分析として、フィラー(「えーと」「あのー」など)、言い淀みなど、話し言葉の特徴的現象にタグを与えている。文字化データの規模は、収録データ全体で約 100 万語に達しており、そのうち、日英通訳者の発話は約 22 万語、英日通訳者の発話は約 38 万語である。

同時通訳では、通訳機がどのような内容の音声をどのようなタイミングで出力するのが重要となる。通訳者の発声タイミングを明らかにするために、話者および通訳者の発話をポーズで分割し、その開始時間と終了時間を記録するとともに、話者発話と通訳者発話との間の発話レベルでの対訳対応を与えている。

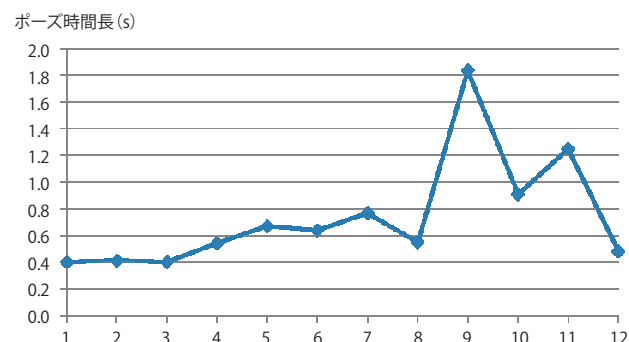
これらの文字化データは、その発声タイミングによって視覚化している。会話音声の視覚化データのサンプルを図-6 に示す。発話に対する発声時間が帯で記されており、グラフの左から、英語話者、英日通訳者の発声を示している。視覚化により、話者と通訳者の発声の重なり具合を観察できる。

* データの利用

音声データおよび文字化データは、名古屋大学同時通



図-6 通訳音声の視覚化

図-7 通訳音声の聴きやすさと平均ポーズ長の関係
(12個の通訳事例を評価の高い順に整列)

	発話頭	発話中	発話末
相関係数	0.14	-0.49	0.24

表-1 フィラーの出現回数と聴きやすさの関係

訳データベースとして、国内外の研究機関に広く配布されている^{☆3}。これまでのところ、システム開発や翻訳実験などの工学的利用のほか、通訳プロセスの解明を目的とした通訳学の分野でも利用されている。

同時通訳データからの知見

同時通訳データベースを使用することにより、通訳プロセスを定量的に分析することが可能となった。これまで通訳者の経験上でしか語られなかった数々の現象や方略のいくつかを、科学的に検証し、解明することができる。以下では、データ分析から得られた知見についていくつか紹介する。

* 同時通訳における聴きやすさ

同時通訳は、話し手の発話に追従しながらその訳出を遂行するものであり、通訳機の出力は、話者の話し方やシステムの処理状況に大きく依存する。このため、同時通訳機が、通常の音声言語システムのように、一定のリズムでテンポよく音声出力することは難しく、聞き手にとっては聴きにくくなる可能性がある。では、聴きやすい通訳音声を出力するために、同時通訳機はどのような方略を用いるべきだろうか。

この問題に対してプロの同時通訳者は、発話速度を適切に調整したり、ポーズやフィラーを挿入することにより、訳出プロセスを制御している。ここでは、ポーズとフィラーに注目し、聴きやすい通訳音声の特徴について

論じる⁶⁾。

聴きやすさの評価を被験者実験により実施した。同時通訳データベースの英日講演通訳データから通訳者の異なる12事例を用いた。被験者は日本語を母語とする31名であり、講演開始後の5～6分までの60秒間の音声を聴取し、聴きやすさを5段階で評価した。

ポーズと聴きやすさ

ポーズを200msec以上の無音区間と定義し、ポーズの長さや聴きやすさとの関係を調べた。図-7に、高評価順に並べた通訳事例の平均ポーズ長を示す。ポーズ長と評価値との相関は少なくなく、図からも、平均ポーズ長が長い通訳事例が低く評価されていることが分かる。また、通訳音声におけるポーズ長のばらつき具合についても、ポーズ長が一定している通訳音声ほど聴きやすいという現象が観察されている。

フィラーと聴きやすさ

通訳音声において長いポーズを回避するために、同時通訳者が用いる方略としてフィラーの使用が挙げられる。そこで、「えー」「あのー」など12種類のフィラーを対象に、出現位置ごとの頻度と聴きやすさとの関係を調べた。通訳発話における出現位置を発話の先頭、途中、最後の3種類に区分した。表-1に出現位置別のフィラー出現回数と聴きやすさの評価との順位相関を示す。発話の先頭と最後のフィラーについては影響は認められないものの、発話中のフィラーの頻度と聴きやすさの評価との間には関係性が観察された。一般に、同時通訳音声におけるフィラーは、通常の講演音声と比べて、発話途中に出現する割合が高いことが分かっており(図-8)、その割合を

☆3 <http://slp.el.itc.nagoya-u.ac.jp/sidb/>

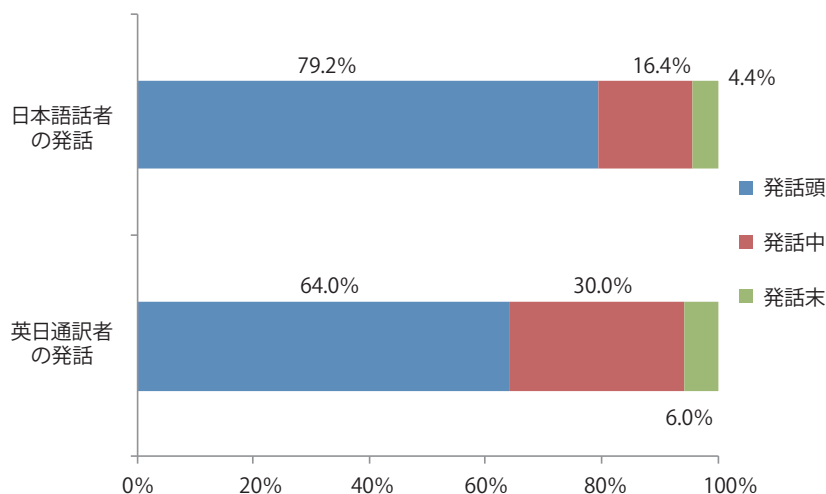


図-8 フィラーの出現割合の比較(位置別)

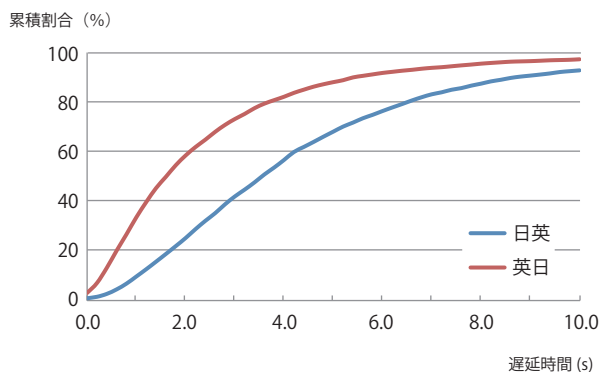


図-9 日英・英日同時通訳の訳出遅延時間

少なくすることが聴きやすい通訳音声を出力するためのポイントとなる。

* 同時通訳における訳出の遅れ

同時通訳において「訳す」という行為は「聴く」という行為が前提にある。しかるに、同時通訳機といえども、その出力は話し手の発声に遅れることになる。では、同時通訳機はどの程度の遅れで訳出すればよいのであろうか。

この問いに対して、同時通訳者による訳出の遅れを参考にすることができる。もちろん、人間と機械が必ずしも同一の性能を要求されるわけではないが、品質と同時性を兼ね備えた通訳を達成するために、プロの通訳者のパフォーマンスは1つの目安になる。

同時通訳データベースの講演データを使用し、対訳関係にある単語間の訳出遅延時間、すなわち、単語の発声終了時刻とその対訳語の発声開始時刻との差を測定した⁴⁾。通訳の方向に関する影響を考慮し、約8時間の日英通訳データから4,468組の対訳語、ならびに、約3時間の英日通訳データから2,629組の対訳語を抽出し使用した。

通訳の方向と訳出遅延

遅延時間の分布として、日英通訳、英日通訳それぞれにおける訳出遅延時間の累積割合を図-9に示す。平均遅延時間は、英日通訳で2.27秒、日英通訳で4.69秒であり、日英通訳の方が訳出の遅れが大きい。実際、訳出の遅延が4秒以内となる単語が英日通訳では全体の8割以上を占めるのに対して、日英通訳では5割程度にとどまっており、その差は著しい。

よく指摘されるように、日本語では一般に主動詞が文末に出現するため、日英通訳では文全体の構造の把握が遅れ、結果的に訳出が遅くなる。この結果は、日本語のようなSOV構造の言語を起点とする場合に、訳出遅延が大きくなることを示している。

単語の品詞と訳出遅延

文構造が異なる言語間の通訳で、その方向性によって遅延時間が変化するということは、単語の種類によって訳出遅延の様相が異なる可能性がある。そこで、起点言語の単語の品詞(ここでは、動詞と名詞)と訳出遅延時間の関係を調べた。紙面の都合上、ここでは英日通訳における遅延時間について述べる。

名詞1,522語、動詞376語の訳出遅延時間の累積割合を図-10に示す。訳出遅延時間の平均は、名詞で1.98秒、動詞で3.97秒であり、動詞の訳出遅延時間が名詞を大幅に上回るという結果になった。この理由として、英語では動詞が早い段階で出現するのに対して、日本語では文末に現れること、また、英語では名詞の文法役割の多くを語の出現位置によって表現するのに対して、日本語では語の生起順序の制約をそれほど受けないことが挙げられる。なお、本稿で紹介した生成技術の妥当性は、このような知見によっても裏付けられる。

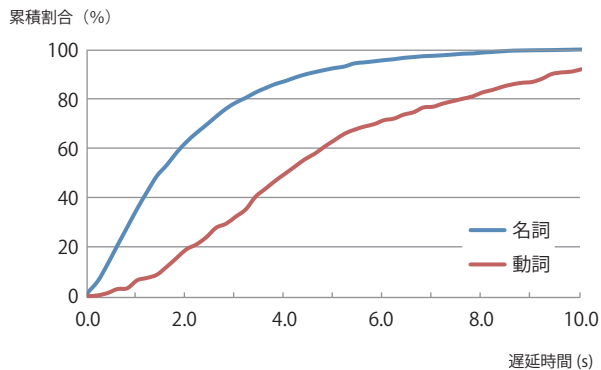


図-10 英日同時通訳における名詞と動詞の訳出遅延時間

将来の課題

本稿では、次世代自動通訳機の実現に向けた研究開発として、同時通訳に関する工学的ならびに科学的成果をいくつか紹介した。しかし、同時通訳技術の実用化に至るまでには、多くの技術的課題が残されている。

まず、本稿では触れなかったが、音声処理技術の問題がある。現状の音声認識は、ポーズなどを手がかりに検出した音声区間に対して、最尤の文字列を出力する方法が一般的である。同時通訳に利用するためには、同時進行的に音声入力を処理し、随時、認識結果を出力する仕組みが必要である。また、音声合成においても、システムの処理状況を考慮して動的に話速を制御するなどの仕組みが求められる。

また、同時通訳の処理単位として、本稿では、「単語」と「節」を用いるアイデアについて述べたが、同時通訳者は状況に応じて通訳単位を動的に設定しており、適切な通訳単位についてはさらなる検討が必要である。また、

同時通訳機の実用化に際しては、訳文の品質よりもむしろ、入力音声への追従性の方が強く求められる可能性がある。その場合、同時通訳者が行っているように、断片的に訳出された個々の表現をどのようにうまく繋いでいくかが重要になる。このような通訳テクニックの全貌については十分に解明されておらず、今後も通訳データの詳細な観察とノウハウの蓄積を継続することが必要である。

冒頭でも述べたように、同時通訳は、音声・言語技術の象徴的応用である。同時通訳機の登場はまだ先のことであるが、実用化の時期は着実に近づいている。

参考文献

- 1) 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).
- 2) 松原茂樹, 相澤靖之, 河口信夫, 外山勝彦, 稲垣康善: 同時通訳コーパスの設計と構築, 通訳研究, No.1, pp.85-102 (2001).
- 3) 大野誠寛, 松原茂樹, 柏岡秀紀, 加藤直人, 稲垣康善: 節境界に基づく独話の漸進的係り受け解析, 電子情報通信学会論文誌, J90-D-2, pp.556-566 (2007).
- 4) 小野貴博, 遠山仁美, 松原茂樹: 大規模音声コーパスを用いた日英・英日同時通訳における訳出遅延の比較分析, 通訳研究, No.7, pp.49-64 (2007).
- 5) Ryu, K., Matsubara, S. and Inagaki, Y.: Simultaneous English-Japanese Spoken Language Translation Based on Incremental Dependency Parsing and Transfer, Proceedings of COLING/ACL-2006, pp.683-690 (2006).
- 6) 遠山仁美, 松原茂樹: 同時通訳における聴きやすさとポーズの関係, 通訳研究, No.5, pp.49-64 (2005).

(平成 20 年 4 月 9 日受付)

松原 茂樹 (正会員)
matubara@nagoya-u.jp

1998 年名古屋大学大学院工学研究科情報工学専攻博士課程修了。博士(工学)。同大学助手を経て、2002 年名古屋大学情報連携基盤センター助教授。現在、准教授。その間、ATR 音声言語コミュニケーション研究所客員研究員、情報通信研究機構研究員。専門は、自然言語処理、音声言語処理、デジタル図書館。