TR-I-0182

ＡＴＲ自動翻訳電話基礎研究シンポジウムの開催（別冊）
ATR Symposium on Basic Research for Telephone Interpretation

江原暉将　梅田哲夫*　竹澤寿幸
Terumasa EHARA Tetsuo UMEDA Toshiyuki TAKEZAWA
鹿野清宏**　森元　逞　飯田　仁
Kiyohiro SHIKANO Tsuyoshi MORIMOTO Hitoshi IIDA

1990.7

Abstract

ATR Symposium on Basic Research for Telephone Interpretation (ASTI) is held on two days long from 11th Dec. 1989. Open house of ATR is also held on 13th Dec. 21 persons from foreign countries and 103 persons from Japan attended the symposium together with 50 ATR people. The discussion is devided 6 technical sessions and 2 panel discussions. Integrated processing of speech and language is the main theme of the symposium. Constraints from language models are useful to make the perplexity lower. Linguistic knowledges are extracted from language database to be used for speech recognition and/or machine translation. Machine translation techniques of spoken dialogues are also discussed. Dialogue based approach of machine translation between human and machine is proposed.

ATR Interpreting Telephony Research Laboratories
ＡＴＲ自動翻訳電話研究所

---

*　現在　ＮＨＫ放送技術研究所
**　現在　ＮＴＴヒューマンインターフェース研究所

付録２

ＡＴＲ　シンポジウム

講演記録

Opening Session

葉原　耕平（ＡＴＲ）

樽松　　明（ＡＴＲ）

Session 1

川端　　豪（ＡＴＲ）

予測ＬＲパージングを用いたＨＭＭ連続音声認識

Wayne Ward（カーネギーメロン大学）

音声認識のための知識の予測的活用

新美　康永（京都工芸繊維大学）

音声認識システムにおけるアイランドドリブンパーサー

Session 2

森元　　逞（ＡＴＲ）

Spoken Language Processing in SL-TRANS

Robert C. Moore (SRI International)

音声処理と自然言語処理の統合

Stephanie Seneff (ＭＩＴ)

音声言語への適用のための自然言語システム

Session 3

Stephen Levinson（ＡＴ＆Ｔベル研究所）

音声表記からの連続音声認識

中川　聖一（豊橋技術科学大学）

言語学的知識の役割と音声言語理解のための構文・意味パーシング法

Frederick Jelinek (IBM Research)

統計的方法による言語翻訳

Session 4 (Panel Discussion)

藤崎　博也（東京大学）

Renato De Mori (McGill Univ.)

Victor W. Zue （ＭＩＴ）

白井　克彦（早稲田大学）

鹿野　清宏（ＡＴＲ）

Session 5

Kenneth W. Church

パーシング、単語の連想および典型的な述語と引数の関係

久野　ススム（ハーバード大学）

Identification of Zero-Pronominal Reference in Japanese

日高　　達（九州大学）

Computational Linguistics for Pattern Recognission

Session 6

飯田　　仁（ＡＴＲ）

Intention Translation Method: A Spoken Dialogue Translation

System Using a Lexicon-Driven Grammar

Harold L. Somers（ＵＭＩＳＴ）

対話翻訳の新しいアプローチ

Bernald Lang（ＩＮＲＩＡ）

形の整わない入力処理に関する生成的見地

富田　　勝（カーネギーメロン大学）

ＣＭＵにおける音声翻訳に向けての研究

Session 7

Christian Boitet（ＧＥＴＡ）

音声合成および対話ベースの機械翻訳

Walther V. Hahn （ハンブルグ大学）

ＡＩにおける自然言語対話研究のパラダイム

Wolfgang Wahlster(サールランド大学)

対話システムにおけるアンティシペーションフィードバック

中村　　孝（大阪大学）

Parsing Utterance Mechanism Based on Black board Model

Session 8 (Panel Discussion)

長尾　真（京都大学）

天野　真家（東芝総研）

田中　穂積（東京工業大学）

Harold L. Somers（ＵＭＩＳＴ）

飯田　　仁（ＡＴＲ）

森元（ＡＴＲ）　皆さん、おはようございます。時間ですのでシンポジウムを始めさせていただきます。本日は当ＡＴＲシンポジウムにご参加いただきましたことを歓迎いたしますとともに、厚くお礼申し上げます。ではまずオープニングセッションから始めたいと存じます。〔開会の〕あいさつとご発表は日本語・英語間で同時通訳いたします。お手持ちのレシーバーを、日本語には１チャンネルに、英語には２チャンネルにお合わせ下さい。では最初にＡＴＲ副社長で、ＡＴＲ自動翻訳電話研究所会長でもいらっしゃる、葉原博士から開会のあいさつをいただきます。葉原博士お願いします。

葉原耕平（ＡＴＲ）　皆さんおはようございます。日本語でスピーチをさせていただきます。本日はＡＴＲ自動翻訳電話基礎研究シンポジウムを開催いたしましたところ、このように多数の方々に御参加いただきましてまことにありがとうございました。

このシンポジウムは自動翻訳電話に関します最新の研究成果を討論するために企画いたしたものでございます。

自動翻訳電話と申しますのは新しい概念でございますので、多くの関連分野の研究が必要でございます。特に音声情報処理と自然言語処理の統合処理が重要でございます。本日は国内はもとより、北米及びヨーロッパから多くの参加者をお迎えいたしまして、この新しい研究分野に関しますシンポジウムを開催できますことを心より嬉しく思う次第でございます。

本日の参加者の中には、ＡＴＲについて必ずしも十分に御存じでない方もいらっしゃるかと存じますので、最初に私からＡＴＲについて簡単に御紹介申し上げます。

ＡＴＲはアドバンス・テレコミュニケーションズ・リサーチ・インスティチュート・インターナショナルの略称でございまして、日本語では株式会社国際電気通信基礎技術研究所、17文字に及びますが、略称ＡＴＲインターナショナルと申します。1986年の春に産・官・学の御支援を得て、設立されまして、通信分野の基礎技術、先進技術の研究開発を行うことがその目的でございます。

ＡＴＲインターナショナルはその傘下に４つの研究開発会社を持っておりまして、これらを支援いたしております。

ＡＴＲ自動翻訳電話研究所はこれら４つの研究開発会社の中の１つでございます。このほかにＡＴＲ通信システム研究所、ＡＴＲ視聴覚機構研究所、ＡＴＲ光電波通信研究所の３つがございます。したがいましてより詳しく申し上げますと、ＡＴＲという名称はただいま申し上げましたＡＴＲインターナショナルという親会社と４つの研究開発会社の総称でございます。

ＡＴＲにおきます研究項目は一言で申しますと、人間思考的な見地に立つものでございまして、究極の目標は将来人類の福祉に寄与することでございます。

ところでこれらの研究開発会社の資本金の70％は政府関係機関でございます基盤技術研究促進センターから、残りの30％は民間企業等から出資されて成り立っております。

基盤技術研究促進センターと申しますのは、新しいプロジェクトごとに設立されました研究開発会社に投資をすることによりまして基盤技術の研究開発を推進する仕組みとなっております。

基盤技術研究促進センターからの出資によりまして既に70を超えるいろいろなプロジェクトが開始されておりますが、ＡＴＲの４つのプロジェクトはこの中でも代表的なものでございます。

我々の研究所は1986年の春に設立されましたが、ことしの３月にはこの関西文化学術研究都市の精華、西木津地区と申しますところに最初に建てられたこの建物に移ってまいりました。基礎的、独創的な研究を進めるのにふさわしい環境かと存じます。

またＡＴＲはこの新しい関西文化学術研究都市におきまして中核的な役割を果たす使命を担っていると自負いたしております。現在研究者の数は外国からの方々十数名の客員研究員を含めまして約 180名でございます。

ＡＴＲは現在満４歳に近づこうといたしております。人間に例えて申しますと最も言語学習の進む年代に差しかかったところでございます。皆様方のお力添えによりましてこの子供を立派に２言語話者あるいは多言語話者に育て上げていきたいものと念願いたしておる次第でございます。

ＡＴＲ自動翻訳電話研究所での研究のビジョンは、技術の進歩によりましていつの日にか世界中の人々が自分自身の言葉でお互いにスムーズにかつ快適にコミュニケーションができるような技術手段を提供していこうということでございます。このような自動翻訳電話を実現するには、３つの大きな要素技術が必要でございます。すなわち音声認識、機械翻訳、音声合成の３つでございます。これら３つの要素技術に必要となる重要な基礎研究はいま佳境に入りつつあります。このシンポジウムはＡＴＲ自動翻訳電話研究所が主催を

いたしまして、ＡＴＲインターナショナルが協賛で開催をしているところでございます。このシンポジウムが研究成果を交換し、また日本と海外の国々からの参加者の親睦を深めるよい機会となることを心から望むものでございます。

最後にこのシンポジウムが開催できるようになりましたのは、参加者の皆様、特に遠路はるばる海外から参加してくださった方々、また御多用の中、アドバイザーをお引き受けいただきました藤崎先生、長尾先生のおかげでございまして、心から厚く感謝申し上げる次第でございます。

最後に私がいま話しておりますことは同時通訳の方々によりまして英語に翻訳されているかと存じますが、自動翻訳電話の技術が少しでもこの同時通訳の方々のすばらしい通訳に近づくことを期待するものでございます。ありがとうございました。

○　ではつぎに博松明博士から基調講演を賜ります。博士はＡＴＲ自動翻訳電話研究所の社長でいらっしゃいます。博松博士、お願いします。

博松明（ＡＴＲ）　座長そしてお集まりの皆さん、ありがとうございます。私の話の目的は電話通訳に対する音声処理および言語処理における研究の展望をご披露申し上げることと、当分野における諸問題を論じることにあります。また音声言語の通訳に対する要件の概要についても言及し、ついで最近のＡＴＲの研究状況を概括し、さらに今後一層の努力を要する領域について論じて、私の議論を締めくくらせていただきます。

自動電話通訳システムは、話された対話を話者の言語から聞き手の言語へ自動的かつ同時に変換しようというものであります。これはもちろん世界各国の人々に対して、言語の障壁を克服しコミュニケーションを円滑にするのに役立つでしょう。このようなシステムを作り出すには、まず大きな構成技術の開発を要します。それは音声認識であり、機械翻訳であり、音声合成であります。こうした各個別のサブシステムは、つぎにこの自動通訳電話システムを形成するのに統合しなければなりません。自動通訳電話システムは異なる言語を話す話者間で双方向の音声言語通訳をします。音声処理と言語処理の関係は極めて重要になります。

音声言語通訳システムに対する要件はビューグラフに記述します。高性能認識と翻訳が達成されねばなりません。入力信号、出力信号ともに音声でありますので、プリエディットもポストエディットもありません。これは音声言語通訳システムの出力が高度な頭脳をもって、二つの異なった言語を話す話者が互いに困難なくコミュニケーションできるようにしなければならないということであります。

第二に、使用者、つまり実際の対話の参加者や聞き手は通訳された内容を聞いてその意思の意志を理解するのであります。人間と機械との間の相互作用という観点から見ますと、すと、当システムは機械処理を通して人間対人間のコミュニケーションを可能にするわけるわけです。人間が積極的に割り込んで入ってくることを考えますと、従来の機械翻訳シ翻訳システムよりも豊かな能力をもっているものでなければなりません。

第三に、音声言語通訳システムが扱わなければならない文章の種類は、書きことばを扱う機械翻訳システムで処理するものとは非常に異なっているということがあります。話しことばでは各センテンスは短く、センテンス構造も特別に複雑ではないのですが、話された対話というのは省略や照応表現が含まれています。またこうした対話は構文的に形の悪い表現も多く含まれます。しばしば文法上の誤りあるという特性を扱うのに、音声言語通訳システムは、決して完ぺきではありえない音声認識装置が起こすエラーやあいまいさと格闘しなければなりません。避けられない認識上のエラーやあいまい性を容認するにあたって、パーシングアルゴリズムを有効に使用するには、音声認識に続く言語パーシングプロセスが必要になるものと思われます。

第四に、音声言語通訳システムに対する要件というのはリアルタイム動作でなければならばならないということであります。聞き手がシステムからの出力をまさにその場で待って待っていることを考えれば、システムが当該センテンスを翻訳するのに許される時間は数間は数秒以下のものであります。さらに前に述べた事柄に応答するのに、逐次応答が必要が必要になります。演算時間が指数関数的に増大するのを避けるためには、変換アルゴリルゴリズムが決め手になります。また話されたセンテンスを音声メールボックスのようにバッチモードで翻訳することもできるでしょう。しかしながら、そうした装置は通置は通常の電話通訳では極めて限られたものになると考えられます。

第五に、電話通訳システムの目標として一般に認められておりますものが、無制限領域でのユニバーサルな対話であるとしましても、現在の実現の可能性の高い目標は、タスクを特定化した分野に限ったシステムであります。話しことばのあいまいさを解消するものは領域知識であると予想されます。制限領域での予測を用いた効果的な処理が期待されます。

６番目にはこのシステムの使用者が単言語の使用者であろうということであります。す

なわち話し手はターゲット言語を知らず、聞き手もソース言語を理解していないであろうということであります。この状況は一方で、誤解を防止するために通訳の精度に対して厳しい要求を課すことにもなります。自動翻訳電話システムがまったく新しい概念であるため、システムの全体設計は性能レベルを考慮して決定されることになるものと思われます。ここでいう性能レベルとは人間的要素と同様、構成技術個々からも規定されるものであります。システムの使いやすさ、つまりユーザーフレンドリであることも無視してはなりません。音声に加えて画像や文章テキストをも扱えるマルチメディアターミナルを用いれば、通訳電話システムのぎこちなさも償えるのではないかと期待されます。

つぎ、お願いします。この図は一つの案としての通訳電話システムを示しております。音声認識と言語解析の間のギャップを埋めることを目的としています。ここにあるのがキーになる要素で、音声認識と言語翻訳とから生じる音声・言語の統合処理であります。言語ソースモデルという概念に基づき、この処理では音声認識装置に対してトップダウン予測を実行します。ボトムアップの結果には複数候補が含まれますので、言語的制約と各種知識情報に基づいて、候補を絞らなければなりません。他の分野に関しましては他の研究機関とも連係をとりながら研究を進めております。大語彙の連続音声の認識では、まず可能な程度にまで音韻が認識されついで連続語および句（文節）が認識されます。ＨＭＭ音素モデルでの改良点がいくつか導入され連続音声認識に応用されています。音声認識能力を向上させるため、知識を読み取るスペクトログラムを用いたセグメンテーションやニューラルネットワークによる音素認識が研究されております。

不特定話者の問題への効果的なアプローチとして、話者適応が取り上げられています。話者適応に対するスペクトラムパターン学習の手段として、ひとつの有望なアプローチはコードブックマッピングがあります。このアルゴリズムでは音声認識システムに依存しない一般的な話者適応が可能になります。話し手の特性に適応するのに、20から30語もあれば十分でしょう。

音声処理および言語処理の統合はＡＴＲで幅広く研究しています。音声認識についてはつぎのような分野で問題があります。上から５つの候補の句認識率は95％前後になるでしょう。句認識が不確定だと候補が多数生まれます。パープレキシティが増大することはリアルタイムデータ処理にとっては厳しい条件であります。パープレキシティとは認識装置がある時点で一語に選択を絞らなければならない平均語数のことです。現在のところ音声言語システムが扱い可能な能力をもとうとすると、言語の制約がやむをえないように見えます。

ＨＭＭ音素モデルはＬＲパーサーという予測パーサーと一体化しています。私どもの研究では、音声入力におけるつぎの音素を予測し、全般的確率を推定して確認します。この統合アルゴリズムはパープレキシティの高い大語彙に対して有効で、かなり効率よく処理されます。認識個々については、連続音声を句ごとに分離して認識する方法を採用しています。言語処理システムでは最適候補を選ぶのに構文・意味論的知識を使用できる機能が必要になるでしょう。音声認識からの候補数を減らす方法としては、日本語の話しことばの解析を行なっております。この方法では日本語の係り受け関係を用いています。この情報を用いると、音声認識出力から可能性の高い候補が選択できます。特定の領域では知識処理といったアプローチも研究中であります。知識ベースを適用すると音声認識から候補数を減少させられることがわかりました。

語と語の連想関係については、電話通訳システムは文脈中の意味を理解できることも考えうるでありましょう。私どもの研究ではおもにつぎのような問題の解決に注意を向けております。すなわち表現のあいまい性はいかにして除去できるか、代名詞をいかにして識別するか。状況によって変化する話者の理解をいかに正確に表現するか。つぎに話されるであろうことばをいかに予測するか。話し手の話題や陳述の変化にいかに対応するか、といったことがそれであります。対話プロセスの広範な説明に加え、対話モデルも機械翻訳には有効です。日本語の音声対話に特異的な言語上の主要現象は言語学的観点から研究してきており、いずれはコンピュータ上で実現可能な談話対話モデルとして構築することになろうと思われます。ゼロ代名詞、敬語、否定、発話意図の範囲に関する研究テーマも進行しております。発語センテンスのあいまい性を除去するため、表現の使用について実際的な制約を設けるには、こうした制約を使って選択したもっともありそうな解析候補から抽出しなければなりません。日本語の対話解析への現在のアプローチは、型付き素性構造についてレキシカルな構文文法フレームワークと、解析順序が制御可能なパーサーに基づいております。

機械翻訳電話対話システムを考案するにあたって、解決すべき問題の一つは、思考の奥を流れる意味や話し手の意図をターゲット言語に正確に翻訳するにはどうすればいいかということであります。対話を理解、翻訳するのに、多くの研究ではプラン認識モデルに焦点を当てています。対話参加者の心的状態における制約を用いた文脈処理のための演算モデルも研究中です。現在言われております翻訳方法の目標は、本質的には意味トランスフ

ァー的アプローチに基づくものであり、二つの翻訳プロセスに特徴付けられるものであります。一つは、要望、約束、あいさつといった意図を話の中から抽出するものであり、もう一つは発話の中の命題を変換するものであります。解析、変換、生成用情報の総合記述として素性構造を採用しています。素性構造を効率的に扱うための方法を、目下研究中です。

ロバストな翻訳を確立するため、実例本位のパラダイムを研究してまいりました。これは現在の対象に類似した例を見つけ、これを新たな問題解決に採用するという過程で構成されます。この方法は目下音声認識装置の出力のあいまい性除去および連結名詞グループの適切な翻訳に応用しているところです。ＡＴＲでは実験的な音声言語翻訳システムを開発して、音声処理、言語処理の統合に固有な主要問題を確定しつつあります。詳細は本日の発表の中で言及がある予定です。音声言語にとって効率のよいアルゴリズムは必要なことがはっきりしました。

機械翻訳プロセスの終端で現われる音声出力は規則による音声合成に基づくものとなりましょうが、この規則では形態素情報を含む言語学的情報をテキストとともに利用します。音声の明瞭さ、自然さの主な要因は、音声合成単位の選択と韻律ルールの制御が適切に行なわれていることです。長さの異なる各種複合音声単位を用いた音声合成を探求しております。

一人の話し手から他の話し手へ、声を変換することにより、音声合成の個化が実現できました。これに必要な技法は、声質情報に関与する要因を抽出すること、および声質を制御する方法であります。

自動通訳電話による会話を成立させようと思えば、音声認識、機械翻訳および音声合成技術のレベルを向上させるための広範な研究を行なわなければなりません。今後の研究の方向をつぎの点に振り向けています。

大語彙の連続音声を認識する能力をさらに高めること。私どもの目標は 3,000語です。音素認識性能を向上させるために、音声に関連する知識を応用する現在のアプローチを総合するスキームを策定すること。不特定話者認識においては、大規模音声データベースを用いて、大語彙連続音声認識に適用可能な方法を探ることになるでしょう。ピッチ，音声強度および句境界からの情報との関係といった韻律的な情報を用いて、句認識のためのアルゴリズムの精度およびスピードを上げなければなりません。しかしながら、効果的な情報を扱うためには、細心の処理を必要とするでしょう。というのは、韻律的な特徴は日本語の話しことばでは特に信頼性のあるというものでもないからです。

音声処理、言語処理の総合においては、発話に対する語のレベルを予測するスキームを研究することになるものと思われます。センテンス構造のヒューリスティックの統計的文法特性を導入することについても検討を要するでしょう。　いくつかの語からなるある特定の系列の推定確率という形で入力に対する統計的制約を用いて、パープレキシティを低減することも考えています。対話構造のような高度な情報をもっと利用することも研究対象になるものと思われます。日英間の音声言語翻訳システムのプロトタイプも、本研究プロジェクトの中で公開できるものと思っております。言語そのものに関する知識および領域に依存した言語外知識は音声言語通訳の各観点に対する共通ベースとしてまとめられるでしょう。

機械翻訳においては文法やレキシカル辞書を高度なレベルに充実することが、大語彙翻訳を扱うのに求められるでありましょうし、文脈処理に基づく深い理解による翻訳を発展させる道も探らねばなりません。問題への挑戦機運は、大語彙と各種タスク領域に向かって拡張しうる一般方法論に向かうものと思われます。リアルタイム処理への要求を考慮すると、高速演算スキームの研究により、理論的計算言語学とソフトウェアインプリメントの間の従来からのかなり大きなギャップを縮めるようにする必要があります。

音声合成では、ルールによる音声合成を発展させて、会話文におけるより自然な音声品質を得るようにしなければなりません。言語生成における言語学的な情報は、音声合成でのルールの制御として反映するでしょう。異なった言語にわたる音声の声質制御も発達するものと思われます。

音声・言語コーパスの大量データベースも自然言語の広範な複雑さのゆえに、関連する研究活動をさらに促進するのに不可欠となるでありましょう。ただしかし、目標への到達はレベルや技術の向上による段階的なものになると考えられます。領域サイズや固有の領域適用、多言語適用およびマルチスピーカー用途等の観点からすると、システムの拡張性を考慮しておく必要も生じるでありましょう。

この野心的なプロジェクトを考える上でもう一つ重要なことは、国際協力であります。自然言語という各国の言語を、世界各国の研究組織で深く研究されねばなりません。

ご静聴ありがとうございました。

博士、どうもありがとうございました。それではこれでオープニングセッションを終わ

りたいと思います。

Session 1

Integrated Processing of Speech and Language 1

鹿野（ＡＴＲ）　続けてシンポジウムに入らせていただきます。

私ＡＴＲ自動翻訳電話研究所の音声情報処理研究所の鹿野です。午前中のセッションのチェアマンをさせていただきます。

きょうのセッションは特にスピーチとランゲージ、それをいかにしてインテグレートするかということに観点を上げて３つのセッションを行いましてその後パネル討論を行います。ちょうどいまなかなかスピーチとランゲージというのは水と油のようになかなかうまく混ざらなかった。しかしながら方々でそれをインテグレートしてやっていこうという機運が高まっている折でありますので、非常にいい機会かと思います。

講演者の方には多少きついんですけども、15分から20分の間に講演を終了していただきたいと思います。その後５分程度の質疑応答を行いたいと思います。

質疑、応答は日本語でも英語でも結構ですので、どちらでも、多分同時通訳していただけるはずですので、行ってください。

後ここで講演に使われますＯＨＰのコピーを会議の後で送らせていただきます。必ずしも写されなくても大丈夫だと思います。後でＡＴＲの方でコピーして発表者の好意によってコピーいたしまして送らせていただこうと思っております。

それではセッションに入りたいと思います。

川端豪（ＡＴＲ）

予測ＬＲパージングを用いたＨＭＭ連続音声認識

それではＨＭＭ音韻認識と予測ＬＲパーザーを用いた連続音声認識についてお話させていただきたいと思います。

まず基本的な考え方なんですけれども、この方法をＨＭＭ-ＬＲと以後呼ばせていただきたいんですけれども、基本的なブロックが２つあります。１つは予測型のＬＲパーザー。これが言語処理を行います。それからHidden Markov Modelによって構成した音韻照合子、これが実際の音韻認識を行います。

ある日本語の、今回の場合は文節ですけれども、そういう音声のデータをどのようなものが発生される可能性があるかというのをコンテクストフリーグラマーで記述しておきます。そしてＬＲテーブルを生成するためのプログラムというのがあります。それを使ってＬＲテーブルというものに転換いたします。これは後ほど御紹介します。

このＬＲテーブルを用いましてこの予測型のＬＲパーザーはその音声で次に一体どのような音韻が来るのか、またはどのような単語が来るのかというのを予測いたしまして、その予測に従って音声中の音韻を調べると、音韻照合を行うわけです。そしてその結果をまとめ上げまして最終的に最もヒドゥンマルコフモデルですから確率が高くなった音声の候補を認識結果として出力いたします。

まずこの予測型ＬＲパーザーの話から説明させていただきます。

もともとＬＲパーザーという技術はですね、アーティフィシアルな人為的につくられた言語に対して用いられてきた手法です。本来決定論的な方法で文法のあいまいなようなものは全く取り扱うことはできなかったんですけれども、1986年にＣＭＵの富田さんによりましてこれが文法の不確かさ、あいまいさというようなものを取り扱えるように拡張されました。

その基本的なメカニズムというのはスタックスプリティングメカニズムと呼ばれています。これも後で説明します。

ＡＴＲではこれを1989年に音声認識に利用するためにさらに拡張いたしまして、パーザーというよりはセンテンスジェネレーターとしてこのＬＲパーザーを使ってやろうということを考えました。

このプレディクティブＬＲパーザー、予測型ＬＲパーザーと呼んでいますが、これは音声中の単語や音韻を次々と認識の各段階で予測していくことができます。

結局その連鎖でありとあらゆる文章の可能性、文節の可能性を生成できることになります。そして特徴といたしましてはこの予測型ＬＲパーザーを用いる場合の機構の特徴は、この富田氏によって導入されたスタックスプリティングという機構をですね、文法的なあいまい性があったときのみに行うのではなくて、あらゆる音韻の可能性すべてのバリエーションに対してそのスタックスプリティングという操作を行います。これも後で詳しく紹介します。

初めにですね、ごく簡単にＬＲパーザーの基本的な動作を説明しておきたいのですが、これは非常に簡単なおもちゃ、トイグラマーに相当するようなものです。日本語です。センテンス、文章は文節、名詞句と動詞、動詞に展開できる。名詞句は名詞のみであるか名詞と何か、サフィックスであるか、後名詞としてはマメとかアレとか、日本語に対応しておりまして、後サフィックスは助詞のヲ、動詞としてはオクレとクレと２種類しか認められない。非常に小さな文法ですので非常に簡単な処理ができます。これを先ほど申し上げましたＬＲテーブルジェネレーターというものにかけますと、この下に出ておりますＬＲテーブルというものが生成されます。

そこでいま実際にですね、パーザーに、ここにちょっと見にくいと思いますが、ローマ字が並んでいます。ＡＲＥＯＫＵＲＥ、この順番にパーザーに音韻の入力があるというふうにお考えください。

一番初めにですね、パーザーにＡというシンボル、母音のアですね、Ａというシンボルが入ってきました。パーザーはＬＲテーブルを使って何をするかといいますと、まず初めは状態ゼロです。状態ゼロで、実は横に探していくんですね。そうするといま入ったきたアという記号が見つかります。Ｓ２という動作をします。Ｓ２というのはシフト・アンド・ゴーツースティイト２という意味です。シフトというオペレーションはこのようにですね、ある状態、これはおのおのがスタックになっているんですけれども、このスタック上にいま入ってきた入力の記号を積んでいくわけです。そして同じようにして状態２にいきまして、次のものを探してというふうに処理をどんどん続けていきます。そうするとこのようにですね、たとえばＡがスタックに積まれ、ＡＲが積まれ、ＡＲＥが積まれで、これで日本語の単語のアレというものができるわけです。いまこれは状態15にいるんですけれども、状態15におきまして新しい記号が出てきます。状態15にいままではＳ何とかだったんですが、Ｒという記号ができます。これはリジュース・ユージィング・ザ・グラマールールファイルという意味です。先ほどの５番目の文法規則を使いましてそのスタック10にさっきたまっていましたＡＲＥという系列を名詞、ナウンに置き換えてしまいます。このことがスタック10ではＲ５によってＡＲＥがＮに置き換えてしまうということが行われていることがわかります。

このようにしてスタックに単語の可能性をどんどん積んでいきながら、かつときどきそれを縮退させ、リジュースさせてスタックの操作を進めていきます。

そして実はですね、その次はいま状態４なんですけれども、この４でちょっと困ったことが起きます。状態４の項目に実は次にオが来るんですけれども、来たときに２つ動作が記述されてしまいました。これはコンフリクトと呼ばれる現象です。従来の伝統的なＬＲパーザーではこれはもう処理できなかったわけですけれども、富田氏が考案いたしましたスタックスプリティングという動作を導入することにいたしましてこれが解決できます。何をするかというと、ここで２つの動作がさっき指定されていた、シフトIIとリジュースの２つが指定されていたときに、このスタックを２つに分けてしまうんですね。スタックをスプリットしてしまうわけです。そうしてこのスプリットした上の世界と下の世界を全く独立なものとして別々に処理を進めていきます。これによってそのようなコンフリクトが起きた場合でも全く問題なく処理を進めることができ、言語のあいまい性をこの方法によって取り扱うことができるようになりました。というのが富田氏の提案した一般化ＬＲパージングの考え方です。

これに対しましてＡＴＲで開発されました予測型ＬＲパージングというものは、ちょっと動作が違います。一般化ＬＲパージングの場合にはまずある音声の系列が入ってきて、その系列を次々とパーザーが処理していくという動作をしていました。ところが予測型ＬＲパージングの場合には、何かが先に入って来る前に既にこの状態で横一列を全部調べてしまいます。そうすると初めからＡという記号とＭという記号に何か起きる可能性があるということがわかります。これが音韻を予測したということになるわけです。すなわちこの文法に従うある発生は必ず一番はじめにＡかＭのどちかの音韻が来るということがわかったわけです。

そしてそのおのおのについて、たとえば２番にいって次はＲ、９番にいって、次はＡですね、というふうに処理を進めていくことができます。

このような機構は先ほどと同じようなスタックで見てみますと、このようなかなり複雑なスタックになります。

先ほどはグラマーアンビギュイティーのためにここでのみスタックスプリティグが起きていたわけですけれども、予想ＬＲパーザーでは音韻の可能性が分かれた部分全部でこのような分岐が次々と起こっていきます。これが予測型のＬＲパーザーです。

次に、パーザーの話はそれくらいにして、簡単にＨＭＭによる音韻照合部分の説明をいたします。

ＨＭＭというのはHidden Markov Modelと呼ばれている統計的なモデルですが、それを用いて音韻をモデル化します。端的に言うならば子音に対しては３ループのモデルを使いま

す。それから定常的な母音に対しては１ループのモデルを使います。

このHidden Markov Modelの各状態ごとにディレーションの継続時間の制御を行っております。これらの特性によりまして非常に高い音韻認識が可能となりまして、これを用いて連続音声認識が高精度に行えます。

次にですね、ＬＲパーザーからHidden Markov Modelを駆動するやり方について簡単に御説明いたします。

これはHidden Markov Modelにおけるトレリスと呼ばれる構造をあらわしています。これは縦軸の方向にですね、Hidden Markov Modelをどんどん連結していきます。それから横軸の方に入ってきた音声、入力が配置されています。今、あるHidden Markov Modev Modelに対して処理が行われた状態では、この格子のトレリスのここの部分までが計算さ計算されています。各トレリスには、Hidden Markov Modelの確率値がつけられています。

ＬＲパーザーから音韻照合子を駆動するというのはどういうことかと言うと、今ある段階までに求まっているそのHidden Markov Modelの確率、ＣＳですね、これを覚えておきまして、これを先ほどのスタック、これですね、このスタックのおのおのにすべて記憶させておきます。そうしてスタックを発展させる段階で、その一つ前の確率の、１ラインの確率の値を使って、次に予測した音韻のHidden Markov Modelをつなぎ、この確率の値をどんどん更新していきます。こういう方法によって全く中間的なシンボルを介さずに、音声認識が行えますので認識の精度が飛躍的に高まります。

そこで一体このＨＭＭ-ＬＲシステムが、どのくらいの精度で音声を認識できるのかというのを示します。

まずタスクなんですけれども、今回は日本語の文節認識を行ってみました。今回は話者依存型、ある特定の話者が話した内容を認識するというものをタスクいたします。
実験としては、それを４種類、３人の男性、１人の女性について行っています。それからタスクの難しさというものを示すのに重要である文法の記述を行っています。重要なのは、約 1,000単語を取り扱える文法であるということです。

これからタスクエントロピーを計算してみますと17.0、そういう値が出てきます。これは単語認識に換算すると10万単語以上の認識を行っていることになります。

それから音韻パプレキティ、各音韻ごとにどれだけ分岐があるかという量なんですけれども、約６音韻、すなわち次々と音韻を予測していくわけなんですが、その各段階で次に予測される音韻の平均値が約６であるという意味です。これからよく英語で使われていますワードパプレキシティを推定してみたんですが、この場合 100以上になるということが計算されています。

このようなタスクにＨＭＭ-ＬＲ音声認識システムを適用して、実際に得られた文節認識実験の結果がここに示されております。１位で88.4%、５位までで何と99%が認識できてしまうという物すごい精度が出るということがわかります。これはこの４人の平均値です。

というように、ＨＭＭ-ＬＲの説明をしてきたんですが、簡単にまとめさせていただきます。

初めに音韻ベースのＨＭＭ、それから予測型ＬＲパージング、この２つを組み合わせることによって非常に高精度の音声認識システムを構成しました。実際に評価実験の結果、日本語の文節認識に対して88.4%の認識率が得られました。実際にこのＨＭＭ-ＬＲ音声認識システムは、ＡＴＲの自動翻訳デモンストレーションシステムでありますＳＬ-ＴＲＡＮＳの中に組み込まれて使用されております。以上です。

質疑応答

問　　今述べられた実験でのＬＲパーサーに要する演算量について少しお話しいただけませんか。１文あたりどれくらいの演算時間が必要かとか、ハードはどんなものか、といったことです。

答　　はい。計算時間は機械によって異なります。音声認識タスクにはアライアントマシーンを使用しました。それから言語処理ですが、言語処理のための計算時間はそれほど多くありません。このシステムでの計算時間はほとんどすべてが音声認識のためのものです。つまり電話確認装置のためと言ってもかまいませんが。２・３秒の日本語の文節が話された場合、本システムでは10秒かかりません。実時間の２・３倍です。

それから私どもの理解ですと、文脈自由文法の場合だと、ちょうど有限状態オートマトンのように非常に高速な処理が可能です。ほかにご質問は？

問　　最終性能における予測ＬＲ構造の役割について何か考えをお持ちでしょうか。というのは、たとえばもともとのアーリーのアルゴリズムのように一般的な文脈自由のパーシングを可能にする構造は、ほかにもあると思うのです。他の構造で類似の結果が得られる

かどうかといったことや、Hidden Markov Modelから何が生じるか、ＬＲパーシングから何が生じるか、またその性能が何に由来するか、といったことについて実験を試みられたことはおありでしょうか。

答　　残念ながら、アーリーのアルゴリズムについては実験結果がありません。私が日本語の句認識にＬＲパーシングメカニズムを用いた理由は、日本語の句構造が非常に単純で、この方法で十分取り扱えるからです。しかしながら、英語、とくにその格についてはもう少し複雑な文法が必要です。ただ、アーリーのアルゴリズムやＣＹＫその他については認識結果を持ちあわせておりません。ありがとうございます。

○　非常に時間も来ているので、次のセッションに移らせていただきます。どうもありがとうございました。

Wayne Ward（カーネギーメロン大学）

音声認識のための知識の予測的活用

私は Cheryl Young と私とでカーネギーメロン大学の音声理解グループで行ないました研究についてお話ししたいと存じます。私どもが目下取り組んでおります、一種の一般原理というのは、できるだけ多くの利点を得るために、プロセスのできるだけ早い段階で制約を使用したいということであります。また文脈知識を用いれば後続の発話内容を予測できるのではないかと思いますし、こうした内容予測によってつぎの単語予測が可能になるでしょうし、この単語予測をＨＭＭによる認識装置の探索を手引きするのに使用することもできると考えます。私どものシステムで使用している知識ソースは、意味論文法や対話構造であり、これにはゴール、サブゴール、プラン、ヒストリーが含まれますが、さらに領域に含まれる対象およびその属性といったタスク意味論、焦点照応的レファレンス・省略といった一般世界の知識、そしてユーザー知識、つまりある領域のユーザーがもつ知識です。ただ私どもが用いました唯一の知識は、ユーザーが熟練者か初心者かという知識でした、いや、ユーザーです。知識の使用形態はといいますと、先行入力の発話、ゴールおよびプランを理解して、ユーザーのゴールが何なのかを明確にするための推論であり、こうしてデータベースクエリーのあいまい性と一般性を低減しようとするものであります。このシステムはつまりMINDシステムでして、実際にデータベースにアクセスしてこれから答えを返します。知識は概念的な予測を生成するのに用いました。これはゴール・アンド・プラントラッキングで達成し、制約条件と焦点メカニズムを伝播して省略と照応的レファレンスを解決しようとしました。一般プロセスについて言えば、制御フローは、まずユーザーが存在しているであろうと思われる状態のうちもっともありそうなアクティブゴール状態を決定し、後続の発話内容に関する予測で、後続発話に現われそうな概念を含むものを生成させます。これらは層状に生成します。つまりもっともありそうな概念の集合に、つぎにありそうな概念の集合が続き、こうして以下もっともありそうでない概念にまで層をなします。また省略についての予測もあって、それが正当に使用可能かどうかを見ますが、これはありそうな照応的レファレンスに制限することです。

何らかの対象について述べなければ、その対象に言及する代名詞を使用することはできません。それはただ何の意味もなさないということです。つぎのステップを予測するにあたっては、「現ゴール状態完」といったメカニズムを用います。「完」ならユーザーはつぎのゴール状態に移ることが考えられます。ほかにどんなゴールとプランが現在アクティブであって、潜在的に不完全であるかは、さらに追及する必要があるでしょう。続いて、つぎに可能な状態を決定します。今不完全なゴール状態にあるなら、強制停止もしくは現ゴールまたはプランを保留しておくこともできます。もし現ゴール状態が終了なら、親ゴール状態もしくはプランを求めることができます。このときにはまた、保留中のゴール状態を再開することもできるし、ユーザーはどの時点でも質問の種類を明らかにするよう要請することも許されます。われわれのシステムでは意味論文法を用いましたが、意味論文法は分割ネットワークにまとめてあり、このネットワークは部分的に拡張してあります。部分的拡張というのは、われわれの領域にある一定のノードのことです。領域とは、話題にのぼりましたリソースマネジメントとか船積管理、つまり船はどこに出航できるか、積み荷は何か、また船がかかえる問題のことであります。拡張しないノードもあります。それは予測を実行したいと思っているノードです。たとえばある人が目下関心を抱いている船の集合がそうです。ですからわれわれの文法というのは、いくつかの方法で分割されているわけです。

一つは意味論の構文レベルによります。したがって各センテンスのためのネットワークはわかれています。たとえば船舶への損傷に関するものがそうです。船舶の損傷について質問するあるいは話をするときの場合をすべてまとめてひとつの特定のサブネットワーク

にすることもあります。修理回数についてたずねるときの場合は別のサブネットワークになります。船の位置についてはさらに別のサブネットワークといった具合です。これらも単語レベルの意味論によって分割します。さきほど説明しましたように、たとえば船名、港名、船の能力といったものを含む特定のノードは前もって拡張しませんでした。ノードやネットワークも照応や省略ごとに分割します。この分割メカニズムは予測を効率的に活用しようとするためのものです。仮に一定の概念だけが用いられると予測した場合には、一定の変数フィラーだけが関心の対象となります。すなわち照応と省略が適当ではなく、サーチの対象をつぎに現われそうなサブネットワークにのみ限定することができます。したがってわれわれの総合的一般プランは文法を概念ごとにサブネットワークに分割することなのです。これら概念とは領域内にあるもので、文法内のこうしたサブネットワークに対してインデックス化しています。続いてつぎの発話内容に関する予測の層状化集合を生成し、パーサーは予測サブネットと有限状態サーチでのフィラーとだけを用います。

われわれのシステムではベースとしてスフィンクス認識装置を使用しています。これは時間同期ビームサーチを行ない、文脈自由トライホンで構成するHidden Markov Model単語モデルを使用しています。　文脈依存トライホンは単に、先行・後続文脈にある特定の音素をモデル化したモデルにすぎません。単語遷移を手引きするのに有限状態ネットを用います。認識装置が一つの単語の終わりに来ると、有限状態メカニズムにアクセスして、この単語のつぎに来うる単語を決定しようとします。そして予測されたサブネットだけが伝達されることを許されるのです。また予測された現フィラーだけが変数ノードを拡張することを許されます。たとえばだれかがこんな質問をしたとします。「どんな船がシナ海にいますか。」するとデータベースは特定した船舶リストを返し、以後この船舶集合に関する質問はこの応答リスト上の名称に限定されることになります。これがこの種のシステムから抜け出す制約条件の一例です。最初の対話はＣＧ18「グリッドゥリ」が２日のうちに真珠湾を出発する、で始まります。ＣＧ18は特定の船級で、「グリッドゥリ」は船の名前です。この任務には表面レーダー、３－Ｄレーダーおよびソナーを要します。「グリッドゥリ」はＳＰＳ 483Ｄレーダーが動作しないことを報告してきます。これは基本的にはいわゆる日常報告であることを意味します。この船はある一定の任務についており、問題が生じています。これがシステムに対する設定でした。この時点でユーザーはこの問題に対する解決を探し出すよう求められます。この船を修理する必要があるか、それともほかに使用できる代わりの船があるかどうかを決定しなければなりません。

こういう場合あなたならどうするでしょうか。ここでのユーザーがつぎに実際に言ったことは「「グリッドゥリ」のＥＴＲを示せ」でした。ＥＴＲとは修理見積時間に対するアクロニムすなわち頭字語です。ここで仮に予測を一切用いないで文法を全部使用したとすると、つぎの発話の始まりの単語として 532個もの可能性がありました。しかしゴールとプランを用いれば、このサーチすべき単語数を、発話の始まりの語として 194個にまで低減することができました。さらにユーザーの経験、ユーザーは熟練者か初心者かを考慮すると、この文頭の語としてサーチする語数は66にまで減少しました。最初に認識された単語〔ＳＨＯＷ〕が与えられると、可能な単語数はそれぞれ 385、31、27個となり、最後の単語については、それぞれ 151、59、６個となりました。これによりわれわれがこうした予測メカニズムを用いて得たサーチの低減という考え方を理解いただけたのではないかと考えます。話しことばではわれわれは、タスクの中でどれだけ制約条件を得たかを示すのに、しばしばパープレキシティを使用します。パープレキシティとはあらゆる時点における可能性の数の幾何平均だと考えることができます。このタスクつまり船舶とその問題点を扱うＤＡＲＰＡリソースマネジメントタスクでは、語いは 1,000語でした。われわれが用いたテストセットは、システムを学習するのに用いたものとは完全に独立したものです。トーンデータベースから10個の対話を取り出しました。トーンとはタスクオリエンティッドの自然誘導発話に対するアクロニムすなわち頭字語を言います。

カリフォルニアの海軍海洋センターの人々は実際こうした仮説的シナリオを実施し、こうしたタスクを実行したことも記録されています。われわれはこれらを行なった人々から実際のシナリオを３件入手し、転記してユーザーにそのセンテンスを読み戻してもらいました。これは読んだものです。自発的な発話ではなく、もとの本来のユーザーからの実際の転写として読んだのでした。われわれはつぎに７件のパラフレーズしたシナリオを加えました。つまり異なった表現を用いて同じことをやったわけです。タスク全体を見ますと、仮に文法全部を用いたとすれば、総合的なパープレキシティは 279でした。文法全部と申しますのは、いかなる種類の予測も用いていないということです。単にゴール・アンド・プランレベルの予測を用いるだけでも、このパープレキシティは31に減少し、さらにユーザーが初心者か熟練者かという予測を適用するとシステムの総合的パープレキシティは17.8にまで下がりました。同一のタスクに対して10人の話し手を起用しました。８名は男性、２名は女性で、トーンシナリオの対話は、ここでもまた１話し手あたり20文としましたので、全体としては 200文になります。テストセットのパープレキシティについては、

われわれが実際に文法のみを使って試験した発話集合に関するパープレキシティであり、242を数えました。層状化予測を用いると18.3としました。繰り返しになりますが、層状化予測とは、最大限に予測した概念集合の中に受け入れ可能なパースが見つからなかった場合、後戻りして予測度の低い集合を使い、同じストラテジー受け入れ可能なパースが見つかるまでこのストラテジーを繰り返すというものです。予測を用いなかったときの試験に対する単語認識精度は82.1%で、予測を用いた時には96.5%となりました。単語認識精度とは1からエラー率を差し引いた値ですが、ここでエラー率というのは置換回数、挿入数、欠落数をそれぞれ加えたものです。ここでもう一つ重要な点はエラーの性質です。挿入エラーの大半は単純なもので、挿入・欠落エラーは冠詞の「the 」でした。これは文法があらゆる時点でこの冠詞の「the 」を包含するのも欠落するのも自由にできるようになっているからです。これは何の差異も生じるものではありません。置換の大半は「its 」とすべきところを「his 」としたもので、ここでも問題はやはり代名詞ですが、これはどれも非常に類似したものと受け取られやすいものですし、システムにとっては構文上取り立てて大きな違いではなかったものであり、文法上は相互に使用が許されてもおりました。

発表は以上です。何かご質問があるでしょうか。

質疑応答

問　　層状化をどういうふうに行なったか、少し説明していただけますか。

答　　はい、かしこまりました。この点については簡単にしてしまってすみません。時間内にあらゆる論点に立ち入ることはできなかったものですから。この問題につきましてはスライドを持っておったのですが、やり過ごしてしまったのです。あらかじめお断り申し上げておかなければならないのですが、層状化のもつ危険性は、実は層状化アルゴリズム自身にあるのではなく、誤ったパースを拒否するような認識をする点にあるのだということです。

層状化の方法としては、単に段階的に制約条件を弛めていくだけです。ですからもとの層は制約条件をすべて使っており、全知識ソースが可能なのです。第1層がその下にありまして、これは単にユーザーの制約条件を弛めることによって生成しました。したがってユーザーが初心者か熟練者かを知ることから生じる制約条件はすべて破棄したのです。その下の層はゴール制約条件を弛めたものですが、これは単にゴールツリーの中のレベルを上げてその後に続き得るものはいずれも許容することによって達成します。こうして構文上の制限を緩和したわけですが、これは基本的には、文法の中のどのネットでも用いよ、とするゴール制約条件を弛めることでもあります。これが動作しない場合、全単語システムに後戻りすることになっています。そしてどの単語も他のどの単語を追跡していいわけです。認識された文字列はパーサーに渡すのですが、このパーサーは、内容語がそこにある限り、全単語出力をかなりうまく扱えるものでなければなりません。現実の機能語制限はあてにしませんでした。つまりそれは一つのキーワードであり、言わばドリブンパーサーのようなものとも言えますが、そんなに単純でもありません。層を生成する際の現実的な問題は当該のパースがよくないということをどうやって認識して、別の層に移行する必要があるかどうかを決めるかということです。これにはわれわれは非常に単純なメカニズムを用いました。それは制限度の大きい領域であるのでこの当該領域を抜け出すというものです。それから概念ですが、予測は概念に基づいており、概念それ自身や船名等はまず混乱することはありませんでした。またここでのアルゴリズムは単に2層のスレッショルドです。これは一定のスレッショルド基準を通過した総合的パースで、発話内の単語はいずれもスレッショルド基準を通過します。私どもはこの種の拒否能力をよくするために多くの努力を傾けておるところでございます。

問　　今のお話は彼が述べられたこととどこか共通するように思うのですが、それは領域を構文的に制限し、省略で可能なことあるいは消去できることを構文的に制限すればするほど、メタリンギスティックな質問あるいは明確化に対して困難を生じるからだと思われます。そしてこうした質問や明確化はこの種の問題の対話で中間体となることが非常に多いわけです。これは同じように処理できるものでしょうか、あるレベルからつぎの……

答　　明確化の質問は文法では別のサブネットになっており、いかなる時点でも使用できるようにしておりました。したがって明確化に関するような質問にはいかなる制約条件もあり得たわけです。ことばを変えて言いますと、これらのサブネットはあらゆるレベルでいつでも発生し得ると予測されていたのです。

問　　先生は制約条件の全レベルを代表するのに有限状態ネットワークをご使用になった

と、承ったように思うのですが、有限状態ネットワークよりも複雑なモデルを使用するのに、先生の方法を拡張するにはどういう方法があり得るかといったことについてお考えはおありでしょうか。あるいはそういう必要はないと考えられますか。

答　　そうですね、私どもでは現在他のバージョンのパーサーを持っています。実はＡＴＲの北さんがインプリメントされたＡＴＲのパーサーと非常によく似たＬＲパーサーを持っているのです。それを現在評価中であります。私どもでは有限状態ネットを使用し、またＢＢＮがトップエンドを生成しつぎにこれをある種の格フレーム状のメカニズムで処理することを提案しているものに非常によく似た統計的文法も持っています。こうしていくつかの異なったことを試みているわけなんです。実のところ実際に評価し終え、結果が出ている唯一のものは、意味論文法から生成した有限状態ネットです。

○　ほかに質問ありませんか。日本語でも結構ですので、まだ十分あります、５分ぐらいまだありますので。

問　　東京大学の藤崎です。ただいまの非常に面白い、興味深いお話、どうもありがとうございます。これはすでに発話と言語、談話知識の総合に立ち入っているわけですが、先生のお仕事はすでに発話と言語、談話知識の総合ですね。しかし、それでは、第３レベルを総合すればするほど、システムは特殊化します。これの将来の方向は、えー、このリソースマネジメントタスクから行きますとですね、私が考えますに、その、ＤＡＲＰＡのつぎのゴールは、もっと広い適用性のあるより広いシステムのようなものに近いと思われるんですが、先生はこの知識処理と言語処理をどういうふうに分離なさろうというのか、あるいはこれは非常に強く結びついているのでしょうか。

答　　私どもは２つのまた別のタスクを行なっております。内部タスクとしてはオフィス管理タスクを行なっております。スプレッドシートや音声計算機、それに音声メールシステムを持っておりまして、今おっしゃられたように新ＤＡＲＰＡコモンタスクがオフィシャルエアラインのガイドタスクなわけです。そしてそうです、私どもは技法、基本的にはこれらの技法を用いて、これらのタスク全部をやろうとしています。それから意味が異なるという理由で他の領域に移るときには、新たな意味論文法やフレームあるいは意味論構造を作り出すということを、間違いなくやらなければなりません。領域意味論の置き場所については現実的な必要を感じませんので、意味論が領域によって変化することに関してはそれほど心配しておりません。心配なのは領域移行のもつ労働集約型の特性であり、半自動的に何らかの世界レベルの意味論知識と英語の構文知識および新たな領域から非常に高速で別の新たな領域へ移行するための新領域からの例に関する手頃なコーパスを得るツールを手に入れようとがんばっているところなのです。われわれは、それは非常に敏感な何かだと思います。それゆえにいくつかの新しいタスクを実行しようと努力しているわけですし、こうした新システムすべてにおいてこれら技法を適用するつもりです。

ありがとうございました。

○　もう一つ質問を受けようと思いますけども。どうぞ。

問　あのですね、予測、ダイアログの予測でうまくいっているというのでおもしろかったんですけども、このトピックスが変化した場合ですね、バイザウェーとか何かそういうような言葉のときには、どのように、予測がうまく、きつくなると、かえってそういうようなトピックが変化したときには認識がうまくいかなくなるような気もするんですが、それはどのようにして対処されているんでしょうか。

答　　わかりました。一つにはわたしは、今お話になったキューフレーズとしての「by the way 」のようなもののほうに話題が向いてしまいがちなのですが、私どもの文法には、確かにそのような性質の句はございます。しかしこれらが正しく認識できるという自信がありませんし。さらにこのような事柄は当面単に層状化アプローチだけでとらえていくつもりです。すなわちわれわれの予測が正しくないことが明らかな場合には、正しいと思われる一般的なレベルに後戻りすることになるでしょう。われわれは実際、層状化アプローチからは離れつつありまして、より総合的な確率特性をここでは使用する努力をしているところです。それは、私はこのことをするにあたって正しい道筋ではないかと思うのです。ですから現在の対話内にとどまる総合的確率もあれば、新たな対話に移る確率もあるし、もし事実キューフレーズを検出できるものがあれば、それはその対話の性質を変えるか、あるいは何らかの明確化制限対話に入る確率を変えることになろうかと思います。それで先生の質問に対する回答としては、私どもが本当にやろうとしていることは、より単純な

確率を使って、層状化フレームワークに対抗するより総合的なものを手にし、こうした確率を変えるのに、われわれが見つけられる知識はすべて用いるということです。だれかが何かを言えることはあり得ない、などというつもりは毛頭ありません。

Wardさんどうもありがとうございました。

○　次に移りまして、次は京都工芸繊維大学の新美先生にお願いします。

新美康永（京都工芸繊維大学）

音声認識システムにおけるアイランドドリブンパーサー

島駆動パーザーというのはですね、音声理解システムの中でよく使われておりますラン島駆動コントロールという方式がありますけれども、その方式における言語処理プロセッサーと、そういうような意味で使っております。

それで私のきょうのお話はですね、一般的にＣＦＧないしはDefinite Clause Grammarで与えられたタスクの文法が与えられましたときに、それを機械的に島駆動パーザーに変換する方式というのをお話し、プロローグのプログラムに変換する方式というのをお話したいと、そう思います。

それでまず初めに音声理解システムでよく使われております２つの制御方式というのについて、簡単にお話をいたします。

それで一つは、音声の、この図は、こちらが時間で、こちらの方向に音声を処理していく途中で種々あらわれてきます仮説、そういうのを書いてあります。それで上のこちらの方の絵はですね、音声を初めから、そのスタートのところからずっと後ろの方に順番に処理をしていく、普通レフトライトコントロールと呼ばれておる方式を描いております。

それでこのシステムでは、ここに点線で書いておりますのは、普通は単語の系列でありまして、まだ文章になっていないもの、そのうちの一番点数の高いものを一つ選びまして、ここのところに言語処理のプログラムが介入いたしまして、この後に幾つか続き得る単語というのを予測いたします。

そしてその単語が単語認識部によって認識されて、幾つかが認識されて、この後ろに新しい単語がつながって、新しい仮設ができると。そしてまた、この全体の仮説の中から、から一番点数の高いものを選んで、同じような方式でやっていくと、そういうものであります。

それに対して、きょうお話しようとしております島駆動コントロールというのは、その考えております単語というのが、この初めからずっとつながっているわけではなくて、途中から切れていてもいいと、そういうものであります。ですからこういう場合ですと、こちらの方向に今申し上げましたような予測をして、それで認識をすると、そういうこともありますし、あるいはこちらの方向に対して予測をして、認識をするということもございます。

ですからこういう部分的な単語列をこちらに延ばしたり、こちらに延ばしたりしながら、音声のカバーする範囲を、全体をカバーするまでその処理を続けていくと、そういう方式であります。

その場合にアイランドと申しますのは、この部分的な単語列と、そういうものを一つ一つをアイランドと、そういうふうに申しております。

それでこの２つの方式を比べてみますと、その言語的な制約条件を課すという点から考えますと、レフトライトの方式の方がはるかにそういう制約条件は強いことになります。しかしこの方式ですと、例えばこのあたりで非常に音声的にあいまいな部分が出てまいりますと、そこから先にこの方式で進んでいくというのは非常に難しいと、そこんところで止まってしまったり、その部分で非常にたくさんの仮説というのを調べなきゃいけないと、そういうことになります。

それに対しまして、アイランドードリブン方式では、もっともらしい単語、アイランドというのを考えていきますので、もしもこの辺で非常にあいまいな部分がありますと、これですと、こちらからこの辺まではうまくいくと。そのほかのところは、こちらがうまくいくというところで、この辺のところは一番後まで残すと、そういうふうな制御方式がとれますので、非常に品質の悪い音声にとっては、この方式はいいということになります。

それで今申しましたアイランドパーザーというのが果たすべき役割というのは、どういうファンクションを持っていなければいけないかというと、３つあると思います。

一つは、単語というのを、まずその与えられた音声全体からもっともらしい単語というのを幾つかスポットいたしまして、それから種になるアイランドというのをつくります。それが一つのファンクション。

もう一つは、与えられたアイランドに対して、その右とか、左とかに、先ほど言いましたように接続し得る、そういうワードクラスというのを計算すると、そういうこと。

それからもう一つ、左から右方向へ進むコントロール方式にはなかったことですけれども、途中で２つのアイランドというのが時間的に接近してきた場合に、その２つが結び合わさって、一つの大きなアイランドになり得るかどうかということをチェックすると、その３つの機能が必要であります。

これから話はContext Free Grammarに限りまして進めたいと思いますが、ある文脈自由文法が与えられたときに、今申し上げましたアイランドパーザーというのは、大体どういう考え方でその３つの機能を果たすかと、そういうことについてお話をして、その次にそれを機械的にプロローグのプログラムに直すにはどうするかと、そういうことについて簡単にお話したいと思います。

それで一般の文脈自由文法というのは、ここに書きましたように、非終端記号からなるルールと、それから非終端記号を終端記号に書き替える、こういう２つのルールとして表現することができます。

それでまず与えられた文脈自由文法に対して３つの関係というのを定義いたします。

一つはレフトリンク、あるいはライトリンクという関係で、もう一つはパート・オブ・スピーチと、そういう関係であります。それでレフトリンクと申しますのは、今C→C1,C2,--,Cnというルールがあったときに、このＣ１というシンボルとＣとの間にはレフトリンクの関係があると。あるいはこのCnとＣというのには、このライトリンクの関係があると、そういうことでありますが、もう少し詳しく述べますと、レフトリンクの（Ｃ，Ｇ）というのは、Ｇというシンボルがこちらに対応しまして、それでこれからルールの適用によって幾つかのストリングが出てきますけれども、そのストリングの一番左端にこのＣというシンボルがあらわれると、そういう関係があるときにレフトリンクが成立すると、そういうふうに申します。ライトリンクについてもまた全く同様で、ここんところがライトに変わったというだけであります。

それからパート・オブ・スピーチというのは、こういうルールが、要するに直接ターミナルシンボルに変換されるような非終端に対してパート・オブ・スピーチと、そういう関係が成立するということにしておきます。

この２つの関係を使って、まずあるアイランドの左とか右に出てくる可能性のあるワードクラスを計算するにはどうするかと、そういうお話であります。

それで今ある場所、こちらはずっと音声がこうあるといたしまして、ある所に単語Ｗというのがワード・スポッティングという操作で認識されたと、そういたします。それからここにありますようなこんなルールというのは、Ｃというのが、Ｃ１、Ｃ２、Ｃ３と、そういうので書き替えられるルールがあると。そうしますと、まずこの単語が与えられて、その次にこのこちら側に、右側にどういう単語が来るかと、そういうことを予測する場合ですが、この場合は甚だ簡単でありまして、このＷというのが書き替えられる、こういうルールによってこのＣ２というのを見つけて、このＣ２を右辺に含んでいるようなルールを持ってきまして、ここにこういうパート３ができるわけですが、そのこれの左側にあるシンボルを取りまして、これからレフトリンクの（Ｘ，Ｃ３）、パート・オブ・スピーチのＸという関係を満たすようなワードクラスというのが、ここのところに来ると、そういうことになります。

これは非常に簡単にできる場合でありますが、次に今の話が少し進みまして、この下にＷというのがあったわけですが、品詞のレベルで話をいたしまして、この今こういうルールが適用されて、それでその次にこのＣ５という単語クラスが予測されたと。これが認識されて、Ｃ２、Ｃ５という島ができていると、そう仮定します。このときに、次にこのこちら側に来る単語クラスを予測したいわけですが、そのときにどうするかと言いますと、先ほどの場合ですと、Ｃ２というのを含んでいる任意のルールというのが適用できたわけですが、今度の場合はＣ５というのを使っていますので、このときに既にこのＣ３とＣ５の間にレフトリンクのＣ５、Ｃ３という関係があると、そういうことがわかっておりますので、その拘束条件がこちらに伝播してきます。

ですからＣ５というのを右辺に含んでいるようなルールのうちで、こういうふうに真ん中に含んでいるのはだめでありまして、これの一番、Ｃ５というのを一番右側に含んでいて、かつこの左辺のこのシンボルと、Ｃ３の間にこういう関係が成り立っている、すなわちこれとこれの間にこういうルールとして結ばれている関係が存在する、そういうルールだけがここの適用、となります。

このパージング方式というのは、基本的にはボトムアップのパージング方式というのを使っておりまして、したがいまして単語予測というのは、本質的にトップダウンの要素であります、トップダウンオペレーションでありますけれども、そのトップダウンだけですと、そのルールの中にリカーシルなルールを含んでいる場合には、無限ループに陥ると、そういうことがありますけれども、我々の場合には、それをレフトリンクとか、ライトリンクという関係を使って避けている、そういうことであります。

次にパーザーを構成するためにアイランドというのをどういうふうにして表現している

かと、そういうことについて簡単にお話をしたいと思います。

アイランドというのは、ここに書きましたように三つ組トライグラムで表現されておりまして、一つはアイランドを構成しているワードクラスであります。それからあと２つは、そのアイランドというのを構文解析しましたときに出てきます構文木に相当するものですが、それをレフトパージングヒストリーと、ライトパージグンヒストリー、そういう形で覚えておきます。

それで例えば、この場合のＣ２、Ｃ５という先ほど出てきました島に対しては、どういうふうにして表現するかと申しますと、ルールの中にですね、こういうルールが与えられましたら、そのルールの右辺の非終端記号の間に、こういうふうなアイデンティファイヤーと称するものを、ルール全体でユニークに決まるアイデンティファイヤーというのを挿入いたしまして、それを使ってルールのどこまで適用したかと、そういうことを表現しております。それでこの場合ですと、右側のこちらをこういうふうにして、右側のヒストリーはＣ５とＣ３と、それからこの後ろについているＩＤ３と、そういうものが書かれております。

それに対して、こちら側は、このＣ２と、それからここに出てきますアイデンティファイヤーの１、そういうもので、この３つのリストとして表現しておる、そういうことになります。

それで、こういう解析履歴を残しておくというのは、なぜそうしておくかと言いますと、音声の場合は、単語列が与えられたときに、一々それを解析し直して、その右ないしは左の方向にどういう番号が来得るかということを計算するのは、しょっちゅう行わないといけませんので、それを一々解析しなくてもいいと、そういうためにパージングヒストリーというのを残しております。

これはルールというあるルールが与えられたときに、それから今申し上げましたようなことを行いますパーザーというのを自動的に生成する、自動的というか、機械的にルールから機械的に構成するわけですけれども、そのときにできてくるプロローグのプログラムというのを示したものであります。

なぜそれが、こういうのを書けばいいかということをお話しますと非常にややこしくなりますので、まずとにかくコンテキストフリーのグラマーが与えられますと、先ほどから申し上げているような機能を果たす単語予測のプログラムというのが自動的に生成され得ると、そういうことであります。

次に島駆動パーザーのもう一つの重要な機能であります２つの島が与えられたときに、それをどういうふうにして結合するかと、あるいは結合のテストをどうするかと、そういうことについてお話したいと思います。

これは２つの島、Ｃ２とＣ３、それからＣ４、Ｃ５というのが与えられておりまして、それぞれの解析履歴、パージングヒストリーがここに書かれている、そういうふうにします。それでそれとパーズしたときのパーズ３というのは、こういうふうになると。この場合は、こちらのこのルールと、この解析木と、この解析木というのを、こういうふうに、このＢ１というのをこの下に持ってくると、そしてそのＢ１からＣ３とＣ４を出すと、そういうふうな形にしてやりますと、ここんところがうまくつながると、そういうふうになっているわけであります。

こういうことを調べます基本的な条件として、解析木の間にどういう構造が成り立っていればよいかと申しますと、まず一つは、ここに書きましたように左側の島の右端と、それから右側の島の左端とが、こういう一つのルールから出てくるこういう２つのノンターミナルシンボルによって構成されていると、こういうことがまず必要条件になります。こういうことが成り立っておりませんと、その初めからこの２つはつながらないと、そういうことがわかりますので、まず第一番初めにこういうことを調べます。そのためには、こちら側の島のこちら側の解析履歴をたどって、こういうＩＤというのを見つけます。それからこちら側の島の右端の解析履歴をたどって、やはりこのＩＤというのを調べて、これがこちらから来たやつと、こちらから来たのとを、同じＩＤに到達すると、そういうこと、そういう条件で、まずこの必要条件を調べます。

それが成り立っていますと、あとはお互いにそのほかの部分がですね、うまく解析木の中に入り込むかということについて調べるわけですが、その入り方といたしましては、これは先ほどの絵と同じものでありますけれども、片方の解析木のある部分が、片方の解析木の下に完全に入ってしまうと。こちらの部分木として表現できる、そういう条件が一つであります。

それからもう一つは、こちらにありますように、こちら側の部分と、こちら側の部分が完全に一つの部分からできてきて、同じ立場で、どちらも、こちら側がこちらの下に入るとか、そういうことではなくて、対等の立場で同じ一本のルートに結びつくと、そういうふうな条件が必要であります。

それから木を結合する場合には、この関係か、この関係か、どちらかしかありませんか

ら、それが、これが成り立って、その次にこれが成り立って、これの後にこれが成り立つ、そういうようなことを順番に調べていけばよろしいわけですから、条件としては、この２つというのがリカーシブに成り立っているかどうかということを調べていく、そういうことになります。

それで今お話いたしましたようなことは、与えられた文脈自由文法とは関係なくて、簡単なプロローグのプログラムで実現できると、そういうことでありまして、これがそのプログラムでありますが、まずこれで２つの島を調べるというプレディケートでありまして、こちら側の左の島と右の島の島の番号ということで、そこから解析木の情報を取り出しまして、部分的にパーズをして、ここんところでＩＤとありますのは、先ほど必要条件のところで申し上げました、両方から来たときに同じＩＤに到達するかという必要条件を調べて、あとそれがうまくつながるかどうかということを、このジョインというプレディケートで調べる。このジョイン幾つかありますけれども、これは先ほど申し上げましたこの３番目のやつが、両方が同じ条件で一つの解析木になると、そういうことを調べるもので、こちらの２つがですね、どちらかの解析木というのが、片一方の解析木の中に完全に埋め込まれると、そういう条件を調べている。このあと残りは一つの解析木というのが片方の解析木の中に埋め込まれたときに、反対側の木に対して、その拘束条件が伝播しますので、それについてその伝播を伝えると、そういう役割をしております。

以上で大体私の話は終わりますけれども、きょうお話いたしましたことは、ある文脈自由文法が与えられたら、それから機械的にアイランドパーザーというものを構成することができると、そういうお話であります。

質疑応答

問　先生のパーサーの複雑性について考えをお持ちでしょうか。この種のパーサーの理論的挙動のことなんですが。通常の文脈自由のパーサーにおいて、ｎの３乗ですが、このアイランドドリブン方法はどうでしょうか。

答　私の方法はたぶんＣＹＫのものと同等だろうと思います。と言いますのは原理がボトムアップパーシングに基づいているからです。

問　しかし先生はこのパーサーについてまだ数学的解析を与えておられませんね。

答　ええ。

問　第２の質問は対話翻訳についてです。先生は増加パーシングスキームをお好みのようですが、それはセンテンスを終わりまで聞く前に部分構文解析をしようということですか。先生のパーサーでこれは可能なのでしょうか。これで増加的にそれができるとは思えませんでした。話し手のセンテンスが完了するのを聞き終わる前にパスを始めることができるのですか。それならこれは真に増加的なパーシング方法だということになります。どうもよくわからないのですが。

答　島駆動パーシングはアイランドドリブンに対して与えられます。パーシングの前に完全なセンテンスが与えられなければなりません。そうしてときにはここでパーズするわけです。ですから完全なセンテンスが必要です。

Session 2

Integrated Processing of Speech and Language 2

森元逞（ＡＴＲ）

Spoken Language Processing in SL-TRANS

ＡＴＲ自動翻訳電話研究所の森元です。

ＳＬ－ＴＲＡＮＳといいますのは、私どもの研究所で現在実験を進めてます日英音声言語の翻訳実験システムです。

大まかな構成をこの絵に書いておりますけども、まず入力された日本語を、先ほど報告のありましたＨＭＭ－ＬＲで音声認識いたします。その後、ちょっと上を後で説明しますけども、認識された結果を解析し、それから英語のコードに変換し、英文を生成し、最終的に英語のスピーチとしてDEC TALKを使って出力いたします。

それでこのシステムにおいて、もう一つコンポーネントがありまして、これはこういうコンポーネントを制御するため、及び後で説明いたしますけども、この音声認識と、それから言語処理を結合するための処理を分担しております。

それで本日は、このシステムにおけます音声言語処理、特に音声認識と、それから言語処理、これをどのようにつないでいるかといったところについて御説明したいと思います。

今までの発表にもいろいろありましたけれども、音声認識をやる場合に非常に重要な問題は、音声認識をやるときに音響的な処理だけで音声認識をやると、非常にあいまいさが残ってしまうと。したがいまして何らかの言語情報を利用する必要があるわけですけども、問題点としましては、どのような言語情報を用いたらいいか。例としましてはシンタックスセマンティクス、または単語の共起関係とか、いろんなレベルの情報を使うことが考えられます。

それからもう１点は、どういうふうにそれを使ったらいいかということです。

それで、これが我々のシステムの大きなジェネラルスキマを示しております。日本語で入力されました音声をＨＭＭ－ＬＲで日本語の文節単位に認識します。このときに、先ほど川端の報告がありましたように、日本語の文節の構文情報、シンタックスが使われます。このＨＭＭ－ＬＲの出力は、各文節に対して幾つかの候補を出すということになります。したがいましてＨＭＭ－ＬＲの出力は、文節に対してラティスを構成するデータ構造となります。この文節ラチィスのデータ構造を入力としまして、その次に日本語の文節間のかかりうけ関係というものを使いまして、フィルタリングを行います。

例えばここの絵ですと、３個ずつの候補が出てきてますが、それに対してフィルタリングを行うことにより、候補数を絞るという処理を行います。それで最終的には、まだ幾つかの候補が残っておりますので、この候補に対して、さらに厳密な文としての構文、意味関係のチェック、それからセンテンス構造のヒューリックスを用いて最終的に正しい文を選択するという処理を実現しております。

それでちょっと補足的な説明になりますけども、ここの部分でも構文情報、文節の構文情報が用いられている。それからここでも当然ながら文としての構文情報が用いられている。問題は、この両者の関係をどうするべきかという問題があります。すなわちその構文情報及びそのバージングメカニズムを全く両者に対して独立に設けるか、それとも共通して使うかといった問題があります。共通して使うというのは、一見非常によさそうに見えるんですが、問題がありまして、まず処理単位が違うと、音声認識と言語処理で処理単位が違う。それから処理の目的がやはりどうしても違うと。処理の目的と申しますのは、音声認識で言語情報を使うというのは、音声認識の制約として言語情報を使う。それから言語処理では、最終的には文としての意味を取り出すといった目的に使いますので、多少目的が違う。それから両者のモジュラリティーの問題があります。

我々のシステムは、先ほどの説明にもありましたけども、入力音声を各文節ごとに独立して音声認識するという方法をとってます。したがいまして、その利用している構文情報も当然文節構造をベースとした文法を利用しているということになります。

言語処理の方の文法ですが、これは全体のやはり意味を取り出す必要があるということで、我々の場合ですとＪＰＳＧ、これはＨＰＳＧの日本語バージョンなんですが、これを用いています。こちらの方がより正しい意味が取り出せるだろうということでＪＰＳＧを用いています。

先ほど述べましたけれども、処理単位が違うということは、音声認識の場合は、文節単位ごとに処理が行われるということで、トリー構造はこういう構造になります。ここがない、これ全体がセンテンスなんですが。ＪＰＳＧをベースとした場合のシンタックス構造がこういうふうになってしまうということで、どうしてもやはりこれを一緒にすることはできない。特に問題となりますのは、ここに点々で書いておりますけども、この構造から

この構造をコンストラクティブリーにつくり出すことができないというところが最大の問題です。

したがいまして、現在のところ我々のシステムでは、音声認識と、それから言語処理は別々の文法を使っておるということを行っています。

次に音声認識のＨＭＭ－ＬＲについては、先ほど川端の方から報告がありましたので、その後の係り受けを用いたフィルタリング処理以降について御説明いたします。

係り受けですが、ここに書いてますように、日本語の文節間の関係、ここに幾つか挙げていますが、例えばプレディケートとその格要素の関係、それから名詞・名詞の関係、それからプレディケートに対する副詞の関係、こういった関係がかかりうけの関係ですが、こういった関係を用いて複数個出てきた文節候補の中から、正しいものと思われるものを取り出して、それ以外のものを捨てるという処理をやっております。

このために係り受けディクショナリーというのをつくっております。これはＡＴＲで収集してます会話文のコーパスですが、そのコーパスから取り出したものです。かつこのかかりうけ辞書の中には、コーパスの中に出現したその頻度も同様に提示されております。この辞書と、それから実際のラティスをマッチングすることによって、フィルタリングを行うわけですが、マッチングの方法としましては、イグザクトマッチではなくて、３レベルからなるマッチング、ベストマッチング方法を採用しています。

まず一番最初はベストイグザクトマッチを行いますけども、その次には標準形によるマッチングだとか、セマンティクフィーチャーによるマッチングを行います。

同じレベルで複数個のかかりうけがあったとしますと、その場合にどちらかを選ぶ必要があるわけですけども、その場合には、ここの式で書いております式でスコアを計算しまして、よい方を選ぶという方式をとっています。

この式ですが、Ｆは頻度です、かかりうけディクショナリーに書かれている頻度、それからＤは、２つの文節間の距離、それからＳは音声認識スコアです。

例えば入力文がこういう入力文、３つの文節からなる文章とします。これらがおのおのの文節に対する候補です。これに対しましてかかりうけフィルタリングを行うことにより取り出された結果がここに示しています。

その次に言語解析にいきまして、先ほど説明しましたように、厳密な構文的、意味的関係がチェックされます。それでもなおかつまだ複数個の候補が残った場合、ここに示しますようなペナルティの計算、評価をやることによりまして最終的に１個のセンテンスを選択するという方法をとっています。このペナルティですが、ここではＮというのはシンタックストリーのノードの数、それからＮＵは必須格のうち、どれだけ省略されているものがあるかという数です。それからＳＸは音声認識のスコアです。

ここに例を示しております。ある入力に対しまして、一番目の文節に対してこの候補が出てきた。それから２番目の文節に対してこの候補が出てきたとします。この２個、２個ですから、合計４個の文としての候補があるわけですが、このいずれの文章も、構文的、意味的には正しいものになっています。したがいまして、その次の処理としてペナルティの計算をやると。まずノードの数の計算をやりますと、このペアが一番ノードの数が少ない。それから同様にこのペアが一番必須格の省略の度合が少ないということから、このペアが選択されるということになります。

それと先ほどのこの式でシンタックストリーのノードの数、それから必須格の省略の度合というのを考慮しているわけですけども、これは別の言葉で言いますと、よりシンプルなセンテンスの方がより優先的に選択されるといったことをあらわしています。

それで最終的なシステムとしての性能ですが、現在実験を進めておりますところの全体の中間段階ですが、これまで得られました実験値をここに示しております。

入力は特定の話者です。入力した文の数37、それから全体の文節数は83、平均的な一文あたりの文節数は 2.2です。それでＨＭＭ－ＬＲですが、第一位に対する正解率が87%、これは文全体としての正解率を計算しますと 、74%ということになります。この値をもう少し高めるためにＨＭＭ－ＬＲでは上位５個をとるということにしています。結果は上位５個で96%ということになります。上位５個をとるということは、一方文としての候補を増やしてしまうということになります。５個とることによって、平均的にですね、一文当たり、一つの入力文当たり28.7の文候補ができてしまうということになります。しかしこれを文節フィルタリングを通すことによって、平均的に大体 1.5個の候補に絞ることができると。すなわち文としては平均的に 2.4個の文候補に絞ることができる。最終的にこれを言語処理することによりまして、37文のうち34文、すなわち92%のセレクションレートをえることができております。

簡単に報告をまとめます。音声認識と言語処理の新しい接続方法について報告しました。今後の課題ですが、さらに統計的な情報を導入すること、またはもっと高度のインフォメーション、例えば対話構造などのインフォメーションを導入することによって、さらに認識率を高めることを考えてます。以上です。

質疑応答

荻野（筑波大学）　筑波大学の荻野と申しますが、この方式、いろんなところに数式が出てきまして、ウエイトがかけられるということですが、最終的なＳＬ－ＴＲＡＮＳのこの92%という率は、それぞれのウエイトなどは、恐らく最適なものにしたときにここまで行くという話でしょうね。そのときにですね、ですからもうこの方法では、これがいわば限界だと思いますけれども、なおかつ翻訳できない、認識できないものが３センテンスありますね。これはどういうものなんでしょうか。それをお聞きしたいと思います。

森元（ＡＴＲ）　３センテンスですけれども、いずれもですね、音声認識の段階でもう既に候補の中に入ってないといったものです。この方法ですと、ボトムアップ的に文を選択しますので、候補の中に入ってないと、それを訂正するということができませんので、どうしてもそれ以上はできないということになってしまいます。したがいまして、将来的な課題ですが、さらにそれを訂正するとかですね、それからこの方法ですと、ＨＭＭ－ＬＲのとこは音韻までの予測はできてますけども、さらに上位の情報を使って、先ほどのウェンワード先生の報告にもありましたように、例えば対話構造を使って単語を予測するとかですね、発音を予測するとか、そういったところまで持っていく必要があるんじゃないかというふうに思っています。

荻野（筑波大学）　今ＨＭＭ－ＬＲのところの問題だという話でしたけれども、今トップ５個を選ぶと96%までうまくいくという話でしたが、例えばそういう失敗するというのは、５個じゃなくて何位ぐらいなんですか。６位とか７位なんですか、それとも10位とか20位なんですか。

森元（ＡＴＲ）　ＨＭＭ－ＬＲがですね、レフトーライトで、左から右にやっていくんでですね、頭で失敗しますと、もうそれ以後全部バーストエラーが起きてしまいます。そこの問題を一つ解決する必要があるんですけども、ちょっとそれはまだ研究している段階です。

○　ありがとうございました。ではちょっと残念ですが時間になってしまいましたので、次の講演に移りたいと思います。

Robert C. Moore （ＳＲＩインターナショナル）

音声処理と自然言語処理の統合

これはＳＲＩの複数の人々による共同研究です。 My Murveitは音声研究プログラムに従事しており、本システムで用いる音声認識装置を主に担当しております。 Mary Dolrymple と John Bearは人工知能センターにおりまして文法を書き、Fernando Pereiraと私とは同じく人工知能センターでパーサーについて研究しております。 Fernando は私とともにパーサーの初期バージョンを研究しておりましたが、数カ月ほど前ベル研究所に移籍しました。皆さんはすでに音声認識におけるある種の言語的制約についてお聞きになっていらっしゃいますが、それは音響的な情報だけを使っておりますと認識精度が十分に得られないからに他なりません。本プロジェクトでのわれわれの主な関心は自然言語上の制約を音声認識に適用する際の効率に取り組むことで、今日まで来たわけですが、こうした制約が、割合に複雑な言語モデルで表現されるとき、われわれの場合、単一化文法なのです。われわれが開発しました方法は、ダイナミックグラマーネットワークと呼べるようなものです。アプローチとしては、先に申しました自然言語パーサーを使用するにあたって単一化文法に基づいて状態遷移ネットワークを増加的に生成しようとするものです。この状態遷移ネットワークはHidden Markov Modelに基づく音声認識装置を用いれば標準のダイナミックプログラミングアルゴリズムで使用可能なものであります。

これがわれわれのシステムの全体構成の概要です。私どものHidden Markov Model音声認識装置は他の多くのシステムと同様二つの主要部に分れております。一つはHidden Markov Modelワードプロセッサーです。これはあるいはワードマッチャーと呼んでもいいかも知れませんが、特定の単語のＨＭＭモデルに基づくフォーンを持っております。ＣＭＵでのシステムと同様私どものＨＭＭは主にTriphoneに基づいています。このＨＭＭシステムのもう一つの構成部は有限状態文法プロセスです。これによってＨＭＭワードマッチャーがある単語モデルの最終状態に至ったときに、仮説がPruneされなければ、文法プロセッサーつまり有限状態文法プロセッサーは、有限状態文法にしたがってつぎに来る可能性のある

単語が何であるかを計算しなければなりません。さて標準ＨＭＭシステムではこの有限状態文法は処理が始まる前に表現されますが、われわれのアプローチではこの有限状態文法は処理の進行とともにまさにダイナミックに計算されるのです。皆さんはすでにお気づきだと思うのですが、だれかがもっと複雑なモデルを一般に状態遷移ネットワークに翻訳しようとしたり、文脈自由文法や文脈自由よりも複雑な何かをお持ちの場合、状態遷移ネットワークは無限になるはずです。それでこの問題を扱うにあたってのわれわれのアプローチは、音声認識装置が生成する単語仮説が必要とする状態遷移ネットワークのフラグメントだけを増加的に計算しようというものです。

こうすることによって音声認識に対する標準H-MMアプローチにおいてダイナミックプログラミングアルゴリズムの効率を利用することができ、また一つには自然言語パーシングでの余分な処理をできるだけ少なくもできるのです。このアプローチが遭遇する大きな問題の一つは、連続音声認識における単語境界の不確実性の扱いです。たとえばＬＲ、失礼、予測ＬＲ法に関する発表で、予測ＬＲパーサーはある単語について可能なエンドポイントの確率ベクトルを扱わなければならないとのことでしたが、このアプローチですとパーサーは単語がどこで終わるかという問題にはまったく関与しなくていいわけです。それはまったく認識サーチだけに限定されます。パーサーがしていることは状態遷移アークを生成することだけなのです。ですからこれが動作する仕方というのは、まずある仮説単語が一定の状態で終わりますと、情報が単一化文法パーサーに伝わり、パーサーがつぎの状態はどうなるか、今の状態から出てくる可能性のある遷移は何であるかを計算するということになります。このようにこれは一つの単語が終わる最初のときにだけ起こることでありまして、その単語がつぎのフレームまたはその後のフレームといった具合に終われば、認識装置が使用する、いや必要とする情報はすでにその状態遷移ネットワークに記憶してあるわけです。追加情報を求めてパーサーに戻るということは必要ありません。ではこれがどうして可能なのでしょうか。状態遷移アークを計算するというのは、普通パーサーが行なうこととは考えられていません。

これが動作する仕方は、状態遷移ネットワークにある状態は実はパーシング構成集合に対する一つのタグであり、これを自然言語パーサーが追跡するというものであります。つまりパーサーはこうしたタグを作り出してある一定のタグに対し、これに対応するパーシング構成が何を行なうかを示すインデックスを持つわけです。パーサーはこのタグを音声認識装置に渡し、音声認識装置がこれを、失礼、音声認識装置が終わった単語が、パーサーに戻される状態で始まったということを検知したとき、パーサーはこうして作出されたテーブルの中でタグが対応するパーシング構成が何を行なうかを調べます。これには仮説単語を用い、タグに対応するパーシング構成の一つ一つを前に進めます。音声認識装置が仮定した単語ごとにこれを進めるのです。結果としてできたパーシング構成の個々に対して新たなタグが生まれ、この情報を音声認識に戻します。ですからこの技法には、標準ＨＭＭに基づく音声認識アーキテクチャーに対して何ら実質的変更を加えなくともよいという利点があるのです。そしてわれわれはこの利点をまさに活用しているわけです。と言いますのは、私どもはＳＲＩにおきましてまた別のプロジェクトを持っておりまして、そちらでは標準ＨＭＭに基づく音声認識を大量語彙を用いてリアルタイムで処理するための専用ハードウエアを構築中なのです。この新アーキテクチャーを開発するにあたって、ハードウエア設計に対して必要な変更は、単にこの専用マシンの状態遷移テーブルの中の状態遷移追加アークについて増加的入力つまり書き込みを許容するだけですんだのです。

ですからこれはハードウエアにとっては最小限の変更でしかなかったわけであり、われわれはほどなく自然言語処理ができるこの総合アーキテクチャーをサポートするリアルタイム専用ハードウエアを手にすることになるでしょう。このアーキテクチャーの最初のインプリメントは、私どもで去る２月に行ないましたが、Ｓｕｎ４上で当システムを走らせたところ、本日すでに皆さんがお聞きになられましたＤＡＲＰＡリソースマネジメントの試験材料からいくつかのサンプル文をとり出して、これらについて走らせましたら、センテンス当たりの平均パーシング時間は約２分でした。これら試験材料を用いて行なわれた研究で私どもが存じております唯一のものでは、単語ラティスパーシングアプローチを用いて行なわれておりまして、この研究に従事しておられた方々からお聞きしたところによりますと、シンボリックＬｉｓｐマシンでセンテンス当たり２～３時間かかったということでした。確かに私どもではやや高速のハードウエアを使用しておりますが、おそらくこの適用に対しては３ないし４のファクターで、われわれは確実にわれわれが知っております他の結果に比べまして大きな高速化を達成しつつあったのではないでしょうか。

さらにもう一つ指摘させていただきたい点がございます。それは今の特定の場合ですが、われわれは構文的制約のみを含む文法を用いておったということ、そしてこの試験材料では文法なしと比較してエラー率がほぼ１／３に減少したということ、またある特定のアコースティック単語モデル集合では、われわれの音声認識装置は82.6%の単語認識率を達成していたのですが、構文的制約だけを用いてもこの試験ではこれを88.4%にまで向上させ

られましたが、これはエラー率にでも約１／３の減少です。で、これがわれわれの初期インプリメントでありました。その後の何カ月かにわたってパーサーと総合アーキテクチャーに関する研究をかなり行ないまして、改良パーサーと全く同一の文法、全く同一の単語モデルによる全く同一のタスクのためにわずかに修正を加えた総合アーキテクチャーとを用い、パース時間をセンテンス当たり約２分から12秒に短縮することができました。私どもはこれを顕著な改良と考えております。それでこれからお話し申し上げる内容は主に、この改良がどこから来たかという点であります。この 131秒から12秒へ至った改良には、実は２つの主要な要因があります。一つは11というファクターでの高速化です。この高速化のほとんどはアーキテクチャーをわずかに変更したことです。プレディターというよりはむしろフィルター……。初期インプリメントではわれわれはパーサーのイメージをまさに文字通り状態遷移ネットワークとしてとらえておりました。ですからある単語がもたらし得るパーシング構成の集合を計算したそれぞれの時点で、われわれは単語予測の集合をもまた計算していたことになります。その後の研究ではこのアーキテクチャーを変更して、パーサーが現実には生きた状態、つまりある単語が受け入れ可能で、その状態に対してあるパーシング構成にある全部状態に対して全単語を予測するようになりました。

それからこの変更がそれ自体で約５のファクターで高速化を実現していることもわかりました。この変更の以前では計算時間の80％を予測に用い、わずかに20％がパーシングに割り当てられているのが現状だったのです。ではこの高速化の原因はどこにあったのでしょうか。われわれが構文的制約を用いている限りでは、実際アコースティックスのほうが構文論より、つぎに来る単語が何であるかを予測するプレディクターとしてはるかに優れているでしょう。それで私どもの音声認識装置を見てみますと、アコースティックに可能な単語仮説は可能性のある語彙のわずかに数パーセント、そうですね、５％に満たなかったのに対して、構文論だけを用いても、平均すると少なくとも全語彙の半分は可能でした。すなわち 1,000語の語彙に対する言語モデルは、わずかにほぼ［？］のパープレキシティを有していた……［it］はそのパープレキシティを 700くらいに減少させただけでした。これは Ken Church がこのテーマ関する発表の中で何度も指摘した点であります。ですからもっとも制約の大きい知識ソースを最初に適用する通常のヒューリスティックを用いますと、アコースティックを用いつぎに構文をフィルターするほうが、構文を用いてそれからアコースティックでフィルターするよりも、効率的であることがわかりました。もう一つ行ないました変更はパーサーをスタックベースのものからむしろ状態ベースのものに変えることでしたが、これが何を意味するかはつぎのスライドでご説明いたします。

時間の制約がありますので、パーシングアルゴリズムの詳細については多くを述べることはできませんが、こんなふうに表現できるでしょうか。われわれが採用しました初期アルゴリズムは予測ＬＲパーサーとみなすことができます。ここではＬＲパーシング状態はプリコンパイルされるというよりはむしろダイナミックに計算されます。複雑な単一化文法を用いると、ＬＲパーシングテーブルをプリコンピュートすることは、計算論的に単純には可能ではありません。一般的に単一化文法ではＬＲパーシングテーブルは無限になるはずです。私どもの文法の場合、たまたま結果的にパーシングテーブルが有限となる文脈自由文法と同等となるような文法を書いたわけですが、それは極端に大きくなるものと思われます。パーシングテーブルの大きさを推定するいい方法を持っていませんが、比較的単純な文法のためのはじめの頃の実験で、われわれは文脈自由文法の翻訳では生成される可能性のある、明らかにノンターミナルなシンボルをすべて計算したことがありまして、それは１万のオーダーでした。一般的に言って、ＬＲパーシング状態の数は文法サイズに対して指数関数的な関係にあります。ですからわれわれがかりに何万ものノンターミナルなシンボルを文法中に持っており、ＬＲパーシング状態がそれの指数関数であるとしますと、それは巨大なテーブルになることがおわかりいただけるでしょう。ここでキーになりますことは、パーシング構成、失礼、パーシング構成が何において成立しているかを考えますと、それはスタック、しかも入力のある初期セグメントを解析する可能なカテゴリーのスタックであります。

このことはまたＬＲアプローチでもいえることです。改訂版パーサーではわれわれの当初のシステムがそうであったような、そしてＬＲシステムが現在われわれの当初のシステムがそうであったような、そしてＬＲシステムが現在そうであるようなスタックベースであるよりはむしろ、われわれは効果面では位置ベースであるようなパーサーとなるものに変更しましたが、その位置は入力信号中の回数（times)ではありません。これが単語ラティスパーシングのようなアプローチでは大きな問題となります。さらにこの単語ラティスパーシングでは入力信号のそれぞれ異なった各位置が異なった仮説としての扱いを受けることになるのですが、われわれはむしろこのパーサーの位置として状態遷移ネットワークに対して生成される状態を用いました。ですから一般に、別のパスを通って同じ状態に到達できるからという理由で、全単語仮説は一般に何らかのネットワークをもたらすような生成の仕方をされるのだと考えられるのであれば、それは一つのツリーだと、つまり音声

認識装置が生成する単語仮説のツリーだとお考え下さい。われわれのスキームではこのツリーのノードはそれぞれ状態遷移ネットワークの中のある状態に対応します。ですからわれわれが行なったことは本質的には、パーサーを書き換えて、単語仮説の入力仮説のツリーの中のこうしたノードをかなり標準的なレフトコーナーパーシングアルゴリズム中の位置として使用しようとしたことなのです。さてわれわれの元のパーサーすなわちＬＲスキームではスタックよりもこうしたノードのほうが少なかったわけですし、一般にこれらノードの一つ一つはいくつかの異なったスタックによって到達可能でありますから、われわれは改訂版パーサーではデータ構造をはるかに少なくできたのです。なぜなら元のパーサー中のスタックいくつかに対応して、新パーサーでは単一の状態を持つからです。ですから結果として、はるかに少ない数のデータ構造を生成できたわけであり、パーサーは２を越えるファクターで高速に動作するようになりました。

さて去る２月から現在までの間に、いくつか他の変更も加わりました。時間の都合上そのうち一つだけお話し申し上げたいのですが、それは文法の適用範囲を大幅に増大したという点です。元々のタスクは、後の測定でＤＡＲＰＡプログラムで用いられているコーパス、つまりリソースマネジメントコーパスの36%しかカバーしていないことが明らかになった文法を使って実行していたのです。われわれはこの文法を拡張して、現在では91%の適用範囲を持つに至っております。

それではこの結果についてですが、パース時間は 3.5のファクターで増加しました。これは不運なことですが、がまんできないほどのものではありません。一つ非常に重要なことがあるのですが、それは文法の複雑性や大きさにたいしてこの種のアルゴリズムをどのように数量化できるのかを測定しなければならない点です。そして理論だけで行なうことのある解析では単により劣悪な数字しか出てきません。このことが実用上何を意味するのかを明らかにすることは大変です。しかしこの場合も経験的にはパース時間を少し、過剰ではなく、犠牲にすれば文法を向上させられることがわかりました。

それでは最近の成果を少しご披露して発表を終わらせていただきます。これは別のセンテンス集合に関するものです。二つの異なったプルーンスレッショルドでのパース時間は、センテンス当たり約１／２分から１分のものです。

ここでもまた非常にパープレキシティの高い文法に対してエラー率で約１／３の改善が見られるわけです。つまり文法の複雑性の上昇とパーシングアルゴリズムとの間には、以前ずっと適用範囲の広い文法を用いていたときに比べて２から４のファクターでの改善を行ないつつあるのです。

ここで今後私どもで行なおうとしている計画について少しお話させていただこうかと思います。私どもは究極的にはリアルタイム性能に関心を持っておりますので、パーサーのスピードをまだ向上させるために多くのことをなさねばなりません。

この分野で計画していることは一定の単一化ステップをプリコンパイルする可能性を探ることでして、これによってシステムが動作中に単一化ステップをコンパイルしなくてもいいようにしたいわけです。もう一つ興味を寄せておりますことは、文法の中にもっと制約を入れたいということです。文法と申しますのは単一化文法です。これは非常に強力なフォーマリズムですので、構文的制約は言うに及ばず多種類の意味論的制約を表現することができます。ですからわれわれが実験してみたい一つが、音声認識用の文法に意味論的制約を含ませることということができます。またこうした解析上の構文的制約や意味論的制約と統計情報とを結合する方法も探っております。といいますのは、これを音声認識における一つの制約ソースとして無視しないことが非常に重要だと思うからなのです。

ありがとうございました。

質疑応答

問　先生は最初に使われた単一化文法が本質的には文脈自由文法だとおっしゃいました。

答　いえそんなことは申しませんでした。

問　最初のですが。

答　いいえ、申し訳ありませんが。何かと取り違えていらっしゃるように思うのですが。ただ同等なものではあります。

問　それが私の言わんとするところです。それで仮に先生のおっしゃった文法の複雑性を変えれば、適用範囲を広くできますね、第２の文法でですが。これもまた文脈自由文法と同等だったのでしょうか。というのは私の質問はサイズをスケールに表すと一方でパーシング時間が落ちるということになりはしないかという点にあるからなのです。もちろん

文脈自由文法と同等ではない完全単一化文法を実際にお持ちの場合の話ですが。

答　　答えは「イエス」です。第２の文法でもまだ自由文法と同等でした。しかし私は、本来的に非文脈自由である単一化文法を使ってもパーシング時間は悪影響を受ける、とは必ずしも考えません。その理由はパーサーはそのことから何ら有利性を引き出そうとするものではありませんし、事実そんなこともなかったということです。文法が文脈自由相当だという議論は基本的には、全素性値の大きさに対してたまたま一つの境界線があったと帰納することによって一つの証明となります。英語ではネスティングされるギャップフィラーの係り受け関係の最大数は２のようです。これが英語の文法において無制限な情報量に対してもっともらしい議論のあるところとして、私が知っている唯一の場所であります。経験的には２より大きな値は決して必要ないようです。それで文法には物事を無制限に増やせる力を持ったメカニズムを利用したのですが、経験的事実として、これらがこんなふうには使えないという結果に終わったのです。こうしたことから私は、本質的に非文脈自由文法であった何らかの構造を仮にお持ちだとしても、それが逆のインパクトを受けることはないだろうと思います。

Stephanie Seneff（ＭＩＴ）

音声言語への適用のための自然言語システム

これが私の発表のテーマです。ではさっそく本題に入らせていただきます。まず私どもで最近開発いたしましたＭＩＴのVOYAGERシステムを紹介いたします。これはこれだけで一つの独立した音声言語システムを構成しております。ユーザーはこれに向かって話をしますと、システムが応答します。システムからユーザーに追加情報とかそういったものを求めるときのためのダイアログも含まれております。このシステムの対象はナビゲーションアシスタンス（行き先案内支援）です。つまりある地域の二つの地点間の方向をたずねることができるのです。また電話番号、レストランで食べられる料理や住所とかそういったこともたずねることができます。出力はグラフィックスの形で出てきます。

地図の画面表示があり、合成音声と文字の形での表示もあります。ここにいくつか例を示しました。この領域でのセンテンスはもっとも近いホテルからＭＩＴはどれくらい離れているかというものです。「日本食が食べられるレストランでハーバードスクエアの近くにあるものを知っているか」となっています。語彙は小さく、 300語以上、 300語をわずかにこえる程度です。基本的には今のところおもちゃに毛がはえたくらいのシステムですが、完全な統合化を探求するには非常に有効です。さて私の本日のトピックの「ＴＩＮＡ」ですが、これはこのシステムの自然言語構成部で、ご覧の通り自然言語は等しく重要な役割を２つ持っています。音声言語システムではその一つが認識装置のための制約を与えることです。この役割がどれくらいうまく行くかは要因の数によります。要因とはたとえば、システムがどれだけうまく意味的制約を取り入れられるかといったことです。理想を言えば、構文的にも意味的にもきちんと形が整ったセンテンスだけを受容するようにしたいわけですし、こうすることによって制約条件を最大限に活かすこともできるでしょう。もしこれができれば、標準的なバイグラムあるいは単語ペアモデルから得られるよりも多くの制約を有意に得られるでありましょう。そうすればもちろんバックエンドが代表する意味をも付与しなければなりませんが、これには応答を得るためのいくつかのファンクションコールが必要になってきます。

では私どもの設計理念についてですが、それは認識と自然言語の間で完全な統合化を促進するようなシステムを開発したいということなのです。そしてわれわれの観点は、仮にわれわれがこれまでに解析してきたこの単語シーケンスが与えられたときに、文法の方でここに続き得る実際の単語が何であり、それらに付随した確率がどれほどなのかを判断できればよしとする、ということです。そこでわれわれは文法がこの種のことをできるような設計をしました。そして構文、意味について均一な扱いを実現しました。それはどういうことかと申しますと、パースツリーの中に意味ノードがあるということであり、意味ノードだけでなく構文ノードもあって非常に似通った方法で扱われており、さらにシステムが構文的制約を扱うときと非常に似通った方法で扱う意味的制約もあるということです。したがって全体プロセスはずいぶん単純で、構文と意味とが完全に統合されます。

私はこれまでに何度か「T Node」について話してきましたが、活字になった記事もあります。もっとお知りになりたい方は論文の参考文献に出ております。今日は制御ストラテジーについてはあまり多くをお話しいたしません。ただそれは開発者が与えるひとまとまりの文脈自由ルールとして始まるとだけ申し上げておきます。これらルールは自動的にネットワーク構造に変換されますが、このネットワークは効率的なインプリメントにつながる膨大なルールの共有化を必要とします。

ネットワーク内のアークは一群の例からそのまま自動的に学習される確率を持つように

学習されます。それからルールというよりはむしろノードのレベルでエントリされる構文的・意味的制約があります。将来的にはこの問題についてもう少し詳しく突っ込むつもりですが、基本的にはこれは重要な点だと思います。こうした制約ですね。ノードは一つの主語のようなもの、主語ノードと言ってもいいでしょうか。ノードは、ボトムアップもしくはトップダウンサイクルを行なっているときにシステムを通して出てくる素性に適用可能な制約を持っています。そしてこのノードはこれら素性をつぎのノードに渡していく中で、素性に変更を加えることができるのです。しかしこれは特に問題ではありません。この変更を行なうときにそれが実際どのルールにはまっているのかということはどうでもよいのです。つまり一種の文脈依存制約メカニズムが適用されているのです。文脈自由ルールもそうです。ですからシステムはノードを気にしませんから、ずいぶん単純になります。主語ノードは様々なルールに現われます。制約条件を適用するときには自分がどのルールの中にいるのかは気にしません。たとえば主語ノードは情報を子ノードに転送するときには格をノミナティブに設定するでしょう。続いて子ノードは全代名詞を制御してノミナティブな格を持つようにするでしょう。すると「me」や「them」の代わりに「Ｉ」や「they」を得ることになります。そしてこのことはこの主語ノードがどのルール中にあるかとは関係なしに起こります。

つぎにここでパースノードの中身は何かについてほんのもう少しだけ詳細に立ち入ることにします。まずパースノードにはもちろんパースツリーの内部に向かうポインターがあります。ですから自分がどこにいて、親、子、兄弟姉妹はどれかがわかっています。第２に、前述しました、将来使われる、異なった目的をいくつか持つ目的語を指すバックポインターが二つあります。これらはもともとギャップメカニズムを扱うのに導入されたものでしたが、意味的制約を適用するのに非常に便利なものでもあります。それでわれわれはこれをこの目的にも用いました。

ではこれらがどのように動作するかについてほんの少しだけ説明します。もちろんこれはこの時点で手に入れることができる意味情報にしたがって設定される一連のビットも持っておりますし、意味データを代表するビットも別に持っています。重要なことはコンパニオン文法を指すポインターを持っているということですが、コンパニオン文法は、ノードがこの時点でツリーに導入できる各種制約を指定します。

こうした制約は情報がつぎのノードに渡されていくにしたがってここにありますこれらエントリすべての変更に相当します。もし文法ノードに制約がなければ、これは実際によくあることなのですが、その場合には行なわれることはこの情報をすべてそのままつぎのノードに、トップダウンサイクル、つづいてボトムアップサイクルともに、すべてそのままつぎのノードに渡すだけとなります。これは非常に単純なメカニズムです。

はい、パースノードの一例がこれです。これは非常に単純なセンテンスで、「ＭＩＴはどこですか」ですが、いくつかの点を説明するのにこれを用いることができると思います。まずこれらノード自身はすべて、これらはパースノード、センテンス、Ｑ主語、ここでＢＥ質問、これらすべてです。これらのうちいくつかはターミナルで、ここの場合はポイントはパースノード以外の実際の語彙にあります。文脈自由ルールがあるのがおわかりいただけるでしょう。というのはセンテンスはＱ主語に行き、ＢＥ質問がこれに続き、ＢＥ質問は主語ごとにファイルをリンクしに行き、これに述語付加詞が続きますが、システムはこれらルールを本当は知らないでいます。その必要はありませんし、動作とともにこれらルールを作り出しているのだということも知りません。知っているのはただノードだけなのです。

さてもうひとつ指摘しておきたいことがあります。それはこれらノードのいくつかは意味的であり、ツリーの下のレベルに下がるにしたがってより強く意味的になります。私はこれがいい方法だと信じていますが、それにはいくつか理由があります。これは意味論文法とは異なります。なぜならこうした高度な一般ノードがありますし、これら構文レベルでの複雑性を表現することができます。だれも各意味カテゴリーごとにそれぞれこの複雑性を表現していきたいとは思わないはずです。そんなことをすると組み合わせが多すぎてパンクしてしまいます。ここでは構文レベルにとどめておきたいところではないでしょうか。これでもこうした関係は表現できますし、例えばアウトトレースの低い方のレベルにある意味ノードを予約しておくだけです。

さてこれらの用途の一つとして確率指定があります。といいますのはこれは非常に大きな意味があるのです。興味の対象が今の場合例えば、「where is」という、こんな語順が与えられたときに言わんとするところだとしますと、「ＭＩＴ」がつぎの語として選ばれる見込みはどれくらいでしょうか。本システムがこの見込みを表現する方法は、ここにこうして示しましたパスに沿った条件付き確率、Ｐ１、Ｐ２、Ｐ３、Ｐ４の積がリンクの確率となります。このＰ１がリンクの確率です。「to be 」という語はＢＥ質問の文脈での主語が続きます。もちろんこの確率はかなり高いのですが、「it」や「there 」といった他の可能性もあります。そしてこうした確率はすべて親子の間の確率、つまりその主語が

ある場所から始まる確からしさ、その場所がある学校から始まる確からしさ、そして最後にその学校がＭＩＴで始まる確からしさを表します。ですからこうした条件付き確率で「ＭＩＴ」が続く確からしさを表現することは理にかなっているわけです。さらに指摘しておきたいことは、この場所の確率は、親が何であるかにかかわらず、ある場所のエントリすべてに対して計算されるという点です。つまりセンテンスの中の異なった構造上の役割を横切る確率のプールがあるということなのです。もちろんこれらがまったく同じ確率になることはないかもしれませんが、データを平滑化して、まばらなデータという問題を回避するのにはいいメカニズムだと思います。

それではトレースメカニズムについて簡単に触れてみたいと思います。このセンテンスにおいて「where 」という語があり、文頭に位置しておりますが、これは実際にはこのＢＥ質問における述語付加詞にあったこの自然な位置から引っ張り出されたものです。ですから「Is MIT in Cambridge 」があるとすると、それは「Is MIT where」と同じ構造となります。しかしもちろん英語ではＷＨで始まる疑問文で質問するときには、ＷＨの語を前に引き出しますから、文法の仕事の一つがこれをその自然な位置、ここ、に再挿入するということになるわけです。本システムでこれを行なう方法として、現焦点とフロート目的語を用います。この特定の場合にここで起こることは、文法ノードがそうするように指示するからですが、このＱ主語が、どうすればフロート目的語の位置に自分自身を入れられるかがわかるということです。実際このノードを現焦点スロットに入れてみます。すみません。これが情報を全部ＢＥ質問に渡しながら……。ＢＥ質問はアクティベーターと呼んでおりますノードです。これが行なうことは、これの残された兄弟姉妹が何であろうと、現焦点に何があろうとそれを取ることを認識し、それをフロート目的語スロットに移すことです。ですからこの時点ではこのＱ主語はフロート目的語の位置に入っているでしょう。一度ノードがフロート目的語の位置に入ってしまえば、どこかに吸収されねばなりません。なぜならノードはトレースになったからです。このフロート目的語を吸収できるＢＥ質問の下のここにあるノードを見つけられなければ、このパースは失敗ということになります。

この場合は幸運なことに吸収体であるこの述語付加詞があり、これで一致をチェックします。チェックするものは意味的一致で、ノードがあればこれを吸収できるということを認識します。こうしてこの付加詞を受け入れます。それからこの主語を示し戻して、入力文字列からは何も取り出す必要がありません。

では意味的制約ですが、すでに申し上げたように、現焦点とフロート目的語は意味的制約を与えるのに非常に便利だということがわかりました。そして私は実際に一つの原則を決めようと試みました。つまり私はこれら２つのノードを使用することしかしない、というものです。この原則を守るにあたっては問題解決を求めてパースツリーを振り返るまいと思います。私は自分の制約がすべて後戻りすることなしにこのノードに局部的に適用され得ることを望みます。ですから基本的にはこれらは第２オーダーのメモリーシステムを可能にします。これらはこのセンテンスで先に述べた２つの中心概念を示し戻すのです。例えば「What street is the Marriott on?」というセンテンスを持っているとします。これは私がここの「on street 」と呼ばれるリードに到着する時点での話ですが、こうしますと私は Marriott という現焦点にエントリを持つことになり、「What street 」というフロート目的語にエントリを持つことになります。こうして実際これらの意味カテゴリーがこのセンテンスに対して適切かどうかを確認し、このセンテンスは適切なので生き残ります。

もし「What bank is the Marriot on?」というようなセンテンスを持っているとしますと、このトレースは意味カテゴリーとして「street」でなければならないということを認識します。銀行を受け入れることはできませんから、パースは失敗することになります。同じようにここでも、「What street is Cambridge on?」ですが、この時点で「on street 」ノードはあるカテゴリー地域にある現焦点を持つことはできないということを承認するでしょう。地域は路上にはあり得ませんから、このパースも失敗します。それでこれがわれわれがこの２つの目的語を使って意味フィルターをかける方法なのです。さて現焦点は動詞が提示されるときにはいつも必ず一つの主語を含んでいますから、この時点で主語・動詞の共起関係を実現することは非常にたやすいのです。そして例えば動詞のインターセクトがその主語が「street」というカテゴリーであることを要求します。そうすると説明しましたように、ノードはただこれらの制約をこれら２つの局部的に使用可能なパラメータに適用するだけなのです。

さてこのパーサーには生成能力があり、これは非常に便利だということがわかりました。特に文法のデバッグや形の悪いセンテンスを見つけて、修正やドローイングボードに戻ることによりこれらを文法から取り除いたり、間違いが何なのかを検出したりするのに役立ちます。また特に意味の変則が現われて、これら不都合を除去するのにどういったフィルターを用いればよいかが、多くの場合非常にはっきりします。生成されたセンテンスのい

くつかについて、ここに何例かあります。これらは連続して生成された５つのセンテンスですが、これらはほとんどの部分、構文的にも意味的にも正しいということがおわかりいただけるでしょう。

Do you know the most direct route to Broadway Avenue from here?

Can I get Chinese cuisine at Legal's?

I would like to walk to the subway stop from any hospital?

これは少しおかしい。なぜなら「from any hospital 」つまり「どの病院からでも」となっています。

Locate a T-stop in Inman Square.

What kind of restaurant is located around Mount Auburn Kendall Square of East Cambridge?

さてわれわれはこのパーサーを最近評価いたしました。特にこのバージョンのパーサーですが、これはあるセンテンス群についてVOYAGER領域で動作するものですが、このセンテンス群はコンピュータとの対話形式でシステムを用いて実際の主題から作り出しておいたものです。ループ中にある係りを一人裏の部屋に用意しておきまして、彼がバーバチムで人々がしゃべることをキー入力します。つまり皆さんが認識装置をお使いになるときの問題はわれわれにはなかったわけです。ただ問題だったのはこの人たちのターンオーバタイムが非常に速いということ、それはもちろん当然あり得ることだったわけですが、そしてその人たちはできる限りその領域内で自然発生的な文を作り出したのです。
われわれはシステムがどういうものか概要を説明し、システムが知っている特定の対象物やレストラン、ホテルのリストを与えました。もちろん彼らはこのリスト外に出ることもありましたが、大半は自分たちが知っている対象物の語彙内にとどまる努力をしてくれました。ただ言語を用いるときの使い方は自由にしていいことになっております。

こうして約 5,000のセンテンスを集めたわけですが、これらセンテンスのサブセットを作って、 3,000以上の大きな学習文と 560の独立したテスト文に分けました。この 560センテンスについては私は見ておりません。こうして私はシステムを２度評価したのです。最初は初期VOYAGERシステムを使いました。これはデータ収集に用いていたものです。これらテストセンテンスを評価してからつぎに学習センテンスを見ましたが、これらセンテンスに現われる用法を反映したルールをいくつか加えて、カバー率を拡張するようにしました。それから先へ進んで、テストセンテンスを再評価し、状況を改善できたかどうかを見たのです。私の評価にはカバー率とパープレキシティの両方を含みましたが、パープレキシティは２つの方法で測定しました。一つはある特定の単語に続き得る単語はいずれも等しくありそうだと仮定することでした。単語はすべて等しくありそうだというわけです。このことについては説明しなかったですが、両方の場合についてこれら 3,300センテンスでこれを学習したのです。ですから学習文でこれら確率を学習して、確率を組み入れたときにパープレキシティの計算で用いた確率を得ようとしたわけです。

そしてこれがその結果です。テスト文については、前述しました初期システムはカバー率拡張以前のものです。これには確率をまったく考慮しなかった場合に20.6というパープレキシティがありましたが、確率を考慮することでパープレキシティは 7.1に減少しました。つまり確率によってパープレキシティにファクター３の改善が得られ、センテンスのカバー率は69％となります。つぎに私は文法を改良しましたら、当然のことながらパープレキシティは20から27に上がり、そしてここでは確率を考慮してわずかに１だけの上昇で、8.3となりましたが、カバー率はうれしいことに７％上昇して76％になりました。これは本当に喜ばしい結果でした。というのはユーザーは実際しばしば領域外に出たために、絶望的なセンテンスサブセットが現われて、まさか取り扱えるとは思わなかった、つまりバックエンドがこの種のセンテンスを扱えるとは思わなかったのですから、これはかなり満足できる数字なのです。

さて私は学習文も見ておりますが、これは推定値がどれくらいのいい線を行っているかに興味があったからです。つまり学習文でどれくらいいい確率推定値が得られるかということです。それでわかったことは、この場合、確率いや失礼、パープレキシティは25で確率なしではやや低くなりますが、確率ありのときとほぼ同じと言えるでしょう。ですから今の学習文での確率については特に何か学習的なものが得られたとは言えないと思われます。またカバー率の方はややよかったのですが、ほとんど同等の78％でした。

さてここにありますのがわれわれの今後の計画です。まずＴＩＮＡが私どものシンボリックのＬＩＳＰマシンに乗っております。認識装置は私どものＳｕｎ上にありますから、この状況では全面的な統合化には大きな問題です。

最近やっとＳｕｎのためのコモンＬＩＳＰパッケージを手に入れましたので、研究所の方ではＳｕｎに自然言語システムを移植する仕事をしているところです。これができればわれわれは全面統合の実験を行なえるようになります。スピードを上げるために文法をＣに移植することも必要かどうかまだわかっていません。現在でもかなり速いのです。今の

ところシンボリックＬＩＳＰマシンで部分的にでもリアルタイムで走っています。一般にＳｕｎに移植すると高速化できます。これにはデターミニスティックな入力を用います。それでわれわれはノンデターミニスティックな入力でなんとかして全面統合にもって行こうとしても、スピード的には大丈夫だろうと思っております。統合についての第一歩ですが、当面われわれは単に今あるシステムがあるだけです。認識装置は単一出力を出し、パーサーに渡されます。ここで認識が間違っておりますと、パーサーは通常失敗です。しかしわれわれはすでに認識装置でＮベストアルゴリズムを行ない、Ｂｉ－ｇｒａｍ文法を用いてわれわれが今手にしつつあるＮベストアルゴリズムを制御する方向で、対策を講じてあります。私はジェネレーターを使ってＢｉ－ｇｒａｍ文法を開発してきました。それでもしこのモデルをお持ちであれば、文法は確率を持ったモデルを持つことになります。これらの確率を使ってセンテンスを生成することができます。何千、何万というセンテンスを生成して、そしてＵターンしてこれらのセンテンスをＢｉ－ｇｒａｍモデルを生成するのに使います。

われわれはこのアプローチを採用していますが、Ｎベストアルゴリズムで使用できるＢｉ－ｇｒａｍモデルを手に入れることに関する限りでは、動作するようです。ただ私の最終目標はこれなのです。私はＮベストアルゴリズムが長期的な解決になるとは思いません。全面的統合を実現するのが私の望みであり、それをベストファーストサーチとシステムを実際に駆動するパーサーを使ってやりたいわけですが、これは一度単語シーケンスを与えられば、パーサーがつぎに来る単語が何か、それらに関連する確率はどうかを教えてくれ、認識装置が働いて、アコースティックスに対抗してこれらの単語を採点する（score)、という意味においてです。以上が目下私が従事しておりますことがらです。

これで発表をおわります。

質疑応答

問　ただ今のご説明で、このシステムにおいて分離システムから完全総合システム移って行く様子を伺いまして、実に美しいシステムだと思いました。私がお伺いしたい点は、知識のことなんですが、特定知識を一般知識へ組み込むのに、先生のプログラムの場合どういうふうにおやりになったのか、そこのところろご説明いただけると、おっしゃるポイントが広げられるように思うのです。今はVOYAGERをお持ちですが、その適用を広げるにはどうすればいいとお考えでしょうか。

答　おっしゃることは確率の問題ですか？

問　そうです。

答　私の考えでは、このシステムはかなり移植性が高いと思うのです。私はすでにＴＩＮＡといくつかの異なった領域に移植しておりますが、移植時間は相当速いです。これはその方法にあると思うのですが、私は文脈自由文法を用いておりまして、これは完全にクリーンに書いてあるのです。これらルールには制約はただの一つもありません。Ｘはただ ＡＢＣに行きます。ですから書くのは簡単なんです。そして高いレベルでは構文的です。意味情報を加えたいときには、それをとにかくシステムの中に入れます。一番適切な方法はパースツリーの下の方のレベルに入れてやることだろうと思います。こうすることによってパースツリーから直接ファンクションコールへ、そしてバックエンドへと翻訳する非常に均一な手順が得られるわけです。こうして一度新しい意味概念をパースツリーの下方レベルにエントリーしてしまえば、センテンスレベルの構造において名詞句の単位すべてに直接利用できるのです。ですから実際にはやりません。これが現われる異なったパターンをすべて明示的にスペルアウトしなければならなくなるのは、意味パースのようなもののところでではありません。意味ノードが構成成分の低いレベルに存在するに過ぎないということが重要だと思います。つぎに続く構成成分に行き着く前に、主語や直接目的語といったノードを持つレベルまで上がらねばなりません。ですからわたしは移植性は高いと思うのです。

問　それでは私の最初の質問ですが、現在のところそれはセンテンスベースで動いているわけ……

答　センテンスベースで動いています。

問　談話レベルの制約とかそういったものをもっと考慮すればどうなりますか。

答　確かに、ある談話レベルのスタッフをシステムにインプリメントしています。バックエンドでそれをしています。そして当面のわれわれの見方としては、照応参照やこの類に利用できる目的語をバックエンドに維持しよう、そしてこれが目下やりつつあることでもあるのです。フラグメントは許容しています。これには何ら問題はありません。一定の手順を決めていますので、それに沿えば、ある人が何かあいまいなことを言いますと、機械はそれをはっきりさせるために、質問をもって応えます。こうして機械は新たな情報を獲得し、一度このセンテンスが完全だとわかれば、システムは応答することができます。私どもではこの種のものをインプリメントしていますが、実言語世界とは反対に、バックエンドの中へ保存することが主になります。

答　たいへんありがとうございました。

問　私のグループでは特定の動詞と特定の目的語、たとえば「drink 」と「water 」とかいったものの間の制約に関心を持ってきたように思うのですが、この種の制約を先生がやられたＰ１かけるＰ２といった方法で確率をかけ算して表現することはできないでしょうか。

答　はい。今のところ私の見るところでは、動詞とその目的語とは通常互いに隣同士の位置にあるわけです。その間に割って入る構成成分はありません。

これが目的語を意味的に積極的に表示してみたい条件なのです。別の明示的な言い方をしますと、この「サーブ（仕えること）」は「食品」に従われています。これが後で確率の問題へと導かれるかどうかわかりません。これは私がただ今用いておりますアプローチですが、この対極には目的語に「サーブ」するという言い方があります。そしてつぎに意味フィルターを使って、それが食品であることを決定します。で、まあ私の個人的見解なんですが、意味カテゴリーに対して意味フィルターを用いる時にはトレードオフがあると思うのです。ただし、一つの関係を有するこれら二つのノード間距離がどれくらいかということと、文脈自由部分による接続を表現するのがどれくらい簡単になるかということとを扱わなければならないとしたらの話です。その対極にはこれでは複雑すぎるという考え方がありますが。ですからケンブリッジにあるどのレストランで中華料理が食べられるかと言ったとしても、レストランと「サーブ」とが、その関係を表現する文脈自由ルールで結合されているというふうには申し上げたくありません。主語・動詞の場合はフィルターリングで行なわなければなりませんが、動詞・目的語の場合は自信がありません。といいますのは、動詞・目的語は通常互いに隣同士の関係にあるからです。ですから私は、これを文脈自由部分で直接表現なされるのではないか、さらにそうすればそこからよりよい確率表示が得られるのではないかと思いますが、フィルターが確率を妨害するという意味ではおっしゃることは正しいでしょう。確率は、ある意味では、フィルタリングがないことを前提にしていますので、フィルターと確率を混ぜ始められると問題が生じかねません。フィルターは単に排除作用があるだけですから。

Session 3

Integrated Processing of Speech and Language 3

皆さん、ＡＴＲのすばらしいランチョンをいただいて、第３セッションへようこそお越しくださいました。テーマは引き続き、音声と言語の統合処理です。最初の論文は音声表記からの認識です。著者はＡＴＴベル研の Stephen Levinson 、 Ljolje 両博士です。それでは演者の Levinson さん、お願いします。

Stephen E. Levinson （ＡＴ＆Ｔベル研究所）

音声表記からの連続音声認識

座長、ありがとうございます。発表を始めるにあたり、まずこの研究を手伝っていただいた Andre Ljolie さんと大変なご尽力を賜った同僚の Laura Miller さんに感謝の意を表したいとおもいます。

この図は、これから皆さんに概要の一端をお見せしようとしておりますシステムのブロック図であります。記録された音声はスペクトル分析を通りますが、これは自己相関でありＬＰＣケプストラムおよびデルタケプストラムへの変形であります。

音響的特徴のベクトルが10ミリ秒ごとに得られます。つぎにストカスティックモデルであります音響的音声モデルに基づいて、音声デコーダーが音響的特徴ベクトルを音声単位列に翻訳いたします。われわれが使用しております音声単位は47個ありまして、概ね英語の音素に対応しています。この音声単位はつぎに語彙アクセスプロシージャーに入力され、このプロシージャーが辞書に基づいてこれらを単語仮説へ翻訳し、単語仮説は入力をパーサーに形作り、パーサーが正式文法に基づいて標準語彙表記を実現します。さてこのブロック図は特に際立ったものではありません。多数の認識システムがこの一般カテゴリーに入ります。

これからお話ししようと思いますインプリメントにつきましては、特殊なことが二つあります。まずこの音響的音声モデルは語彙と文法から完全に独立しています。　これには音声構造と言語のサウンドパターンに関するストカスティックモデルがあります。第２の特徴的なことは、音声デコーダーが一度この音声単位列を生成しますと、システムの残りの部分は元の音響的信号に対して全くアクセスしないという点です。この時点では音声単位とその持続時間だけがシステムの残りの部分に伝送されます。

言語の音響的音声構造をモデル化するのに用いるストカスティックモデルはHidden Markov Modelの特殊形です。Ａｉｊが単に、一つの状態から他の状態に移る確率となるときの状態遷移マトリックスはマトリックスＡです。各状態そのものは音声単位と一対一対応しています。一つの状態は一つの音声単位であるわけです。したがって非常に乱暴な意味での遷移確率というのは言語の音声構造をモデル化することになります。持続時間のモデル化はγ関数によって行ないますが、γ関数には二つのパラメーターがあります。いわゆるアイダパラメーターとヌーパラメーターがそれですが、この二つで持続時間の平均と標準偏差が決定されます。つぎにスペクトル所見そのものは多変数Gaussianでモデル化しますが、このGaussianにおいて所見は、ある成分から他の成分への相関とともに、これらとパラメーターを形成しますので、状態依存分布の一つ一つの平均と共分散でもあります。このモデルには、合わせますと全部で19,000のパラメーターがあります。

ところで音声表記のプロセスは非常に単純です。といいますのは各状態はphoneと一対一対応しておりますから、音声転写を実現するのにしなければならないことは、単に一つの所見集合が与えられたときの最適状態シーケンスを決定するだけでなのです。そしてこれはここに示しましたビタービアルゴリズムによって行なわれます。時間Ｔにおいて状態Ｊにある可能性最大値をこのリカーションから計算します。それは全先行状態Ｉに対する最大値であり、ある先行時間τに状態Ｉにある確率かける遷移確率かける時間の長さτかけるτ所見の確率の可能持続時間τすべてに対する最大値でもあります。したがって前進においてはαを前進方向で計算し、そしてダイナミックプログラミングで普通行なわれておりますように、αの配列を通って後戻りし、その状態と持続時間シーケンスを再構成します。ですから計算の終わりには、状態個々の識別ができ、どれくらいの時間その状態にいたかを知ることができます。これが語彙アクセスプロシージャーに渡す情報で唯一の情報となります。

さてこの語彙アクセスプロシージャーは、文字列対文字列編集問題以外何ものでもありません。音声デコーダーから測定した状態シーケンスとアクセス中の特定の単語の辞書発音との間の距離を計算するのはアルゴリズムに他なりません。　そしてこれはこれら局部的制約をもったダイナミックプログラミングアルゴリズムによって行われます。Ｑここにあるのがいわゆるシンメトリックな制約です。この水平アクセスは文字列への挿入に対応

し、この垂直アクセスは正しい表記からの削除に対応します。この対角線はＱＪへのＱｋの代入を表しております。それでこれら局部的制約が与えられたとき、全可能距離の配列を計算するダイナミックプログラムがこれになります。これは対角線に代入にかかわる距離を加えたものです。これは対角線に代入にかかわる距離を加えたものの、水平に削除にかかわる距離を加えたもの、および垂直に挿入にかかわる距離を加えたものの最小値です。そうしてこの計算が終わりますと、配列横列のＭＮ番目のエレメントは文字列対文字列距離の最小値を含んでいることになります。

さてこれで私がまだ皆さんに申し上げておらないことは、いかにして二つのシンボル間の距離を測定するかという問題だけになりました。この説明のために、音響的音声モデル自身に目を向けます。ＱＩ、ＱＪ二つの状態のスペクトル距離と継続時間です。スペクトル距離はローバーメトリックという数学的プロシージャーを利用するのですが、これは単にそれぞれの平均間の距離にすぎません。このτＩはそうすると分布となり、ピュータティブシンボルの継続時間となります。間の代入距離には二つの係数があります。挿入と削除のためのメトリックスは非常に単純なヒューリスティックでインプリメントします。すなわちこれは特定のホーンに対する代入サイレンスのコストが与えられるということです。

このパーサーは通常文法に基づいたもう一つの非常に標準的なダイナミックプログラミングアルゴリズムです。この通常文法では語彙単語Ｖによって状態ＲとＳがリンクされています。この二つの状態はHidden Markov Modelの状態と混同してはならないものです。これらは完全に分離しており、構文構造において単語を合わせるノードを表しております。ここでＳＴのＰを時間Ｔにおいて状態Ｓにある可能性の最大値としますと、ＳＴのＰはこのリカーションに従います。これはある等値（equivalence)クラスのプロダクションルールに対する最大値であり、ある先行時間において状態Ｒにある可能性かける単語ＶがシンボルＴ－τとＴの間に来る確率の先行シンボルすべてに対する最大値でもあります。この等値クラスとは、これに対してこの最大値を計算するものであり、単に左側にＲ、右側にＳをもったプロダクションルールすべてを含んだ集合にすぎない。

音響的音声モデルを学習させるために、ＤＡＲＰＡデータベースからの約 4,000文に依拠しました。これには 109の男女両方の話し手が含まれ、話し手当り約36文となります。また先に述べましたように47個のフォーンのような単位があります。これらデータはすべて、シーケンシャルＫ平均アルゴリズムという最適テクニックを用いてセグメント化しラベル付けを行ない、モデルのパラメーターを単に標本統計のみを用いるだけで推定しました。

こうしてある任意のPhoneのセグメントをすべていっしょにし、これらの平均値、分散値、継続時間を単に数値平均として計算しました。この学習ではデータポイントは全部で約４千８百万個あり、先に述べましたように、モデルのパラメーターは 191,000個です。ですから少なくとも、これらパラメーターを推定するのに使用するデータとしては十分な量だと言えます。

さていよいよその結果なんですが、結果は２種類についてお話しさせていただきたく思います。まず最初のブロック図を戻りましょう。われわれの関心は音声表記の精度にあります。で、われわれができることと言いますと、音声デコーダーから出てきますこのシンボル列を見て、これがどれくらいいいのかをたずねることです。さて実際には、これを測定することは容易ではありません。というのは、このシンボル列と比較する相手は何かというのが問題だからです。これを本当に発話と比較されるでしょうか、それとも標準の発音と比較されますか。答えは非常にむつかしいのです。

それで私のやりたいことは、オーディオテープを再生することによって、この転写がどれくらいいいかを実感していただきたいのです。はい、このテープですが、これはこんなふうに作成いたしました。一人の話し手、それは実は私自身だったのですが、その話し手が６個のセンテンスを話します。するとそのセンテンスはスペクトル分析と音声デコーディングを受けます。この音声デコーダーからの出力を取り出しますと、それがシンボルの名称であり、それらの継続時間に粗ピッチ輪郭を加えたものとなります。そしてこの情報を音声合成装置へのテキスト中に入れたのです。　こうすることで音声合成装置へのテキストの結果を聞くここができます。近似的にですが、あー、少なくとも音声デコーディングがどれくらいいいかの感覚をつかむことができます。

さてこれに関して興味深い特徴があります。音声単位は47個ありますので、これには約 4.5ビット必要になります。また継続時間は約40フレームよりは少ないので、これにたぶん 5.5ビットを要するでしょう。ピッチ当り10ビットを割当てて、全体にセグメント率をかけますと、この情報を転送するのに必要なデータ率は１秒当り 120ビットをわずか越える程度となります。ですからこれでわかることは、ボコーダーのビット率は極めて低いということであり、この転写をお聞きになるときにご注意いただきたいのがまさにこのことだというわけです。今同時通訳を聞いていらっしゃる皆さん、ちょっとヘッドホンを外してみていただけますか。そしてこのテープを直接聞いてみて下さい。よろしいでしょうか。

これでこの合成の性能を実感していただけると思います。

How many cat two casareps are there for submarines
that are in back strait?
Where there more than 85 vessels at sea on 31 ofOctober?
When is the estimated time of arrival in home port
for the Sherman?

どうでしょう、これで1秒当り 100ビットのボコーダーがどんなものかおわかりいただけたかと思います。

われわれは非公式な実験を行なって、このボコーダーの有用性を試してみました。私は同僚の何人かにこのテープを、オリジナル付きはもちろんオリジナルなしでも聞かせてまして、彼らにこのボコードされた音声の語彙表記をしてくれるようにたのみました。こうして発見しましたことは、彼らが単語の内約75個を正しく確認し得たということです。われわれはこれら同じセンテンスをとって、これとこれを使用するシステムにデコードするように命じましたところ、システムは単語の内86個を正しく認識しました。つまりシステムの方が人間より、実際ずっと賢いということになります。もちろんこれは単に非公式な試験でしかありませんでしたので、われわれはこれが実際どれくらいの性能を示すかを見るのにもっと研究をいたしました。それでこれが音声デコーディングの結果です。

これが皆さんにお見せしたい結果としては最後になるんですが、この単語精度の結果なんです。われわれはＤＡＲＰＡの試験データについていくつか試験を行ないました。全部で３つです。まずある特別な学習文に目をつけました。これは 109人の話し手それぞれから二つの文を取り出して作成したものです。ここでは78.1%の単語精度が得られました。さてここで単語精度とは正解数から挿入数を引き、削除数を引きさらに代入数を引いたものを指します。それで89年２月の 300文の試験文を試してみました。これは話し手が異なる。　話し手が異なれば文も異なります。ここで72%をわずかに越えています。また別の 300文について同年10月にも試験を行ないました。ここでも話し手が異なれば文も異なっており、結果は76%をわずかに越えました。

これについて言うべきことが二つあります。一つは、非常に明るい展望をもったものですが、こうした値の差がひどくはないということです。つまりシステムが正しく行なっていることが何であっても、システムはそれに対して相当にロバストだと言えます。良くない点はこれら数値が従来のシステムの能力と比べると競合性がまったくないということなのです。事実、ＤＡＲＰＡの研究で今のと同じデータベースを使っている人たちは、この種の結果を得るのに何ら文法的制約を用いる必要がないのです。ですからわれわれはこの問題に関してはまだまだ先が長いということなんですが、ただこのおもしろいアーキテクチャーとシステムの利点、特に音声表記能力とはいずれは有効だということがわかるものと期待しております。ありがとうございました。

質疑応答

問　　たとえば、実験でエラーが起きた場合、どういうエラーが起きたか教えていただきましたらありがたいと思います。

答　　実験でのエラーというのは何を指していらっしゃるのかよくわからないのですが。

問　　ミスレコグニッションですね、あのう、たとえば文章でどういうようなケースたとえば文頭におきるか、あるいは認識率のわるい音素があって、そういう場合にエラーが生じやすいとか、そういう現象についてもし現況をお聞かせ願いましたらありがたいと思います。

答　　わかりました。ありがとうございます。すみません、先程はちょっと理解をまちがえてしまいました。それは非常におもしろいご質問です。システムの性能はバイモーダルなのです。つまりとびきりいい仕事をすることがあるかと思えば、恐ろしくひどい仕事をすることもあります。事実、このことを定量化しようとしますと、エラーの２／３はセンテンスの１／６に由来します。つまり先程のセンテンスでは認識装置の失敗は壊滅的なのです。また事実、後に戻って、特定の語彙表記をもたらした音声表記を見てみれば、これが本来は起こるべきものではなかったということができるはずです。なぜなら音声表記は非常にいいのです。ただ語彙表記が完全にまちがってしまっているわけです。今のところなぜこれが起こるのか理由はわかっていません。おそらく挿入、削除、代入距離の選択に関係しているのであろうと考えられます。はっきりさせるにはもっと研究が必要でしょう。ただ音声表記は非常にロバストなことは確かです。

問　先生は実験を機械、人間両方を用いて行なっておられますが、機械つまりシステムが起こすエラーと人間が起こすエラーとの間に何か相関はあったでしょうか。

答　そうですね。ちょっと考えさせて下さい。いいえ、ありませんでした。人間の聞き手は、音声表記が悪いと誤りを犯し、これが良好なときにはいい結果を出しました。機械を使っている場合には、音声表記のエラーと語彙表記のエラーとを相関づけることはできないと思われます。そしてこのことはまたわれわれが直面している問題を別の面から示唆するものでもあります。語彙アクセス部分のどこかに、何か間違いがあるのです。そしてただ今申し上げましたように、それが何かはわかりません。

問　語彙への音素エントリーは単に基本発音だけですか、それともリダクションインターバルをつけられるのでしょうか。

答　単語個々に対する語彙にはエントリーはただ一つだけで、それは単語の引用形です。またわれわれは音韻学的なルールや単語接続ルールあるいはそういったものをモデル化するような試みはいっさい行ないませんでした。

それでは２番目の論文ですが、タイトルは「言語学的知識と音声言語理解のための構文・意味パーシング法」です。著者でかつ発表者であられるのは豊橋技術科学大学の中川教授です。

中川聖一（豊橋技術科学大学）

言語学的知識の役割と音声言語理解のための構文・意味パーシング法

時間が余りありませんので、私の細かい話は抜きにして、基本的な考え方というのを、音声理解システムについての基本的な考え方について述べてみたいと思います。

一つはですね、タイプされた入力に対する言語の解析方法と、音声入力に対する解析方法とは、どういうような点が違うかという、まずその基本的な考え方なんですが、例えば自然言語の文の集合が、この波線に囲まれた文だとします。そしてあとタスクの面がもしあるとしますね、認識対象、あるいは言語の質問回答でもいいんですが、そういうようなタスクの面に対して、まずシステムは文法を持たなければいけないということですね。その文法によって生成される文の集合というのは、こうなったとします。これの大小関係はいろいろあると思うんですが、しかしこの場合においてですね、タイプによる入力の場合ですね、普通自然言語学者がよく想定するのは、タイプによる入力、入力に誤りがないという場合、その場合にはですね、例えばここのあたりの文というのは、自然言語的にはおかしな文ですけども、それは余り被害がないわけですね。入ってくるのはここだけですから、入ってくる文の扱う文はここだけですから、自然言語の文だけですから、もしシステムがですね、このような文を受け付けるようになっていても全然被害がないということです。

ところが音声による入力の場合、これは音声そのものは正しく文に仕上げていると仮定してもですね、認識装置が間違うということで、こういうような文もですね、候補として出てきます。まともにここへ全部出てきます。だからもしここがですね、受け付けるようなシステムであれば、文法がですね、そしたら認識誤りになってしまうということでですね、音声の場合、特にですね、この面積を少なくしておく必要があるということです。それが根本的な違いであります。

そのためにどういうような文法を使うかというか、アプローチが完全に変わってきます。そういうなので、一つはこういうような非文をリジェクトしたいというのが一つなんですけども、そういうような意味とかですね、最近ではストキャステックなモデルとして、Ｂｉ－ｇｒａｍ、Ｔｒｉ－ｇｒａｍとかいうようなものが使われています。それはですね、この真ん中の辺が確率が高く生成されるとか、そういうようなモデルなんですけども、だからある意味では全分野をカバーするというんで、午前中もありましたけども、カバーレッジは大きくなるわけですけども、ここら辺確率が小さくなると、そういうようないろんな問題があります。

それでですね、またちょっと話が変わりますが、どっちのシステムがいいか、２つのシステムがあったとした場合に、どっちのシステムがいいかという、一つのシステムはですね、すべての文章は必ず一単語だけ間違うというようなシステムがあるとします。もう一つはですね、入ってくる文章のうちの半分は正しく認識できると、あとの半分は２単語以上間違ってしまうと、そういうようなシステム２つあれば、どちらのがいいかということになると思うんですけども、単語の認識率で出せば、Ａの方が精度がよくなります。そういうような意味では、ディクテーションシステムなんかでは、こちらのアプローチの方が

好ましいかもわかりません。人間が介在しまして、１単語間違うたら、そこだけ直せばいいと、そういうような問題になりますから。しかしこれをですね、質問応答システムとか、そういうような音声理解に使うという意味では、正しい文章が多い方がいいと、１単語だけ間違ってしまったらもう意味がないということですね、極端に言えば。そういうような意味で、Ｂと、こちらの方がやっぱり音声理解とか、そういうような意味にはいいということです。

このシステムはですね、こういうようなシステムは、例えばＴｒｉ－ｇｒａｍベースのモデルとかですね、そういうなにをこちらのシステムに対応すると。Ｂというのは、普通行われているシンタックスとかセマンテックスなベースのようなシステムは、こちらのシステムになるんではないかと。しかし両方ともいい面を持っていますから、将来的にはやっぱりコンビネーションがいいんであろうということで、例えばストキャステックシンタックスグラマーとか、そういうような、これは午前中でもSeneffさんが発表された、そういうようなのに近くなってくると思いますけども、こちらの方が主流になってくるんではないかという気がします。

それでですね、結局一言で言えばですね、音声理解とか認識で言語モデルは何のために使うかというのは、最終的にはですね、もちろん意味を理解できないとだめだということなんですけども、認識という意味で言えば、サーチスペースをリダクションするということですね。認識率を上げるということが直接的な目的ではないかと思うわけです。そういうような意味で、パープレキシティというメジャーが使われているんですが、そういうようなのをいかに少なく小さくするか、パープレキシティというのはエントロピーのですね、２のエントロピー乗ですね、べき乗なんですが、そのエントロピーを小さくするかというのが問題になるわけです。

例えば、これはちょっとまた余談になりますけども、シャノンがですね、有名な論文を発表しまして、例えばアルファベットですけど、これはアルファベットのＺｅｒｏ－ｇｒａｍだったら、１アルファベット当たりのビット数は 4.7ビットとかですね、Ｔｒｉ－ｇｒａｍまでアルファベットの３次組ですね、Ｔｒｉ－ｇｒａｍで考えれば 3.3ビットぐらいになるというような発表をしてます。

これはアルファベットが単位なんですけども、最近の音声の言語モデルとしては、単語を単位とするＴｒｉ－ｇｒａｍというのが使われています。それがいいかどうかというのは、先ほどちょっと言ったんですけども、例えばこういうなのをHidden Markov Modelであらわそうということもできるわけですね、情報源をHidden Markov Modelで表現しようという。そうするとですね、これは我々最近やった結果なんですけども、例えばですね、これはいろんな言語の話なんですが、大体ビットとしては１アルファベット当たりですけども、2.7ビットとか、それぐらいになります。そうするとＴｒｉ－ｇｒａｍよりももっといいモデルであるということになりますね、Hidden Markov Modelは。だからこれは一例なんですけど、Hidden Markov Modelというのは、かなり潜在的なパワーを持っているんではないかというのを改めて最近わかったんですけども。

そういう関係で、これからちょっと紹介したいのは、パープレキシティとですね、文の認識率とは、どういうような関係があるか。パープレキシティが幾らになったかという話はきょうもありましたけども、それが文認識率でどの程度反映されるかというのをはっきりとらえておこうということなんです。そのためにシュミレーションでいろいろとやっているわけですけども、それをちょっとやってみるとですね、これに仮定を設けるわけですね。簡単な文ですけども、インプットがですね、単語に今、単語単位のパープレキシティを考えるとします。ある単語に対してですね、正しいカテゴリーのレファレンスパターンとのマッチングはですね、こういうような分布をしているとします。こちらですね、こちらの分布しているとします。そして誤ったカテゴリーとのマッチングのスコアですね、これはこのように分布しているとしますね。こちらのスコアの方がいいというメジャーになっているんですけども。こういうなのですね、カテゴリーが増えれば増えるほどですね、これがこういうような分布していますから、カテゴリーが 100カテゴリーの認識の場合だったら、ここから、この分布から 100個スコアをとってくるというシュミレーションですね。ここから１個だけとってくるわけです。どちらの方が大きかったかによって、正しく認識されるかどうかというのは決まるわけですけども、そういうようなことでシュミレーションをやるとですね、こういうような関係が得られます。これはボキャブラリーサイズなんですけど、パープレキシティですね、パープレキシティ、こちらが認識率です。大概こういうような関係が出てきます。いろんな曲線が出てきますけども。

だからこういうような曲線からですね、例えば 100単語に対して、パープレキシティ100 に対してですね、例えば60％ぐらい得られるシステムであれば、パープレキシティが10ぐらいになればですね、認識率が90数％になると、そういうようなことなんですね。

だから言語情報を使わなかったら、例えばＤＡＲＰＡでやっているようなリソースマネージメントシステムであれば、ボキャブラリー 1,000語ぐらいですね、 1,000語。それで

例えばワードの認識率が30%とか、これぐらいであってもですね、パープレキシティが20ぐらいになれば、単語認識率は90近くなると、そういうような関係になります。

だからですね、パープレキシティを減らすというのは、単語認識率が上がるわけですね。そして言語上、完全に除いたときの、あいまい性を除いたときの尺度というのはパープレキシティですから、だから連続の音、文の認識率というのは、その単語が10単語からなっているとしますね、文章が。そうするとこの90%の10乗というのが文の認識率になると。連続単語の認識と同じようなタスクになるわけですね、あいまい性を完全に除くから。そういうような意味で文の認識率が推定できるわけです。

だからまとめて言えばですね、こういうような関係になります。例えば要するに求めたかったのは、認識装置が与えられると、それとボキャブラリーサイズが与えられると。そうしたときにパープレキシティを計算してですね、それでそのタスクの平均文長なんかわかれば文の認識率が推定できると、そういうような方法なんですけども。

こういうので実際のシテスムにこの方法を適用すると、まずまあまあうまく合うということなんですけど、例えばこういうようなシステムでやっているわけですけども、これはUnixに関する質問回答システムをタスクと想定してですね、ボキャブラリーのサイズが500単語ぐらいというようなやつに対してですね、こういうようなシュミレーションをやると。文の推定率を出すということをやってみますと、単語が 521単語です。それでパープレキシティ10とか、19とか、50とか、いろんな文法のですね、複雑性によってパープレキシティ変わってきますけども、そういうので実際の実験値はどの程度になったかというのと、推定値がどの程度になったかというのとを比べてみますとですね、今の方法で推定した文の認識率と、実験値とは非常によく合うというような結果が得られます。

そういうような意味でですね、パープレキシティを減らしたらいいという話ですけど、どれだけ減らせば意味があるかというのがはっきりしてきたということなんです。

もう一つ有名な方法として、SPHINXでカーネギーメロン大学で開発されたSPHINXを評価してみようということです。SPHINXであれば、これを例にしたら興味あると思うんで言いますとですね、音素単位の認識を考えるとですね、英語は44音素があります、大体ですね、44音素。SPHINXのですね、連続音声中の音韻認識率と、文節の認識率というのは、大体74%と報告されています。このタスクに対してですね、音素当たりのパープレキシティは 1.6ぐらいです、換算してみますと 1.6。だから先ほどのこういうような関係のグラフを使うとですね、44音素に対して74%であったと。そうするとカテゴリーが 1.6になれば、どの程度の認識率になるかというのは、先ほどの曲線から求まります。それが99.1%です。

それで1文章当たりの音素の数は48音素です。だから99.1%の48乗すれば文の認識率が得られます。76%、これが推定値です。実際の報告では70%ぐらいと言われています。これは割と合っているんであろうと思う。

もう一つは、単語単位で測定した場合を紹介しますと、単語は 997単語ですね、リソースマネージメントタスクは、語彙数が 997です。 997、言語モデルがないときは 997単語の連続単語の認識ですね。 997のときには、大体ですね、単語の認識率は72%という報告があります。パープレキシティは、文法を使ったときはパープレキシティ20です。これはＢｉ－ｇｒａｍを使ったモデルですけども、20使ったときにはどうなるかといいますと、だから 997で72%ですから、20単語では何%になるかと、また推定しますとですね、これが96%になります。パープレキシティ20に対する単語認識率はですね、96%に相当しています。だから96%の８乗、８乗というのは、１文章８単語からなっていますから８乗です。そうすると大体74%ぐらいなんです。いずれにしても74とか、76とか、実際の70%の値に近づくというので、こういうような方法でかなりうまく推定できるということでわかってきたというのでですね、音声認識ではパープレキシティを小さくするというのが重要であるということが数量的にわかってきたわけです。

そういうような方法からですね、じゃあ人間並みのシステムをつくろうと思ったら、どの程度の精度が要求されるかというのを逆に考えてみますとですね、シャノンの研究ではですね、人間がですね、テキストを読むときに、文脈情報まで使って、１アルファベット当たりのエントロピーを計算するとですね、大体パープレキシティに換算すると、２ないし 2.5ぐらいと言われています。エントロピーで言えば、１ビットないし 1.3ビットですね、１アルファベット当たり。意味情報、シンタックスすべて使った場合の話ですけども、文脈情報も全部使って、きょうの午前中のですね、ワードさんがやっていたああいうような談話の知識とか、そういうのをすべて使った場合の話ですけども。パープレキシティとしては、音素当たり２ないし 2.5ぐらいであります。単語に換算すると、ワードユニット当たりのパープレキシティは 100ぐらいである。要するに人間並みの言語情報を使えばですね。そうしたときにですね、例えば音素の話で今やりますと、１文章は大体30ないし50音素からなっているとしますね、そうします。そうすると文の認識率、理解率でもいいんですけども、90%以上の精度を得たいとします。そうするとですね、音素の認識率は99.6ないし99.8%ぐらいが必要とされます。すなわちパープレキシティ２ないし 2.5に対し

て、音韻認識率は99.6%ないし99.8%を要するということです。

ですから英語の場合は、実際には音素は44カテゴリーあります。日本語の場合25種類ぐらいありますね。だからパープレキシティ２ないし 2.5、ここで99.8%ぐらい必要なんですね。だからずっと逆算して40音素ぐらいだったら、大体90%ぐらい、日本語の場合25としたら95%とか、大体ですね、ちょっと近似がありますけども。いずれにしても人間並みのですね、システムをつくろうと思ったら、まず音響レベルでは、音素レベルでは90%以上まず必要であろうということが考えられます。このエバリエーション法ではですよ。

それともう一つは、人間並みの言語情報を使わんとあかんということですね。そういうような仮定のもとです。だからもちろん人間の場合は、タスクはリミットされてないですから、我々は普通ワールドモデルでやりますからね、実際にパープレキシティ 100ぐらいは実現可能です。ただしＤＡＲＰＡではパープレキシティ 200を目標にしているといいますが、それはもう成功しないと私は思っていますけども。だからですね、我々実際やっているのはパープレキシティ10とか20ぐらいです。今ＤＡＲＰＡでやっているのはパープレキシティ60ぐらいでやっていますから、まあまあリーゾナブルな範囲だと思いますけども、 100以上超えたら非常に難しい。だから 100以下にパープレキシティを絞らないとだめであろうということなんですね。

だからダイナミックに知識を用いて絞っていくというのは、非常に重要だと思うんで、きょうもいろいろ話題に出たということですね。それとストキャスティクを入れると、言語モデルに。そういうようなのが方法としてはいいんではないかという気がします。　その前にですね、歴史を振り返ってみると、歴史はこういうようなことを教えています。すなわちカーネギーメロン大学はHEARSAYシステムとHARPYシステムを最初つくった。ハーピーシステムは非常によかったということですね。最近ではですね、同じカーネギーメロン大学のANGELシステムとSPHINXシステムをつくった。やはりSPHINXクスシステムはすぐれていたということです。

またＢＢＮでは、昔はHWIMシステムというのをつくっていて、最近はBYBLOSということで、こちらの方が圧倒的にいいというので、こういうようなことはどういうようなことを意味しているのかと言ったらですね、やはり入ってきた音声に対して、最もいい単語シーケンスとか、そういうようなものね、シンタックスやセマンティックスのコンストレイントを満たす最もいいワードシーケンスを探したいと、それのアプローチですね。成功したのは、HARPYにしろ、BYBLOSにしても、SPHINXにしても、そういうアプローチがやはり妥当ではないかということだと思うんです。

それは我々自身も確かめています。これはSPOJUS-SYNOと称している方法、これはシンタックスとかセマンティクスを文脈自由文法で表現してですね、最もいい単語シーケンスを求めようというシステムです。それに対してSPOJUS-SEMOというやつは、セマンティク主導型のやつなんで、格文法とか、依存文法なんか使ってやっているわけですけども、文の認識率を比較すると、相当差が出てきます。同じ単語の認識までは同じ全くシステムを使っています。しかし言語レベルを変えるだけで、これだけ違うわけですね。どちらも最もいい単語シーケンスを求めようとしいるわけですけども、依存文法とか格文法だけでは、やはり言語モデルとしてよいモデルはつくれないということを示しています。

もう一つの話題としてはですね、そういうような意味では最もいい単語シーケンスを見つけるというのが、やっぱりパーシングメカニズムになると思います。計算量とか、そういうようなもの、まだいろいろあると思いますけどね。そういうような意味ではレフトーライト型と島駆動型と、いろいろあるわけですけども、それはどちらがいいかというのは、結論的から言いますとですね、全くどちらも同じ精度を得ようと思うたら得られます。時間を無視すれば、サーチスペースを大きくすればですね、島駆動型であっても、レフトーライト型でも、全く同じ精度が得られます。

しかしですね、計算量が限られるとか、サーチスペースが限られると、そういうような条件のもとでは、どちらがいいかと。そういうような比較もする必要があるということでやったことがあります。そのほかいろんな人がやっています。例えばワードさんも同じ研究でやっておられた。それはですね、ネットワークグラマーとレフトーライトグラマー、ネットワークグラマーとかＴｒｉ－ｇｒａｍね、これとの比較をやっているとか、そういうような研究をやっていますけども。私はですね、この文法をですね、文脈自由文法であらわしたときに、島駆動型とレフトーライト型とどちらがいいかというのを比較したわけです。これはレフトーライト型のタイムシンクロナスなパージングメカニズムですけども、詳細は省きますけども、こういうような方法でやる。

もう一つは、島駆動型です。これは午前中に新美先生が発表されたパージングメカニズムとほとんど一緒なんですけども、多少違うのは、このアイランドの長さ、長さに同期して分析をしていく、解析を進めていくという点が違うんですけども、いずれにしてもこういうような方法で比較しました。

実際にはこういうようなタスクですね、このボキャブラリーとしては 100とか 250とか

、500とか、これぐらいのボキャブラリーのタスクに対して、このパージング方法を比較したわけです。そうするとですね、先ほど言ったように、両方とも最適な単語列を見つけたいというのは、趣旨は一緒なんで、文法も全く同じように使っていますから、文の認識率はほぼ同じです。しかしパージング単語が非常に違うということですね。特に島駆動型の方が時間がかかると。すなわちアイランドから右にも左にも制限がないわけですね、中ぶらりんですから、フリーですから、だからそういうようなので予測される単語数が大きくなるというので処理量も増えるということです。

だからそういうような文の認識率に限って言えば、ほぼ同じです。ただですね、島駆動型のよい点はですね、これはまたタスクが違うので、ほぼ同じという結果を示しているんですけども。しかしですね、よく言われることなんですけども、入力文のですね、ヘッドの部分ですね、ヘッドの部分がノイジーな場合では、レフトーライト型では、最初こけたら最後までこけるというやつですね。そういうなんでは、そういうようなときには島駆動の方がいいであろうと言われていますけども、それを実際にやってみたわけです。　すなわちヘッドの部分がですね、非常にノイジーな入力文に対しては、どちらがいいかとしたら、これは明らかに島駆動型のパーシングメカニズムの方がいいということです。これは常識の線に合ってますけども、そういうような結果を得ました。

以上まとめますと、私の言いたいのは、音声認識で言語情報の役割というのは、サーチスペースをげる、認識率を上げるということですね。　それと、どの程度下げたら文の認識率はどの程度上がるかというのも、ある程度めどがついたということ。それと基本的なパージングというのは、最適な単語列を見つけたいと、そういうような基本方針でやるべきではないかというわけです。

そういうので島駆動とか、レフトーライト、両方ともいい面も悪い面もあるんですけども、いずれにしてもそういうような方向がいいということと、先ほど言ったようにパープレキシティを下げる方法としてストキャステックな文脈自由文法とか、それがすべてではないですけど、そういうような方向が正しい方向ではないかと、そういうような気がしています。以上です。

質疑応答

問　　藤崎です。これはおもしろいです。言語が情報を表現するメカニズムに対して非常に興味深い洞察を与えてくれます。これはおそらく先生のご議論の中心課題ではないと思うのですが、パープレキシティについてまたシャノンの結果の拡張について、先生はパープレキシティを音素ごとに述べられました。私の見解では、言語学的表現の情報内容あるいは英語の言語学的表現のコーパスは、単に全発話ないし全内容の平均であるべきでないと思うのですが、仮に発話には新しい情報がいっぱい詰まっているか、あるいは多数の新トピックを扱うものとすると、そこにはもっと情報が必要になってくる。これに対して、トピックが徐々に発達してくるとおっしゃるのなら、1マルコフ状態当り1音素当りの平均情報は低くていいということになりませんか。これは私の経験から得たことですが、フィニンガンスのウェークとイソップの寓話の情報内容を確かめたことがありまして、1語または1文字当りの情報内容は2対1の違い以上だということがわかりました。この点について先生のご意見をおうかがいできますか。

中川　ここで言った方法は、あくまでも近似の方法でありまして、続けばいろいろ出てきます。　しかし近似としてはいい線をいっているんではないかということです。そういうようなことですね。これからですね、もう少し厳密にやろうと思うたら、急に難しくなります、非常に。

問　　ベル研の Ken Church （ケン・チャーチ）です。1単語当りのパープレキシティについて約 100とおっしゃった数値なんですが、どうも人々がどこへ行き着けると想像しているのかについてのもののように受けとめています。そしてそれは、究極的には、思うのですが、それはほんとうに大きな機会と一つのやや異なった問題を指し示しているのではないでしょうか。つまりテキストの圧縮です。 100というパープレキシティは1語当り約6ビットに対応します。これは先生がおっしゃった1文字当り3ビットより少し良好です。ＵｎｉｘのＯＳで今日用いられている最善の圧縮テクニックより5倍はいいと思います。5というファクターはまさに追求の価値があります。何か〔数値〕お持ちですか。それとも単に観察にとどまるのでしょうか。

中川　パープレキシティ60と 100の場合が余り差がなかったということですか。

これ、私の結果が、60と 100とでは、余り変わらないですね、文の認識率はそれほど。

だからそんなに疑問というか、何か知らなかったんですけど、外れてないと思いますけど。

本セッションの最後の論文になりましたが、タイトルは「統計的方法による言語翻訳」です。演者はＩＢＭ研究所の Fred Jelinek 博士です。

Frederick Jelinek（ＩＢＭ研究所）
統計的方法による言語翻訳

私どもはここのところ、やはり音声認識で採用している私どもの方法をいくつか言語翻訳にも試してみるべきではないかと考えています。言語翻訳を統計的に行なうにはもちろん、どこかで統計を推計しなければなりませんが、われわれの所では非常に幸運なことに、カナダ人が何人かいまして、フランス語も英語も両方しゃべれますし、またカナダではどちらかの言語が他よりまさっているなどということは毛頭ありませんので、カナダの議会は、英語で話されたことはことごとくフランス語に翻訳し、フランス語で話されたことは英語に翻訳することを徹底いたします。

つまりわれわれはこうしてフランス語・英語間で翻訳されたセンテンスについて莫大な資産を持っていると言えます。これはとりもなおさず、音声認識に関する推計学的アプローチを開始するためのデータベースであるわけです。もう一つフランス語を選んだ理由は、フランス語と英語はボキャブラリーのモジュールとしては現実的には同一であると考えられるからであります。ここにコーパスの Hensardからの例が一つあります。これをご覧になられると、語対語訳でもかなりうまくいくということがおわかりになるでしょう。フランス語はここで見ますかぎり、 Hensardトランスレーターが行なう純正な翻訳であります。私が行ないましたことはこうしたものではありません。本国におります者たちも同様です。これが公式文書がどんなものかを示したものです。ご覧のように本質的には語対語の語対語訳になっていますし、一定の固定表現もあります。これは別問題なので下線を引いてあります。これらは辞書を引いても必ずしものっていない固定表現なのですが、翻訳システムはもちろんこれらを出力する能力がなければなりません。

それでわれわれは自らに向かってナイーブな人間になるように宣言しまして、ナイーブな翻訳メカニズムを持とうとしているわけです。このやり方の誰のとも異なっております点は、私どもではピボット言語やそれに類したものをいっさい使わないということなのです。われわれは直接フランス語から英語へ行きます。また英語からフランス語へ行くのではなく、なぜフランス語から英語へ行くのかという点にも多くの理由があるのですが、その一つは英語の話し方はよく心得ておりますので、フランス語についていいか悪いを判断するより、英語の方で善し悪しを判断する方が確かだということなのです。

つまり基本的な考え方はこうです。ソーステキストがあるとします。これを一まとまりの固定の言い方に分割します。これらを語として考えてもかまいませんが、「ne pais 」のような否定がありますと、それが一つの固定の言い方になります。語のブロックとなるものはいろいろあるでしょう。フランス語における pase composc ……全体は一つの固定の言い方となるでしょう。といったことがいろいろあります。ですから考え方としてはソーステキストを一まとまりの固定の言い方に分割することであり、つぎに何らかの用語集を用いてその固定の言い方を他の言語に翻訳し、ついでこの言語の文法を用いて語を並べ替えます。こうして目的の言語における一つのセンテンスを得るわけです。もちろんこれは非常にナイーブなアプローチでありますが、これでどこまでやれるかというのを見ていくことにしましょう。

ここにありますのは一般的アプローチです。実際われわれの出発点、少なくとも私の専攻はコミュニケーション理論ですので、われわれは有名な、そう、シャノンの図と言ってもかまいませんが、それでモデル化しているわけです。すなわちわれわれはフランス語から英語に翻訳しようとしているのです。そしてフランス語から英語に翻訳しようとしておりますので、われわれの取るべき見方としては、英語を生成するソースがあるとしなければなりません。英語をフランス語に翻訳する一種のチャンネルもありますが、フランス語を英語に訳し戻すデコーダーが必要です。

All right not the right thing that a Frenchman first
thinks of his sentences English naturally.

彼らには直接フランス語で話していると仮定しましょう。いずれにしてもこの種のデコーディングを行なうためには、これはフォーミュレーションとしてはまちがっていないと思います。で、われわれのねらいはこのモデルのパラメーターを推定したいということで、このデコーダーを構成するためには、デコーダーがこの観察フランス語を与えられたとして、これがもっともありそうな英語を見いだせる最大の可能性となるでしょう。それで私がやらなければならないことは、ソースのモデルを持つということ、すなわち、ある英語のセンテンスがフランス語の基になる確率はどうなるか、そして英語のセンテンスがこ

のフランス語のセンテンスに変形できる確率はどうなるか、ということです。

このモデルを得るためには、私はある推定プロシージャーと、 Heisan コーパスでのフランス語・英語センテンス間で合わせたデータについて推定を扱わなければなりません。この英語の言語モデルは簡明です。これは単に英語を了解することから得たものです。別の言い方をすると、このペアから得たのではないということになります。これはどんな種類のものであれ英語の大型コーパスから得ることができます。われわれの場合、もちろん、これらは単にわれわれのＴｒｉ－ｇｒａｍ言語モデルであり、われわれが日頃音声認識でずっと用いているものです。ですから問題は翻訳モデルを構築することなのです。その翻訳モデルは非常に単純なものです。

そこにはブロックへの分割が含まれます。そしてこれらブロックは翻訳されたブロックです。ただしフランス語から英語への翻訳の話ですが。

こんなことを言って申し訳ないのですが、われわれのモデルでは英語が一次言語なのです。私の考えようとしているやり方ですと、英語のブロックがフランス語のブロックに翻訳されそれからフランス語のブロックを並べ替えます。ですからこのアプローチの基本パラメーターは、つぎのものになります。ある特定の英語の単語が一つのフランス語に翻訳される確率がわかっていなければなりません。そうしてつぎに欲しい情報はその英語の単語がｎ個のフランス語に翻訳される確率です。というのは、はじめにある英語のブロックが一つ一つ別々の単語だからです。フランス語のブロックはいくつかの単語の集まりですが、われわれはプリミティブなので、単一の単語からしか出発できないのです。つまり単一・多数語翻訳ということになります。

それでは、私の全体確率モデルですが、Ｅに関してコンディショニングをしておけばよかったですね。私の全体モデルで特定の語「Ｅ」がこれらｎ個のフランス語を生成する確率は、英語の「Ｅ」がｎ個の語を生成する確率かける、このブロックで個々それぞれの語を生成する確率の積となります。ここでもう一つモデルが必要になります。それはひずみモデルです。すなわちここで英語からフランス語でのシーケンスに移るシーケンスを取り出すメカニズムが必要なのです。

もちろん本質的には、フランス語はほとんど英語と同じだという考え方がありますので、私としてはフランス語の位置をほぼ保存する、英語のセンテンスでの元の位置からの位置をですね、保存するモデルが欲しいのです。つまりいわば一つの弾性的なメディアといいますか、ある種の位置保持機能が必要なのです。しかしもちろん、それは絶対的な位置ではあり得ませんので、その位置からはずれる確率が登場してくることになります。で、ひずみ確率をいくつか使おうというわけなのです。ここで二つ注意しておきたいことがあります。まず第一は、ある英語の単語がたとえば仏語にはない冠詞とかだったりすると、英語の単語が完全に消えてしまうということがあり得ます。それでこの場合、ある英語の単語が与えられたときに０個のフランス語を生成する確率は０より大きくなるということがあります。第二に、英語からは出てこないフランス語がいくつかあるということです。これらは自然発生的に生成したものに違いないのですが、ここでわれわれのデータは直訳ではないということを思い出さねばなりません。われわれのデータは翻訳者がカナダ人が好むであろうと思うように翻訳した結果なのです。ですから英語からフランス語に行くとすれば、そのフランス語はフランス人の耳に訴えるものであるべきでしょう。つまりそれは英語ではなくフランス語の構文であり、英語国民ではなくフランス語国民の考え方でなければなりません。それでどこか自然発生的に現われるフランス語の語があるのです。

それではつぎにこれら三つの確率を推定することをやってみなければなりません。というのは先程申し上げましたように、私のモデルからすでに手にしております英語の確率を推定する。…… そしてそれのやり方としては、私はこれら推定する必要のある確率に対して初期値をいくつか仮定します。そして英語とフランス語のセンテンスのペアを取りまして、可能なアラインメントをすべて試みます。アラインメントといいますのは、どのフランス語がどの英語から来るかということです。そしてこれら確率がどれくらいになるかに関する私の初期推定を与えまして、これらアラインメントの確率を計算します。この今の私が申し上げたアラインメントを用いた場合ですと、

John が Jean を produce,beat が est battu を produceするのですので、確率はこのようになります。これは John が一つの単語を produceする確率かける John が Jean を produceする確率かける、この言わば John が現在第６番目の位置にあるという理由で第６番目の位置にシフトされる確率かける、「does」が完全に脱落する、ありませんでしょう、確率つまり「beat」が est battuに入る確率等々であります。これが私のモデルなのです。

そして私がやっておりますのはこの推定プロシージャーにおいて非常に単純なアプローチです。私の最初のモデルだけを使ってもっとも可能性のあるアラインメントを見つけ、これが実際に起こったことだという視点に立って、これらすべての確率に対してカウンターを得ます。そして適当なカウントをこれらカウンターに入れるのです。で、この今の場

合ですと、カウント１をこの

John が Jean になるところに入れ、カウント１を

dog が chien になるところに入れ、というふうにやって行きます。

また、カウントを一つ追加することにもなるでしょう。それはフランス語の位置第６が英語の位置第１からくる確率に対してです。これは、私の場合 8,000万語になる全コーパスを通じて行ないます。つぎに全確率を角度正規化し、コーパスにもどり、これをもう一度やります。それからしばらくしてくり返しをやめます。

これが私の推定値です。こうした推定値は何なのか。どんなふうに見えるでしょう。まず用語集をお見せしたいと思います。このことはつまり、この私が得たものは、このプロシージャーから得るさまざまなおもしろい語に対するさまざまな確率であるということです。「the 」という語は 0.936の確率で、フランス語の１語を生成し、確率 0.064で消失する。これが何かを生成するときには、「le」が40％で「la」が20％となります。つまり皆さんが予想されたあたりではないでしょうか。ここ以外では現実の数字があります。この語は必ずしも、二つのフランス語に翻訳されるとばかりは限りません。時には消失し時には一つのフランス語になります。英語の「not」のマッピングの結果もっともなりやすいフランス語は ne pais non rien です。

それではつぎに、いくつかまた興味ある語を見てみましょう。それらは一種の互いに鏡に写した像の関係にあります。左側には「still 」という語がもっとも「encore」に翻訳されやすい確率があり、反対側には単位対照としてその反対の確率があります。

ところで今これを行なったときの語彙数は 9,000語でした。ある語がある語に翻訳されたときに、その語がこの 9,000語になかった場合、それは未知語としました。これがそこにある星印の意味です。 9,000語のボキャブラリーにない語はすべて未知語とします。こうしてこの高い確率が得られるのです。これをご覧になることはできますが、この問題で皆さんを煩わせることはしたくありません。それからもう一つ、いや、これでいいのです。そう、これ以上はもってきておりません。

つぎに興味があることは試験文を使うことです。つまり英語、フランス語両方に存在し、学習の間機械が一度も見たことのないセンテンスを使って、どんなアラインメントがもっともありそうかを見る、ちょっと見てみることです。ここにいくつか例があります。さてこれは、これら２文の語のくり返しを行なった後でもっともありそうなアラインメントです。相当にいいです。「not 」は「ne pais 」になっていますし、「implemented 」は「mises en application」等々となっています。これはかなり大きな希望を与えるものではないかと思います。

ここには輝かしいセンテンスがあります。われわれが非常に好んでいるものです。とても長いのです。お読みになれるかどうかわかりませんが、特にわれわれが誇らしく思っている点は、この「starred 」という語が文中ではこんな前にあるのに、「Marquees de un asterisque 」と翻訳されています。それからこの方法はもちろん翻訳には成功しないとしても、真の句ボキャブラリーつまり真の句辞書を得るためのすばらしい方法であります。

これらはどの辞書にも見つかりませんし、またどの辞書をもっても始めたことがありませんので、すべて等しくありそうなことなのです。では今、表示をお示しして、われわれは翻訳実験をやりました。この翻訳実験はここ、ここのこの後ろです、に記述しております。英語の語彙は英語の全コーパスに見いだせる語彙です。フランス語の語彙はフランス語の全コーパスにある語彙です。言語モデルはＴｒｉ－ｇｒａｍです。そしてわれわれはやりました。テキストの 3,000万語についてパラメーターの抽出を行なったのです。おわかりいただけるように、これはやや計算的に高くつきます。しかし当社は39台の大型コンピュータを持っておりますので、これはできるわけです。

そして皆さんに申し上げたいことは、これは単に試験テキストだけでは終わらないということです。つまりテキストはただセンテンスを取り出すだけでこれをまったく見ないというのでは見ることはできません。われわれは実際チェックしなおして、試験テキスト中のセンテンスはどれ一つ学習テキストの中に現われないことを、完全に確認しようとしました。これは少々危険なことです。というのはこれらは、あー、この試験では短いセンテンスを扱っているからです。短文は何度もくり返されるということはあり得ます。「shame 」のようにです。これがわれわれの結果です。これはまだ世界をあっと言わせるものではありません。その結果のいくつかをこれからお目にかけましょう。ここにあるのが文です。これらは完ぺきに翻訳された文です。 これを消しましょう。するとお読みになれるでしょう。私が完ぺきと言うときには、これがわたしの翻訳だということをご理解下さい。これが Hensardデータに実際に現われた翻訳です。翻訳してみますとイディオマティックだということがおわかりいただけるでしょう。機械にとりましてはイディオマティックな翻訳は相当にむつかしいのです。

それでは何が正しいものかご覧いただきましたので、これから正しくないものを見ることにいたします。センテンス個々の解析にご注意をいただきたく思います。これがそのセ

ンテンスです。フランス語は

en true dialavicve proverts meioan.

です。ヒューマンつまり議会の翻訳者が最初のセンテンスを翻訳し、われわれは２番目のセンテンスで行ないます。これがその解析です。これが示すところは、失礼、ところでこれはログ単位での言語モデルの確率です。もちろん確率は常に負ですから、負であればあるほど、確率は低くなります。この赤点をつけたところでは、勝者は高い確率を持っている者です。これで言語モデルは機械翻訳の方が好きで、翻訳モデルは機械翻訳モデルが好きだということがおわかりいただけたと思います。特に誰が何をより好むかというのは、翻訳モデルがここで獲得することですが、それはひずみ確率に由来します。

事実人間の翻訳は語対語訳ではありませんから、「se one」という語を、たぶんわれわれが好まない位置へもってきます。ここのこの位置です。ここでおわかりになれるのは、「on」が「one 」から来て、［through 」が「find」から来てというふうになっているということです。これがフランス語センテンスでのこの語の元の位置です。ここでは「suivant」はどこか別の所に置がねばなりません。順序が整っておりませんので、ペナルティを払います。ここにもう一つ言語モデルは人間の翻訳の方を好む場合が、翻訳モデルは機械翻訳の方を好むという例があります。ここでもまたこれは主にひずみ確率、ひずみ要因のせいなのです。もちろんわれわれにはひずみ要因が問題なのははっきりしています。このひずみ要因は間違っています。というのは、仮にある何または節を取ってそれを英語のセンテンス中のある位置から他の位置へ動かしたとしますと、このことはもちろんいつでも起こっていることですが、そうするとこのモデルによって私はこのペナルティをこれらの単語のそれぞれに対して払うことになります。明らかに私はこのペナルティを一度だけ払わねばなりません。このシステムをつぎに具現化するときには、われわれはそんなふうにやるつもりです。ここにまた別の反対のものがあります。ここでは言語モデルは機械翻訳を好み、翻訳モデルは人間の翻訳を好みますが、機械翻訳が勝ちます。これはすばらしい翻訳です。まさにわれわれの今の混沌の時代には極めて優れたものであります。ご覧のように統計的方法は知識と語彙を抽出し、カナダ議会のデータベースの奥にあるものが何であるかを調べるのに、非常に優れたものであります。これで私の発表を終わります。ありがとうございました。何かご質問がありますか。

質疑応答

問　エジンバラ大学の Pete Whitelock と申します。あなたはシステムを 1,000時間、8,000万語について学習させられたのに、達成精度は40%だとすると、質の改善の望みはどこにあるのでしょうか。

答　多くありますよ。まずわれわれＴｒｉ－ｇｒａｍ言語モデルを追及し続けることは論を待ちませんし、われわれが持っているものもＴｒｉ－ｇｒａｍ言語モデルしかないわけです。われわれが持っているものはまさにそれなのです。ですからまず、対応する事柄の翻訳を試す文法を何か持たなければならなくなるでしょう。対応のないものはだめです。第２に、われわれはもちろん今よりも優れたひずみモデルを持つ必要性もあります。われわれのひずみモデルにとって何が問題かはすでにお話しいたしました。第３に、単独の英語の単語から複数のフランス語に行くことは防御不可です。われわれは明らかに英語の表現からフランス語の表現へ行かねばならないのです。これがわれわれの立場であり、この三つのステップが近い将来にわれわれのプログラムに入れて、目に見えるようになるものです。われわれはこの研究が今ご覧になったように、計算的に極めて集約的であると思っていますが、仮にこれを粗く行なえば、つまりプログラムを組むときに、それを組みたいように組めばです、これはまったく不可能になるでしょう。ですから主に記憶上の節約をするためにおびただしい数のプログラム上のトリックがあるのです。というのは先程の言語モデル確率をすべて検索する能力が要求されるわけですから。あなたにそれができなければならないのです。このアラインメント作業は、カウントを増やすために最善アラインメントを見つけようとする場合のものですが、もちろんｎの階乗です。ここでｎはセンテンスの長さです。明らかにわれわれは何らｎの階乗を行なうのではありません。われわれはヒューリスティックを持っています。しかしそれらはここでもまたわれわれが提供し得ると考えている最善のヒューリスティックではありません。ましてやわれわれが考え得る最善のヒューリスティックでもありません。

問　しかしあなたの分野、あなたの分野の大半は、あなたが改良を行なおうとしておられるところでしょうが、それらはいずれにしても生成文法を組み込むということの下にくくられてしまうように思われます。

答　そうですね、私の考えでは、この方法の強さは特別な場合を何も考えなくてよいのに反して、通常の翻訳プロジェクトでは考えなければならないという点にあると思います。つまりそうしたプロジェクトではたえず語彙にこれやあれやを加えなければならないのです。ところがこれ、このアプローチではそれをすべてあっという間にやってしまいます。何も成功するとかそういうことを言っているのではありませんが、できるところまでやってみようというつもりではおります。そうして振り返ってみて、いくらか文法を制御できるかどうかを見ます。明らかに翻訳文法はとるに足りないもの、違いますか。つまりアメリカに何百とあれば、日本にもおそらく何百とある。したがってそれら文法にアクセスするのは、われわれの単純なプロシージャーを使い果たしたときだろうと思います。

他に何かご質問は。それでは当セッションの演者、お三方にお礼を申し上げて、閉会といたします。

Session 4

Future Direction of Integrated Processing of Speech and Language
(Panel Discussion)

鹿野清宏（ＡＴＲ）

ＡＴＲの鹿野です。　まず最初に、現在ＡＴＲ、自動翻訳電話研究所だけでなくて、視聴覚機構研究所も一部入りますけども、その分野で、音声認識関係でどういうことを今やっているかということを簡単に紹介させていただきます。

まず最初に、よくこの図面を使っていますけども、これかなり古いんですけども、我々は自動翻訳電話、例えば日本語から英語というものをこのようなブロックで現在いろいろ考えております。このためにはやはり不特定話者の音声を扱わなきゃいけないと、しかし不特定というのは非常に難しいであろう。そのために話者適応というような話者の適応化とか、そういうことを考えようと。

それから２つ目に認識方法として確率的なアプローチというのは非常に強いんですけども、必ずしもこれだけでどこまでできるかというのはなかなか自信がない。したがってやはり知識ベース、特徴ベースか、あるいはニューラルネットというようなものもコンバインして、並行して進めていこうというのが音声認識の立場です。

それから次にこの部分が言語モデルの場合です。従来機械翻訳では、どちらと言いますと正しいテキストが入ってきて、それを解析して、うまくいけば翻訳したいというような考えでやっております。しかしながら音声認識では、きょうもずっとそういう議論があったように、次にどういう単語が来るかとか、どういう音韻が来るかというようなことをプレディクトするということが非常に重要です。わざとここにソースと置きましたのは情報理論的立場から、やはり言語のモデリングというのをしっかり考えようという意味で言語ソースモデリングということに重点を置こうということを考えております。モデルとしてはＴｒｉ－ｇｒａｍとか、あるいはニューラルネットというのは、統計的なモデル、あるいはきょう午前中川端が話しましたような文法ベースで、そのうちに、さらにその中に確率を放り込んでいこうという、こういうパーザーベース、２つのモデルを考えております。

これで認識をいたしまして英語の翻訳して、規則合成を出して、あわよくば発声者の声質に声を変換して出そうというのが概要です。これらの要素技術について、いま現在研究を行っております。

まず最初に連続音声認識について、どのようなことをやっているかということを簡単に紹介したいと思います。

御存じのようにHidden Markov Modelというのは、非常に強いシステムでして、我々のところでも、まだ文節認識というレベルで評価いたしますと、ほかの方法よりも非常に強くて、やはりいい結果を出しております。

それからもう一つ、あと３つアプローチがあるんですけども、一つ目は学習ベクトル量子化をしたフィンランドのコホーネン先生と非常に近い話なんですけども、それを使った方法を行っております。それからニューラルネット、Time Delay Newral Networkというニューラルネットワーク的なアプローチを行ってます。

それから最後に、Victer Zue先生のところでもよく行っておりますような特徴ベースと、それからリーディングのプロセスというのをしっかり理解して、そのプロセスをエクスパートシステムを書きながら、さらにその中にいろんなものをつかんでいこうということを並行してやっております。

しかしながら、やはりこれそれぞれの方向だけで進めるということには、なかなか無理がありまして、あるフェーズでは、それぞれのメリットを総合するということが非常に重要だと考えています。例えば学習ベクトル量子化の場合ですと、これにHidden Markov Modelというものを統合していこうということを現在やっております。学習ベクトル量子化は違う言い方をしますと、マトリックスベクトル量子化というように、音声のダイナミックスというのを中に入れることができて、従来のフレームワイズの普通のHidden Markov Modelよりも、あるキャパシティーがあるだろうと。その中になおかつニューラルネットワークが持っているようなコレクティブラーニングの能力もありますので、それを生かしてやっていこうと。実際これはかなりいい線まで今いっていると思います。

それから次にニューラルネット、もちろんニューラルネットを単に識別だけ

で使うのだけでは不十分でありまして、やはり音声のどこに音韻があるかということが非常にあいまいであると。そういうシフトが起こってもできるような方法を開発しようというので、時間軸についてのタイムシフトインバリアンスを持ったようなネットワーク、ニューラルネットというのをやっております。これも音韻スポッティングとか、例えばＤＴＷと組み合わせまして、連続音声に持っていこうという途上にあります。

それから最後に知識ベース、なかなか知識ベースだけで音韻を認識するというのは非常に難しくて、最初はそれをやろうとしたんですが、なかなか難しいと。比較的簡単に書けるのは、どこに音韻があるかというセグメンテーション知識というのは非常によくわかると、そのプロセスを知識ベースで書いて、しかしながらあとどういう音韻があるかということは、なかなか難しい。

例えばフォルマントのローカスというのは、なかなか見つけるのが難しいと。そういうものが現在はＴＤＮＮ、ニューラルネットワークを使って音韻の識別をすると。２つをコンバインして、知識ベースとして、システムをつくっていくことを考えています。

こういうシステムが実際ありましても、言語モデル、モデルというのがないと、実際は何を言ったかというのは認識できないわけですね。これは耳だけで聞かれてもなかなか頭がなきゃやはり音声は認識できないというわけで、そういう言語モデルの研究も初めております。

一つはカーネギーメロン大学の富田先生がやられました一般化ＬＲパーザーというものを使いました。これはある種の予測ができると、そういう意味でソースモデリングと非常にふさわしいし、なおかつ文脈自由文法が使えると。そういう意味で、それを利用しております。それがこの話です。これは午前中川端が話した話です。その中にやはりモデリングと言うからには、テキストデータベースから、いろんな統計量をとって放り込んでやりたいと。そしてパープレキシティを減らして、精度を上げたいという意味で、ストキャスティックＬＲパーザーというのを今開発しております。

これでかなり精巧、認識精度は上がっております。この場合テキストデータベースというのを使っていきます。

しかしながら、よく日本語を見てみまして、多少アメリカに比べてテキストデータベースというのは規模が小さいわけですね。なかなかまだ十分な量がないと。

もう一つのアイデアは、言語の中の単語より下のレベル。例えば我々は多分日本語の無意味な単語を言われても、それを何が言ったかというのはかなりわかる。人の名前とか、何かは、ある程度わかります。それはやはり日本語の音韻の続く規則をある程度知っているからだと。そういう意味で音韻のシーケンスをあらわすのは、音韻ソースモデリングということも考えていこうと。だから単語以下の所にも言語情報はいっぱいあるんじゃないかということを考えます。

そういう観点から言いますと、基本要素が音韻とか、シラブルとかというふうになりますので、テキストデータベースは少なくとも、かなりの量が集まります。カテゴリの数も非常に少ないわけですね。 100ぐらいか、音節で 100ぐらいしかありませんので、そういうのでこういうモデリングの問題を統計的にどうやってやったらいいかということを今やっております。これがうまくいけば、音韻タイプで当たってみたのを実践したいと。現在はこれとＨＭＭと組み合わせて音韻認識率を出してますけど、大体90%程度の音韻認識率が得られております。

最初知識ベースとこれを組み合わせまして、例えば音韻タイプライターを実践したいということを考えております。

今の話はすべて話者依存の話をいたしまして、不特定のことは一切含まれておりませんで、不特定ってなかなか難しいわけですね。不特定の問題というのは非常に難しくて、我々の方では話者適応ということを考えています。その中には人の音声の空間というのは有限個のものであらわしてやって、その空間をほかの人の有限個の空間にマッピングするという、コードブックマッピングという考えで進めております。ただしディスクリートなもので表現しますと、非常に粗くなりますので、その間の補完をファジーという概念で補完してやろうということで、連続性を保つようなことをしております。

それでそれがかなりうまくいっておりまして、例えばボイス、人の声をほかの人に変えるというのは声質変換にも適応できますし、あるいはバックグラウンドの雑音の適応化にも使えるであろうと。かなりそれが進みましたので、次

は本当に不特定でやりたいと。不特定のキーとしては、なかなか余り持ってないんですけども、一つはここに話者モデルというようなことが書いてあるんですけども、実際に入力の入ってくるところの時系列を、今まで独立値としてかなり扱ってきた面があります。それをもう少し時間的な相関というのをうまく統計的に取り上げてやって、そういうのを放り込んでやって、どういうんですが、その空間というのを自動的にスペシファイするというのは、非常にまだ実際に研究をやっておりませんので、やってないことはないんですけども、少ししかやっておりますのではっきり言えませんけども、そういうことを使ってやっていこうということを今考えております。

そのほかニューラルネットの分野でも、ここら辺にノンリニアマッピングみたいな方向もありますし、いろんな有望な方向が出てきつつあります。そういうのも試して、不特定に挑戦したいと思っております。

もう一つ我々のところで重点を置いていますのは、日本語の音声データベースというのをかなり重点を置いて集めております。ここに書いてあるのが、大体我々のところでやっている主なデータベースです。最初は特定話者用の大規模な単語音声データベース、5,000 単語以上になりますけど、それにラベルしたデータベースを随分集めました。そのあとに連続発声、音韻バランスした連続発声というものの連続音声のデータベースを、また10名ぐらい集めました。現在は研究の進展に伴いまして、不特定の音声、これを集めている段階です。これが20人程度集まった段階になっております。そのほか将来としては雑音のデータベースとか、あるいはプロソディーも将来使いたいと思います。ピッチのデータベースをつくるとか、あるいは多少分野が違いますけども、音声合成のデータベースというのをつくっていこうと思っております。

以上簡単にもう１回まとめますと、一つは連続音声認識では、できるだけいろんなアプローチをとってやろうという方針をとっております。先ほど言いませんでしたニューラルネットの中でも、必ずしもＴＤＮＮだけじゃありませんでボルツマンマシンもやっておりますし、あるいは非線形予測モデルもやっております。そういうものをコンバインしながら一つのモデルをつくり上げようというのが一つのねらいです。

それからもう一つの面は、言語モデリングというのをテキストデータベースを使って、それをしっかりつくりたいと。この部分では、Jelinekさんたち、非常に影響されてるんですけども、そういうものを日本語でもしっかりやりたいということを考えております。

それからもう一つは、話者適応ということでやってきたんですけども、いよいよやはり大語彙連続の不特定ということも基礎研究として、これからやっていきたいと。その手がかりとして話者マルコフモデルのような時系列間の相関をうまく使ったような方法というのを考えていきたいということをやっております。また後で違うことを話させていただこうと思っております。以上です。

藤崎　鹿野さん、ＡＴＲにおける音声認識の現状とプロジェクトに関して、たいへんあり　がとうございました。パネリストとフロアの皆さんの紹介をお願いするよりも、まずRenato De Mori教授に教授の最初の発表をいただきたいと思いますが、簡単にご紹介しておきますと、 De Mori教授は1941年、イタリア・ミラノにお生まれになり、ミラノ工科大学で教えてこられました。現在はカナダの方で、 McGill 大学のコンピュータサイエンス学部の学部長をやっておられまして、ご専門は推計学的言語モデルと接続モデルの応用でいらっしゃいます。またＰＡＭＩパターン解析および機械知能さらにコンピュータ音声と言語に関するＩＥＥＥトランザクションの編集者でもいらっしゃいます。 Renato さん、お願いします。

Renato De Mori (McGill大学)

Cache-Based Language Model for Speech Recognition

藤崎先生ありがとうございます。皆さんよろしくお願いします。まず最初に論じたい項目は、この一枚に要約してございますが、つぎのような問題に関するものです。今、推計学的言語モデルを使いたいと仮定してみましょう。これは数値決定基準を適用したい、その他の理由からです。さてわれわれは今、われわれの推計学的言語モデルに対して良好な確率を収集してはあるが、ある単語がある点にある予想は一般的、静的確率だけでなく、最近の過去において言われたことにも依存する可能性があるという事実を考慮に入れたいとしましょ

う。これら確率モデルにおける事実を考慮にいれる最初の試みについては、近々発行される論文に記してありますが、つぎのスライドに要約しています。これに関して簡単なバックグラウンドを説明させていただいて、それから議論や結論に入りたいと思います。

われわれが興味を抱いておりますことは、前に一定の単語列が与えられたとき、つまり解析状態が与えられたとき、ｉ番目の単語がＷである確率を計算することなのです。これは発話の部分として表されますので、ｇｊは発話の一部に対してそのままとなり、Ｗが関連した発話の全部分にわたって変動する合計によって近似することができます。ですからこれは何も目新しいことではありません。これは文献にも発表されております。例えばDerovault , Merialdoがそうです。われわれの役割はこれをどういうふうに計算するかにあります。発話の部分ｇｊが与えられたとき、単語Ｗの確率をどう計算するか、です。かいつまんで申し上げますと、こんなふうに計算することを提案したいのです。二つの組み合わせとして、Ｋｎ，ｊかけるこの確率が起こる統計的頻度、プラスＫｃ，ｊかけるキャッシュメモリーの内容となりますが、このキャッシュメモリーとは、時間ｉにおける発話ｊの一部としての単語Ｗの頻度を累積したものです。というのは、ある日におけるキャッシュメモリーは各瞬間瞬間に更新されるからです。Ｋ、ＪおよびＫｃ，ｊといった係数は最大見込法によって推定します。

われわれはこのアプローチを用いてＬＯＢコーパスによる実験を行ないましたところ、キャッシュメモリーを採用したことによりパープレキシティが３倍減少するのを観察できました。もちろんこうした結果は少々楽観的すぎるきらいはあります。というのはＬＯＢコーパスはこの頻度に関係する確率を細かく推定するには十分な単語を含んでいるとは言えないのです。ですから実際の実験ではこのアプローチを採用する利点は、今の場合ほど大きくないでしょう。ただそれなりの実験はそれなりの結果を出しています。

藤崎　Renatoさん、ありがとうございました。これは改良、言語モデルの改良に関係したものです。もちろん言語モデルはトピックや文脈以上に関係したものですが。つまり De Mori教授の論点は、言語モデルを使用する際には、文脈や話し手や、最近に扱ったトピックやらに対して適応性がなければならないということでしょう。これは言語モデルがどのように改良されるかということ、どうすればよりよい応用ができるかということについての新しい考え方でした。

ではつぎに白井教授に最初の発表をお願いすることにします。白井先生は1939年中国にお生まれになり、早稲田大学を卒業され、現在同大学で電気工学を教えておられます。先生はこの20年来、音声認識にたずさわってこられ、特に音声合成プロセスおよび自動音声認識とともにそのモデル化、またロボットによる音声インターフェース等にご興味をお持ちです。早稲田大学はロボットに関して非常に強力なプロジェクトを持っておられまして、先生は単に音声認識そのものにご関心がおありだというにとどまらず、大学の方では人間・ロボットの強力なインターフェースをお持ちです。それでは白井先生お願いします。

白井（早稲田大学）

きょうのお話は、たくさんいろいろあったわけですけれども、何らかの方法で制約を入れると、ここではストラクチャーと書きましたけど、を入れることによって計算量を減らすというか、広い音声認識、あるいは翻訳というようなことに結びつけていくためのアプローチがどういうふうにあるかということ。過去にどんな努力がされてるか、ここに大分前から、昔のものから書きました。そういう整理的なモデル、あるいは音響的なモデルとか、あるいはperceptualとか、linguisticも、これはたくさんいろんな分野があろうかと思いますけども、それからここにちょっとマジックで書くことになりましたが、、会話的モデルというのは言われていますけれども、現実には余り研究データがないという問題があるわけですね。やはり会話的モデルをつくるためのデータを集めるということが非常に重要だというのも、これは将来志向の研究のベースとして絶対に必要であると、ここをやらなきゃいけないんじゃないかと。

それから従来セマンテイカルモデルとしてこういういろんなことがあったかと思います。そういう問題の中で比較的成功しているものということで、例えば私自身が司会者の紹介にありましたけど、こういうモデルをつくって、モデルベースで、モデルをつくるとどういうことが起こるかと言うと、こういうモデルからジェネーレートされる例えば音声ですね、これは普通音声合成という

わけです。その逆を我々解かなきゃいけないと、それで非常に大きいここに制約がある。

そういう制約のもとでのインバースプログラムを解くんだと、そういうアプローチというのは、非常にもっともらしいし、非常に強力なアプローチではあると。だから私も非常に過去に憧れていたし、今でも憧れていますけれども、かなり一生懸命やっているんだけれど、非常に大きい人間の持っているこのスピーチのバラエティーというものになかなか対応しにくいという事実があると。要するに非常にバラエティー大きい話者の個人性を初めとして、方言もあるし、文法も、非常に人によって違うし、大体語彙が違うという問題があると思うんですね。

そういうものが個々にしか扱うことができないという問題が、実際はすべてのこういうモデルと書いた方がいいかどうかあれですが、いろんなエレメント、ファクターが非常に緊密に結び合っていて、きょうの話でも音響、あるいは音響的な認識と、それからパーサーとを結びつけるというような努力が非常にされているわけですが、とにかくいろんなものを結びつけないとなかなかうまくいかないんだということが現実としてある。

それじゃどうなってるんだということなんですが、将来であるのか、あるいはカレントトレンドであるかというふうな気がしますけれども、要するにシンプルでないということですね。シンプルでないとすりゃ、要するに簡単な原理原則でもって、こういう原理でシステムを全部貫いて動かそうということは、それが原理だと言ってしまえば原理だけれども、非常に複雑なものをつくらなきゃいけない。そういうものを人間の力でもちろん我々よく考えて、その中身を分析してですね、できるだけシンプルな原理に基づいてやるべきであるということは事実だと思いますけども、いずれにしろどうも例外というのが常にある。大体こういうものは、ちょっと例外の量が多くなると、これは既に全部質的な問題になるといいますね。そういうことがあって、要するに現象論的なモデルを我々つくる、有効につくる道をつくらなきゃしょうがない。そうするととにかくデータがたくさんなきゃだめだ。それからそのデータをいかにしてスピーディーに処理するかという努力に皆さんなってきているんではないかと。それはそのアプローチとしての確率的であり、あるいは情報理論的であるというアプローチが非常に成功するということにならざるを得ないという気がいたします。それが一つ。

それからもう一つは、これはさっき言ったように、非常に結びつきが強いということから、一つの大きい全体でトータルにとらえるということが非常に重要じゃないかという気がするんですね。きょうのお話でも、やっぱりコンポーネントのとらえ方になるんですが、というのは、我々の今の力でできることは、そんなに大きくはないということかもしれませんが、要するに音韻的なレベル、それからシンタックスというのも結びつけようという考え方ですけど、そういうことでいけるのかどうかということですね。

やはりもうちょっと大きい会話なら会話を対象にして、その会話の中のいろいろなステートというのをもうちょっとアブストラクトにとらえるということが必要なんじゃないか、そういう記述方法というのがないと、なかなかうまくいかないんじゃないか。やっぱり入ってきた文章を我々一生懸命、とにかく何か正しく認識しようという意識が余りにも強いという、私はやはり気がするんですけどね。

そういう立場から、どんなことができるのかというのは、まだ私にも答えはあるわけではないんですけれども、やはり会話の場面というのを、できるだけたくさん集めて、その中で起こってくる現象をいかに近似をしてステートに分けていくか、区分していくかということかなと思います。

それと、そこで発せられるさまざまな表現という間の相関関係というものをやはり密接に追っていかないと、なかなか役に立つような認識装置、あるいは翻訳というようなところまでつながってこないのでないか。それは一種のコロキアルな表現をできるだけきちっと集めるとか、そういうようなことにも通ずるのかもしれないし、いろいろですけど。

テクニックは、今私が申し上げたことのほとんどあれなんですが、一番最初に簡単な構造をいろいろ組み合わせるということを当然やられていると思います。だけどこれだけでいくだろうかなというのが、ちょっと私にはよくわかりません。

それから一番最初に書いたそのストラクチャー、効果的なストラクチャーを探せということは、こういうシンプルな構造の組み合わせによって、その構造

が適切なところに使われれば、それなりの組み合わせで、大きい構造を全体につくることが確かにできるであろうという気がしますが、もうちょっと音声の中に含まれているさまざまなストラクチャーというのは、非常にいろんなものがあるという気がするんです。

ですからいろんなバラエティーを持ったストラクチャーというのを、効果的なストラクチャーというのを探すべきだということであります。

例えば文法というのは、我々書き言葉から来ているんですけれども、書き言葉から来ている文法というストラクチャーは、確かに非常に有効で強力であるけれども、余りにも強力過ぎるという面もあるし、もうちょっと違ったストラクチャーを探さなきゃいかんのではないかと。

あとは自己組織化、これはさっき言った情報理論的な方法も含めて有効な方法を考えないと、なかなかいかないです。

あと高速サーチについては、後でまた少しお話できたらと思います。

藤崎　白井先生ありがとうございました。

それではつぎに、ＭＩＴの Victor Zue 博士から、言語処理および音声処理の将来についてこれまでなされてきたことにつき、広くお話をうかがいたく存じますが、その前に30と１秒ほどお借りして先生のご紹介をさせていただきます。 Zue教授は1946年、中国でお生まれになり、ＭＩＴで学位をお取りなられ、以来20年間ＭＩＴにおいでになります。主に音声認識の各分野で研究をおやりで、スペクトログラフの読み取りにおけるご専門では有名であられますが、他にもタイムデータベースや音声データベースでも非常にすばらしいお仕事をされておられます。これら分野のほとんどにおいて、ことにスペクトログラフの読み取りとデータベースの非常に精密なラベリングでは、今、世界で、他の追随を許さないのではないかと思います。ただ目下のところはＭＩＴコンピュータサイエンス研究所で音声認識システムグループのヘッドをしておられます。では博士お願いします。

Victor W. Zue （ＭＩＴ）

藤崎先生ありがとうございます。今日われわれは音声と言語の統合処理についてたくさんの興味ある発表を耳にしました。研究が世界中で活発に行なわれていることは明らかであります。ＭＩＴでもわれわれはある時期音声研究に従事したことがございましたが、近年は音声言語システムの開発に取り組んでおります。私の予測を提示するのはむつかしいことです。お集まりの著名な皆さん方を前に、いわゆる言語研究の今後の方向を予測するなどとは、まことに僭越かと存じます。そこで私の希望としては、この後数分をいただいて、私どもの所での研究課題について私の個人的見解を申し述べ、、つづいてわれわれが重要と考えかつこの５年ないし10年に追及したいと思っております点のいくつかを指摘させていただきたく思う次第です。また自然言語についてほとんど何も知らない一人として、できれば音声認識に焦点を絞って話を進めさせていただきたくも思います。私の話の中ではたぶん五つのトピックスに触れることになりましょうが、それぞれにだいたい１分をあてる見当でまいります。

私が重要と考えます第１の事柄は、音声信号の変異性を理解することです。今や音声信号が大量の変異性、音響的変異性、音声学的、音韻変異性、といったいろんな変異性を示すことはすべからく知られているところであります。そして私の考えでは、これら研究分野のうち、われわれが追及してみたいものがいくつかあると思います。例えば、聴覚的モデル化ではわれわれは、人間の聴覚についての知見に基づいた信号表現を組み込むことにより、実際に音声認識性能を改善することができるということを知っています。そしてこれまでのところまだわずかに１年だけの研究でしかありませんが、私は、例えば、われわれはこのモデルの改善を続け、また耳も二つ組み込んで、音の位置確認といった問題の解決に資し、局部的妨害に対するシステムのロバストネスを高めるべきだと考えています。音声学的モデル化も重要な問題であり続けるでしょうし、最近では一種の文脈依存音形（Phone)に懸念を抱いている方々もいらっしゃいます。

しかしわれわれは、例えば、音声学的ディメンションの変異は、ただに局所的な文脈よりずっと一般的であることを知っています。またもっともっと洗練された音声学的モデル構築する試みをすべきことも明らかでしょう。人が違えば話し方も異なります。皆さんはこの種の変異を語彙表現の形でとらえたいと思っていらっしゃるのではないでしょうか。実際ほとんどの方は非常に単純な

語彙的表現つまり一単語一発音というものを得る努力をなさっておられますし、そうでなければ2パスプロセスで発音ネットワークを得ようとされております。

最初はルールで発音ネットワークを生成し、つぎにこれらアークを適当な確率でポピュレートするわけですが、私はこうしたプロセスをどういうふうに結合し、これら発音ネットワークを生成する自動的方法を達成するのかを、われわれは恐らく考えてみた方がいいと思います。

最後に私はテンポラルなディメンションは非常に重要だと考えます。De Mori 教授も述べておられますように、先程のご発表でですが、セグメントのディメンションでの期間的モデル化でも、発話速度を扱い始めています。それでこれらが変異性を理解するというこの分野での問題のいくつかであります。

ではつぎの問題として語彙アクセスの方に移らせていただきたいと思います。もしこのスライドを10年前にお見せしていれば、現在の状況の中で適切なのとまったく同じくらい適切であったでしょう。そして今から10年後、このスライドをお見せする機会があれば、それよりまだ適切なものとなりましょう。語彙を表現するのに使用する単位はどんなものでしょうか。平坦な構造を用いるべきではありません。今日では、顕著な特徴から音声学的セグメント、あるいは音素、音節、単語等へと変化する階層構造を用いるべきだと信じる人たちが増える一方です。こうした単位が何なのでしょう。私には、何か一まとまりの単位があるとは考えられません。これら単位について利用すべきある組み合わせはたしかにあります。しかしこれら本質的に異なる単位からどういうふうに根拠を組み合わせられるでしょう。

仮に高頻度の語を持っているとしますと、それらの語については多くの学習データをお持ちだということになります。そして当然ながらそれらに多くの注意が向けられます。でなければ、たとえば音声学的セグメントを使う方がいいということになりかねません。もっとも重要な疑問はいかにしてこの制御ストラテジーが語彙アクセスのためになるかという問題です。今朝、左右性、アイランド・ドリブン、その他さまざまな語彙アクセス法の種類について発表を聞きましたが、もちろん関連問題は、いかにして音声認識を自然言語処理に結び付けるかという点にあります。

つぎにお話ししたい第3の分野は自発的発話に関係します。コンピュータとの対話中においては、人間はセンテンスを読むことはいたしません。自発的発話にはあらゆる現象が含まれます。たとえば偽りのはじまり、躊躇等がそれです。皆さんがいかに大きなボキャブラリーをお持ちでも、ユーザーは非常に創造性が豊かであり、新語を持ち出し、さらには非文法的な構造に取り組まさせられることにもなります。アコースティックな理解、音声学的理解というのは、言語的理解とともに、自発的発話現象の理解では非常に重要です。最近われわれは音声言語システムの研究を始めましたが、自分たちのシステムの性能が余りにも悪いことで、しばしば力のなさを思い知らされ、時には恥ずかしくなることさえあります。

もう一つ、音声言語システムについて議論を始める、つまり音声と自然言語を統合しようとする際に、今後重要性を増すと私が考えております分野は、韻律論を扱わなければなりません。ここにありますのは、韻律論が音声学的レベルでいかに音声理解を助けるかを示したいくつかの例です。たとえば強勢・非強勢アロホーンを用いて、音声学的認識の精度を高めることができます。語彙アクセスについては、一定の語の強勢パターンを知ることにより、これもまた語彙アクセスへの別の情報源として利用することができます。われわれが研究しておりますデータベース問い合わせ型の問題では、センテンスがWH型の語尾が下がるイントネーションを持った疑問文か、それとも yes・no型の語尾が上がるイントネーションを持ったものかを予測できる能力は、構文解析にとって大きな助けになるでしょう。まあもちろん強調があれば、意図するところが何なのかをわからせるのに役立つイントネーションパターンを使用することもできます。

最後に、もう一つ指摘させていただきたい点があります。それは非常に明白なことなのですが、音声言語システムを開発するためには、われわれは単に多くを知らなければならないだけではなく、自分たちが何をしらないかも知らなければならないということです。私たちはいつまでも、マイクに入ってくる音が何であれ、それが有効な発話であると前提していることはできません。発話はたくさんある種類の音の一つを形成し、様々な種類の音を理解しようとします。それが動的な音なのか、あるいはまたドアを閉めるときの音や犬の鳴き声、たとえば二人の人が互いに話しているときの音のように静的な音なのかという

ことは、非常に重要な問題となるでしょう。

そして思いますに、われわれは音声を多種類の音の一つについて適切な展望の中に置くまでは、有効な音声言語システムを開発できないであろうということです。またもう一つのわれわれが何を知らないかを知るという問題は、新しい語を見分けられるかどうかという能力です。われわれにとって新語が出てくるということは重要なことだと思いますが、これが出てきたときに、われわれはそれを取り扱うことができないといけません。

最後に全体を一つの展望にまとめさせていただきますと、私の考えでは、音声言語システムには、まさに三つの部分があると思います。まず音声認識問題を扱わねばなりません。これは信号を語の集合に変換するということです。この問題はわれわれの自然言語処理と密接に関係がありかつこれと完全に統合されるべきものです。そして最後に応用領域が適切でなければなりません。これらは関係し合っているのです。また私は変異の問題にも立ち入りました。語彙アクセスの問題にも立ち入り、自発的発話の取り扱いですが、これは音声認識とだけではなく自然言語処理にも影響を及ぼす問題です。それではこの辺りで終わらせていただきます。

座長　博士、今後の問題やわれわれがこれから何をなすべきかといった問題についてじつにすばらしい精確に的を得た、筋道の通ったご発表ありがとうございます。ではパネリストの皆さん、各ご発表に関するフロアからのご質問に連動して進めてまいりたいと思います。

何かご意見、ご質問は。パネリストの方々はそれぞれ喜んでご意見を加えて下さり、ご回答下さるでしょうし、短かなコメントもいただけるものと思います。それではよろしければ……ございませんか。はい、どうぞ。筑波大学の板橋教授。

板橋　デモリ先生にお伺いしたいんですが、先ほどキャッシュメモリーを使った動的な確率、情報のお話をされましたが、そのキャシュメモリーの適当なサイズというのは、どういうふうに決められるのかについてお尋ねしたいと思います。

またその適当なサイズは、どのくらいのものとお考えなのか。

答　通訳が途中から入ってきたので、ご質問の趣旨が理解できているといいのですが、とりあえず答えてみます。

問　私の質問はキャッシュメモリーの大きさについてです。

答　実際の大きさは固定されてまして、200でした。すなわち発話それぞれの部分に対して200個の位置があるということになります。適切な大きさを決める、あるいは十分であろうと思われる大きさより大きくすることはおもしろい問題です。近い将来においてこの問題に集中することは考えていません。別の問題に集中したいと思っておりまして、それは同じ概念を単に一つの語に対してだけ用いるのではなく、階層にも用いるというものです。ですから実際の実験では200でした。以上です。

問　その定式化は発話の文法部分に使われる言語の短期的性質だったのでしょうか。また先生はセンテンスのパターンすなわちスタイルを定式化されたという言い方をするとき、それは適切でしょうか。そうした語を用いるとか、発話の部分を用いるよりむしろ、短期的統計が発話の一部になっており、先生はセンテンススタイルの短期的性質をつかもうとされているのだと思うのですが、これは正しいでしょうか。

答　その通りです。われわれが行なったことは単に、概念が有望であるということを示す試みにしかすぎません。われわれはそれを敷衍したいと願っています。たとえば対話に対してですとか、対話の履歴だとかそういったものに対してです。真の問題は、統計的確率を得るための統計を集めるのに必要な現実に手に入るデータベースが、われわれがこれまで行なってきたこと以上に進むのに十分豊かでないという点にあります。

問　Zue 博士におうかがいしたいのですが。

座長　Ken Churchさん、どうぞ。

問　博士に対する私の質問は、前の語彙アクセスのスライドの前だったと思うのですが、そこで博士はそれを10年前にも見せることはできただろうし、これから10年後にも再び見せることもできるであろう、とのことでしたが、楽観的に言って、おうかがいしたいのは、われわれは進歩しているのか、ということなのです。

答　ええ、そうです、そう思います。たとえば私の話の中で、わたしは人々が表現を認め始めているという事実を述べました。フォーマティカルな表現は階層的でなければなりません。あなたはそれをなしとげた最初の人の一人でいらっしゃいますから、当然ご存じのはずですね。

問　たぶん座長もまた Zue博士に少し質問をしてよいかと思うのですが、これをより階層的な語彙にする際、語彙部分をも適応性のあるものにする必要があると感じないでしょうか。つまり以前より前に現われる一定のグループの語に対して一定の話し手またはトピックスを必要とする場合ですが。どう思われますか。

答　ええ、それは間違いなく非常に有効な何かになると思います。ＣＭＵではたとえば Wayne Wardさんと Cheryl Young さんが、その辺りのことをどうやるかについて発表されました。

座長　はい、どうぞ。 Wahlster 博士。

問　あなたは新語の問題について述べられましたが、われわれが何を知らないかを知るという下りについて、あなたはこの下りの中に新語を分類されています。ある言語における多義語の問題もそうです。たとえばドイツ語の学術雑誌でドイツ語と英語を混同したとします。あるいは学術的な話の中ででも。また英語においても多くの名字が日本語や中国語その他に由来するという問題もあります。ではここで健全な分類すなわち新語の分類とは、まずこの語が他の言語から他の音声システムといっしょに借用されたものであることを見分けなければならないということを意味します。この特異的な問題、すなわち自発的発話においてある言語から別の言語への切り替わりを認識するという問題を扱うこのトピックの下に、何かアプローチはあるのでしょうか。

答　はい、それは明らかです。それは非常に重要な問題ではありますが、私はそれよりやさしい問題の解決に努力するのが、まさにこれからの10年だというふうに考えています。

コメント　優先順位はあなたがどこの出身かによると思います。アメリカなのか日本なのかあるいはヨーロッパかということです。確かに、あなたの問題、つまり Wahlstar 教授が指摘された問題はヨーロッパ諸国のかなりさし迫った共通化では特により重要です。それは非常に重要ですが、また日常の新聞や各種文書にはたえず新語が作り出されています。これは辞書では決して見つかりません。これは非常に大きな問題なのです。

問　DeMori博士におうかがいします。文書理解に使用できるパープレキシティ低減技法を使用することに関するお話がありましたが、私はそれを興味深くうかがいました。ご発表の中でかなり普通といいますか、データ圧縮に用いられている技法に非常に類似した技法を本質的にお使いであるように私は聞いたのですが、データ圧縮に関して行なわれていることに関する系統的な研究はあるのでしょうか。また自然言語あるいは音声理解に利用

できそうなものは何なのでしょう。

答　ある程度において「イエス」です。問題はここにあります。統計的に一貫した何かを行なうという問題です。統計的確率をもって。

座長　はい、どうぞ。マイクが来るまで待ってください。お名前をお願いします。

問　ＡＴＲの John Meyersです。 Zue博士が言われた、偽りのはじまりや躊躇といった音声言語システムの問題は興味深かったと思います。私はこれを自然言語の方々にお聞きしたいのですが、どなたかこうした問題に取り組んでおられる方がいらっしゃるのでしょうか。パーシングあるいは自然言語理解の分野でです。

答　そうですね、この自発的でない発話の問題はじつにわれわれの興味をかきたてます。ただわれわれは懸念し始めています。この問題はわれわれが懸念し始めているものなのです。後でちょっとした実演をお目にかけますので、言語学的解析だけではなく音響的なまた音声学的な解析がいかにしてなされ得るのかをおわかりいただけるのではないかと思うのですが。

コメント　もちろん博士はこれを今ここで実演なされるのですが、これは第２のご発表に準備されたものなんです。座長としてはちょっと悩んでしまいますね。博士にではなく、形の悪いセンテンスのくり返しセンテンスや非文法的な変形にです。一方音声言語は、口ごもりや偽りのくり返しといったものを完全に受け入れることができるのです。非文法的ということは文章言語の文法としては極めて滑稽なものです。これに対して何かおっしゃることがあるのではないかと思いますが。何か他にコメントはありませんか。

森元　ＡＴＲの森元です。ちょっとお伺いしたいんですけども、今白井先生のお話にもありましたけども、将来的にはいろんな情報を一つのものに統合していく必要があるというお話でしたけども、音声言語を理解するといった場合に大きく二つあると思うんですが、一つはまずいかに音声認識をするかという問題。それからただ音声認識できてもですね、やはり理解をするという問題が避けて通れないわけで、何らかの言語処理といいますか、理解する過程が必要だと思うんですが、そういう意味で私の個人的なあれでは、非常に音声認識と言語の理解というのは緊密な関係でもって将来構築していく必要があるんじゃないかと思うんですが、その辺のですね、メカニズムとして、かなり将来的にですね、緊密化していくべきなのか、それともやはり音声認識は音声認識、それから理解は理解という問題でですね、とらえていった方がいいのか、その辺について何かお考えありましたら、お聞かせいただきたいんですけども。

白井　私言いたかったことは、やはりそれは統一されるというか、一つのものとして考えられるべきだろうという問題意識をね、要するに音声認識と、それから言語の理解というのを分けた問題とするのはおかしいんじゃないだろうという意味で、それは今までいろんな発展的な研究のいろいろな歴史の問題が一つありますね、これまでの問題のとらえ方として。だけどそれ以外に計算機の能力という問題があると思う。やはりカスケードに順繰りに処理をやっていかないと、非常に膨大な量の計算をやらなきゃならないんですね。その効率を考えると、とてもじゃないけど、今現在やられている以外にちょっと考えつかないというのが現実の問題だと思う。

しかし今後は、もう少し遠った、もうちょっと大きくとらえてもね、要するにサーチスペースを結局増やすということになるかと思いますけども。そういうことが可能なんじゃないかなという意見だったつもりですけども。

白井　非常にいいご質問です。

座長　そうですね、それは認識と理解の統合に関係します。多くの状況では、理解が正しい認識のための前提条件となります。またそれは言語処理と音声処理の統合とも関係します。

Zue 博士がすでに指摘されましたように、将来は部分的にではなく完全に統合されねばなりません。博士は部分的オーバーラップを示唆されていますが、私は博士のことばを文字どおりに解釈するなら、円はほとんど一つになろうと思います。また聴衆の一人として、私は私のこの見方と同じ見方をお持ちの方が大勢いらっしゃるものと考えます。この点に関し Wayne Ward 氏に氏の見方を敷衍していただけるようお願いできますでしょうか。

この問題についてはわれわれは少し前に話をしたのですが。

コメント　そうですね、私は基本的にはあなたが今おっしゃたことに賛成です。それは理解というのは確かに有用で入力の正しいデコーディングのほとんど前提といってよいものであります。これに対する限定詞として、いくつかの役割、たとえばディクテーションのようなもの、がこの種類の領域では近い将来に手が届くようになるとは思えません。ですから私が話をする大半の相手は現実の応用のためのフロントエンドなのです。ここにおいて領域知識を実施することがほんとうに可能になるのです。あなたは実際のところある現実にこの応用にアクセスするのに必要な意味論をおやりですから、プロセス全体を援助するのにできるだけ早い時期にこの情報を使用するのが、まさに有利だと思います。

発言を許されている間に、あと二点についてさっと触れておきたいのですが、一つは自発的発話に取り組んでおられる方々で音声言語システムのための最新のＤＡＲＰＡ共通タスクは、ある応用例のユーザーによる無制約の発話を扱う際のものです。制約はただ領域が強制するものだけです。ユーザーはボキャブラリーを与えられませんし、文法も与えられません。ただ自分自身のやり方で対話することは許され、この特定の分野にあるプレーヤーはすべて、何らかの形でこれを扱い始めることを強制されます。第２点として、私はＺｕｃ博士が言われた、われわれが知らないことは何かを知ること、すなわち拒否能力のある形態に関するスライドにもう一発の弾丸を打ち込みたいのです。システムはいかにしてパースの質あるいは良さやシステムがもたらす認識文字列を認識できるのか。ここでの主要因は、思うに、状況全体の意味論となるのではないか。これはユーザーがやろうとしていることを全体として考えるときに意味をなしたのか、ということなのです。

座長　Fredさん、ありがとうございました。

答　　危険を省みず今の観点を攻撃してみましょう。まず問題は二つあります。その一つはとるに足りない領域です。もちろんそこではあなたは実用的な知識を持っており、いろんなことができます。しかしディクテーションや各種自発的テキストあるいは自発的発話に立ち入れば、われわれはもちろん意味論に関して何の理論もまったく持ち合わせておりません。ですからわれわれが音声認識の前に理解をもたなければならないということは、音声認識をつぎの世紀まで延期しよう、と言うようなものなのです。第２に、あなたと賭けをしてもいいですが、今後５年間に、少なくともパープレキシティに関する限り、文法がＴｒｉ－ｇｒａｍとなることはないであろうということがあります。

われわれはもちろん言語モデルとしての文法を得るように努めているにはいますが、現実は語彙的依存は非常に大きな制約となり、われわれが扱える文法は単に文脈自由型文法だけなのです。これらはまた制約が十分でないことも確かなのです。ですから理解を使わないということが非論理的であるということは、私には、われわれの寿命は 200年もないというのと同じようなことなのです。われわれは単に 200年も生きるようには作られていないだけであるように、われわれは今のところ意味論理論を持ってはいないのです。つまりこれは現実的には、音声認識の基本としての理解を持つようにはならないであろうということです。

座長　はい、これは非常に興味深い問題です。理解が非常に重要なときもあれば、それがそれほど重要でないこともある。つまり状況です。 Wayneさんに何か一言、本当に一言ですが、お願いできますか、 Ken Church さん。ほんの一言ですよ。

問　　はい。フレッドさんがおっしゃった最初の点については他日に譲るとして、第２の点は強調しておきたいと思います。私はこの点に関しフレッドさんはまさに正しいとおもいます。語彙的事項に対する制約やＴｒｉ－ｇｒａｍのような特定の事柄は、われわれが自然言語においてなし得ることはどういう種類のことかといった問題をきれいに払拭してくれるであろうということです。私はわれわれが本当にこの点を極めて真剣に認識しなければならないと思います。皆さんもご存じのように、事実は Bob Mooreが一生懸命に研究を続けてほんのわずかな収穫さえなかったのに反し、Ｔｒｉ－ｇｒａｍははるかに多くを達成するでありましょう。これはまた事実です。

答　　はい。そんなたくさんおしゃべりになるようには期待しなかったのですが、まあいいとしましょう。語の収集は極めて重要であると、私は思いますが、それをＴｒｉ－ｇｒａｍのような三つの範囲に限定しなければならない理由がわかりません。私は Kenさんのお仕事について話すつもりでした。つまりまず構文解析についてある程度話して、つぎにそれに基づいて収集の話をしようと思っていたのです。たとえば「drink 」の目的語として何が考えられえるか。それは水でありあるいは一定の液体でありましょう。木を「drin

k」するものはいません。ですから語の収集情報は極めて重要だということには同意しますが、必ずしもこれをＴｒｉ－ｇｒａｍのようなインプリメンテーションに結び付けては考えません。

座長　Steve Levinsonさんが先で、つぎにあなたということでお願いします。

問　　私はベル研の Steve Levinson です。わたしは Fred さんの２点は自己矛盾を含んでいると思います。Ｔｒｉ－ｇｒａｍはわれわれが現在使用している文法より優秀です。また彼の予想ではわれわれが将来使うかもしれない文法よりもです。その理由は語順構文には制約がほとんどないからだと言います。制約は文法の語彙指定ルールに現われます。そしてこれは意味論に関係するのです。さてなぜＴｒｉ－ｇｒａｍが、いやこのことが効果的に示すことは、Ｔｒｉ－ｇｒａｍを使用することから得られる特別の利点は、まさにＴｒｉ－ｇｒａｍが何とか意味情報をとらえるという利点でもあるわけです。したがって、ｎｇｒａｍと同じくらい初歩的なモデルでさえ意味上の制約の多くをとらえるからと言って、われわれは意味論のためのモデルを実現することはないであろうと言うのは、いささか悲観的すぎるでしょう。私は、われわれはたぶんこのいくつかをとらえることができると思いますし、事実彼の翻訳結果はこれら推計学的モデルが、ある非常に効果的な方法で、自然言語に関し多くの事柄をとらえるという証拠にもなるわけです。意味論がやらないことは関係を明示的にしないということだけです。しかしだれがそんなことを気にするでしょう。意味論が仮に言語の意味の本質をとらえるのなら、それを使わない手はありません。私はもっと楽観的でありたく思います。

座長　フロアを先にしましょう。　お名前をどうぞ。

問　　エジンバラ大学の Pete Whitlockといいます。私は、われわれが今日、意味論が実際どれほど翻訳に必要であるのかについて、非常にいい証拠を見たと思うのです。というのは今日の非常に優秀な通訳者は、きっと彼らは理解していなかったと思われること、そしてそれは私もたえず理解できなかったことでもあるのですが、それらが何であれ、非常にすばらしい翻訳をされたことを見ればわかります。

座長　ただ、その優秀な通訳者は通常あらかじめ原稿をもらってはいますが。

はい、つぎの方。

問　　はい。私は Fred さんの指摘はなかなか的を射ていると思いますが、ただ彼の問題についての問題の一つと私の問題とは異なるとも思います。 Fred さんは、意味にのっとって行動するための意味を理解する作業に関係した視点を無視していらっしゃいます。それは対話によってコンピュータと話をする対話タスク環境におけるわれわれの問題の主要部分であるのです。コンピュータはあなたが何を言っているか理解しなければなりません。それは単なる認識とは対極的な作業です。そしてこの場合、制約を実行する一方で理解をするというプロセスを組み入れることは適当なのではないでしょうか。バックエンドが理解し得ないような文を認識しようと苦労しなければならない理由がどこにあるでしょう。そのことを認識するのにそれを制約することだってかまわないでしょう。また私は完全統合というにはその場合適切だと思います。ですからそれは単にわれわれが取り組んでいる二つの異なった問題が問題なのだと思います。それゆえに強調点が異なるのです。しかし私は推計学的モデルの観念を心から支持します。このモデルが実行されるときは構文解析との関連で行なうべきだと確信します。この構文解析で厳密にテンポラルな意味での左シブリングや右シブリング共起関係に対抗する主語・動詞共起関係といったことをやるわけです。Ｔｒｉ－ｇｒａｍは厳密にテンポラルですが、しかし仮にあなたが文構造をも組み入れることができるなら、あなたは究極的にＴｒｉ－ｇｒａｍより優れた仕事ができるだろうと思います。おそらく５年や10年といった単位ではないでしょうが、究極的にはです。

座長　はい。Fredさんお願いできますか。

答　　はい。もちろんＴｒｉ－ｇｒａｍが意味論をとらえる程度に、意味論は有用ではあります。もしそれがあなたのおっしゃったことの全部なら、われわれは、と彼は言ったのですが、われわれは理解を持たねばならないでしょう。彼の言わんとするところは、われわれはＴｒｉ－ｇｒａｍを使うべきだということですが、それならいかにして私は彼と議論できるでしょうか。

座長　さて、そろそろ時間のようですので、パネリストの方々に各自の見解の２回目の敷

衍をお願いしたいと思います。それでは Renato さんにご自身が重要だと思われる問題、とくに接続について、これまでと違った問題についてお話いただきたいと思いますが、よろしでしょうか。

Renato De Mori

ありがとうございます。私は Zue教授がお話になったことをたぶん受けた形で進めさせていただきます。そしてこの種の素性、単位といったことを認識使用とする際に私にとって関係あると思われる問題をいくつか論じてみたいと思います。コネクショニストモデルは非常によく普及してまいりました。鹿野先生がＡＴＲで行なわれたすぐれた研究についてご発表、要約なさったわけですけれども、白井先生もこのことに言及されました。 Zue教授もまたこのことについて興味ある研究をなさいました。ですからわれわれはこの種の図表に触れる機会がどんどん増えるのだろうと思います。図表と申しますのは、この方向に沿ってわれわれが時間を持てるといったもののことです。われわれは一定数のクラスや素性等々を持っており、各時間フレームに対してわれわれはある程度の根拠、擬確率あるいはコネクショニストモデルが計算した実確率も持つことになるでしょう。

これは非常に単純なニューログラムです。ことばはよくありませんが、ソノラントフリカティブな状態の節といったようなものです。詳細をお見せする時間がありませんが、とりあえず問題は、まず、コネクショニストモデルの価値は何か、です。これまでのところコネクショニストモデルは、変動する継続時間を持つパターンを扱うのが非常にうまいとは言えないようです。 Zue教授もこの問題について述べられ、これを扱うことの重要性を示されました。ですからおそらくこれらモデルはHidden Markov Modelと組み合わせる必要があるのだと思われます。これはあちこちでやられようとし、われわれもまたやろうとしていることです。Hidden Markov Modelを先験確率を計算するのに適用することおよび時間Warpingを行なうことです。またコネクショニストモデルは、たとえばHidden Markov Modelに入力を与えるのにも使えるかどうかです。そのような場合、近い将来人々が出くわす問題はおそらくこの性質のものに違いないだろうと思います。何よりもわれわれはグローバルないし局部的最適化を試みるべきです。つまりコネクショニストモデルのパラメーターと、同時に単一プロセスにおけるHidden Markov Modelのパラメーターとを推定しようというものです。それよりいいのは局部的最適化です。すなわちコネクショニストモデルはコネクショニストモデルでつなぎ、Hidden Markov ModelはHidden Markov Modelでつなぐ試みをするということです。

もう一つの疑問はわれわれはコネクショニストモデルに対する学習を指導型で行なうか非指導型で行なうのかという問題です。　たとえば仮にコネクショニストモデルをとHidden Markov Modelと関連させてつなぎたいとすれば、当然そのつなぎは非指導型のものでなければなりません。しかしわれわれは局部的最適化の方向に進もうとするかもしれません。そうなると今度はそれを指導型で行なうか非指導型で行なうかが問題になります。あと一つ疑問はアーキテクチャーです。最善のアーキテクチャーは何なのでしょう。単に文脈、音響レベルの文脈に関する情報を過剰にとらえないような単純フィードフォワードネットワークなのか、より大きな音響レベルの文を考慮できるが、いつも先験的に決まる継続時間を持つ時間遅れニューラルネットワークに行くべきか、それともボルツマンマシンスキームに由来するリカーラーネットワークを採用すべきか。このことについてもう少しはっきり言わせていただくと0、まずこれらスキームを採用する動機として考えられるものは何か、です。可能な価値として、われわれは入力ウィンドーの形状に入力を持てるということが挙げられます。このウィンドーは、たとえばスペクトルやデータを包括するものです。われわれは時間・頻度領域形状にセルをいくつか持つことができ、こうしたプロパティエキストラクターは、われわれの期待通り、われわれが抽出したいものの種類にしたがって、これらプロパティエキストラクターの各種形状と平行して作用するネットワークをいくつか持っています。そしてすべては出力時の根拠のおそらく少数の「程度」の数として要約できるのではないかと思います。

ではつぎにアーキテクチャーの問題を説明、少なくとも紹介します。時間遅れ神経ネットワークは、遅れ要素と組合わさった入力ないし隠れ層を使用しますので、利用できる情報は単に時間Ｔにおける情報だけではなく、時間Ｔ－１、Ｔ－２、Ｔ－３といった時間の情報も利用できます。ヒドゥンユニット、入力ユニット、出力ユニット上ではあらゆるレベルでフィードバックをかけることができます。私はここにこのスキームをちょっと表してみました。これは遅れ要素です。もしわれわれがこの種のフィードバックをかければ、一定数の問題が起きます。いかにしてネットワークをこれら要素とつなぐのか。問題はそれを行なうアルゴリズムですが、たとえば安定性の問題が出てきます。興味深いことはこの種のスキームを用いると、少なくとも私個人の意見ですが、われわれは良好な音声認識システムのメモリー特性を知る試みができると思うのです。こういうふうにして過去の履歴を考慮に入れなければならないのです。

さてここで一つ質問してみましょう。このようなフィードバックなし一定数の遅れ要素を持つことだけで、良好な音声認識システムのメモリー特性のこうしたプロパティをとらえることができるでしょうか。それともわれわれはフィードバックを必要とするのでしょうか。どうやら答えのない質問に終わりそうです。以上です。

座長　新しい論点をありがとうございました。ここでフロアとあなたとのやりとりをしていただくよりも、白井教授の方でもシステム構成や言語処理、音声処理に適した構造を作り出すものは何かといった点について、先生のお考えを敷衍していただけると思います。

白井　先ほどそのサーチの問題というのをちょっとお話したんですが、とにかく早くやる方法というのをいろいろ考えなきゃいけないということは、今話されたニューラルネットワークというのは、一つの並列処理ということの一つの手段として、もちろん当然使われるべきだと思うんですけども、いずれにしろこういう実験のための環境というのが必要なんではないかということです。

最近の汎用の計算機は非常に速くなってきたので、特別にこういうものをつくる必要はないという意見もあるかもしれない。ただいずれにしろ、非常に普通使われている、例えばN-CUBEとか、そういったシンプルなMassive Parallelというマシーンでは、必ずしも適切でない。すぐもとで計算することができるけれども、必ずしもそれにチューニングしたような形で手間を考えると必ずしも適切でないということです。要するにいろいろなファンクションを持った集合体といいますか、そういうストラクチャーが必要じゃないかということです。

その中でいわゆるロジックの部分と、それから先ほど私が言ったような会話のいろいろな場面といっても、もっとパターンとして取り上げるような、パターンとしての処理の部分というのが並列に動くべきだと。それじゃ、それがどういうストラクチャーなんだと言われると、今ここではっきり言うことができないけれども、そういうことが必要なんじゃないかなというふうに私は思っています。

それからさっきサーチのことをちょっとお話したんですが、私ここに、これは漫画的なんですが、今これはアガという発音なんですが、こういうGは、我々は確かに認識しにくい。これは波形がきれいですから認識できる、どういう方法でもできるかもしれませんが、こういうGが非常に認識率がここで低いという、そういう場合に、前後のアガ、非常に確実にとれれば、そういう条件下で、そういうコンテキストでGを確認するということは、非常にやさしくなるわけですね。普通そういうことでやり方が通常されていると思うんですが、そういう種類の、要するにこれは先ほど、午前中、例えば新美先生のお話で島駆動というような話ありました。ほとんどそれに近いわけですけども、普通のパーサーで、きょうLR、それから島駆動との比較なんていう話もありましたけども、これはとことんやるならば、ずっと最後までやれば、確かに計算のコンプレキシティから言えば同じかもしれません。しかし実際のこういういろんなあいまい性を持ったシステムで動く場合には、私はやっぱりこういうリライアビィリティーの高いところから計算が進むと、それでしかも確率が全部ダイナミックに変わっていくという必要があるんではないか。スタティックな確率で全部フルサーチをやるなら、それは確かにどの方法でも同じだということになるけれども、一種の加速度法が必要です。要するにニュートンラプソンみたいな方法が当然こういうサーチの方法の中でとられるべきだという気がしています。

座長　白井先生ありがとうございました。それではつぎに Zue教授に自発的発話の特性について、お考えを敷衍していただきたいと思いますが、よろしいでしょうか。何か実演が必要でしたね。

Victor W. Zue　はい。それでは自発的発話について何分かちょっとお話しし、データベースの重要性をもとに最終結論へと移りたいと思います。最近われわれはいくつかの自発的発話を収集いたしました。それはユーザーが実際にコンピュータシステムに向かって話したものです。で、われわれはこの話を記録することにして、かなりの発話を収集したわけです。およそ 5,000文にのぼります。さてデータ収集プロセスはこのようなものです。われわれは記録過程の終わりに、被験者に自分たちが作ったセンテンスを読むように頼んだのです。そうしてこれがその 5,000組の10,000文というデータベースです。各文には自発的発話版と言わばそれの読み上げ版があることになります。ここで私がやってみたいことは、皆さんにこれらセンテンスの一部を再生して、被験者があるセンテンスを自発的に作るときにすることと、それを読むときにすることとの間の違いをおわかりいただけたら、ということなのです。これらは異なるはずです。最初の例では、両者はほとんど同じです。

まず読み上げたものをお聞きいただき、つぎに自発的のものとなります。第2の例は両者の間にわずかな違いがあります。これには偽りのはじまりがあり、もう一つの方は、フ

ィルポーズがあります。これはこの人がつぎに何を言おうか考えているからです。こちらのはこれからの10年内に、できれば扱わないですませたいものです。ではこのテープを聞いてみて下さい。

（テープ再生）

I'm getting hungry.

How far away is ...

これは文法的とは思えません。

第３番目のは非常にむつかしいと思ったものでした。この女性は何を言うべきか、まったく気持ちを決めかねているのです。私は思うのですが、人々は自発的発話を見なおさなければならないのではないでしょうか。でないと音声認識システムあるいは音声理解システムを構築すると言う時、それでこの種の問題が解決できると考えるとしたら、それはわれわれが自分で自分をからかうようなことになると思うのです。ここにありますのは、われわれが行なった解析の種類の例です。これはわれわれが収集したデータの約20%に基づいています。これは読み上げられたセンテンスと自発的センテンスに対するデータベースでのポーズ継続時間のヒストグラムです。これは平均値でそちらはこの両者に対する標準偏差です。これは見えないと思いますが、われわれが得た知見の一つで、ポーズは自発的発話では読み上げられた発話よりほとんど５倍大きな頻度で現われたことを示しています。

第２に、それらの継続時間の方が長く、平均値も高く、またこの継続時間もずっと変動の大きい傾向があります。われわれは、様々な発話音のセグメント特性という点についてこれを少し解析してみました。一種の言語構造をも見ることができるでしょう。センテンスの10%は偽りのはじまり(False Start)を含んでおります。ではこのデータベースにおいてです。偽りのはじまり(False Start)について異なった種類を見ようというのであれば、偽りのはじまりの後に後が続くところにも偽りのはじまりがあります。カテゴリーは同じ、後も同じで繰り返しがあり、カテゴリーは同じ、これら２語、新しいカテゴリー、ここ、ここ、あるいは完全なバックアップ。この句はこんなふうに繰り返されています。確率が異なるところには異なったセンテンスグループがあり、音声言語システムはこの種の問題を扱う能力がなければなりません。もちろんわれわれはこの問題に取り組んだ最初の研究者というわけではなく、たとえばベル研の Don Hindle はこの問題に注目した最初の人たちの一人です。

それでは二つのことを述べて話を閉じたいと思います。第１は私にはデータベースの重要性を強調することができない、ということです。大きければ大きいほどいい。多様性が大きいほどよい。この分野では過去10年間に進歩がわずかしかなかった理由は一つには、こうした現象を研究したり、システムを学習させたり、評価したりするのに利用できるデータが膨大であるというのも大きいと思うのです。

第２に指摘したい点は皆さん方のシステムを評価する正式な方法を持つためのデータ収集もまた同じくらいに重要であるということです。今日の音声認識システムでは、われわれは、たとえば単語精度あるいは文精度を報告するのに標準化された手順を持っていると言えるでしょうが、しかし音声言語システムについて話をしようとすると、それは非常にむつかしく、性能評価は未知数です。これは研究上の問題です。仮に私が皆さんに「今何時」とたずねても、答えは一つでしょう。しかしもし「今何時か知っていますか」という聞き方をしたら、答えは「はい」でもあり、「５時25分」でもあり、「ボストンの午前３時」であったりするでしょう。われわれがこの問題をどれほどうまく解決しつつあるのかを評価するものさしを作るのは非常に困難な問題なのです。

座長　Zue 博士、ありがとうございました。最後の論点はほんとうに新しいもので、もちろんもっと詳細に論じなければなりませんが、時間の制約があるのでそれはちょっとできそうにありません。たぶんフロアの皆さんとのやりとりは可能かと思います。それからおそらく白井先生、いやいや、鹿野先生とわれわれとでお答えできるでしょう。また皆さんのお考えをＡＴＲの今後の研究目標と合わせていただいてもかまいません。

鹿野　まず今までいろいろ出てきました問題を、私なりに多少メモをつくりまして、私の考えをそれに対して少し述べたいと思います。

その後に、実際自動翻訳電話という立場で、将来どういうことをやったらいいかということを簡単に私なりに話してみようと思います。

まず言語モデル、これかなり議論されまして、大体コンセンサスはできたと思います。しかしまだ私が考えてますことは、やはり単語をレベル以下にも何かやはり言語情報は絶対あるんではないかと。そういうことをやはりきちんとやれば、オープンボキャブラリーとか、未知語の手がかりが得られるではないかということを少しＡＴＲではやってみたいということを考えてます。それで一つ書いておきます。

それからあとはDc Moriさんのキャシュの考えを使うと、これはやはりロングディペン

デンスの問題というのを統計的にどうやって扱うかという問題で、非常に難しい問題を含んでいます。ユニィフケーションを使っていくという話もそうだし、ドメイン知識も使うというのもそうかもしれません。この中にやはり統計的なものも入れなきゃいけないかもしれません。これをやっていくには、やはりまだまだ言語の意味でもやはり基礎研究が要りますし、あるいはインテグレーションの面でも、まだまだ基礎研究が要って、すぐにはできないという気がいたします。

それからあとZue先生方の音声のスポークンのアスペクトというのは、非常によくわかるんです。音声というのは、やはり書いたものじゃなくて、話してるものであるから、そういう面で扱わなきゃいけない。だけど音声認識をやっている我々は、それをいつも逃げてるわけですね。

というのは、先ほどのZue先生のデモにあるように、非常に難しい問題がたくさんあります。とても今の技術では多分できないと思います。まずそれをやるには、やはりデータベースというものをやはりそれぞれの言語においてまず集めて、それからやはりじっくりできる問題からやっていくということが必要かと思います。

特にまたダイアローグという意味でも、やはりデータベースをとらなきゃいけないと、そういうことをやはりＡＴＲでもできたら近い将来、いつかと言いませんけども、やりたいと思います。

それからこのインテグレーション、書こうと思ったんだけども、なかなか書けなくて、多分今の個々の技術、音声認識、言語処理、あるいはインテグレーションの技術というのを、多分これをまとめ上げても、システムは動かないと私は思います、残念ながら。しかしながら個々の技術を高めるとともに、どこまでできたかということをやはり世の中に知らせるということは、やはり一つの重要な役目だと思います。

そういう意味でますますこのインテグレーションというところで音声の研究者と言語の研究者が協力し合うという、こういうことをやっていきたい、やっていかなきゃいけないと思います。

このあとは、少し音声寄りに寄りますけども、少し言い忘れたことをメモ的に書いてあるんですけども、やはり認識の面でもやはりグローバルな情報、ロングディペンデンシーですかね、そういうものを扱うという技術が全く未熟です。統計的なものでなかなか扱うのは難しいんです。だけどもこちらの方へ挑戦していくということが新しい分野を開くと思います。

これは完全に僕が先ほど言うのを忘れたので、後から１回言わなきゃいけないなと思って書いたんですけども、音声認識をやる場合に、やはりロバストネスの問題というのを無視してはいけないと思います。我々のところの例えば方針ですと、音韻モデルなどはすべて単語発声の音声でトレーニングします。それで連続音声を認識しようというチャレンジいたします。実際単語データで音韻認識してみますと、95%とか、何か威勢のいい数字が出てきます。連続音声で評価しますと、83%とか、84%とか、がたっと落ちてしまうけど、そんなことを構わずに認識しているわけですね。

タスクも違うタスクで多分トレーニングしなきゃいけないし、そういう問題というのをクリアーにして、やはりどんどんアプローチしていかないと、音声認識もどこまで行ったかというのは、なかなか評価できないと思います。こういうロバストネスの問題は、話者適応をやった場合とか、あるいは環境のノイズ、ノイズの問題とか、マイクロホンの違いとかという問題も、やはりチャレンジしなきゃいけない。このあたり、ここは音響的モデルの話は少し飛ばさせていただきます。

それから最後に一つ、理解という話があって、ここで一言言わなきゃいけないと思ったんですけども、自動翻訳電話というのは、両側とにもう人間がいるんです。マンマシーンじゃないんですよね。マンツーマンでコミュニケーションですから、ある意味ではインフォメーションパッシングでもいくんじゃないかと。だから不完全情報でも、ある程度情報いけば、人間がリカバーしてくれる可能性もあると。何も訳さないよりも、何か適当に訳した方が多分情報を扱うと、完璧である必要はないという観点で、自動翻訳電話は逃げる可能性があるという、逃げてるわけじゃありませんが、そういう考え方もあると。特に音声というのは、やはり人と機械とのコミュニケーションというのを狙うんではなくて、多分人と人のコミュニケーションが音声になされているのが大部分で、そのアスペクトというのをやはり考えてアプリケーションを探さなきゃいけないという気がします。

これは非常に最後にトンチ的になりまして、余り中身はないかもしれませんけども、今我々のプロジェクトは、ちょっと７年のプロジェクトの中間にかかっています。次の３年やることはほとんど決まっていると思いますけども、その３年後どうしたらいいかと。この７年で終わるわけじゃないわけですね、いろんな研究として、先をやらなきゃいけないと。そういう場合に、どういうことを考えようということを考えたのを、それなりに考えた図なんですけども、一つは今言いましたように、これはやはりマンマシーンコミュニケーションで、ヒューマンツーヒューマンコミュニケーションなんですよね。そういう面を

よく考えてやはりシステムをつくっていこうという一つの考えです。あとやはり音声ばっかり考えてる必要はなくて、やはりマルチメディアの時代であるから、ビジュアルな情報というのも当然使うべきだということを考えなきゃいけません。

ここに書いたのは、だんだんやはりプロジェクトとしても、このＡＴＲだけで済むわけじゃありませんで、実際システム化のプロジェクトとか、あるいは研究もしなきゃいけませんし、何よりもましてやはりデータベースというのは、やはり音声もやはりテキストデータベースも要りますので、そういうこともやらなきゃいけない。まずますこういうのが広がっていって、さらに国際的研究協力というのはますます重要になってくると思います。

こういう意味ではやはり、こういうようなシンポジウムもまた開かれて、意見を交換すると、国際協力が広がるというのは非常にうれしいと思います。

以上、簡単ですけども。

座長　どうもありがとうございました。これでほとんど全体討論を閉じることになりますが、しかし主催者の許可が必要です。榊松先生、10分ほど延長して、フロアとパネリストとのやりとりをもう少し続けたいのですが、よろしいでしょうか。長い一日でしたが、まだ重要課題が出てまいりました。他に質問あるいはコメントございませんか。どうぞ、Ｗａｌｔｅさん。

問　　自発的入力に関して質問が一つあります。これは私の感触なのですが、形のよい音声入力を認識するタスクはホープレスではないかと思うのです。解析タスク、たとえば自発的入力のです、はあなたが実行することになる書かれたタスクを理解するというタスクに匹敵します。文書テキストにおいて機械に消去あるいは挿入をその順序で含むキーストロークのシーケンス全体を渡すとしましょう。するとそれは自発的発話に対してアナログ的になるでしょう。なぜ単に入力にためらいがあるか形が悪いかを認識して、ユーザーに形式を整えるように頼まないのか。たとえば書いたものを読むときのようなものにです。

座長　どなたでも、Ｚｕｃ博士、鹿野先生。これはＡＴＲの今後のプロジェクトにとって不可欠です。

答　　それではほんの２点ばかり意見を述べさせて下さい。まずもしあなたが本当にミスを犯しそうだとわかっているのなら、それは興味深い代替になります。しかし問題はそれがわかるかどうかなのです。第２に、私はわれわれがどのように音声認識問題と取り組むべきかを示唆するつもりはありません。私はただ、もしわれわれが研究室内条件の発話を扱う音声認識アルゴリズムの改良を続け、それから振り返って、第１の問題にけりをつけてから第２の問題を解決しようとしても、それは単に自らを欺くだけの結果に終わるであろうと思います。

鹿野　日本語で答えさせていただきます。

非常に難しい問題で、何と答えていいかわからないんですけども、一つは人間も必ずしもやはり、かなり自分の声をコントロールできると、その気になればタスクであれば、ある程度やはりきちんとした言葉をしゃべる可能性があるという意味で、全くの自由発声というのはなかなか難しいけれども、その前に何かやはり使えるものがあると。もちろん間違
った入力と何かの問題を扱うということもやはり重要で、その一歩として、研究としてはやはりやる必要があると思います。だからそんなにイチか、ゼロか、そういうなのができる
かできないかという話じゃなくて、やはり技術が徐々に進むので、やはりそんなに絶望的にならずに、長期的視野に立って、やはり研究というのは進めばいいと思います。

座長　それはカウンターですか。はい。それでよければ、つぎのプログラムへ移りたいと思います。あっ、そうでした。Ｈａｒｎ教授お願いします。

問　　むつかしいのはネーティブスピーカーにとってはっきりとしゃべってくれとか、書きことばのように話してくれとか言われたときの理解のしかたなんですね。それはあなたが実際解析できるように質問をタイプしてくれと彼らに頼んでも同じだと思います。実験によるとナイーブなユーザーは何がむつかしくて何がやさしいかについては非常に漠然とした考えを持っているということがわかっています。音声言語についてそうした実験があるかどうか知りませんが、書きことばを使った実験で、彼らは何がむつかしいのかわかっ

ていないときには、よりむつかしいセンテンスを作るということもわかっています。ですから何がいい発音かというナイーブな考えは、大きな声でしゃべるとかゆっくりしゃべるとか、変化があって、それでレファレンス等への困難が生じるわけです。ユーザーへのヒントよりも自信がもてるかどうかわかりません。マイクの前でどう振る舞ったらいいのかという問題は非常にむつかしいでしょうし、系統性のないむつかしさをさらに加えもします。

座長　そのむつかしさについては誰も異論はないでしょう。しかしこの問題を追及するように研究者を威嚇すべきではありません。これは恐るべき仕事ではありますが、いずれは攻撃を受けるものとなるに違いありません。そうお思いになりませんか。もちろん鹿野先生は、たとえば協力的な演者でいらしたと言うことができますし、あるいは単にある種の学習を受けたり、実際にシステムを設計したりなさることもできるので、ナイーブな話し手の側に最小限の圧力をかけることはないでしょう。それでなお正しい応答を得られるのです。ご意見はいかがですか。

鹿野　非常に難しい問題で、特にコメントはないんですけども、やはり一つは人間がどのようにやはりすれば認識してくれるかということを示すようなシステムというのもやはり今言われて気づいたんですけども、重要な観点でそういうことがわかるようなシステムを研究するということもやはりこれからは重要になるんじゃないかと、非常にためになるコメントだと思います。

座長　他にご質問、コメントはありますか。Ken Churchさん、どうぞ。

コメント　これは単なるコメントにすぎませんが、われわれは何も世界が抱える問題をすべて今この場で解決しなければならないというものでは、必ずしもないでしょう。そしてたぶんわれわれがこれ以上は手に負えないという場合があっても、たとえば Zue博士がここ10年は扱わずにすませたいと望んでいるような問題などがそうですが、われわれはまだ牛乳からクリームを分離するように、おいしいところだけを取り出すことはできるでしょう。すなわち応用の類です。J. Wilpon さんだったかがやられているような、クレジットカードの種類を見分けたりする小さなボキャブラリーとか、その他何でもやりたいことでいいじゃないですか。それでもどうにもならなければ、オペレーターのところへ行ってみてはどうですか。何かそんなことでもすれば、今の場合なんとかなるのではないでしょうか。つまり誰かが偽りのはじまりで始めたら、その人にユーモアを言ってやって、「彼はやっていない、少なくとも数パーセントやらないだろう」という希望をもつとかですね。

座長　はい。お名前を先にすみません。

コメント　Harold Somers といいます。UMIST Manchesterです。他の機械化や一般大衆が使用するもの、彼らがそれらを使うために学ぶこと、新しい行動様式を学ぶときのもの等々のコンピュータ化からもたくさんの証拠が得られると思います。われわれが、音声言語の場合ですが、われわれが、人間が理解するのと同じような仕方で音声を理解したり、人間が理解する種類の音声を理解するようなシステムを構築する努力をしなければならないというのは必ずしも正しいとは言えないと思います。人間は学ぶということができるわけです。おわかりいただけるでしょうか。

座長　あなたのご指摘はなかなか鋭いと思います。たとえば昔ですと人々は手書きしか持っていなかったが、現代はこのようなまったくつながりのない不自然な印字に完全に慣れてしまっていますし、文字という文字はすべて画一的で連絡がありません。ただ、美しい手書き文字しか見たことがない人たちが初めてこれを見れば、おそらく非常に奇妙で解読不能と写るでしょう。しかし現代のわれわれはこれを完全に受け入れている。

　他にどなたか。ご質問はありませんか。はい、それではこれがたぶん最後になるでしょう。どうぞ。

田窪　神戸大学の田窪です。さっきのフォールススタートとか、スポンテニアススピーチの話ですけども、単なる言い間違いとか、繰り返しとかではなくて、今僕がっているような「あのう」とか「そのう」とか、「えっと」とか、日本人ならよく使う言い方、それと英語だと「Well」とか「You Know」とか、「Oh」とか、ポーズフィラーみたいではあるけども、一応機能はあると、そういうふうな要素の研究とかいうのは問題にはならないんでしょうか。ＡＴＲの人にこれを言ったら、それは全部ノイズとして扱うから我々の研究領域には入らないというふうに言われたんですが。

座長　そうですね、 Zue博士はいかがですか。鹿野先生のお答えでもまだ否定的な感じですが。

答　　あなたは正しいと言っていいのではないでしょうか。これらの言語外要素、つまりフィルポーズとかこんなふうな躊躇とかですが、これらはすべて確かに情報を伝えているのです。たぶん一つのセンテンスの構文構造としてですね。ただ実を言うとわれわれにはわからないのです。われわれにはわからないという理由は、収集データが十分ではないという点にあります。したがって、こうしたことについて統計的に意味のある見解が引き出せないでいるのです。たとえばこの自発的発話のデータベースではわれわれは 5,000文を収集したのですが、仮にこれらセンテンスの10%が偽りのはじまり(False Start)を含んでいるとすると、例文はわずか 500文に減ってしまいます。あなたがこのセンテンスの言語学的構造のどこでこうしたことが起こっているかを明らかにしょうとされるのなら、もっと多くのデータを必要とするでしょう。またこれら自発的発話現象から、本当に情報が伝わってきそうだと思われるのなら、これらがどうして起こるかを理解することが大きな助けになるでしょう。

座長　座長の特権を利用して、２・３コメントをつけ加えさせて下さい。偽りのはじまり(False Start)はおそらく話し手の心の状態を示しているのだろうと思います。そして非常にすぐれた翻訳ではそれが何らかの形で翻訳されねばならないのでしょう。「躊躇」もたぶん表現されるべきですし、これら言語外要因および言語平行要因すべてもやはり伝達されるべきものです。すると問題はわれわれがいつ研究を開始するかです。そしてたぶん非常にすぐれた音声圧縮装置でもこれはなし得ないでしょう。というのはこうした発話が作り出した実際的状況および非言語学的環境もまた供給してやらないといけないからです。これについてはどう思われますか。たぶん Zue博士。

答　　はい。あなたに同感です。ただそれでも、理解と希望的観測との間には非常に重要な違いがあるとは思うのですが。また単に推測するよりは理解をしなければならないのだと思うのです。それは一見自明なようですが、時としてそうではないようなこともあります。

座長　先生、多少短くしていただけませんか。もう最後で時間がね……。

田窪（神戸大学）　まだそういうふうな研究は言語学の方でも余り進んでないんですけども、最近はそういう「Well」とか「Oh」とか、「あのう」とかというのは、ディスコースマーカーズとしてプラグマテック、ディスコースストラクチャーを研究している言語学者は非常に研究していると思います。しかし言語学者がやることですから、データが多少少ないということで、工学者の方にたくさんデータを集めて研究をしていただきたい。

座長　言語学者が協力できる、ちょうどいい分野があります。もちろん大型の音声・言語コーポラが言語学者にとって利用できるようになるにつれてですが。楽しみにしていて下さい。これは今後を展望するのに本当にいい分野です。さて、パネリスト同士、またフロアとパネリストの間でまだまだ討論、議論はおありでしょうが、残念ながら時間切れとなりました。結論や要約というより、ただ、活発なご討議やご意見の発表等をいただいたパネリストの皆さんおよび質問者の皆さんにお礼を言わせて下さい。本当にありがとうございました。これで第１日目を終わります。

Session 5

Statistical Methods of Analysis

おはようございます。

ＡＴＲ自動翻訳電話研究所の木暮です。

今日のセッション５、 Statistical Methods of Analysisのセッションを始めたいと思います。

Kenneth W. Church （ＡＴ＆Ｔベル研究所）

パーシング、単語の連想および典型的な述語と引数の関係

この論文は、それぞれ分野の違う４人の著者が書きました。私はコンピュータサイエンスを研究しているKen Churchといいます。William Galeは統計学者です。Patrick Hanksはイギリスの辞書出版会社、Collinsの辞書編集者です。Donald Hindleは言語学者です。これからコロケーションとコーパスをベースとした言語学的研究のまとめをし、ＡＣＬ　ＤＣＩの宣伝をしようと思います。

ＡＣＬは計算言語学の学会でＤＣＩはその分科会です。その分科会は希望者全員に利用できる大きなテキストを作ろうとしています。それについての冊子もあり、それはさしあげられます。私たちはなんとか 200万語を収集しテキストをつくりました。冊子は英語ですので、翻訳することができます。誰か翻訳したい人がいらっしゃればどうぞ。そのテキストには基本的に二つの面があります。ひとつは動詞の順序です。まあしかし今は説明しないでおきましょう。ここで主に述べたいことがらは、ここに記してあります。これは計算言語学において、コロケーションが重要な役割を果たさなければならないということです。それがするよりももっと大切な役割です。理想的には、長期目標ですが、制限を受けないテキストの意味分析ができるようになりたいと思います。これは当然膨大な辞書を含みますが、大きな辞書を手作りすることは、費用もかさみますし、私にはＡＴＲのようなことをする余裕もありません。ちょっと羨ましい気持ちです。そして、私たちはどうやってそれを行なうのか知らないので、それは成功しないでしょう。そして知識表現技術はたいていとるにたらない領域に応用されますが、制限されないテキストには応用されません。

さて私にはＡＴＲのようなことをする余裕が無いので、不十分なままの見方をします。それは言語学的にも、計算言語学においても古い見方です。言語学で私が言いたいのはこのようなことです。つまり、最初は非常に引用しやすい陳述であるということです。統計的な見方だと要約を使うでしょう。たとえばこれ、意味を近似する相互情報量のように簡単な要約統計学です。

さて、コロケーションが何を意味するのかここで少し述べましょう。ときどき私は同時生起という言葉を多かれ少なかれ同種の事柄に対して用います。Don Hindleが使いたがりそうな例をあげますと、つまり、目的語と動詞のあいだには、ある関係が存在し、それはなんらかの理由により、他のものよりも、その傾向が強いのです。コロンビア大学の卒業生フランク・スマジャの観察によりますと、ある種の形容詞は他の形容詞よりもある名詞と強い結びつきをもつ。それで、 strong tea(強い紅茶) は powerful tea(力強い紅茶) よりもずっとよい。 strong も powerful も同じような意味であるにもかかわらずこうなのである。

固定表現について話しましょう。こういったものは複合語です。英語において一つの単語は簡単に見分けられると言われます。御存知のように単語間にはスペースがあります。単語は文字の連続でそれはスペースにより区切られています。これが日本語や中国語になると単語は複雑な概念を呈します。しかし、思うのですが、そんなに明らかではないにしても私たちは同じ問題を抱えています。そして、固定表現があります。イディオム表現があり、心理言語学で単語連想と呼ぶものがあります。

これらはさておいて、近似するために相互情報を使う方法として、たとえば最後のことですが、 1,500万語のＡＰニュース電信を見、単語５つからなるひとつの窓、四角の窓を見、二つの単語がどれだけの頻度で、相関して同時に現われるかを調べるとします。それが、これのすることです。そして、ごらんのとおり、リストのてっぺんに来るのは本当にそれらしく思え、リストの下の方に来るものにはあてはまりそうではありません。さて、ここで電話会社で働くとします。するとなにか四角の窓のかわりにもっと興味深い窓で、何をするかといえば、これはまたDon Hindleの示唆なのですが、パーサーを働かせます。たとえば４千４百万語のキストに対するパーサーですが、これで動詞と目的語をひろいだします。そして、非標準の窓で、同じ統計量を計測します。そしてまた気がつくことですが、リストの上位はとても興味深いのです。リストの下位の方は調べもしないでしょうが、実際それらはおもしろくないでしょう。

さて、このようなテキストの見方をどのように応用したらよいでしょうか。あきらかに、認識に応用できるでしょう。音声認識への応用は、短期的には、その分野の他のものに比べて、あまりよいとは思えません。しかし、すべては多かれ少なかれ、同一のパラダイムを受け入れるでしょう。それは、未知の単語の入力系列があり、雑音の多い回線を通り、ある単語の出力系列を得るような物です。単語の出力系列は観察可能です。原理的に可能な全ての入力を探索し、言語モデル、雑音の多い回線モデルで高い確率をもった生成物を見るという標準的な方法を使って、入力の単語系列が何であったか推論することができます。その組合せを見ることにより、この与えられたシナリオに対してもっともらしい入力単語系列が発見できます。今日の、音声認識装置はたいていこの方式で動いているようです。そして、もちろん同じ議論がここでもあてあはまります。そして私たちはこの概念を使うことができます。少し違うかも知れませんがこれをn-グラムと呼びましょう。この目的のためだけのもっとも純粋なn-グラムとでもいいましょうか。また、昨日あきらかになったことですが、テキストの圧縮にも同種の議論を使うことができます。そして私たちの言っているのは、文学のなかに散在する数字をざっと調べるだけでも、最良のテキスト圧縮についての「４」の要素を得るために利用可能な技術を使うことができなければならないということです。

さて、機械翻訳ではっきりしていなければならないのは、ここで取り上げているコロケーションがあまり直訳にならないということです。たとえば、「tap a telephone （電話を盗聴する）」をフランス語に翻訳するとします。すると、私は異なる動詞、 lipcnのような動詞を使うでしょう。しかし、扉を tapする（たたく）という場合であれば、もっと tapに近い単語を使うでしょう。ですから、どのように翻訳するかはコロケーションにかかっているわけです。疑問を索引づけするときの情報語義検索とはどのように自動的にそれを、そして本のための索引を作成するかということです。または大きな資料から、どのように自動的に興味深い記事を拾い出すかということです。

そしてまた、コロケーションの考えは興味深い概念にとって役に立つでしょう。
さきに言いましたようにワードプロセシングについては、英語で単語を同定することは取るに足らないことではありません。中国語や日本語ではことにそうです。しかし、何が単語であるかを同定するのに同種の議論を使えると思います。私の同僚のリチャード・スプローグはこれについて研究していました。いま私が、そしてもちろんCollinsの同僚もですが、最も興味を持っているのは、辞書編集を短期間に行なうためにこの考え方をどのように使えるかと言うことです。

ここ数年の間教師あり学習法について多くのことを耳にします。Hidden Markov Modelであるとか、ニューラルネットであるとか、ハードアルゴリズム等があります。これらは教師ありでなければ使えないというわけではないが、たいていは教師ありで使われます。これは決してよい方向であるとは思えません。これらの統計を信じすぎてはいけないと思います。どのみち完全ではないからです。同僚のPatrick Hanksに仕事で骨を折らせたくはなく、ただ私は彼の仕事が速くなるようにしたいのです。そこで私は、教師なし学習ではなく教師あり学習のことを考えます。

さてここで、この考え方にどのようにして教師信号を導入することができるかについての戦略について少し話しましょう。いいですか、ひとつは、私たちが関心を抱いている問題について窓の大きさを調節することが可能であると言うことです。窓の大きさが違えば異なる事実に焦点があたります。他にできることは、ある問題について寛大であるため、私が述べたようなパーサーで、または音声のラベル付けのほんの一部分でコーパスを前処理するということです。また、私たちが関心を持っている問題点をきわだたせるデータだけを得るために適当な資料採集、つまりフィルタリングをするのも可能です。そして最後に、異なる質問を得るために異なる統計を使うことができます。これこそが、認識応用のために適当と信じられる統計量です。これは辞書編集における質問よりも適当です。つまり、ある単語または単語の組を本に入れるか入れないかを決定する場合、そして、たとえばこれをイディオムと呼ぶか呼ばないかを決定する場合、どの様な信念をもって、これはよい機会であるという無益な仮説を拒否できるかを問うのがよいでしょう。そうです。これは信頼できる区間についての陳述で、これは可能性比率についての陳述なのです。そして、目的が異なると統計も異なります。

さてこのスライドは窓の大きさ調節のいろんな効果を示します。ある表現はとても固定的で、他のあるものはそれほどでもないということをお目にかけたいと思います。固定された表現についていいますと、これは単語５個のなかに出て来る「bread an a butter」の間がどれだけ離れているかを見ています。私のコーパスでこれらの単語は必ず正確にまた例外なく２語の距離にあります。これは別に驚くべきことではありません。固定した表現というものは多様性が極端に乏しいものです。固定の度合の少ない表現は多様性に富んでいると言えるでしょう。

ところで５単語の中に「男」と「女」が見いだされるときには、男が女よりさきに出て

来る傾向があるようです。このコーパスは疑いなく性差別主義者なのです。しかし、かなり多様性のあることもあります。おもしろいことに数の不一致があるときには多様性に富む傾向があるのです。おそらくこれは硬直性が少なく結びつきが弱いのでしょう。さて辞書編集者ならば、固定した表現を、固定の度合の低いものから区別したくなるでしょう。それでいろんな目的のためにこの統計に興味を持つでしょう。これは異なる目的に異なる統計を使って調節した例です。

さてこんどお見せするのはこれがいいでしょう。小さい窓です。たとえば１語あるいは２語です。この場合には１語です。この場合には２語です。それに対して、これにはたぶん対数的に減衰する重みづけのなされた大きな窓またはその他のものを使った方がよいでしょう。さてこれでは異なる語彙の制約はまた異なるということを示しています。つまり、「refrain from」はコンピュータ科学者が固定されているほどにも固定されているのです。これに対して「keep from 」にはもう少し多様性があります。なぜなら間に論題となるものを入れなければならないからです。この２語が隣合うことはありません。そして、「coming from 」はどこかその中間です。つまり、それは若干任意の論題なのです。
さてこれは、関心のある何かを目立たせるための音声のラベル付けの部分を使う例を示しています。この例で私は前置詞と関連している動詞と、不定詞を示す「to」と関連する動詞とを区別したいと思います。それで私の行なったことは音声のラベル付けの部分をコーパスの上で走らせて、統計量を測定し、そしてある区別の得られることがわかりました。Abstractは「that」をとる動詞と「from」をとる動詞との区別を示しています。機会があればそれについて考察することをお奨めします。

あきらかにあなたたちは、単なる統語論的範疇ではない他の物で前処理したいでしょうが、もし私に、意味論的範疇を定めることができればあなたたちもそれで前処理したくなるでしょう。私が議論したかったのはコロケーションは階層とは別物ですし、分類法とも違います。これらの関係が把握するものはコロケーションの把握するものとは異なると思います。それでたとえば、議論の組と動詞の組とを考察してみると、このような行列を得ることができ、この行列が果して簡単に分類法で述べられるかどうかを問うことができるでしょう。とてもできそうにないと私は思います。実際これらの事柄は、太平洋はおろか、大西洋を越えて翻訳されることもありません。そしてまた手作業で階層づけする事が、非常に費用のかかることだとしても、制限されないテキストにそれが成されたことはありませんでしたし、それを自動的に行なうという試みも奨励されましたが、あまりかんばしいものではありませんでした。

私は他の認識応用について取り組んで来ました。ここにＯＣＲの問題があります。これはあなたたちが得るようなテキストの一部を示します。これは私たちが見つけた誤認識のうちのあるものの項目索引です。さあこれを考察してみましょう。私がＯＣＲを好きな理由は音声回線よりもずっとよいからです。ＯＣＲの誤認識率は 1,000字中、数文字です。ところが音声回線はあきらかにこれよりずっと劣っています。もっとも私は音声回線も、より困難な音声問題について作業するに当たり、よい方法論を開発する上で助けになると思います。

この問題を考えてみましょう。たとえば、「farm」と「form」の間で混乱があるとします。それがこの文脈あるいはこの文脈にあるとします。私ならコロケーション、いやＴｒｉ－ｇｒａｍでさえこれらをうまく区別できると主張できます。統語論はあまり役に立ちません。「これは名詞だ」と言ったところでそれほど有用ではありません。そして私は多くの場合この例が当てはまると思います。つまり、簡単なことをかたづけてしまうまえに、エントロピーを減少させるため、統語論的パーサーを使用するなということです。ここで私は少しややこしくなった同じ議論を述べようとしました。これまで何回も統語論的パーサーを使おうという試みがなされましたが、あまり成功しませんでした。制約のあるところに統語論の居場所はないので、私は主張するのです。これは私の直観ですが、制約は語彙の問題です。

私は信じるのですが、統語論は究極的に重要なものであるとは言え、それは間接的に重要なのです。統語論のそれ自身による制約というのは答えにはなりません。心理言語学者なら誰でも知っていることですが、Uni-gramとBi-gramの確率性は統語論的な要素を完全に圧倒してしまうのです。心理言語学でこれらを研究しようと言うのであれば、これらを制御しなければなりません。この議論はAbstractの中にありますし、いまはそれに立ち入る時間もありません。結局統語論を使う方法は、コロケーション要素を破壊してしまうと思います。問題はこの名詞とあの名詞がどのように結びつきやすいか、またはこの名詞はどの程度あの動詞の主語になりやすいかということです。これらは、音声系列のなかのこれらの系列をとらえるよりも、はるかに重要な要素です。

結論になりますが、私が主張しようとしてきたのは意味論を使うのを避ける方法です。それは周辺の研究です。私は意味上の分類を行なうのではなく、単語の統計的な頻度をもとに単語を分類したいのです。私は実に、古い、「古い」と言いたくはないのですが、言

語学的に、また計算言語学の古い考え方の流儀には多くの有益な点があり、それに立ち戻って考察することが必要であると思います。参考文献はいくらでも見つかります。たとえば、チョムスキーの先生は大いにこれを唱道しました。そして、これは分類法よりもはるかによい方法であると主張したいのです。制約は単なる階層ではない。これが私の言いたいことです。

質疑応答

問　　統計的な計算をする場合、ある特定の分野のテキストに限定する意図をお持ちですか。あなたの場合、分野も異なり、種類も異なる文章を、同じように処理されているようですが、それでは統計として意味を成さないのではないですか。たとえば、単語というものは、数学や電子工学では通常の使い方と非常に異なって使われるでしょう。もし、テキスト分野を、例えば数学に制限すれば、他の単語に対してコロケーションはとても明確になるでしょう。電子工学についても同じことが当てはまると思います。しかしもしどの分野のどの文章も同じような平均的な方法で扱うのであれば、統計的なコロケーション現象はなくなるでしょう。

答　　だいたいにおいてそれは妥当な指摘だと思います。私はそれを「教師あり学習」対「教師なし学習」という観点で議論したいと思います。つまり、ご指摘のように、多義性というものや、資料採集において人間の手が加わる等の多様な問題があります。それである目的のためにはいろんなジャンルからいろんな片寄らないコーパス、たとえばブラウンコーパスや、ＬＯＢコーパス等で採集したほうがよいでしょうし、他の目的のためには一ジャンルからのみ採集した方がよいでしょう。金融関係の話とスポーツ関係の話とを平均してはならないし、医学と数学とを平均してもいけません。これでコーパスはとらえられる限り、語用論的なものとなるであろう。私は供給者がラベル付けした方法によって採集しました。つまり、ＡＰはＡＰ自身の方法でラベル付けするのです。ウォールストリートジャーナルであれば他の方法を取るでしょう。私の持っているハーバー＆ローの本は著者の方法でされています。ＤＯＥの概要はキーワードによってされています。さて私は、この異なるものを全部平均しようとは、そうすることに意義が見いだせない限り、そうしようとは思いません。あなたが多分お持ちの、これらの異なる質問のそれぞれに対してラベル付けしたような資料に喜んで取り組みたいと思いますが、私たちはそれぞれ自分たちの物をもっているのです。

問　　もうひとつ質問ですが、日本、とくに九州では辞書編集と日本語についてかなり研究しています。英語の辞書についても研究しています。そこにはなんらかの関連性、分類学上の階層のようなものがあります。これらは分析の過程で内容から自動的につくられます。そこで私の聞きたいのは、もちろん完全ではありませんが、米国やヨーロッパにもその方向の研究はあるでしょうか。

答　　はい、そのAbstractの中の参考文献にあると思いますが、マーチン・ショネラとあと二人簡単な名前のロイ・バード、そしてジョージ・ハイホーンズによって、ＩＢＭの仕事について成されたものがあると思います。1983年、ＡＣＬからの参考文献だと思います。そして最近、いや私が現在調べているのですが、もうすぐ発刊される本のなかの記事のなかで、その半分はあなたが言われたことについて書かれています。ほとんどは米国の研究で一部ヨーロッパでもなされています。それが日本にもあると聞いて驚きはしませんが、日本にもそういう仕事があるとは知りませんでした。

司会　２番目の講演者はハーバード大学の久野先生、タイトルは Identification of Zero - Pronominal Reference in Japanese という題です。

Susumu Kuno（ハーバード大学）

Identification of zero-Pronominal Reference in Japanese

私の話は全く統計と関係ありません。そして謝らなければならないのですが、私の概要は、論文の結論なり焦点を表わしておりません。概要を書くのが早すぎました。

話したいことは、日本語でのゼロ代名詞の違いをどうやって認識するかという問題です。どんなパーシング問題、機械翻訳等について作業している人であろうと、このいらいらさせられる問題と取り組まなければなりません。日本語のゼロ代名詞の振舞いについては、これまで３つの提案があったと思います。そのうちの一つは吉本氏によるもので、私はヨ吉本氏のおっしゃったことを言い替えているようなものです。そしてこの場の議論には無

関係なすべての条件を省略しているようなものです。吉本氏は、ゼロ代名詞が基本的に、単語をパーシングする話題または特別な話題、つまりそれ自身についての文を導入する特別な話題により、先行詞と、接尾辞を持たねばならないＮＰとを持っていると仮定しておられます。亀山氏は「性質を共有するという制約」を提案しておられます。その制約とは、隣接のふたつの文中のふたつの代名詞は、談話中の同じ話題に関したものでなければならない、つまり主語が最初で非主語が続くという下降優位順序中の次の性質の一つを共有しなければならないというものです。これもまた亀山氏が述べたことの言い替えに過ぎません。亀山氏には別の提案もあり、ゼロ代名詞を一つだけ持つ一文で言われていることですが、もしその文が完全なＮＰ、先行詞を持つものなれば、可能な先行詞と強音との間のあらかじめ決められた優位順序は、話題が先であるというものです。吉本氏と亀山氏はふたりともこれらの原理を他の原理で補わなければならない場合を議論されています。私はこれを議論しようとは思いません。そして、この一般化が日本語におけるゼロ代名詞の非常に大切な側面を把握していることは認めなければなりませんが、この一般化には問題もあります。

ここで文（４）を調べてみましょう。「太郎が花子に何をしたのですか。」「キスをしたのです」と答えることができますが、この答で太郎と花子は脱落しています。太郎も花子も話題として目立ってはいません。ですから吉本氏の一般化ではゼロ代名詞がそこで現われるということを説明できません。そして亀山氏の「性質を共有するという制約」をＢに適用することもできません。なぜならその制約はゼロ代名詞を含む隣接の２文章だけに適用できるものだからです。ここにはゼロ代名詞を含む文章は一つしかありません。亀山氏の提案するもう一つの制約は、ゼロ代名詞が話題を先行詞とし、次に主語等を先行詞としやすいということですが、Ｂにはあてはまりません。なぜなら、始まりの文がゼロ代名詞を一つ持つときにのみこの制約の適用が可能であるとの条件があるからです。この文にはゼロ代名詞がふたつあります。私たちはこういうことを説明しなければならないようです。

文（５）を調べてみましょう。「太郎が花子に次郎を学校で紹介したのですか。」「はい、学校で紹介したのです。」同じ問題があります。この文（４Ｂ）と文（５Ｂ）は、話題としてとくに目立たない主語（直接主語でも間接主語でも）のＮＰに帰するゼロ代名詞があっても一向にかまわないことを示しているようです。しかしこの仮定も文（６）では矛盾します。文（６）は、（Ａ）「山田はコーヒーを飲む」です。（Ｂ）「田中も飲む。」日本人研究者が、（Ａ）の続きとしての（Ｂ）に対してどんな反応を示すか私にはわかりませんが、この意味が、田中もコーヒーを飲むというのであればこの文はひどいと思います。「田中も飲む」は、文（７）に見るように、ゼロ代名詞が主語の位置にある文の目的語として機能しているのか、ゼロ代名詞を文（Ｂ）中に持つ主語なのかという点に関してあいまいです。文（７）Ｂの意味は、「はい山田もコーヒーを飲む」であるはずです。

文（８Ｂ）の意味は「山田はコーヒーを飲む」です。「田中も飲む」、すみません、（７）の意味は「山田は飲む、田中も」です。実際には疑問です。文（８Ｂ）は意図された解釈ですが、それでもまずい文です。他にも一連の文があります。「太郎は花子が好きだ。」「次郎も好きだ。」ここでも状況は同じです。「次郎も好きだ」は統語論的にあいまいで、「太郎は次郎も好きだ」という意味での文（Ａ）「ゼロ代名詞－次郎も好きだ」となり、また（Ｂ）として「次郎も－ゼロ代名詞－好きだ」ともなります。しかし、２つは「次郎も好きだ」という意味で、おそらく文（Ａ）を意味するのです。田中は次郎を好きで、文（Ｂ）の解釈をするのはとても難しいのです。それでこれはゼロ代名詞が目的語を指すのは非常に難しいことを示しています。しかし、先に私はゼロ代名詞で目的語を示すのがとても簡単であると言いました。間接的であろうと言外による暗示であろうとです。以前のものです。それで８Ｂはひどいと思います。文（９Ａ）はどうでしょうか。「山田はコーヒーを毎朝飲む。」文（９Ｂ）「田中もコーヒーを毎朝飲む。」この文は私にとっては完全です。なぜ、文８Ｂはひどくて、文９Ｂは大丈夫なのでしょうか。私はこういう問題を解きたいのです。私は２通りのゼロ代名詞があると仮定しました。第一の分類は真のゼロ代名詞です。第二の分類は見せかけのゼロ代名詞です。見せかけのゼロ代名詞というのは偽りの代名詞であって、機械翻訳の問題等を解決するため仮定できるものなのです。

文（10Ｂ）は真のゼロ代名詞の一例です。「山田は酒乱だ。」「だから誰もパーティに呼ばない。」このゼロ代名詞は山田という人を指しています。山田という話題が先行詞です。私は基本的に吉本氏の、ゼロ代名詞はＮＰで特徴づけられる話題を先行詞として必要とするという一般化に賛成します。そしてまたこの要件に当てはまらない別種のゼロ代名詞もあります。そしてこの見せかけのゼロ代名詞と呼ばれる別種のゼロ代名詞については、背景となる知識が必要です。私はこの見せかけのゼロ代名詞が、談話の省略過程により生成されると提案します。

日本語には談話から回復可能な要素を省略するのにふたつの方法があります。一つは（11）のように動詞を焦点とする談話省略法（Discourse deletion Strategy)で、焦点のあ

たる情報を残し、他のものをすべて省略します。焦点が文中の動詞でなければ動詞を残し、文を省略します。一方、連結詞省略法に焦点を当てて、焦点となる情報を残し、その他のものを省略します。「だ」「です」を加えて文構造をつくります。

文（13）では同じ疑問が起こります。13Aは「太郎が花子に何をしたのですか」です。答は「太郎が花子にキスをしたのです」です。これについてはしなければならないことがたくさんあります。それでは談話省略法の一つを適用してみましょう。文（B）では焦点と動詞省略法を用いています。「キスなのです」です。「キス」が答の焦点です。そして答の文構造を保持するため、文の焦点ではありませんが、動詞「したのです」をつけ加えます。一方、「キスする」に焦点を当てることもでき、連結詞「キスです」をつけ加えます。どちらも文の答です。

同じように、「太郎が花子を次郎に学校で紹介したのですか」があります。「学校で」が質問の焦点です。「どこで太郎が次郎を花子に」または「花子を次郎に紹介したのですか」答は「はい、太郎が花子を次郎に学校で紹介したのです」です。焦点は「はい、学校で」にあります。そして文（B）に見られるようにそれを保持し、動詞をつけ加え、「はい、学校で紹介したのです」となります。または文（C）に見られるように、連結詞をつけ加え、「はい、学校です」とすることもできます。ここでの要点はこれらの答が得られるのは太郎、花子、次郎をゼロ代名詞で置き換えることによってではないということです。

ここでの提案はボード上の省略であり、代名詞化ではありません。もし、これらの答にゼロ代名詞化が含まれているというのであれば、文（15）ではゼロ代副詞と呼ばれる過程があると言わなければなりません。文（A）のような副詞化です。「誰が一生懸命フランス語を勉強していますか。」「花子は勉強しています」はまともな答だと思います。これは「花子はフランス語を勉強している」ではなくて、「花子は一生懸命フランス語を勉強している」という意味です。省略する前の構造を書きたければ、いわゆるゼロ代名詞の大きな効果だけでなく、ゼロ副詞の部分も再構築しなければなりません。そしてこの過程には終わりがありません。次のように仮定する方がずっと筋が通っています。つまり文（B）の「花子が勉強しています」は、焦点となる花子の保持と動詞の繰り返しを包含していると仮定することです。なぜなら、文の答が必要で、その他の物は、「一生懸命」や「フランス語」の代名詞化の過程を経ることなく一気に省略されたからです。

ここで制約があります。談話省略についての制約があり、それは（16）に示しています。これを私は「省略過程のつっつき順序」と呼んでいます。重要でない情報から省略し、重要な情報を最後に省略します。これは辻褄が合います。種々の刊行物で私はやり遂げてきました。

そして（17）ですが、（16）の流れは他の多くの言語と同じようにクロスランゲージの英語に当てはまります。（17）は情報の流れ原則です。強勢や形態論理学的に目立つ焦点要素を持たない文の要素は、普通重要でない情報を先に、重要な情報は後という順序で配列されています。この原則も、それぞれの言語が免れられない統語論上の制約構造を条件として多くの言語に当てはまります。（16）と（17）を一緒にすれば、日本語文の焦点位置である（18）が得られます。すみません。（17）を日本語に適用すれば、日本語で動詞の位置は文末に固定されているので、日本語文での焦点位置（18）の原則を得ることができます。日本語では文中、動詞のすぐ左の要素がその文の焦点です。動詞自身が焦点でない限り、それがいちばん重要な情報になります。

さてここでこれらの原則を用いて、ゼロ代名詞のややこしいふるまいを説明できます。「太郎が花子に次郎を学校で紹介したのですか。」「はい、学校で紹介したのです。」これはうまくいきます。「学校で紹介したのです」の「学校で」が焦点だからです。他のことはすべて重要ではないのです。この省略パターンは省略過程のつっつき順序原則に従っています。（20）の「山田はコーヒーを飲む。田中も飲む」はどうでしょうか。動詞のすぐ左の位置の省略が焦点位置です。それが最重要情報であるはずですが、それは省略されました。省略場所の左にきているものは情報の流れ原則に従うと重要ではないのです。この20Bは省略過程のつっつき順序原則に従っていません。（21）では省略場所が「飲む」のすぐ左ではないのです。Bでは「毎朝」に焦点があります。省略が本当の焦点位置で起こらずに、他のところで起こっているのでこの原則をまだ破っているのですが、それほど重大な原則破りではないため、ペナルティが起こらないのです。同様に、文（22）の「誰が一生懸命フランス語を勉強していますか」「花子が勉強しています」では、答の焦点は「花子が」にあります。それが「誰が」に対応しているからです。質問の焦点は「誰が」であり、「花子が」は答の焦点です。情報流れの原則は、WH問答があるときには適用できません。それで省略のつっつき順序原則（Pecking Order of Deletion Principle)が適用できないのです。当然の如くにそれは適用されて「花子が」が最重要情報であり、「一生懸命フランス語を」が重要でないと承認されたので、ペナルティなくその原則に従ったのです。

さてまえにも申しましたように談話の省略により生まれた一種のギャップにはまったく

代名詞がありません。しかし、これらの答の省略前の構造を再構築するため、分析を試みなければならず、ゼロ代名詞、ゼロ副詞等があるかのように思わなければなければなりません。これらは生成可能な若干偽りの要素であり、言語学的な基盤に立てば多分正当化されない要素なのです。さてどうやって見せかけのゼロ代名詞的な指示対象を同定するかが問題になります。見せかけのゼロ代名詞はもとのＮＰと同じような順序に従い、同じように機能しなければなりません。それが省略のつっつき順序原則に反すればペナルティが生じます。

（23）で最も大切なのは何でしょうか。見せかけのゼロ代名詞の先行詞としての話題となるＮＰには何の制約もありません。一方（24）は真のゼロ代名詞の指示対象および英語における「彼」「彼女」等のように真のゼロ代名詞の同定として話題となるＮＰを必要とします。日本語では音声的な理解を持たないということがただ単に生起したのです。さて話題となるある種のＮＰまたは談話に導入される指示対象にＮＰとＣがなければなりません。これについてはすぐ説明します。

（25）はゼロ代名詞に対する非焦点制約です。ゼロ代名詞はそれが現われる文において非焦点要素でなければなりません。それで話が通ります。もしゼロ代名詞が焦点であれば、それは強調されなければなりませんが、ゼロ代名詞を強調することは不可能です。文（26）をとばして文（27）に行きます。「花子が太郎をたたいた。泣きだした」です。これは形の良い談話ではありません。これの意図された解釈は「太郎が泣きだした」です。文（Ｂ）は文（Ａ）に続いていません。当然これはおかしな談話です。文（27）と文（28）を比較してみましょう。「花子が道で太郎にあった。汚い服を着ていた。」文（27）と文（28）にはきわめてはっきりとした対照があると思います。文（28）は形の良い談話であると思います。「彼」というゼロ代名詞の意図された先行詞は太郎ですが、話題として目立ってはいません。しかし、文（27Ａ）と文（28Ａ）との違いは、後者が太郎を談話に導入する点です。この違いこそ、私がゼロ代名詞指示対象の条件を述べることで描き出そうとしていることです。

手短に亀山氏が、ご自身の性質共有仮説を正当づけるため議論した４つの談話を議論します。「ローザは誰を待っているのですか。マリーを待っています。夕食に招待したのです。」これらのゼロ代名詞を、指示ギャップ位置としてとることができますが、例えば文（Ｂ）中の最初のゼロ代名詞が見せかけのゼロ代名詞であるか真のゼロ代名詞であるかはわからないのです。なぜなら、同定の過程でそのどちらをも調べなければならないからです。この方法はペナルティシステムに基づいています。もし与えられた仮説が、私の確立した原則のどれかに反するなら、その選択は間違っているのです。それでま文（た29Ｂ）に対して真のゼロ代名詞があると仮定すると、その先行詞は明らかに話題として目立つＮＰなのです。そしてそれは動詞前辞の位置にありません。ですから、真のゼロ代名詞に対する非焦点制約を冒しているわけではないのです。それで制約に反することは何もないので、ゼロまたは満点というように開始点を受け入れるのです。これには問題がありません。そしておわかりと思いますが、Oiが談話関係により、生成されたというもう一つの仮定があり、それは見せかけの代名詞であり、ゼロ代名詞を生成するのです。それが真の代名詞として、または見せかけの代名詞として同定されるかは問題でないのです。

文（29Ｃ）に行きましょう。第一仮説はO1とO2が見せかけのゼロ代名詞であるということです。ここでこれらは、「夕食に」があるため、動詞前辞の位置には現われません。省略のつっつき順序原則に対する違反はありませんので、結論としてはゼロ代名詞１がローザでゼロ代名詞２がマリーなのです。なぜなら談話継続ゼロ代名詞、見せかけのゼロ代名詞は引き金となる文に対応しなければならないからです。そしてペナルティはありません。この解析の妥当性はゼロで、「ローザはマリーを招待した」がこの文の解釈であると結論づけられます。

第二仮説に進みましょう。O1とO2は真のゼロ代名詞で、O1は目立つ話題を指し示すOiです。ですからこれにはペナルティがありません。これは話題制約にしたがっています。O2はマリーです。マリーは話題ではないので、この文は「マリーを待っているのです」を導入しますが、マリーを談話には導入しません。ですから話題制約原則と話題先行詞原則に反します。それでマイナス１のペナルティが与えられます。そして「ローザはマリーを招待した」には２つの派生があります。ひとつの派生、以前のものはゼロの完全な得点があるのでよりよいのです。

さて私たちにはもうひとつの相互参照仮定「O1がマリーを指示する」が必要です。非話題制約が冒されています。O2は目立つ話題のOiを指示します。これらを合わせてマイナス１からスタートすることができます。「マリーはローザを招待した」はマイナス１となり、拒絶されます。これもよりよい得点の派生があるからです。

細かく調べている時間がありません。まだ３つの談話がありますが、それは省略します。これらは後であなたたちが受け取る資料の中にあります。まったく同じ手順で各派生の得点づけができますし、選択された派生はゼロ代名詞指示対象の正しい解釈を表わしていま

す。そしてこれが言語学的音声解析の例であり、私たちが目下取り組んでいる技術プロジェクトにおいて正しく機能するでしょう。

質疑応答

問　20番の文章はまったく問題ないと思います。例文20はまったく大丈夫です。

答　文を読んでもらえますか。

問　「ジョンはコーヒーを飲む、ビルも飲む。」そして、2番目の文章ではコーヒーの条件付きです。

答　次のはどうですか。「太郎は花子が好きだ。次郎も好きだ。」
次の文については忘れましょう。
いいえ、これを調べたい理由は、この現象について話したかも知れませんが、まるでそれがAll or Nothingの現象であるかのように話したかも知れませんが、実際は、確率論的な蓋然性を含む現象であるととらえなければならないのです。私はそう感じます。

問　話の部分がコーヒーを飲むかティーを飲むかということであれば、例文20番はまったく問題ないと思います。あなたが給仕するとき、どちらをこしらえるか決定するためにコーヒーかティーの選択をしなければならないときに、そう3人の人がいて、コーヒーがいいのかティーがいいのか決定するために、その場合、「山田はコーヒーを飲むし、太郎も飲むから、コーヒーを入れよう」

答　しかし、このような文脈の場合、話題としてすでにコーヒーがあります。

問　それが私の言いたいことです。文番号28の下の質問について、27番と照らし合わせてなぜそれが問題ないのか考えてもらえますか。
それは「毎朝」ですか。
例番号28の紙がありますか。

答　どれですか。27と28ですか。わかりました。

問　私の質問を言います。まず次の質問です。「私」とか「あなた」等の人称代名詞を省略することについてどう思いますか。私の話したいのは、日本語では「私」または「話者」そして「聞き手」をつねに削除できるということです。中国語ではかなり制限されています。これをあなたが説明された原則で説明できるでしょうか。それとも他の原則が必要ですか。

答　「話者」そして「聞き手」はつねに談話の文脈の中にいると言えます。それらは一種本来的によく目立つ話題なのです。これが、それを取り扱う一つの方法です。

問　なぜ中国語はその場合違うのでしょうか。中国語でもこういったゼロ代名詞を許していますが、「私はあなたを愛している」と言うとき、日本語では「私」と「あなた」を省略できますが、中国語では言わなくてはなりません。

答　私はその質問に答えられるだけ中国語をこれから勉強しなければならないでしょう。しかし、あなたが質問したこれについては、私のいいたいのは28番が太郎を談話に導入するということです。これは吉本氏が議論した構造と非常によく似ています。「太郎について話をすれば。」これも太郎を談話に導入しますが、人や物を談話に導入する方法は他にたくさんあると思います。

問　ありがとうございました。

司会　どうもありがとうございました。
このセッションの最後のスピーカーは九州大学の日高先生で、タイトルはComputational Linguistics for Pattern Recognition というタイトルでお願いします。

日高　達（九州大学）

Computational Linguistics for Pattern Recognission

九州大学の日高と申します。

私の今日の話はですね、確率文法をですね、扱うときにどういうふうなことを考えてやらなきゃいけないか、特にパターン認識においてですね、確率文法を設計したり、それから確率文法へ情報を渡す、そういうときにですね、どういうことを注意しなければいかないというようなお話をしたいと思います。

音声認識または文字認識、そこではですね、２つのフェーズを持っておるわけですね。一つはそのパターン認識のフェーズと、もう一つのこの文章がその中でしゃべられているわけですから、言語　プロセッシングをやらなきゃいけない。ところがこれを評価するときにですね、パターン認識はストキャスティックスが評価基準になるわけですね。ところがこちらはその言語理論。言語理論というのは、非常にはっきり言いますとでね、その文であるか、文でないということをはかる基準なわけです。そうしますとですね、こちらの評価基準とこちらの評価基準がかなり違う、そこでどういったことがはかられているかというと、この文法というのをですね、この確率化しようとね、いうことがはかられるわけです。

まず簡単な例からいきたいと思いますが、例えば文字認識ですね、ここにおいてまず升目の中にその文字が書かれる、文字が書かれる。それでそういった場合の話をしたいと思います。

まずその線図形がかかれるわけですね。そうしますと、この線図形がどの文字であるかということをそれぞれ判定する、判定するということになるです。そこが文字認識なわけですね。ところがこの文字列、これは文がしゃべられているわけだから、これをセンテスである必要性がある。

そういうふうなことを考えますとですね、時系列が入ってきたと。そのときにですね、そのときにこれが書いた人がどういった文を書こうと思って書いたかという文字ですね、文字列。これをその推定するそのこの確率を最大にするようなＣ１からＣＮを求めること。これが問題になるわけですね。

これをですね、ベイズでこれ展開しますと、ここの式はここと同じようになります。この分母の方ではですね、ＣｉＣＮに関係ありませんから、これをマキシマムにするようなＣの列を求めるということは、その分子をマキシマムにするＣの列を求めるということになるわけですね。　しかもですね、ここのこの部分を見てみますとですね、これは例えばそのブロックスタイルで、そのつまり楷書で文字が書かれている場合はですね、これはこういうふうに独立、お互いに文字同士が独立であるいう仮定が使えます。

ほぼこういうふうに分解できる。これを見てみますとですね、書く人がＣｉというカテゴリーの文字を書こうとして、具体的にはＸｉという文字を書いたと。図形を書いたと。書く確率なわけですね、これはどういった情報かと言うと、文字認識ですね、一文字認識で使うその評価基準なわけですね。

実はですね、ここで注意しなきゃいけないのは、その一文字認識でやるのは、これとこれが逆転していましてですね、Ｘｉというある文字が書かれたときに、それを書いた人は、何を書こうとして、Ｃｉのキャラクターを書こうとして書くということになるわけですけども、ここに出てくるのは、その関係が逆転していることにちょっと注意していただきたいと思います。

しかしいずれにしてもですね、このファンクションは文字認識システムがですね、それ自体内部で持っているその函数であります。

それからもう一つ、ここですね、ここは何を示しているかというとですね、このＣｉＣＮというセンテンスが発生する確率といいますか、指摘する確率を指しているわけですね。

そうしますとですね、これは文法というのはですね、その文法にかなっているかどうかじゃなくて、このセンテンスが発生しやすいかどうかを判定する基準でないと困ると。そうあってほしい。そうすると結局この評価基準がですね、そのまま使えなくなるというわけですね。

それでこの文法なるものを確率化しようということでなってくるわけですね。ここのこれをMaximumにするということは、この文章をMaximumにするということと同じですから、したがってその評価基準は、こういうふうにこう変わるわけですね。

ここで確率文法というものが出てくるわけですけども、まず確率文法というのは一体何かと。皆さん方は恐らく多くの方は知ってはるかもしれませんけれども、例えばそのストキヤステックなＣＦＧですね。文脈自由文法、これは書き換え規則にですね、確率がついている。どういう意味の確率がといいますかと、今ノンターミルＸがですね、αｌに書き換わるときに、αｌに書き換わる確率がＰｘｌがあるというふうなことを示しているわけです。

そういうわけで、このＸからの書き換えの規則のその確率ですね、付属している確率を全部足しますと、これは１になると。それを使いますと、例えばＸからずっと書き換え規則を使いまして、ツリーができて、こういうセンテンスが発生したとしますと、その次の

発生確率をＰ１×Ｐ２×Ｐ３というふうにこう提示するわけですね。その各書き換え規則同士はですね、互いに独立であるという仮定があるわけですけども。

そうしますと例えばこれはこのツリーが発生する確率なわけですけども、このセンテンスが発生する確率はどう考えればいいかというと、このリーフのセンテンス、を生成するツリーを全部発生する確率ですね、Summationになるわけですね。これがＣ１、Ｃ２、Ｃ３というセンテンスが発生する確率ということになるわけです。

この場合はですね、このセンテンス、つまり文字列を推定することだけを考えているから、それでいいわけなんですけども、例えばその言語理解をやりたいと、例えば何ですかね、音声を聞きましてですね、それから文字理解をやりたい場合には、その文字列が出たってしようがない。つまり文法構造が出なきゃいけない。ツリーが出なきゃいけない。

そういう意味では、この書かれた文字からですね、この一番確率がMaximumになるツリーですか、ツリーを推定するということになるわけです。もちろんこのツリーのリーフはですね、ＣｌからＣＮとしますと、つまりこれをMaximumにするようなそのツリーを推定するということになるわけです。

そこで、どういうふうなバージンアルゴリズムがその確率文法に対して、バージンアルゴリズムをつくるかということになる。昨日、新美先生のところでも若干お話になりましたけれども、アイランド方式といいますか、それに今あそこではそう呼ばれたんですけども、それに近い方式でブレッドスファースト方式でやるわけです。

例えばこれをですね、母音列とします。まずセンテンスをですね、声が入ってきて、例えば今度はその音声認識の話をいたしますとですね、音声が入ってきますと、これからそれを音声を母音列に直すいう操作があるわけですね。そこで認識のフェイズで何が行われるかというと、このその列の、ＩからＪまでの列はＷを話すときにですね、Ｗという単語がＸｉ＋ｌからＸｊまでの音韻列として発生する確立がＰなわけですね。それをまず検出してもらう。

これがその音声認識の前段階のですね、パターン認識のその仕事なわけです。それが入ってきますと、つまりこういうものですね。こういう確率、Ｐというのはこういう確率です。それが入ってきますとですね、次には言語レベルでは何をやるかといいますと、これはちょっとＸと書いてますが、これはＺに直してください。これＺ。このＷという単語をですね、しゃべるときに、この音韻列のＸ１＋１からＸＫの列で発生する確率がＰであったというわけですね。

それからＺからＷが発生する確率ですが、これがそのＰ’としますと、これをこういうアイテムですね、パーズリストにこう蓄えていくわけです。パーズリストはどういうことを意味しているかといいますと、パーズリストはこういうアイテム蓄えていくわけです。ｉとｊはその生成数です。Ｘはノンターミナルでというのはある確率を示しているわけですが、このＸからＸをルートとするですね、親とするそのパーズツリーがある。最後にはそのこういう音韻列の発生するとこがいくわけですけども、それもそのツリーはいっぱいあるわけですね。これを発生するツリーはいっぱいあり得る。その中で最大の確率を持つやつをここに蓄えておくわけです。

そういうことをする、そういうアイテムを今から全部つくっていくのがこれなわけですね、例えばまずこういうアイテムから出発しましてですね、この２つから、このアイテムをつくることから出発をしまして、２つのそのパーズツリーをくっつけていきます、くっつけていきます。こういうルールがあると、これからくっつけていきます。それで前に(ikz)の、この(pikz)というのが前に蓄えてあったとすると、これ掛け、これ掛け、これですね、その大きい方をここへ幾つも蓄えていく。前のやつが小さかったら大きいやつに書き換えていくということですね。このつくったアイテムをこの確率が小さければそれは捨てて、つまりこちらを消して登録しないということをやっていますと、最終的にはですね、例えばこういったアイテムがありますとですね、これが何ですかね、０からＮまでのこの音韻列をですね、発生するセンテンスの親になるノードなわけです。その後はこのパーシングツリーはどうやってつくるかというと、ＰＬをですね、こうつくってきた過程を、まずこれから逆過程をこうたどって、ツリーをつくっていくわけですね。それアイランド方式とまあ言えるやつで、例えばこれはＣＹＫの方式でもアーリーの方式でもこれに乗せてやっていくことができます。ダイナミックプログラミングの方式なわけです。

じゃですね、次の問題は、じゃその確率文法はどうやってつくるのかと、どうやって設計するのかと。何か初めから与えられてもどうもその頼りない、信用できないということに思われるかもしれません。確かにそれはある意味では意義があることなんですけれども、じゃその一番大事なことはですね、確率的な文法がどうやって設計されて、設計された文法がどういう意味を持つのかということがですね、一番、つまり統計的にどういう意味を持つのかということが一番大事なことだと思うんですね。それはその文法というのはやっぱり集められたデータからですね、統計的にこうつくられるべきものなんですね。

それをまずそのつくり方をですね、お話したいと思うんですが、まずその文脈自由文法

があります。これを確率化しようとするのは、書き換え規則にこの書き換えの確率をつけることだということなわけですね。すべての書き換えにその書き換えの確率をつけることだ。

そこでですね、いろんなデータを集めてきまして、文を集めてきまして、人間がパーズをやりましてですね、パーズツリーをつくります。パーズ・ツリーをですね、たくさん集めて、例えばＮ個を集める。そうしましてあるツリーにですね、つまりここの中のサンプルに、その書き換えが何回あらわれるかというナンバーをこれで表すとします。Ｎ（Ｔ，×→α）であらわすことにします。つまりこのツリーの中にこれが何回あらわれるかという、その回教をですね、あらわすことにします。

そこでまずこれはですね、サンプルを集めてきたわけですから、最初イニシャルステートとしては、このサンプルツリーが発生する確率を等確率におきます。それからこの確率をですね、推定していくわけです、推定して。このやり方はですね、音声認識におけるHidden Markov Modelのやり方と同じなわけです。

そうしますと今度はこのグラマーが一応できます。この確率において一応グラマーができますと、このデータのツリーの確率は、そのグラマーよって推定されるわけですね、推定される。それはもう既にこれとは違ってきてるわけですね、違ってきている。

その新しいエスティメーションのＰｉ、つまりＰｉが発生する確率を使い分けて、さらにグラマーをつくりかえていくわけですね。この計算に乗せてつくりかえる。これをぐるぐるぐるぐる何遍も収束するまで回すわけです。収束することはですね、何でどこで保障されているかというと、これはその収束することはですね、Baumさんがこれはレギュラーグラマーについて証明しているんですけども、それは拡張できましてですね、文脈自由文法にまでそれは拡張することができます、その定理は。　何を言いたいかといいますとですね、このサンプルツリーが発生する確率はですね、どんどん文法をこう新しく回して、この確率を新たにつくっていくに連れてですね、つまりＫ回目のサイクルでつくられた文法よりもＫ＋１回のサイクルでつくられた文法の方が、このツリー全体をですね、発生される確率が大きくなるというわけです。これはしかもこの確率の総和は必ず１より小さくなりますから、これはいつか収束していくわけですね。結局収束したやつをですね、確率文法としてデザインする。

そうすると、これは何を意味しているかというとですね。このツリーを発生させる、つまりサンプルツリーを発生させる確率を極大にしているわけですね。つまりこういうやつはたくさん発生する、つまりこういうやつをたくさん発生するようなグラマーにしてくださいよという、その条件を満足しているわけです。

ここで一つ問題なのはですね、じゃそれは実際はどうかと、つまりこれをツリーをどれぐらい集まればできるんでしょうかというわけですね。

例えばですね、このグラマー数というのは、この書き換え規則の数というのはですね、これは数百だと言われていますけれども、しかしそれはこちらがノンターミナルであるときの話であって、その名詞からですね、名詞の単語を出すような書き換えというのは、結局ボキャブラリーの数だけあるわけですね、名詞の数だけある。数万あるわけです。その数万の単語が全部出てくるようなですね、このツリーを全部集めるというのは、もうそれは不可能なわけですね。ですから結局このルールは、これがターミナルシンボルである。つまりワードである場合には、新しい手を使わなきゃいけない。つまりもう少しこの数を小さくするようなですね、処置をしなきゃいけない。例えばＴｒｉ－ｇｒａｍを使いましてですね、例えば造語モデルといいますか、語を造語するモデルをですね、新しくつくって、そしてその確率を推定するようなものをやる必要があるわけです。

今日全体でお話しましたことはですね、パターン認識でですね、音声認識で、こちらに送る情報というのは、こちらで考えられる、例えば一番可能性のある単語ではないと。むしろこの関係がですね、パターン認識ではこれがこう逆転しているわけですね。ところがＣｉというキャラクターを書きたいときに、ＸＹというその図を書く、その確率ですね。例えば音声認識ではあるワードをしゃべりたいとしたときに、具体的な音韻列としてある音韻を発生する確率、それの高いものを渡すというわけですね。そこが一つ、それから確率文法というのはですね、集められたデータからつくられるわけですけども、そのツリーをつくりましてですね、正解のツリーをつくりまして、それがそれからつくり出していく、そして文字の確率文法というのは、そのツリーのその集合を発生するその確率を最大に、極大にすることなんだと。

それからもう一つは、それだけでは話はすまないで、例えば確率文法をつくるためには、ワードをですね、生成する機構をですね、新たに導入しないと、それは書き換え規則が物すごく多いわけですから、幾らデータ集めてもその追いつかないという３点のお話で、この話を締めくくりたいと思います。

質疑応答

問　これは質問というよりは簡単なコメントと受け取ってもらいたいのですが、明確でよく書かれたプログラム構造をありがとうございました。一つか二つ指摘させて下さい。尾関和彦博士、たぶんお気づきになっておられないと思いますが、尾関和彦博士は確率文法を初めて研究された方です。見事な仕事と音声認識または言語解析の路線をもっておられます。しかし、博士の行なったのよりも効率のよいコード化法により、これについてよい仕事ができます。つぎにたぶんこの形式化は非常によく、直接音声認識に応用することができるでしょう。しかし、当然認識装置または音声認識には雑音が多く信頼性の低いものです。しかし、これを若干信頼性の低い入力または文法的でない入力等に応用することは役に立つでしょう。そこでご意見を伺いたいのですが、これをどの様に拡張すれば良いでしょうか。

日高　たくさんの入力にですね、ノイズを含んでいるそれから文法にかなわないようなですね、文法に乗らないような、その文がくるときにどういう手を打つかと。それはですね、例えばある単語の発音は、こういうふうにも発音できるんだという変動をですね、音素的なシステムのところに取り入れておくと。それから挿入を取り入れておくと。ところが完全にですね、文法に、それに固有の操作をしながらですね、文法に合うように、こう近づけていくわけですね。そうしませんとですね、文法に全然合わない記号列にしますと、これ確率はゼロになってくるわけですね。ですからその完全に文法から外れた文章ですね、少々の手直しでは済まないような文章をしゃべられたときには、これはもう手が打ちようがないと思います。もっと文法の質を落としまして、もっと文法をAmbiguousにして、大きくしておくと、そういうことでしょうね、そういうことしか今はちょっと私は考えられません。

それからもう一つはそのセマンテックスを使うんだということがあります。私はこれは文法とセマンテックスとは相当違う。文法というのはやっぱり文字列または記号列のその構造である。ところがセマンテックスというのは、その上に載っている情報なわけです。ある主張なわけです。これはやっぱり僕は確率じゃないというふうに思っていますんでですね、確率では扱えるとこと、それからロジックで扱えるとこ、やっぱりいろいろ違うと思うんですね。それをミックスしてうまくやれる、完全にうまくやれる方法があるかというと、やっぱりそうはいかないだろうと思っております。よろしいでしょうか。

問　藤崎先生が言われたですね、尾関さんの研究というのは、かなり前、ちょっと私はっきり覚えてないんですけど、今日発表されたのは、音声認識では、既に確立されている話だと思うんですけども。それとですね、Doctor Jelinekが先週の国際シンポジウムで、あいまいな文法のときにですね、最適なツリーを求めるのじゃなくて、そのツリーの和の確率を求めるというアルゴリズムを発表されたんですけども、先生のお考えではMaximumなツリー一つの確率を求めるがいいのか、ある部分ではそれの方が簡単だと思うんですがね、認識率で言えばやっぱり確率の和の方を選ぶべきであるのか、どちらとお考えでしょうか。

日高（九州大学）　和というのはセンテンスの発生する確率を言われるわけですか。

問　一つのセンテンスに対していろんなツリーがありますね。

日高（九州大学）　それはそのですね、考え方次第でありましてですね、つまり認識だけでよければですね、もう文に換わらないんですね、ところが文法構造を出さないと今度はセマンテックスが出ないんですね。つまりそのサブジェクトがどこか、動詞が何と、目的語はどこかということが決まらないとそのセマンテックスが出ないですね。

そうしますとその正しいツリーを出すということが大事になるわけです。

それからたくさんの一つのセンテンスに対しましもね、冗長度がありますから、いろんなツリーが一応できるわけですね、もとのグラマーから。ところがそのほとんどが間違っているわけですね、人間が見れば。意味を考えれば。だから間違っているやつのツリーの確率は落とさなきゃいけない。それで答えの正しいツリーからですね、グラマーを獲得しなきゃいけない。そこがその問題である。よくセンテンスだけからですね、ツリーを自動的につくってやってあるんですが、それは私はいけないと思ってるんです。ただそれは認識だけをお考えになっている。ところがその後に続くのは意味解析をし、トランスレーションをしというわけですね。そのためには正しい構造を出さなきゃいけない。それがもう一つの主張であったんですけども、言うのを忘れておりました。

司会　時間もないのでコーヒーブレークの方でディスカッションお願いしたいと思います。

日高先生どうもありがとうございました。

どうも、いろいろ会場の方の不手際で、いろいろ御迷惑おかけしたことをおわびしたいと思います。

次のセッションは45分から始めたいと思いますので、よろしくお願いいたします。

Session 6

Translation of Spoken Dialogue

飯田仁（ＡＴＲ自動電話翻訳研究所）

Intention Translation Method: A Spoken Dialogue Translation System Using a Lexicon-Driven Grammar

ＡＴＲの飯田です。日本語でお話させていただきます。

Intention Translation Methodということで、ＡＴＲが目指しております自動翻訳電話の対話の翻訳部分、そこについてお話をいたします。

これは昨日も既にお話がありましたが、音声認識を行って、それから言語処理への橋渡しの総合処理を行ういうことを済ませて、その後に対話の翻訳をする、ここではNADINEと称しておりますけど、その翻訳部分の話をここでは行います。

ここでは書き言葉のそのテキストを翻訳するというのではなくて、あくまでもDialogueを翻訳の対象にします。ということで私どもは、一応トピックを限りまして、国際会議の登録参加に関する申し込み、及びそのオフィスからの説明、そういう質問応答の面について、いろいろ電話会話、及びキーボードの会話、そのようなデータを集めて、それらを分析しながら研究を進めてきております。

それでそのような分析を通して、まず従来のその機械翻訳の技術では、なかなか扱うことは難しいのではかいかと思われていた部分。一つは非常に断片的な発話が多いということと、それからお客さんとそれから事務局側との会話ということで、典型的には依頼の表現というのが非常に多いのですが、そのような自分の意図を伝えるための表現、そういうものが多用されます。さらにそういう部分では、婉曲的な表現と言われるような非常にあいまいな、シンタクティックには非常にあいまいな表現、そういうものも随分使われる。それらのところをどのように翻訳を行っていくか、そのようなことが問題になってくるわけです。

それで私どもは、今申し上げましたように非常に断片的な発話であるということから、完全なその文法が与えられて、それでアクセプトできるようなセンテンスだけではなくて、語と語のつながりで断片的なフレーズをつくっていくような発話、そういうものも扱えるようにということで、レキシカルドリブンな解析手法を採用しました。そしてそこでは、基本的にはHead Driven GrammarでＪＰＳＧ文法、そういうものに準拠した形で解析を進めるという方向で進めました。

さらにトランスファーという過程においては、従来そのシンタクティックなストラクチャーの違いを両言語間で吸収するという問題に対して、いろいろな変換規則を用意するということが必要になってくるわけですが、なるべくそのトランスファーのルールは最小限に抑えて、セマンテックスの表現を重視する。

そしてそこからPropositionalなContentと、それからIntentionalなContentということをできるだけ抽出することによって、PropositionalなContentについては、基本的には格構造というもので記述することによって、トランスファーのルールをなるべく軽減していきたいという方向で考えています。

先ほど久野先生の方からゼロ代名詞の話が出ましたが、ここでも対話の中では、発話者、それからそれらの間で、陽に「私」とか、「あなた」とかという表現が用いられることはほとんど日本語においてはなくて、このスライドに示しますように括弧で穴があいている部分、このように省略される用法が非常に多いわけです。

具体的にはどのように英語に対応しているかというと聞き手と話し手においては、Youとかｌとかは問題ないと思いますが、さらに「Could you」とかですね、このような表現も必要とされるわけです。この文末のところの依頼の表現とか、いわゆる意図に関する表現、「願えますか」、「いただけますか」というようなところに対するいろいろ英語における適切な表現、これらが翻訳において一つの課題になる。

それからそれらのものをどのように扱うかということで、先ほどお話しましたようにIntentionalなContentsと、それからPropositionalなContentsと２つに分けましたが、それらを表層的なSpeech Actのタイプということを抽出するということで、ここでは扱っております。

さらにはそのSpeech Actを確定するということは、また非常に難しい問題になるわけですが、ここでは表層的なタイプを取り出すということで処理しております。

ですからここでは「お持ちでしょうか」というような表現、ここではIndirectのリクエストがありますよと、そしてあるPropositionのContentsに対してのYes/Noの答えに対してのIndirect Requestがありますというふうに書きます。

それからPropositionとしては、登録用紙を持つということを、ここのPropositional C

ontentsとしまして、haveという動詞で記述される表現で書きます。さらには「お名前をお聞かせ願えますか」というような表現。ここでは名前を、ここではPropositionとしてはtellというのを使っておりますが、名前を聞くとか、名前を言うとかいうものをPropositionとしております。

時間が余りありませんので、詳しいお話は省略いたしますが、このような大枠の考え方でもってUnification-basedな解析を行います。

再々お話をすることになりますが、Japanese Spoken Utteranceについては、その非常に断片的であるということと、それからIntentionの表現に関しては、大変多様であるということを吸収するために、Unification-basedのアプローチをとりまして、アクティブチャートパーザーの上で、それらの文法を駆動するということを行っております。

それで文法としてはＪＰＳＧベースな文法、それからレキシコンにはシンタックス及びセマンテックの情報を記述しておきます。

さらにそれらの解析が終わった後、省略されている部分が多々あるわけで、それらを省略を補うEllipsis Analysisということを行います。この手法としましては、いろいろな制約を使った手法をとりまして、基本的には、何々したいのですがとか、相手に対しての敬意を表している表現であるとか、そういう情報を使いまして、そういうプラグマテックな情報を使いまして、ここでは話し手と聞き手の二者間での待遇表現、待遇関係が与えられてて、それらがどらちに敬意を払っているか、どちら側に敬意を払っている文末の表現になっているかというようなことを手がりにして省略の解析を行います。

そういうことを、手法を使うことによって、ここではすべてそのような省略部分が解決しているわけではありませんが、「私」「あなた」というような表現のある程度の部分、ここでは二者間の対話ということを仮定しておりますので、大方の部分の省略を穴埋めをするということが可能になっております。

その結果としてパーザーの結果としましては、ここで示しますような素性構造によるそのセマンテックな表現をとります。

ちょっと時間がありませんので、大枠の話はその辺にしまして、基本的な結果だけちょっとお見せしておきます。それでこれらは明日のオープンハウスのときにデモも行いますので、またそちらをごらんいただければ幸いです。

それで「もしもし」から始まりまして、「そちら会議事務局ですか」「はい、そうです」という場合に続きまして、青い括弧で書かれている部分、これは省略されてしまうわけです。それで特徴的なところは「たいのですが」というような婉曲的な表現、そういうものに対して「I would like to」とか、そういう表現を使うとか、それから「お持ちでしょうか」というような表現については、ここでは単に「Do you」という疑問文でありますけど、「あなたは」という省略されているものはyouして補うというようなことを行っております。

問題点としましては冠詞の問題、「the」「a」の問題、それからさらにここでは出てきておりませんけど、「名前をお聞かせください」というような場合の名前、だれの名前かというようなことをきちんとしなければいけない。your nameなのか、his nameなのか、ハーネームなのかわからないというようなのが現実でございます。

簡単でありますけど、時間がありませんので、この辺で終わりにしたいと思います。

Harold L. Somers（ＵＭＩＳＴ）

対話翻訳の新しいアプローチ

皆さん、こんにちは。まず、最初にはっきりさせておかなければならないのは、これから述べますことは、私一人の仕事というよりは、最近始まった大きなグループの仕事だということです。したがってこれからお話しますことがらは、次年度またはそれ以降にかけての私たちの事業計画とでもいった物だということです。

私たちは、マンチェスターでかなりの間、機械翻訳について活動してきました。その結果機械翻訳は、いわゆる典型的な制約不足の結果である、つまり与えられたメッセージというのは、単なる内容メッセージの外見上の内容、命題内容をはるかに超えたものに依っており、特にその前後のコンテクストや話者の意図が翻訳に影響を与え得るということを意味することに気がつきました。この情報は、ソーステキストでは必ずしも明らかにはならないので、その穴をどう補うのかという問題があります。現段階では機械翻訳を行なう人々は、これを行なうため、種々の方法を試みています。例えば、翻訳の単位を通常の文章以上に広げたり、また特に制限言語を利用して、領域特定知識をもとにした理解を導入するためのパラグラフ単位の翻訳を試みたりしています。

私たちが、試みるよう提案するつもりでいますのは、また別な筋書きで、いわば外国語の知識を持った知能的な秘書の役割を果たすようなエキスパートシステムのようなものを持つということです。この考えは、著者のいいたいことを粗いスケッチから始め、ターゲ

ットテキストを練り上げるのに必要な情報を対話的に集めようというものです。これはたいてい翻訳者を目的とする、従来の翻訳者用ワークベンチと対照的です。そういったワークベンチではシステムとユーザー両者が、ソース言語とターゲット言語どちらについても知識を持っています。つまり、そこではユーザーとシステムの間に、どちらが良く知っているのか、知識はどちらに備わっているのかということについて摩擦がしばしば生じるのです。私たちは、翻訳者用ワークベンチの計画が人間と機械の間での最良最適の労働分配なのかどうか今までかなりの間疑問に思ってきたのです。これは目新しい考えではありません。従って提案された計画では、知識の分配はたいへん明確です。つまりシステムは翻訳について知っており、ユーザーは、自分の言いたいことが分かっており、そのやりかたではユーザーは多分単一言語の知識しか持たなくて良いのです。ユーザーはターゲット言語について何も知らなくてもいいぐらいです。そしてあるテキストを翻訳するのに必要とされるすべての知識は、とにかくテキスト自体の中に存在するという暗黙の前提がある従来の機械翻訳とこれはむしろ対照をなすものなのです。

さてここで私は、対話の翻訳と対話ベースの翻訳の対比をせねばなりません。これまで申し上げましたことすべては、これら二つの活動について当てはまりますが、以下これら二つの違いを述べたいと思います。

対話ベース翻訳とは、ユーザーつまり、ソーステキストの著者との対話をベースに、ターゲットテキストが生成されるような機械翻訳だと言えます。それはユーザーが２言語対話に参加するのを助けるようなシステムである対話翻訳とまったく同じであるというわけではありません。

さて私たちのシステムはこれら二つの内の前者、つまり対話翻訳を例示しようと試みるものです。その研究についての一般的背景は、以下のようなものです。

つまりそれは皆さんが熟知しておられるＡＴＲグローバルプログラムの一環で、つまり対話翻訳についての研究なのです。それは私たちが、Manchesterで関わってきた他の多くのプロジェクト、とくにBritish Telecomeの音声翻訳への「Phrase book Approach」から考えを借りています。さて、ではPhrase book Approachでは、簡単な句を翻訳する試作システムが提供されており、それは検索可能な多くの一連の句をしまっておくことで働くようなシステムです。これは音声認識システムです。これら一連の句はどちらかというと、例語認識の粗い手法で検索されます。もちろん必ずしも音声理解の見地からではなく、言語的意図で粗雑なのです。このアプローチの利点は翻訳の出来映えがたいへん良いと言うことです。翻訳の可能な限りの完全さにおいて完全と言ってよいでしょう。不利な点は柔軟性が無いのです。探している句が句集本になければ、何も表わすことができません。ですからまあそれはどちらかというと私たちの内の何人かが今回来日するに当たって買ったかも知れないような種類の句集本なのです。

もうひとつ影響を与えたのは、約一年前に、ここ日本で出された成果です。それにつきましては、本日の午後お聞きになると思いますが、これはOn-lineのキーボード会話翻訳システムと関係しています。このシステムはＵＮＩＸのＴＡＬＫのようなオンラインキーボード会話機能を使い、そしてその機能は、ＵＮＩＸの意味でのトークの２言語トークを可能にする翻訳システム生産を通じて流れています。

この場合私たちはＡＴＲプロジェクトでのような目的指向の対話、すなわちもっと正確に言えば、一方に会議参加者、もう一方に会議主催者のいるような対話についてのみ話題をしぼっていることに注意して下さい。いま皆さんはこの場面についてよく御存知のことと確信いたします。いま扱っていますようなタイプの対話では、それで随分ちがってきます。音声対話ではなくキーボード対話を扱っていることも強調されねばなりませんが、音声言語と文章言語の違いにも関わらず、真の音声対話と文章対話の間には、いくつかの類似点があることがわかっています。

しかしこれら二つの対話が、機械翻訳システムを経るとき、いろいろな面でときどきひどい歪みが出て来るのです。対話が参加者側に伝わる前に、対話を明確にするための人間と機械の相互作用があるべきであるという考えを私たちは提案しているのです。言い換えれば、ユーザーはシステムにスイッチを入れ、自分自身の対話への寄与を明らかにするためにシステムと相互作用を持ち、それから対話は送信され、翻訳され、そして返事が返ってきて、たぶん向こう側では対話を左右し、助けるこのシステムの、逆の働きが存在するのです。

最後にこれはまだ研究の背景の話ですが、様々なタイプの明確化の対話のあることがわかりましたので、それについてもう少し詳しく説明したいと思います。いま私たちがやろうとしているのは、これらすべてのアプローチを結び、広げ、その結果ターゲットテキストを明らかにするのに、必要な情報を集めるため、ユーザーと相互作用するような一つのシステムをつくり出すことです。そのシステムには古い決まり文句、つまり句集本アプローチにあるようないくつかの翻訳済みの決まり文句もありますが、それはまたもっと柔軟性のある対話をも与えます。そのシステムは対話モデルを用いており、これについてはこ

れから少しだけ述べますが、その背後にある本質的な考えとは、それが翻訳について確実である何かへとユーザーを誘導してくれるということなのです。

20分で多くを語ることはたいへん難しいのですが、もしも覚えていただくべき考えが一つあるとするなら、これがそれです。つまりその考えでは、システムは、それが上手に翻訳できるものが何かを知っており、従ってシステムはあなたをユーザーとして翻訳可能なことを言うように後押ししようとします。あなたが本当に言いたいことはどうでもよいのです。いいですか、従って要は機械翻訳対話では伝えたいことを本当に知っているような対話参加者を私たちは持つわけです。これはテキストの原著者ですらない翻訳者がユーザーであるかも知れないような従来通りの機械翻訳のような物ではありません。

他の相互作用について何か述べるともうしましたが、私たちの扱っているようなシステムでは、四つの異なるタイプの相互作用が存在するようです。

まず一つ目はいわゆるユーザーとユーザーの客観レベル対話というものがあります。これは皆さんが今試みている対話、つまり翻訳しようと目指している最高水準の対話です。

二つ目は対話の中で何が起こっているのかをお互いに尋ね合う、ユーザーとユーザーのメタ対話です。これは対話が翻訳されている過程でよくあることです。というのは通常の対話にはよく当てはまりますが、対話の相手に最後に言ったことが何を意味しているのかを尋ねる「繰り返していただけますか」とかそんな類のものです。歪曲の起こる可能性があるため、私たちは、たぶんそこで誤訳や悪訳が起きているだろうから、その過程を明確化しようとするメタレベルの対話を、対話が機械翻訳システムを経るときに増やすことができます。それは客観レベル対話とは異なるものです。

三つ目としては、客観レベルの対話の進歩に関連したシステムから始まったのかも知れないユーザーとシステムの対話を私たちは持っています。もっと厳密に言えば、それはユーザーにより、思い当たるあいまい性をなくすことで、言い換えれば、正にユーザーが次に言いたいと思っていることを見つけだそうとすることです。

最後に私たちは何が起こっているのかを明確化したいというユーザーの要求に関して、ユーザーに依って始められたよい例であるユーザーとシステムのメタレベル対話も持っていると言えるでしょう。

さてここで、オンラインシステムについて当然起こって来る主な問題、つまりこれら四つのタイプの対話をどうやって区別するのかという問題、とくに対話が、コンピュータか他の参加者のどちらにむけられているのか、または翻訳するのかしないのかどちらを目的にしているのか、このふたつをどうやって知るのかという問題があります。そしてこの問題に刺激された、これらのシステムとの対話の場合には、さらに問題の源を知るのが難しいという問題が加わります。自分の言いたいことを相手が理解していないのが原因なのだろうか。翻訳がまずいからなのか。それともシステムがあまり役に立っていないからだろうか。このように問うことになります。

そしてこのユーザーとシステムとのあいまいさを無くすタイプの従来の対話ですら、従来のＭＴとは異なることにご注意下さい。というのは思いだしていただきたいのですが、ユーザーは単一言語しか理解しないのですから、ターゲット言語についてユーザーに質問することは不可能なのです。いまのところ私たちのとっている方法は、実にあらかじめ翻訳された断片のいくつか、つまり上手な翻訳ともうすでに分かっているような翻訳の対話部分を意味する、これはあらかじめ翻訳された断片のペアーのことですが、それを利用することなのです。これは文全体や、フレーズ全体や、あらかじめ準備されたテキストを含みますが、しかしまたその他、様々な要素のはいる余地もあるだろうと考えられます。これらは人名、数、年代などといったもので、これらが必要なことはまったくあきらかです。

下記を参照して下さいと言いましたが、もうすぐお話することはもっと込み入った話になります。たいへん急ぐようですが、このことをちゃんとお話する時間が本当にないのです。ソース言語の表現が、コンテクストに左右されるいくつかの断片のペアになるかも知れないことがわかりました。従ってここで、私たちはその翻訳関係を、単なるペア（二つの組）ではなく少なくとも次の三つからなるものと定義します。すなわち、ソース言語の表現とターゲット言語の表現、それからそれが生じるコンテクストの中で、コンテクストを表示するような何かです。たとえばある与えられたソース言語の表現は、異なるコンテクスト中では、そのようないくつかの異なる翻訳の断片ペアーになるかもしれません。（たとえば）英語で「OK」は何かを説明しているときならば、日本語の「わかりました」の意味があり、一方同意していることを意味する「いいですよ」の意味もあります。以上です。オーケー話題を変えましょう。

従って対話翻訳の任務は、まずこれら三つからなる適切な組を探し出し、次にコンテクストに従って、適切なターゲット言語の表現を探し出すことです。これらソース言語の表現は、必ずしもあらかじめ準備されたコンテクストである必要の無いことにご注意下さい。それらは、言語的に様々な質、複雑さを持つテキストテンプレートかも知れません。

従って、第一番目の任務、つまり三つからなる組を探し出す任務は、パースするような

従来通りの作業になるでしょう。従って私たちはその領域をとりまいている、そのような考えを利用することができます。そして、与えられたソース言語の表現についての異なるコンテクストがきっかけとなり、異なる明確化対話が生じてきます。

このシステムについての、二つの選択可能な場合が想像できます。一つは典型的なメニュー駆動型で、システムが主導権を維持する場面です。相互作用によってシステムが次の適切な対話の断片、言い換えれば次の適切なＤＦＲ（Description of Functional Role）を選んだり、探したり、そしてまた、一連のソース言語表現を提供します。従って言い換えると、それはメニュー駆動型だということなのです。そのシステムには対話モデルがあり、その対話のなかで人を誘導します。

もう一つの選択としては、意志伝達の目的が確立されたもっと混合性の強い主導的システムを私たちは持つかも知れません。システムが対話プランを作り上げて、その次の叙述について提案するのはユーザーなのであります。システムはその叙述が、その対話プランに当てはまるかを確かめ、そしてもし当てはまらなければ対話になんとか迫るためにユーザーに表現を修正するように依頼するかも知れません。私たちはシステムが「オーケー、わかりません。それはあなたの言いたいことではないでしょう。私の対話プランの中にありません」というような場面も考えています。

さて従来の機械翻訳の話題に入りましょう。品質については保証の限りではありません。とにかくやってみましょう。もちろん対話モデルが、複雑になることにご注意下さい。対話モデルの専門家の仕事について、私たちはとても詳しく見ているわけですが、特にメタ対話を除いて、コンピュータは対話の相手ではなく仲介者に過ぎないので、私たちはすでに自分たちの扱う種類の対話は従来の対話モデルとは異なることに気づいています。従って、対話モデルの従来の仕事に関しては一部分しか関係がないということになります。その多くは関係を持っていますが、すべてではないということです。

最後に少しだけ補足しておきますと、ソース言語表現とターゲット言語表現との間の対等な関係は与えられておらず、むしろそれはその情報内容とは別に与えられているということ、そして、従来の合成的翻訳技法は充分でないということとをもう一度強調しておきたいと思います。局所的な従来通りのやり方が、語義よりもより有益である例をいくつかあげておきたいと思います。

これらは日本で電話をかける際の作法から翻訳された例です。これらのコメントのうちどれでもよいので、ひとつを用いて英語で電話をかけたならば、受け手は「なぜこんなことを言うんだろう」と首をかしげることでしょう。私は特に（日本語の）「お電話するつもりではなかったんですが」という表現が気に入っています。それに対して、英語圏の人々なら、「ああ、では番号違いですよ」と答えるかも知れません。

さて今度こそ最後ですが、従来通りの合成的翻訳の要素で、許されるものがいくつかあると申し上げました。もうすでにそれについては述べました。さらに私たちは生成することを提案します。また生成の要素もありますし、言語要素もあります。なぜなら、そのシステムはメッセージの確認を生成し、そして多分言われたことを明確化するためそのメッセージを再びパラフレーズする事が必要になって来るからです。

ここに対話の実物模型のようなものがあります。今日の午後にもう一度お話する機会がありますので、その際にお見せしようかと思います。私たちのやっていることを見るのに興味をもたれたかも知れません。もしよろしければ午後お見せします。

質疑応答

問　富士通のローケン・キムです。機械翻訳についてはまったく知識が無いのですが、もしもおっしゃったシステムやＡＴＲシステムについて私が正しく理解していますなら、ユーザーはターゲット言語についてまったく何の知識も持ち合わせていないと仮定されるわけですね。もしもある程度知識があったらどうなるのでしょうか。それから、もしも知識があれば、システムの計画体系は変わるわけですか。それともう一つ質問があります。もしあなたの意図とシステムの意図が一致しない場合、どうなるのですかどのようにしてそれを解決するのですか。

答　まず最初に申し上げなければならないのは、ＡＴＲプログラムの一部であると言いましたときに、それはＡＴＲが行なっていることのコンテクストの中にあるという意味だということです。ご承知の通り、私たちはものに応じて、少しづつ違うアプローチを実際は用いています。従って、もしその考えがお好きであれば、駆動力はＡＴＲのものと同じです。

次はもしもユーザーが日本語をいくらか知っていたらどうなるかということでした。たぶんそういう人たちがこのシステムを使うというようなことはないでしょう。しかし、もしそういう人がこのシステムを使うならば、一つの利点は、システムの犯す誤りをよく

理解し、つまり誤りがわかり、その種類もわかり、誤りの源がよくわかるので、誤りについて忍耐強くなるということです。しかし、そういう場合でも私には別の計画体系は考えられません。

三つ目は何でした。ああそうですね。もしもユーザーが対話モデルからそれていって、でもシステムがライトを点滅させながら、オーケーの通知をしたら、でしたね。それはもう普通の翻訳をしているに過ぎなくなります。結果は保証できません。ひとつ素晴らしいスローガンでもって私の提案を締めくくりたいのですがよろしいでしょうか。それはつまり、これは他人の考えから借りてきたものだということです。この種のシステムは従来のＭＴが、何が不可能なのかを示すのと違って何が可能かを示してくれるのです。

Bernald Lang（ＩＮＲＩＡ）

形の整わない入力処理に関する生成的見地

私の話しますことの題は今の通りです。この題を選びましたが、言いたいことをいくらか発展させましたもので、これが本当に適切な題かはちょっとわかりません。別の言葉に言い換えてみましょう。私の本来の仕事は、一般的な自然言語処理と言うよりむしろパーシングなのです。ですから、このトピックについて何も知らない人の素朴な見地について話してみましょう。

私が考察してきた問題に、lattice parsing、形の整わない入力、構文上のあいまい性を無くすこと、そしてもうひとつは、この会議の間、ずっと見てきました統計上の技法のいくつかを使って処理されるのですが、そういう妙な構文である、いわゆる構造上のズレなどがあります。

本来はこれらの問題は、すべて次のふたつのこととつながりを持っています。一つは極めて統語論的な意味での翻訳のパーシングと認識についての様々な技法、もう一つはむしろ確率論的アプローチにもとづいた最適化技法です。

まずパーシングの問題に入りましょう。私の見てきたこの問題のほとんどは、有限状態オートマトン、いや有限状態の変換器というかまたはプッシュダウンマシン、これらの組合せが、本質的な問題なのです。単に文脈自由プッシュダウンオートマトンだけでなく、私の言っているものすべては他の言語学的な形式主義すべてにうまく拡張されるのです。それはまたある種のプッシュダウン装置により認識されます。私がこれらのことについて話しているのは何故なのか少し動機づけするために、これまでやってきたことを振り返ってみます。

私は文脈自由パーシングについて調べてきました。そして、ここにＳという文章があり、この言語の文法からパーサーを得ます。興味深いのは、この文章に対して可能なすべての、またはほとんどすべてのパースを与えるような、チャートパーサーといったものです。文章があいまいなときには、普通一つのパースではなくパースの森を得るのです。そしてその森で充分なシェアリングと、そのシェアリングの正しいコード化をしようとすれば、本質的に発見されることはパースの森が文法により提供されるが、その文法はとても奇妙な文法であるということです。その文法では一つの文章しか作れません。その文章は作業を始めた当の文章なのです。その文法の中には、このＧという言語文法によりこの文章に与えられた構造が保存されているのです。ですから、本質的にはチャートパーシングにより、与えられたある文章に、ある文法を特殊化できるのです。これがチャートパーシングの行なうことなのです。

さて皆さんもチャートパーシングが、パーシング単語ラティスに一般化されることを御存知ですので、お聞きしますが、単語ラティスとは一体なんでしょうか。単語ラティスは実際作れそうな文章の集まりのように見えます。まあ本当の文章もあり、パースできない文章もあるのですが。それらは言語には属しません。しかし文章の集合があり、そのうちのどれがこの言語に属し、その構造は何であるかを調べようとして、チャートパーサーを応用しますと、すべてに対する文法を得ることができます。文章はこのラティスの中にあり、文法的に正しいのです。この得られた新しい文法は、もともとの文法の構造とまったく同じ構造をその文章に与えます。

さて私の仕事は、形の悪いインプットの問題を考察することでした。文章のある部分が失われたときにはいったい何が起こるでしょうか。御存知のように電話の会話ではこれがしょっちゅう起きます。そしてここに非常に簡単な文章があります。「ジョンは？？？メリーを見る。」このなかに何があるのでしょうか。わかるわけがありません。数え切れないほど当てはまるからです。この疑問符は未知の言葉を示しています。もし失われた部分が長ければもちろん好きなだけの言葉が当てはまります。まあこれは限られていますが。

さてこれをまたチャートパーサーでパースできます。そして同じ計画または手順が応用できます。そしてパースの森が得られて、たいていそれは無限になるでしょう。というの

は、それが無限言語を与えるであろうと言っているのです。無限のパースですが、それは何もかも、この文章を構成できるすべての方法を告げるでしょう。

さてこれらすべてを一般化してみましょう。本質的に言ってチャートパーシングで、できることは有限状態装置をパースすることであり、チャートパーサーの使用でパースの森が得られるということです。もし有限状態装置で生成された言語が、ええ、これだとして、パーサーがLという言語のGという文法によってつくられたものであるなら、パースの森は新しいG′という文法になり、それは本質的にはGという文法ですが、K̄およびLの共通部分のために特殊化されているのです。しかしその構造を持ち続けており、それがまた良い点なのです。でなければ結果がおもしろくないものになるでしょう。これについては御存知のことと思います。ここにはなにも奥深いことはありません。そして、その共通部分に対する新しい文法を得られるという事実はここ30年ほどの間に知られております。しかし、私たちの記録アルゴリズムはよいもので、チャートパーシングを行い、構造も保持します。

さてこれのもう一面は最適化問題です。最適化のために私たちは、普通いろんな方法であらゆるところに確率性を、オートマトンの変換であるとか、文法規則であるとか、パースされているストリングだとかに当てはめます。これは特殊な問題例であり、うまく解決されています。ここで私が話しているのは重みのことですが、これは、確率性として自由な重みであります。これらの事柄について一般的な理論があり、それは級数に基づいていますが、本質的には、一般的なアプローチがあり、そこではRという領域に属している重み値をつけることができ、その領域にはここの加法と乗法のふたつの演算を含む代数特性があるのです。さて、この特性があるとき、値とすべての統語要素との関連によってパースされている文と値とを関連づけられます。確率性という見地から、行なうことは規則、変換、入力その他に確率性を与えることです。そして得られるのはあるパースツリーが正しいパースであるという確率です。

しかしながら、おそらく他の物も計算されるでしょう。私たちの目的にとって興味深い他の領域があるかどうかを調べることはおもしろいと思います。この形式理論でおもしろいことはここにお望みの代数特性があり、その代数特性を手にすれば使用している古典的な演算がすべて役に立つということを、形式理論が告げるということです。またここで行なわれるような計算を行なえば、たとえば、あるストリングに関連する値を計算すれば、ある入力ストリングは正確に、良く知られていて好まれている動的プログラム技術であるということです。

さてここで、言語認識あるいは分析についての私の素人的な見方に戻りましょう。音声から始めます。音声波が来て、音響処理装置がありそこで流れが得られ、その流れは有限状態装置または有限状態装置の集合にたどりつきます。そしてまた別の流れが生じます。こういったことが続いて、ついにはあるプッシュダウンパーサーに行き着きます。この流れは、まあ実際にはいつも流れが得られるとは限らず、ときに、流れというよりは有限状態の構造を生じることもあると考えることもできます。これは本質的にラティスアプローチです。ここで起きる疑問は、そういう順序があるならどうしてこのようにいろいろな装置を構成できるのか、どうあいまい性を処理できるのか、どうやって簡単に適切なデータを得られるのかということです。それを取り扱えるのはいったいどの方法でしょうか。文献はこれまで使われてきた種々の方法でいっぱいです。そして様々な実験も含まれています。文献を読んでいていらいらするのは、その実験を相互に関係づけることが非常に難しいということです。いくつか論文を読んでいて感じたことですが、同じ内容ながら少しだけ設定を変えものを読んでいるような気がしました。それで、いったいこれはどういうことなのかを見通す全体的な見方、そして何時ある戦略を捨て、ある戦略をとるのかということについてより理論的な分析を持ちたいのです。

一つか二つ例を上げましょう。もしラティスパーシングのことを話すのであれば、入力ストリングがあり、それは可能性をいくつか与える非決定論的な装置を通ります。それは可能性すべてを含むこともあれば、確率性にのっとってよい得点をあげるものかも知れず、また入力ストリングと分析を指し示す重みづけの方法かも知れません。ですからここで、得られるものはラティスであり、チャートパーサーでそれをパースしたいのです。これは私が前に言ったことです。得られるものは入力された可能性のあるすべてのものに対して当てはまる文法であり、またその文法を、妥当に興味深いパースについて妥当に興味深い文、こういうものに対して取り除くこともできます。実際はこの問題を異なる方法で扱ってきた人もいます。彼らは一つの入力ストリングを得て、それから入力ストリングを有限状態の変換器と結びつける替わりに、チャートパーサーと結びつけて、チャートパーシング過程の中で、暗黙的にラティスを作るのです。たとえばある論文の中で読んだことですが、それはこういったことを行い、ラティスがそこにあるにも関わらず、決して表に出て来ることはないのです。どれが最良の技術なのか私にはわからないのですが、なぜある技術よりも、他のある技術を選択しなければならないのか知りたいのです。

ここに形の悪い入力の問題があり、これは同種のもうひとつの問題なのです。雑音の多い回線があり、雑音の多いストリングが得られます。そのなかでは構成部分のあるものは不正確ですので、ある不確定性が導入されるでしょう。それで与えられたものとは異なる選択が可能かも知れません。普通この不確定性は、混同行列により導かれますが、どんな有限状態装置によっても導かれ得るのです。例えば、失われたサブシーケンスはたいてい有限状態装置で行なわれます。そして、エラーがあります。実際何種類ものエラーがあります。音声送信が不良だったり、初期処理段階でのエラーであったり、文生成不良であったりさえします。透過性については触れないでおきましょう。本質的にこれらの問題は重みづけされた可能性についてのあいまいさの問題であります。そして、これらに対して、もっと正確に重みづけされたオートマトンの構成で対処したいのです。これがこのWの表わすことなのです。

非常によい形式システムがあります。構成の連想性、構成による有限状態変換器の終結、有限状態変換による、正規言語の終結。そしてまた有限状態装置のついた構成によるプッシュダウン装置の終結という事実に依存します。私がプッシュダウン装置というとき、文脈自由装置のみを指しているわけではありません。私の意味しているのはこれが例えば、Definite Clause Grammarに応用される、またはラベル付けシステムに応用できる、または計算言語学的文学に使われる種々の文法に応用できるということです。

事実、おかしな意見があり、多分もう知れ渡っていると思いますが、それはまず、言語学で使用される形式主義に容易に拡張されるということです。しかし、同時に重みや確率性が特徴として文法の中にみられます。そしてたとえば、「ビームサーチ」が、おそらく公正なアジェンダを追加する代わりに、アジェンダを制御する一つの方法として見られます。これは時々、「ビームサーチ」が不公正であるためには役立つ方法である場合に、すべての解決が求められているとき、そして正しい答を得られるような確率性を持っているものに対しては、より正確であろうとするときに行なわれることです。

ここで私の問は、行なわれていることを何時私は見るのかということです。私はそれを正しく、一般的な枠組みの中で設定されているのを見たいのです。そこで私は問題や解答を比較することができるのです。一般的な枠組みというものは存在するでしょうか。何が抽象的な問題で、何が解決でしょうか。

実用的な解決または理論的な解決を意味しているのです。何が具体的な問題で、それはどのように抽象的な問題に表れるのでしょう。具体的な問題のほとんどは、抽象的な見地からみると同一であると私は予想するでしょう。もしそれが明らかになれば消耗しないで済む精力がかなり増えると思います。ありがとうございました。

質疑応答

問　人が望むことのひとつは、これが古い技術の新たな組合せであるように、問題の一分類の再解釈から現われると思います。この仕事から現われる明らかな例がありますか。

答　現在はありません。私は正直に言いますが、このすべてにまだ慣れていないのです。現在は何が進行しているのか理解しようとしているところです。おそらくもっと議論して、技術を良く理解することが必要でしょう。しかしおかしなことは、ビームサーチを取り上げている論文を全部読んだのですが、そのひとつとしてビームサーチが何であるかを言わないのです。ビームサーチを参照に付すものは一つとしてないのです。それでビームサーチが何であるか考えなければなりませんでした。

私の推測は間違っていました。私にとってビームサーチは、これを付加するというよりは正しい答を与える光線だったのです。可能な解答の光線を加えてみましょう。そんなに良くないもの、良く見えないものがあるかも知れませんがほとんどは正確です。それで「パースをあまり調べないようにしよう」と言うかわりに、計算を省力化できるので、「もっとパースを保持しよう」を挿入します。それで最後にはいくつか解決があり、その中から他の原則を選択できます。これはおかしなことです。なぜなら、これを行なう人がいるのは確かなのに、それに言及する人が誰もいないのですから。しかし思ったのですが、私のビームサーチが何であるかについての誤解はおかしいもので、実際は私の理解も意味をなしたと思います。これが答になるかどうかわかりませんが。

富田　勝（カーネギーメロン大学）

ＣＭＵにおける音声翻訳に向けての研究

どちらの言語で使おうか随分迷ったんですけども、ここは日本ですので日本語でやらさせていただきます。

余り時間の方もありませんので、簡単にＣＭＵでどんなプロジェクトが行われているか

ということを御紹介させていただきたいと思います。

ＣＭＵで言語に関する研究というのは、主にコンピューターサイエンス、それからCenter for Machine Translation、それからLaboratory for Conputational Linguistics、この３つのユニットで行われています。そしてComputer Scienceでは主に音声認識に重点が置かれていまして、CMTでは、機械翻訳はもとよりプラティカルな自然言語処理というものに重点が置かれています。

もう少し理論的なComputational Linguisticsは、このこれはPhilosophy Departmentの一部となっているんですけども、ここで行われています。あともっとがしがしの理論の言語学科はですね、ＣＭＵには言語学科がありませんので、すぐ隣のピッツバーグ大学に言語学科がありますので、そこと協力したりなんかしています。

これが、こんなものをお見せしてもしようがないかもしれませんけども、Facaltyのリストになっています。そして御覧のとおり、一人のファカルティが幾つのユニットをかけ持ちするという状態が当たり前になっいますが、この星印はプライマリーデパートメント、それからこのゼロはアジャンクトデパートメントでして、プライマリーデパートメントの定義はですね、サラリーが出ているデパートメントであると。そして何人か、例えばCarbonellとか私のように、これ見えないかもしれませんけども、星印２つ付いているんですね。これはComputer ScienceとCMTの両方から給料が出ていると。これはもちろん倍給料もらっているというわけではなくて、半分ずつ出ているにすぎないわけです。

この中でも自然言語と言っても広いですけども、音声認識とか、機械翻訳に限りますと、これぐらいになって、それでがしがしの音声から機械翻訳、それから音声と自然言語の融合あたりのいろいろな興味がやられるているということになります。

実際のプロジェクトなんですけども、今日は最初の２つを少し詳しく説明させていただきますが、１つはSPHINXとＣＭＴ、CMTの機械翻訳システムと融合して、音声翻訳をやろうという動きと、それからこれは今日、あるいは昨日随分話題になっています確率的の文法を使うということです。これは対話の翻訳をダイレクトメモリアクセスという一つのパラダイムで行おうというもの、それからコネクショニスト、ニューラルネットワークはAlex Waibelなどが音声認識に応用しているというプロジェクトはここにありますが、最近それを実際の言語解析ですね、つまりイルフォームドな音声言語の解析にニューラネットワークを使おうというとんでもないプロジェクトも始まっていると。

それからこれは松下技研の音声認識装置を利用して、日本語から英語のデモを行ったと。それからこれはMINDプロジェクトで、これはWayne Wardが昨日説明したとおりと。あともちろんたくさん音声に関するプロジェクトや、それから翻訳に関するプロジェクます。

最初にSPHINXＬＲのプロジェクトなんですけども、これ簡単に書けばこうなるということなんですけども、まず英語の音声をSPHINXがごく簡単に言えば、それを英語の文字列に変換にする。その後ＣＭＴで開発したパーサーとジェネレーターがありまして、これはもう３年以上いろいろなプロジェクトで使っていますので、かなり安定で、バクが少ないということですが、これで解析を行って意味表現をつくり、その後日本語の文字列を生成して、最後に音声合成器で日本語の音声を出力すると。もちろんこの翻訳の部分、いろいろな問題あるんですけども、その当面の音声翻訳というものにSPHINXとパーサーとの融合とか、この辺に問題があると思われます。

問題はですね、もともとはSPHINXはＢｉ－ｇｒａｍという比較的単純な文法を持ってサーチを行っているわけですが、実際の解析の文法は、これはオーギュメンテド文脈自由文法によって書かれています。

ですからSPHINXが持っている文法とこちらで持っている文法は全く違う文法になっています。そうすると困るのはSPHINXが出力した英語の文字列というものが、必ずしもここのパーサーで受け入れられない。必ずしもというかほとんど受け入れることはない。SPHINXは95%の文字認識率というのを誇っていまして、これ一見非常に高度な数ではありますかが、95%といいますと10単語の文に対して見ますと、大抵１つぐらいどっかが間違っていると。そして１つでも間違っていれば普通に解析できないわけですね。ですからこれ非常に困るということで、じゃここで解析で使っている文法と全く同じ文法、あるいはほとんど同じ文法でSPHINXのサーチをコンストレイントすれば、少なくともここに出てきたストリングは合っているかどうかはともかく、ここのパーサーでそのまま解析することができるということで、それじゃSPHINXに文脈自由文法を使えるようにしようというのがこのプロジェクトのモチベーションであります。

そして昨日も発表がありましたようにSPHINXがＨＭＭのHidden Markov Modelのシステムですので、ＨＭＭＬＲというものを導入してＡＴＲの北さんとか、それからWayne Wardなどの協力を得てインプリメントし、現在それのテストをしているということになります。

もうすぐこのSPHINXの改良は終わるんですけども、今テストしている段階なんですけども、これが完成すればここの、先ほども言いましたようにこのパーサーとジェネレーションの部分はかなりステーブルなシステムなので、文法開発やなんかは全部オフラインでタ

イプ入力によって全部デバックして、そのドメインに合わせた文法などを全部オフラインでグラマーを書くことができる。そしてテストを行った後に、このパーサーは文法のデバッカーとかも付いていますので、文法などを全部デバックして、よし、ここでいいとなったらば、その完成した文法、及びレキシコンをSPHINXの方にばーんとたれ流してやって、SPHINXはその要するにその文法に従って認識を行うというふうにもう完全にこのSPHINXを機械翻訳システムのフロントエンドというふうに考えております。

計画ですが、今年中にぜひトイグラマーによってテストを終えたい、また多少問題がありましてうまく動かないんですが、今年中に動かすようにしたい。そして来年及び再来年は当分の間、ドメインはOfficial Airline Guideという翻訳するには、ちょっと物足りない、余りおもしろくないドメインなんですけども、当分の間音声グループがこのドメインをメインのタスクしてやりますので、一緒に協力して当分の間はつき合うと、これで。それがうまくいきましたならば、今度はもう少しチャレンジングなタスクであるところの会議登録をやってみたいというふうに考えております。

次に昨日、今日と話題になっております確率文法の解析ですが、ぜひこれも使いたい。つまり今のところSPHINXＬＲのプロジェクトでは、確率的ではなくて、普通の文脈自由文法を使っていますが、ぜひとも確率的な文法を使いたいと。理由としていろいろあるんですが、一つはいろいろな通常でない入力をアクセプトするために、いろんなルールをどんどん足していくと、やはりルールの数が増えてしまうんですけども、どうしてもやっぱりいつも当然使うありふれたルールと、それからめったに使わないけども、とりあえず入れとかないとまずいルールと、差をちゃんとつけないとルールの数だけ大きくなってしまってどうにもならなくなってしまう。ですから本当のモティベーションは精度を上げるというよりは、何かめったに使わないルールは非常に低いスコアをついておいて、本当にめったに使わないようにするというのがモティベーションになっています。

これが確率的文脈自由文法、随分今回議論されましたが、要するにルールの一つ一つにある数値がついていると、この数値は手で書いてもいいし、自動的に学習してもよろしいと。このようなものが与えられたときに、これをＬＲのパーシングで解析するときにですね、一番ナイーブな方法としましては、普通にＬＲのジェネラライズドＬＲの解析を行って、ルールを使ってリデュースアクションをするときに、ここについている確率の数を掛けてやるというのが一番単純な方法ですが、それはもうちょっといい方法があるというので、もうちょっといい方法というのは、これらの確率を最初にテーブルをつくるときにプリコンパイルしておきまして、単にリデュースアクションのときに確率を書き込むんではなくて、この一つ一つのシフトアクションのところに、もう既にプリコンピュートされて確率がへばりついていると。したがってシフトするたびにそのすぐに確率を使うことができる。これによってプルーニングというのか、サーチのカットをするときには、なるべく早くカットすることができるということになります。

この方法はできるだろうというめどが立っていますので、計画といたましては今年、今言った確率付きのＬＲパーシングテーブルというものをフィージビリティーはＯＫということですので、来年ぐらいにそのインプリメントというのは、本当のインプリメント、本当に使えるぐらいしっかりしたシステムをつくる。現在おもちゃのシステムはあるんですけども、本当に使えるようなシステムをつくって将来はSPHINXＬＲなんかでも、SPHINXの中でその確率も読みながら認識を行うというようにしたいと思っています。

余り時間もありませんので、この辺で終わらせていただます。

質疑応答

問　あなたは正しいと思います。この会議を通してずっと、制約が通常ノンターミナルを意味する、確率的な文脈自由文法についていろいろ聞いてきました。ところで、昨日の午後論争があり、私はFred Jelinekに譲歩しました。私はＴｒｉ－ｇｒａｍモデルがそれを随分支持したと思いました。なぜなら、それはノンターミナルに考えられていたのと同じようにターミナルへの制約を表わしていたからです。それは私も論文で強調していたことでした。これはコメントのようなものですがこれについて何かありますか。

答　ちょっと考えさせて下さい。昨日のパネルでは見逃しました。しかし、Ken、あなたには音声認識のバックグラウンドがあります。ですからあなたの主な目的はあなたのシステムの単語精度を向上させることでしょう。それは正しいですか。

問　その点は認めます。

答　はい、私にとって単語精度は問題ではありません。一個の単語が間違っていれば、あと何個間違った単語があろうと気にしません。いいですか。私が音声認識システムを使

用したいのは翻訳や理解など有用なことをするためです。私の見るところ、Ｔｒｉ－ｇｒａｍのような単純な文法を使って単語精度を向上させるようなことは私にとってあまり意味の無いことです。もちろんあなたが、書取り機械を作っているのなら、これは非常に重要でしょう。もしあなたが入力をパースしなければならない、つまり理解しなければならないのなら、どっちみち統語論を用いなければならないこともあるでしょう。そしてどうにか統語論的な文法を構築しなければならないでしょう。それを音声認識の中で用いなければならないことはないのですが、いつかどこかであなたはある種の統語論的な文法をもたなければならないでしょう。その文法があるなら、単純なＴｒｉ－ｇｒａｍ文法よりそんなに有用ではないとしてもその文法を文法制約のために使うべきでしょう。どう思われますか。

問　　あなたがそこで最後に使った言い回しはちょっとひどいと思います。それは「そんなに有用ではないとしても」です。あなたが使うべき物は有用なものであって、無用なものを使うべきではないと思います。

問　　規則のプロバビリティーをNodeに入れておいて、シフトするときに使うとおっしゃったんですけども、何かそういったノーマライズというんですかね、要するにDepthが深ければ深いほど、確率値だんだん欠けていくから低くなるわけでしょう。プロバビリティーのノーマライゼーションの問題についてどうやったらいいですか。

富田（ＣＭＵ）　ああ、欠けていくとどんどん小さくなってしまうということですか。普通はログを取ってやって、足していけば計算量も楽ですし、普通はログ。

問　　そういう問題じゃなくて、一つのストリング、今まで分析したとこですね、そのツリーの構造によって、ツリーのデップスが深ければ深いほど、確率値がだんだんだんだん加わっていくから小さくなると、そういった問題についてどういふうに扱ったらいいですか。

富田（ＣＭＵ）　しかし一つの文に対していろいろな解析があって、それを較べるわけですから、解析している途中にカットはしていくんですけども。

問　　そのときに要するにDepthが深ければ深いほど、確率値、数が深くなるからだんだんだんだん数が小さくなっちゃうわけですね。

富田（ＣＭＵ）　ああ、そういう意味ですか。左から右にBreadth firstにずーと行けば、もちろん多少は深さの不公平ありますけども、ある程度あんまりスコアが広くなったらカットするというようなヒューリステックスを入れておけば、厳密に言えば問題あるかもしれませんが、実際にはそんなに問題あるとは思いません。

Session 7

Understanding of Spoken Dialogue

Christian Boitet（ＧＥＴＡ）

音声合成および対話ベースの機械翻訳

ご紹介ありがとうございます。まずお招きいただいたことを、ＡＴＲの皆様に感謝いたします。ここで話題にしますものは、実は会議の話題ではないことを強調したいと思います。なぜならただいま私たちは、対話や話言葉の翻訳を取り扱っているのではないからです。しかし、もちろん音声合成は、話言葉の翻訳にとても大切です。そしてこれまでそれが強調されたことはありませんでした。みなさんが音声認識のことを話してこられ、それはとても難しいのですが、音声合成も、今後における話言葉の機械翻訳ばかりにでなく、他のタイプの機械翻訳にとっても重要なのではないかということを強調したいと思います。また、音声合成技術が、対話の翻訳、対話をベースにした機械翻訳にとても役立つことでしょう。そこにおいて人はかならずしも対話の翻訳が必要ではなく、コミュニケーションしようとする者、本人との対話により、機械が助けられるのです。そして、最後に強調すべきこととして、私の考えでは、ある種の明確な対話のためには音声生成の質向上が有益かつ必要であると思います。

そこで、いろいろな形の機械翻訳システム、音声合成がとても重要であるようなシステムを検討してみましょう。60年代には、見張り人やスパイのために、蓄積しようと機械翻訳が、始まりました。ここでは利用者は最終利用者で、単言語使用者です。普通、この利用者は国防省の司令官です。その後、この技術を、専門家つまりリライターのための機械翻訳に応用でき、翻訳者を機械翻訳に置き換られると考えられるようになりました。そして、リライターの方法が知らされます。それから80年代に入って、この取り組み方は、制限された形の物だけに有効であるということがわかりました。例えば、天気情報とか取り扱い説明書などです。その後、人々は翻訳者のワークベンチに興味を持つようになりました。そしてまたここでは翻訳者つまり利用者は２か国語に通じ、それ以上に翻訳と両国語の専門家であります。

しかし、ここで私たちは、大量の同種の文書から異種の入力を取り扱います。ここに違いが生じます。そして、いま90年代では、記述のあるいは口頭のコミュニケーションを行なう人のための翻訳の機が熱していると思います。そして、著述者のための機械翻訳と、ＡＴＲが行なっているような、話者のための機械翻訳とに分けたいと思います。ここでは明らかに対話の翻訳が考えられています。そして、もうひとつ別のタイプつまり作家のための翻訳があります。

そしてここでまた、システムの中でまたシステムにより使用される知識に応じてわけられる、ＭＴシステムの別の分野があります。これにはおもに２つの取り組み方があります。つまり知識ベースの機械翻訳と対話ベースの機械翻訳です。これをざっと見てみましょう。翻訳するためにはたいてい３種の知識が必要です。ひとつは言語的な知識で、これは中心的な言語知識とコーパスについての知識、つまり、文書作成環境についての知識とから成り立っています。しかし、言語的な環境というものは、世界のモデルの意味論的、実際的、または状況的な解釈が関わってきます。残るのは実際の、コミュニケーションの、意図的な語用論です。

しかし、実際のシステムの中で、この区分けがあるわけではありません。システムで実施されている知識源により区分けして、言語学的なデータベース、知識ベースのシステムに分けることができるでしょうが、もしあなたが天気情報のような制限された世界にいるならば、意味論的な領域、もしくは語用論的な領域とかちあったり、そういう領域をひきついだりするでしょう。ですからここで、あなたのシステム構造にもとづいて、全体を包括するヒューリスティックなデータベースを所有しているのか、ある領域の知識ベースを所有しているのか、領域およびタスク関連の知識ベースなのか、人的知識ベースなのかを区別しなければなりません。もしあなたが、人的知識ベースを多用するのであれば、それを私は、知識ベースの機械翻訳に対するものとして、データベースの機械翻訳と呼ぼうと思います。

知識ベースの機械翻訳では、実施されコード化された、タスクと領域の知識ベースを目いっぱい利用するという試みがなされます。そして機械翻訳では音声合成が使えます。もちろん、テキストが関係する限り、翻訳は合成されるでしょう。それは明白です。入力と、ここでは技術的な手段、ハイパーメディアやハイパーテキストになる文書作成を行ないたい場合の　　その一部は話されるべきなのですが　　そういった手段に使えるのです。ですから、もちろんあなたはそれを何か国語かで、おそらく話言葉の形で、生成しなければなりません。そして、話言葉の入力も生成しなければなりません。

そして個人的な機械翻訳に適用されるであろう逆翻訳（Reverse translation）。それは著者が端末として席に着き、機械はタスクや領域について何の知識もなく、なぜなら、私たちはそれをみんなに対して欲しいのですから、それは対話に多くを頼らなくてはなりません。個人が一つの言語しか知らない限り、機械の行なうことを制御するための唯一の方法はある種の逆翻訳を行なうことです。対話が関係する限り、明確な語句再表現対話を、システムが取り扱えるものへと利用者を導くものとして、修正対話を、これの正確な解釈とシステムとのコミュニケーションにシステムを導くものとして区別します。そして、音声の質はここでは、修正対話のためには決定的なもので、グレイン、ここでは箱のことですが、それのためにとても重要で、その残りのものに対してはあまり重要でないと思います。音声の質については、もちろん音声の成分から合成が起こるのであり、それについての研究もあるのですが、わたしはそれを固定されたものとみなし、言語学的な分析、語用論的な分析により集中しようと思います。

そこで例として私たちがグルノーブルで個人的な機械翻訳に実施しようとしているある試作品を取り上げます。私たちの意図するシステムをざっと検討してみましょう。まずそれは試作品であり、個人的なＭＴの範囲に入ります。これは非常に言語学的であり、対話に依存し、事実とか規則というような明確な知識ベースのものではありません。そして、私たちの母国語であるフランス語から他の言語へ翻訳したいのですが、もちろん他の国にも応用できます。

それで、入力がハイパーテキストになるようにします。そしてここにはいろいろ長所があります。たやすく利用でき、利用者に使いやすく、テキストの単位が小さいのでタイプ入力でき、システムが相互作用的であるなどです。単なるワープロ以上のものであり、ユーザーはもっとアクセスするように想定されています。たとえば、対話は従来のテキストプロセッサよりも容易です。そのテキスト単位においては印刷が簡単であり、これが複雑なテキスト書式化システムに対する利点になります。そして私たちは、分散的構成を作家のワークステーションとして、マイクロコンピュータで使い、大がかりな計算を、すでに機械翻訳システムや言語関係商品開発の基礎のあるミニコンピュータで行おうと考えています。

また、自由な言語のアプローチに対するものとして、誘導された言語のアプローチを使おうと考えています。制御された言語のアプローチが意味するのは、テキストの各要素がとても小さければ、それをマイクロランゲージに、大きければ制限言語に関連づけようということです。そして、私たちはそれを、少なくとも文法的な観点からは非常に精密に、マイクロランゲージは、文法的に規則、制限文法、制限言語の集まりのようにも見える構成の集まりであると定義されると理解しています。そして、マイクロランゲージからとられた小部分をともなうある表現であると。ここで私は、わかりやすい通常の表現と、文脈自由な力強い型の表現を取り上げました。

ここで、あなたが、タイトルから始めることを考えましょう。そして、説明文のリストがあり、それから、たとえば単なる著者名のようなものがあります。そして、さきに述べた、多くの単位と複数レベルを転送することにより、言語学的な技術を使用し、増大させたいと考えます。そのために、ある言語の種類、つまりシステムが取り扱う言語について、計算され、定義される独立の言葉の意味をつけ加えます。

ここで、この対話部分にやってきました。これは対話構造のため、現在私たちが設計しているものです。非常に複雑なものであるとわかります。この左側にはマイクロコンピュータで私たちが何をしようとしているかがわかります。そしてこちらでは、ミニコンピュータで何をしようとしているかが示され、ここでは以前の仕事を削減することができます。ここで、ひとつの目的はある対話がもちろん著者の制御のもとでなされることです。しかし、実時間でマイクロコンピュータにおいてはあるものが多大な言語学的処理の後で出てきます。私たちは言語学者、著者に２、３秒の間でも待って欲しくありません。それが一つの文章に１２秒かかったとしても使われないでしょう。

ですから、ある型の非同期的な対話もしくは処理を行い、対話のある部分については後回しにしたいのです。そしてここで、最初の部分はたとえばマイクロランゲージであるとか、サブランゲージであるとかの位相幾何学的な型を、単位の各種類、カードの各分野、書庫つまりたとえばハイパーカードでの文書庫に選択することなのです。その上で著者は自由に編集すべきなのです。そして、彼の制御のもとにスペル間違いの訂正であるとか、未知の単語を検知するとか、かなり、地域的である用語の標準化であるとかに関して対話を始めるのです。ある時には航空機と呼び、またあるときには飛行機と呼ぶといった具合いです。そしてあなたが慎重でない言葉を使うと、システムは、置き換えたいかどうかをあなたに尋ねるのです。そして私は様式を整えました。

しかし、ここではとても低レベルで考える様式を意味しています。それは文の長さであるとか、ある種の言語学的な要素の繰り返し、例えば頭字語であるとかです。そして英語において、変化した様式で、決まり文句のように見いだされる、固定された言い方やイデ

ィオムの使用に関する対話の一部を、持ちたいと思います。文書のこの部分、変化した様式はあきらかに固有名詞です。これを一般的な言語学のプロセッサで分析すべきではありません。それは馬鹿げたことです。これがこのように出て来るときは固定した言い方であることを著者に確認するほうがずっといいのです。もちろんもしあなたが、説明的な文章で、様式を変えたければ変えてよいと言うのであれば、利用者は確認すべきではありません。利用者はノーというべきです。そのような場合にはイディオムではなくて、言語学的プロセッサがそれを通常の言語連鎖として分析するでしょう。

それから私たちは利用者自身に単語の意味について尋ねるでしょうし、表層分析を行なうまでは、またその後でも実際何がそれを行なうための最善の戦略であるか私たちは知りません。ですから、それを実験してみたいのです。たとえば、英語でcustoms という言葉がでてきたとします。するとシステムは習慣の customsか空港の税関の customsを意味するのかを問い返して来るでしょう。もしテキストの量が少なければこれはあきらかではありませんし、著者が何を意図しているのかもわかりませんし、統計的な方法も全然役に立ちません。なぜなら、あなたが欲しいのは 100%正しい答えですが、統計的な方法では例えば、そうですね、35%ぐらいの確率なのです。25%間違うシステムなら不必要でしょう。利用者に質問することは無料で、バランスや単語の種類やスピーチの部分について多くの指示を得ることができるでょう。もしあなたがそれを以前にするとしてもです。もしあなたがそれを以後にするのであれば、もちろん表層分析成分により行なわなければなりません。

もしここでミニコンピュータシステムがテキストを分析できなければ、それは利用者に返って来るべきで、「それは制限言語のなかにないので、例を下さい。制限言語を変更したいですか。テキストを変更したいですか。」と、言うべきです。もしあなたがそれを変更したければ、システムは他の時間に分析しようとするでしょう。もしあなたがテキストを変更したければ、変更してやり直すでしょう。もちろん私は例をあげて前置詞の付加であるとか、表層分析後のそのレベルでのあいまい性をなくすことだとかを多くできるようにします。ここに表層で私の意味しているのは、そこから最初のひとつづきの言葉を容易に引き出せる「木」を得ることができるということです。

情報は注釈の形で付加できます。しかし深い情報は、ここで正式な注釈として付加できます。深い分析であなたは抽象木を得ることでしょう。それからの処理は、自動的に翻訳へ進められるでしょう。あるいはおそらく生成したいと思っているパラフレーズをあなたが決定した後で、戻り、著者のため、なされたことを制御する目的で逆翻訳を行なうでしょう。私は例をあげ、あなたとともに音声合成が有用で必要か、またはよりよい結果を与える所を調べてみましょう。

もちろん通信回線に関する対話や、システムとの改良対話についても調べましょう。その対話は「ゆっくりしゃべって」とか「これは命令文ではない」など、システムによって発せられる固定された決まり文句です。そうです、それは録音された決まり文句です。本当の質のよい音声合成の必要はないのです。単純に録音すればよいのです。しかし、明確化が問題となる限り、多様性に富むものになります。こういう方式の音声合成を使うべきなのです。

そしてもちろん言語学的なデータベースから用意された質問というものがあります。たとえば「customs 」です。それは習慣でしょうか。または政府の機関でしょうか。それとも内側、外側から見た壁でしょうか。それは抽象的な障害なので、イタリア語の異なる言葉、 muro parateで翻訳します。それから言い直します。明確にするために言い直しするのです。コンピュータの制御室だとかコンピュータ制御の部屋というように同じテキストを提示します。ここではパラフレージングは極めて明白です。テンプレートまたは完全なパラフレージング、そして話題に関係するものに対する発問が、焦点を合わせます。例えば「ボールの写真」です。ボールが写真に写っているのか、ボールがその写真をとったのか尋ねたくなるでしょう。

言い直しにとって、たとえば音声合成がいいということはあまり明白ではありません。ここに例があります。どこでしたか、ここです。これは明確化対話を記述形式にしたものです。ご覧になりたいでしょう。そしてここに口頭形式の物があり、息継ぎを意味するこれがついています。この場合記述形式のほうが少し明確だと思います。「I saw the (girl in the park). 私は『公園にいる少女』を見た。」「(I saw the girl) in the park in Denters.私が少女を見たのはデンターズの公園だった。」という具合いです。

そしてもう一方の例では、この記述例では違いがそれほどはっきりしないのです。その違いというのは、「The reaction produces (carbon and nitrogen) ditroxide.この反応で『炭素および窒素の』ダイトロキサイドが生じる」と「The reaction produces carbon and (nitrogen ditroxide).この反応で炭素および『窒素ダイトロキサイド』が生じる」との違いです。この記述形式はここでは少し混乱させるものになっています。強勢についてはどちらも同じだと思います。「あなたはこの部屋を予約しましたか。」「あなたはこ

の部屋を予約しましたか。」「あなたはこの部屋を予約しましたか。」たとえば強勢は対話において非常に大切だと思います。

そして完全にパラフレーズすると「この反応で炭素テトラキサイドおよび窒素テトラキサイドが生じる。」または「この反応で窒素テトラキサイドは生じるが、炭素テトラキサイドは生じない。」となるでしょう。

確かに口頭形式の方が簡単です。違いはおわかりになると思います。読んで、すぐ違いを見つけることはとても難しいことです。下線をひいて強勢をつけたとしても、それほど明白ではありません。もちろん「窒素テトラキサイドと炭素」と言うべきではありません。なぜなら、炭素および窒素炭素と理解し得たからです。そして同じことが、よく知られている「the girl in the park with a telescope 望遠鏡を持った（〜のある）公園にいる少女」の文についてもあてはまります。

完全なパラフレーズについては、充分注意を払わなければなりません。なぜなら、異なる解釈でも同じ表層形式を与える場合があるからです。「ボールの写真」「ボールの写真」ボールが写真を撮るのか、ボールが写真に撮られるのかということです。ですから完全なパラフレーズが、いつもよいとは限らないのです。

テンプレートによる解消がおもにPete Whitelockおよび彼の同僚によって唱道されました。このスライドが示すように、選択や相互依存がわかりますので、私はその記述形式またはグラフィック形式がよいと思います。もうすぐ話は終わりです。

発問については音声の強勢が明確であれば、記述形式よりもずっとよいと思います。「あなたはこの部屋を予約しましたか。」これは誰か他の人がこの部屋を予約したことを意味しているのか、私が他の部屋を予約したことを意味しているのか、私はどの部屋も予約しなかったことを意味しているのかわからないのです。ですから、読んだ方がずっといいのです。

そして私の結論は、最終的な質と経済性、入力の最終的な質および翻訳からみて、高品質の合成はもちろんのこと、経済性と同一性が関係する限りにおいては、きわめて高品質の音声合成が必要になるのです。そしてもちろん音声合成の必要性を仮定しますが、このスライドでは、入力や翻訳、そして対話ベースの機械翻訳での対話に音声合成が必要になるということに確信のもてる理由を与えることはできないと思います。ですから、話言葉の対話の翻訳にはきわめて高品質の音声合成が必須なのですが、それはすべての種類のMT、とくに対話ベースの機械翻訳の対話成分に貢献すると思います。

質疑応答

問　　きわめて高品質の合成を私たちが持っているかどうか知りませんが、ベル研究所の私の所属部門は音声合成について研究しております。そこでがっかりさせられることは、私たちの合成について何をすればよいのかが、明白でないことです。ここであなたの主張にもの申したいのですが、それでごちゃごちゃしたスクリーンを消すことができると思います。私の不満は帯域幅がせまいことです。それは古い電信タイプにたとえられると思います。ガチャン、ガチャン、ガチャンと動く古いタイプライターです。1秒間に何文字、何バイト打つのか正確なことは知りませんが、たぶん12ぐらいで、実に非効率的なのです。

答　　よい音声合成をうまく応用できる例は非常に少ないというのが私の言いたいことです。これはそのうちのひとつでしょうし、もっとよいものを開発する誘因となるでしょう。

あなたが正しければいいのですが。

そして、それはもちろん話言葉の対話の翻訳には必要なのです。

Walther V. Hahn （ハンブルグ大学）

ＡＩにおける自然言語対話研究のパラダイム

ここで話をするように招待を受けたとき、私は難しい立場にありました。なぜなら私はふたつの椅子の間にいたからです。それは学問関係の椅子ではなく、プロジェクトの椅子でした。去年私たちは「LOQUI」というプロジェクトを完了しました。それは「ESPRIT」プロジェクトで、自然言語によるデータベースへのアクセスに関するものでした。キーボードによるアクセスだけで、これは私がそこから立ち上がる椅子であり、そして座りたい椅子なのです。来年はＡＳＬというプロジェクトがあるからです。ＡＳＬとは「音声と言語モデルの構造」の略です。LOQUIプロジェクトの結果にはあなたたちが知っているものもあるでしょうから、そういうことを繰り返して退屈させたくありません。

ＡＳＬプロジェクトからはまだ実質的な成果はなにもありません。このプロジェクトが始まったのは初夏だったからです。最初のプロジェクトから私たちが学んだ方法、そしてなぜ今音声言語に進んだかを発表しましょう。これは構造問題のレベルについての言語プロジェクトの過程で発見された欠点をまとめたものです。そしてこれがもっと深い構造問題および音声言語環境での音声言語から始めた理由です。統語論や他の問題、または幾分低い言語学的レベル等については触れないでおきます。対話レベルまたは談話レベルで問題になったものについて話します。

LOQUIプロジェクトのなかの対話ストラテジの作業で行なったことは何かと言いますと、さっき言いましたように Prolog のデータベースアクセスシステムプログラムでした。その特徴のひとつは、英語の文でもドイツ語の文でもタイプ入力でき、英語とドイツ語の答から選択できることです。核心となるシステムは言語のペアと同じです。

私たちはシステムの中で、対話や談話、あるいは談話の全体的なパラメータに結びつきのあるすべての特徴を制御するひとつの権威（authority）を持とうとしました。私たちはそのような基準要素を探しました。普通行なわれるように私たちは最初それを対話マネジャーと呼びました。そしてそのような対話マネジャーの周辺のすべての議論を立証しようとしました。他の物をすべて支配するパラメータを見つけることが望みでした。私たちは他の人がそういったプロジェクトでやると思われることをしました。よく知られているパラメータ、言語学的な議論、照応的な疑問、直接再起単語、さらに形態素的な面すべてを調べました。後に70年代終わりの文献のなかで、生成についての議論が見つかりました。調べなかったのは音声の議論です。これは私たちがキーボードから始めたからですが、原則的にもっと進んだ議論があります。それについては主要なパラメータ発見のため調べるべきでした。意味論については多くの作業を行ないました。それから、それが対話の首尾一貫性の原則であり得たので、意味論的な結合の問題から始めて、話題と焦点のようなもの、あるプロジェクトで私たちが調べた参照問題、スピーチアクトから導かれたものとしては扱わなかったプランニング意図というように調べていきました。スピーチアクトはちょっと違っていました。それについては後で触れます。それから、首尾一貫性、パラグラフの特性、パラグラフに対するテキスト設計議論がありました。実際、音声にたとえられるもの、同等なものがあると思います。例えば、ラジオを聴くとすると、スピーカーから流れて来るのが、ドラマかディスクジョッキーかニュースであるかはすぐにわかります。これの意味するところは、形を埋めることのできる、なにか韻律的な手段があるということです。そしてそれらは独自の形式を持っています。それは、テキスト形式ばかりでなくすぐに認識できる韻律的な形式も持っているのです。

私が「物理的な議論」と呼んでいる有名な議論があります。それは物理的な議論でなければならないということです。話者と聞き手が同じで、同じ時間に同じ場所にいる限り、これは確定されるものなのです。私たちの内の一人が、実はそれはトム・バスターなのですが、彼がやったことはまず、再構成することでした。というか、まずスピーチアクトの問題から始めましょう。私たちはスピーチアクトをストラテジや意図と一緒にすぐに導入するということはしませんでした。別のパラメータとして導入したのです。トム・バスターはまずスピーチアクトのより詳細な分類法を探そうとしました。そこにはスピーチアクトの周辺の他のすべての首尾一貫したパラメータを分類するよい方法を得られると思ったのです。この詳細な分類法には長所があり、私たちは主題の領域にずっと近かったのです。システムの主題領域とはプロジェクト管理でした。そしてLOQUIシステムはこのデータベースへのアクセスシステムであるべきでした。ですから、トム・バスターはスピーチアクトの下位分類を見つけようとしたのです。それはこのデータベースへのユーザー用必要情報ステップに一致しました。この長所はユーザーのやり方に容易についていくことができたことです。

私は欠点があることも認めます。なぜなら、領域を変える場合にはこのスピーチアクトの階層も変えることを考えなければならないからです。不必要なのに領域を変化させたり、例えば、標準的なスピーチアクトやスピーチアクトと包括的なプランとの混合物を使用し

たりすれば、これはシステムのまた別な、変えなければならない部分になったのです。

スピーチアクト周辺の対話の首尾一貫性を分類しようとするバットラーの試みのつぎのステップは、進行中の対話と下位部分を見つけることを可能にするブックキーピング表記法を発見することでした。彼はやり取りに依存して、やり取りと動きとを定義づけようとしました。そこではある動きはやり取りを開始させ、ある動きは終了させるのです。そしてブックキーピングはユーザーを考慮します。それは物理的な接触を意味します。主題の基礎があり、それは総体的な主題であり、この特殊な動きの内容であります。

さてあとになって見つけたのですが、私たちは、いかにある言語学的領域、他の首尾一貫性パラメータを支配する語用論的領域を、どうするかのもっともらしい一つの原則を見つけることができませんでした。私たちはなにか違うものを見つけました。私たちは意味論、語用論、プラン議論についてある程度知識を得ましたが、このパラメータの相互関係についてはほとんど知ることができませんでした。たとえば私たちは対象や概念についてのある世界知識が意図と相関していること、話題と焦点も意図と関係していることを知りました。これはジョシと彼のグループ、またその他の人によって示されました。この結合は言語学的な情報の中でも、世界知識等の中でも見られます。しかし私たちの発見したのは、これが首尾一貫した対話についてのパラメータの異なる分野間の関係として確立されるもののほんの小さな一部分であるということでした。そして私たちはこれらの間で線を描くことを追究できる多くの分野を発見しました。

もう一つの例は今朝Ken Churchが提示したものです。たとえばおわかりでしょうが、単語が一緒に見いだされるテキストの中で、統計的な手法のみにより何とか示され得るプランなのです。電話の事例を覚えておられると思います。当てはまりそうなプランを探している場合には単語分類に戻ることもできます。私はそれを想像できますが、私の知る限り誰も正しいストラテジを見つけることに関係したこの情報を利用しませんでした。それで、これは見込みがあると思います。これが今までのところ私たちの知っていることです。これはある原則を見つけるために利用できる何かのほんの小さな一部分に過ぎないことを私たちは知っています。しかし、一方、他のすべてを制御するひとつの議論を得るための、この単一なものに原因を帰そうとする試みを私たちは疑っています。

このプロジェクトのこの局面の成果は次に述べるとおりです。まず、私たちは重要そうな表現問題をすべて調べました。しかしこれらの問題全部は、多くの論文が他の形式論、他の表記法、他の表現言語について議論しているという事実にも関わらず、重要ではありませんでした。いまのところ、対話問題については有効ではないと思います。しかし、通常仮定するように、よい表現は半分解決したも同然です。

二つ目として、プロセスモデル言語の説明としては同じ道をたどるべきではありません。言語の説明は、明確さにより、あるレベルに分割されますが、対話のプロセスモデルは分割されません。

三つ目として、私たちは、相互結合の構造モデルを構築するための構造の詳細を充分知らないということがあげられます。それについての研究は非常に少ないのです。もっと詳細が必要であり、この制御問題について明確に考察する必要があるということはここでの議論で何回も言われました。

四つ目として、ユーザーモデルを使ってスピーチアクトについて私たちが行なったことを試すのは見込みがありそうです。ユーザーモデルのみに基づいた大きなシステムを実施した者は誰もいないのです。例として、私は通訳をこの方向で行なうことを勧めます。なぜなら、このユーザーモデルの話題から始められたからです。この線に沿って何かを行なうことは見込みがありそうです。なぜなら、ここでわかりますように、単語知識と話者－聞き手モデルを使った言語学的な情報の関係について最もよく知っているからです。

五つ目として、これが五つ目です。これも議論の中で何回も出てきました。この分野のほとんどの言語学的な知識は、この資料からプロセシングモデルが作られたのと同じようには展開しません。例えばあなたたちの作業について言えますが、あなたたちのプロセスモデルの精度は、スピーチアクトに関係するこれら進行中の言語学的な作業の精度よりもかなりよいという印象があります。それは実験的な作業であるばかりでなく、スピーチアクトおよび人－機械のコミュニケーション義務、そしてそれらすべてについての理論構築作業なのです。言語学的理論および実験的な作業から知られたことはすでに人工知能に利用されていると思います。それで、言語学からの新しい行為があるはずです。なぜなら、ＡＩは人－人コミュニケーションという背景では実験できないからです。

最後の点はつまらないものです。貢献した人のほとんどがそれを指摘しました。対話の首尾一貫原則についての私たちの間違った考えの決定的な点は、構造問題から派生したことがわかりました。これらの要因すべてが、これらのパラメータすべてがプロセシングモデルの中でどの様に相互作用を持つのかについて私たちは知りませんでした。それで私たちは次にプロセス交替等の制御構造問題に集中しました。そして私たちは、以前に定義されたそのモデルに基づいたそのようなシステムの構造というこの問題を発見しました。こ

れはまず集中するもので、それからこれらを解決する、つまりこれらの問題に解答を与えようとするものです。

これがなぜ音声言語モデルから議論を見つけようとしたかの理由です。なぜなら、構造問題はもっと納得できる方法で解決しなければならないとまず思ったからです。音声言語結合の研究者のなかで、私たちはこの構造問題が談話にあるのと同じように、よく知られていることを発見したので、テキストの超文節現象としての談話と音声韻律学の間に多くの類似点があります。それゆえ、私たちはドイツの技術省を納得させて、ＡＳＬというプロジェクトを設立することに成功したのでした。そのプロジェクトは今始まったところで、この分野に正確にあてはまる研究という定義づけの段階にあります。音声言語システムの例に関する言語システム構造ということです。ですからＡＳＬの意味は音声言語システム構造または音声および言語モデルの構造です。この定義づけ段階は始まったばかりで、その主要段階にはあと３年から５年かかるでしょう。

私たちの興味はたいていこの「音声言語プロセスの最も柔軟性に富み、充分な結合はどれか」という疑問です。最初の１年で文献中の構造問題と呼ばれるすべてのものを収集し、ここにあるこの疑問に戻り、構造について考察し、どんな言語学的情報が音声認識を進歩させるかについて考察しなければならないということははっきりわかっています。これがここでの主題です。そして音声議論から、言語分析はどんな利益をもたらすかということです。しかし、主要なことは、「このモデルから始めて、現実のシステムをつくる段階に進み、異なった構造を構築するためにはどんな選択があるか」ということです。ありがとうございました。

質疑応答

問　あなたはスピーチアクト階層の分類を示しました。しかし、どうやって成し遂げるかまたはスピーチアクトを表現するかの方法を調べることが極めて重要であると思います。この点についてどう思われますか。

答　私が言おうとしたのはスピーチアクトのアプローチがこのシステムではかなり強力であるということです。しかし私たちはそれを発見したのです。言いたかったのはスピーチアクトアプローチがすべてのものが依存する７字則ではないということです。わたしたちは多くの一種のスピーチアクトの効果の順序づけをしました。あなたがスピーチアクトはこのモデルに対して非常に強力な何かであると強調されるのは正しいと思います。しかし、もしひとつの支配的な根拠を求めておられるならこれはたぶんスピーチアクトではないと思います。そして私たちがやらなかったこと、たとえば彼がやったことは公理といったものを使った実際の完全なシステムでした。これはこの仕事の大きな部分なのですが、これらの下位分類にこれがなされたことはありませんでした。この仕事に対しそれが重要であることは認めます。

Wolfgang Wahlster （サールランド大学）

対話システムにおけるアンティシペーションフィードバック

高品質対話翻訳システムには２つの必要条件があると思います。

一つは、上手な人間の翻訳者はソース言語およびターゲット言語を、流暢に話し、理解するということ。つまり、対話翻訳システムは、２つの分析要素－－一つはソース言語用、一つはターゲット言語用－－そして二つの生成要素－－ソース言語用とターゲット言語用－－を持っていなければならないということです。このことは、一つの対話翻訳システムの設計法則が、これを扱いやすくするため、言語的知識ソースの２方向指向性を持たねばならないということを示しています。

二つ目の条件は、上手な人間の翻訳者は話し手や聞き手の信条、彼の目的や意図、その分野での前知識、同様にこの分野に関して持つ可能性のある誤解などを考慮に入れると思われるということです。このことは高品質対話翻訳システムは、二つのユーザーモデルを持たなければならないことを示します。二つのうち一つは、ソース言語における話し手の発話の意図的な効果を決定するためのモデルで、もう一つは、聞き手側での計画された翻訳の効果を予期するためのユーザーモデルです。

Martin KayとIvan Sagは最近、彼らが「交渉機械翻訳」と呼ぶもののための新しい構造を提案しました。主要な考えはまったく簡潔で、私たちはいつも翻訳を交渉者に依って見いだされた妥協案としてみているということです。この交渉者というのは、インターリンガルベースのシステムの古典的な設計に、新しく現われた要素です。分析する部分も生成する部分もありますが、タスクの協議事項を見失わないようにし、交渉者と呼ばれるこの新しい要素もあるのです。

Kayは単純な属性論理ベースのインターリンガル、この場合分析要素から生成要素へ送られる一組の文字ですが、それを使うことを提案しました。しかし、要点は簡単な翻訳が応用への試みに充分応えられるほど、近いものであるかどうかを交渉者がチェックするというものです。つまり、この構造における交渉者は翻訳過程の中で、他の要素に対し、追加情報を求めることができるのです。たとえば、交渉者要素は、分析要素に対して、さらなる情報や別の情報、バックトラッキングのようなものを求めることができます。また生成要素に対して尋ねることもできます。こうして、対話式機械翻訳ができるのです。あるいは、「台所流し」アプローチをＡＩに用いれば、手助けとなる一般専門家システムや知識ベースシステムを考察する事ができるでしょう。

ここではさらに革新的な対話翻訳へのアプローチを提案します。それは新しい構造を持ち、ある意味では似通ったものですが、ずっと急革新的な見解をとっています。ですからこの話の中で使用する基本的な考えはアンティシペーションフィードバックループの考え方です。最初に、問題の生成の側に対して、この案の簡単なスケッチを描きましょう。

これが私たちのシステムの生成要素であるとします。その時アンティシペーションフィードバックというのは、ある発話を本当に理解する前に、あなた自身の分析要素、つまりシステムの分析要素にそれをフィードバックすることを意味します。つまり、聞き手の知識についての仮定を利用します。この場合それは彼の談話モデルであり、意図であり、目的、そして理解なのです。そして、この前提のもと、その簡単な発話を理解しようとします。もしそれであなたの意図に近似できなければ、また別なループにうつって、その発話を洗練しようとします。

要するに、アンティシペーションフィードバックループシステムは、システムが理解するように設計されている伝達を、受け手側でどう解釈するかを予想するという、システムの理解能力の使用を包含しているのです。

本質的にアンティシペーションとは、話し手の側からみれば、次のような問に答えるようなものであります。「私が聞き手と仮定している信条Ｂに関連するこの発話を耳にしたとすれば、それはわたしに何の効果を及ぼすだろうか」また、聞き手のための機械翻訳でこのような問を発することもできます。「もし私がこの言葉を発したとすると、どの様な信条や目的に基づいて言ったのだろうか」

これがだいたいの考えです。ここで私の提案する高品質対話翻訳のためのＡＦＬモデルを考察してみましょう。さきほど言いましたように、二重の要素があります。ソース言語に対しては分析要素があります。ターゲット言語についてはまた別の要素があります。ソース言語に対しては生成要素があり、ターゲット言語に対しては生成要素があります。ここで、Ｓというあるソース言語での発話があると仮定します。これはまず分析されます。それから意味論的な計画された表現を得ます。つぎに生成器で生成し、予想されるＳ′という発話を「計算します」。それからこの近似プロセスがあり、Ｓ′は私たちが見たＳという発話と本当に近似しているかどうかを比較します。もしこれが真でなければ私たちは分析を見直します。最初のアンティシペーションフィードバックループにはもう一つのサイクルがあります。もしなければそれはオーケーで、意味論的表現αを出力します。これがあると仮定しますと、これを生成要素に与えます。そしてアンティシペーションフィードバックのもうひとつのサイクルがあります。つまりそれはこの場合、私たちがターゲット言語の発話を計画するということです。しかしここでまた、ユーザーモデル２のもとで、それは聞き手についての仮定ですが、機械翻訳が、対話プロセスにおいて聞き手と話し手との間の真の媒介者となるモデルを用います。またもう一つのフィードバックループがあります。すなわち、分析要素がこんどは、計画された発話の意味論的予想表現であるα′を生成するのです。そしてまた近似プロセスがあります。もし応用した文脈に応じてα′がαを近似していれば、この近似がどれだけよくて、どれだけ近くなければならないかを定めなければなりません。それからは、修正過程に入るか、あるいはターゲット言語における発話という最終出力を得ます。

さて、なぜ２つのアンティシペーションフィードバックループがあるのでしょうか。分析過程のなかで私が言った通り、第一のアンティシペーションフィードバックループが利用するのはユーザーモデル１という話し手のモデル、そしてシステムのターゲット言語生成能力です。この場合その能力とは、「どんな信条、目的、および意図にもとづいて話し手はこのように言ったのだろうか」という質問に答えることです。つまり、最初のＡＦＬはアブダクティブな味をもっています。これは本当にアブダクションプロセスです。それは合成による分析の特定のバリエーションとして見ることができます。

そして生成プロセスにおいては、私たちのモデルには第二のＡＦＬがあります。この場合、私たちは、聞き手のモデル、ユーザーモデル２、および、ターゲット言語が「計画された翻訳は、聞き手の信条、目的、意図に関連して聞き手にどんな効果をもたらすか」というような質問に答えるこのシステムの分析能力を活用するのです。つまり、アブダクションのある第一のアンティシペーションフィードバックループとは対照的に、これは演繹

的、すなわち演繹法駆動されているのです。それは分析による合成的な一つのバリエーションとして見ることができます。

さてこの内容の中で、ユーザーモデルとは何を意味するのでしょうか。対話システムの中のユーザーモデルは単に自然言語システムのための知識源です。そこにはそのシステムの対話行為に関連しているであろう、ユーザーのあらゆる局面において明白な仮定が含まれます。ユーザーモデリング要素は対話中にユーザーモデルをどんどん組み立て、その内部で格納、更新、削除、入力するシステムの一部分であります。また、一貫性を保持し、他の要素にユーザーに関する仮定を供給します。

さて、いまあなたは、これはたしかに素晴らしい理論上のモデルだが、実際にはうまく行くのかとお尋ねになるでしょう。私たちは、以前ＡＦＬで、様々な実験を行ないました。そして、本当にアンティシペーションフィードバックループを含む四つのシステムを開発しました。私はそれがうまくいっていると思います。例えば、計画された発話が、あいまいかどうかを調べるとき、アンティシペーションフィードバックループがなければ、その発話は聞き手にとって本当にあいまいではないのか、また言い間違いがあるかどうかを判定することは困難だと思います。もし、単に生成する場合でも、常にとても楽観的な考えを持っているとすると、ユーザーモデルと一緒に分析能力を使うことに依ってのみ、聞き手が困難に陥っているかも知れないとわかるでしょう。

私たちは局所的にＡＦＬを使いました。もし総合的なＡＦＬを使うようにし、つまり、最初に全発話を作成してこのループに入ることとすることは、システムをあまりに遅くさせて、これはよくない考えだと思います。私が、提案し、行なったことはバックトラッキングで、局所的にアンティシペーションフィードバックを行なうことです。それで例として、代名詞の作成にこれを行ないました。計画された発話中に代名詞を生成したとき、私たちは直ちに、代名詞分析能力を用いこの代名詞が聞き手にとってあいまいでないかどうか、また、言い間違いが無いかどうか等を分析しました。省略についても同じことを行ないました。省略を生成するとき、私たちは、省略分析要素をユーザーモデルおよび談話モデルとともに使用し、この省略が本当に再構成されるかどうかを調べました。そうでないと次にもっと省略の少ない構造を生成しなければなりません。これらの過程はすべてデイペンデンシー指向のバックトラッキングに基づいています。

私たちは、決定の道筋を保つ Truthメンテナンスシステムを実施しました。私たちは走り書きからはサイクルをたどろうとしません。この四つの実験で、最初はＨＡＭ－ＡＮＳシステムの省略生成要素です。それから私たちはこのアンティシペーションフィードバックの考えを最近使って、エキスパートシステムのマルチモードインターフェースでの指示身振り生成に使いました。そして他のシステムにも使い、そこでは視覚システムへのインターフェースへの空間描写の生成を扱うVITRAを使います。最後に、新しいプロジェクトで、このアンティシペーションフィードバックの考えを文脈依存メディア選択のため生成側で使用します。そしてこれらすべての場合で、もしフィードバックの正しいレベルを発見するなら、このフィードバックの考えはシステムをそれほど遅くしないことを発見しました。ですから本当にそれを使えるのです。

省略の場合の簡単なフィードバックループの例として、「５人の同僚は３週間休みを取りますか。」という質問を想定してみましょう。システムは判断原理に従って、できるだけ短くなろうとします。最初の試行として、この省略要素はちょっと短すぎるものを生成します。「いいえ　２です」といいます。分析要素の使用によってのみ、これがあいまいだということを発見します。２人の同僚に３週間以上の休暇があるのかも知れませんし、５人の同僚に２週間以上の休暇があるのかも知れないからです。それで、この対応が完全ではないことをここで知ります。意図された意味表現および再構成された意味表現が一致しないのです。ここに二つの可能性があります。分析要素から、システムはより省略の少ない構造を生成しようと決定します。この場合には、知識ベースの状態に応じて「いいえ２週間です」または「いいえ２人です」となります。

さきに申しましたように、このアンティシペーションフィードバック技術は、また私たちのXTRAシステムでも使用されました。このシステムは多重方式自然言語インターフェースで、自然言語をグラフィック指示に結びつけます。そしてここに「ここへ私の組合費を追加していいですか」という入力があります。

さて、私たちがアンティシペーションフィードバック原理で取り組んだ作業は単にそういう生成を行なうことです。システムがある身振り、プランされた身振りを生成するにしたがい、システムはプランされた身振りを、まずユーザーモデルを用いて身振り分析要素にフィードバックします。そして、その身振りのあいまい性が最小かどうかを発見します。最小の場合はこの身振りを実現します。そうでなければ、身振りを見直します。

このシステムが何を行なうのか例を示すために、ここにある会話原則があります。例えば、意図した対象を指示行為で隠さないこと。何も見えなくなるのでこのように指し示してはいけません。できるだけ正確な指示をしなさい。それで、もしスクリーン上のこの書

式の会費概念について言及したいのであれば、まず、この原則にのっとったシステムの内部的試行はこのように鉛筆による指示です。そしてこの発話は、この指示はこのような「これを削除すること」発話とともにあります。そしてここを指示します。さて、身振り選択の例では、（テープ替え）三つの候補があります。しかし意図された解釈はこれでした。ですから指示身振りを変えます。そして異なる種類の指示になるのです。さて入差し指を使っています。これが最終的に出力されるのでしょう。

ここではフィードバックループの他の応用に触れる時間はありませんので、結論を言わせて下さい。高品質翻訳システム、対話翻訳にとっては二つのユーザーモデルが必要であると思います。一つは話し手の発話の意図的な効果を決定するためのモデルで、もう一つは、聞き手側でのプランされた発話の効果を予期するためのユーザーモデルです。対話翻訳におけるこのアンティシペーションフィードバックループモデル、それは分析要素と生成要素をソース言語とターゲット言語の両方に持つが、このモデルは上手な人間の翻訳者は両言語を理解し流暢に話すという事実を反映した物であると思われる。ＡＦＬモデルにでも翻訳は機能的な近似プロセスであるとみられています。なぜなら、処理および対話のなかで先になされた仮定も後の段階で、つまり順次的な分析および生成の過程で見直されなければならないからです。

最後に、ＡＦＬについての私たちの実験はいままでのところ励みになるものでした。私たちはこれを生成要素についてのみ行ないました。しかし、システムをひどく遅くさせることなくあいまい性と指示対象間違いを無くさなければならないことがわかりました。ありがとうございました。

問　　あなたの仕事はとても魅力あるものです。翻訳について私たちが行なっている仕事ではシステムの可逆性が非常に問題になります。そしてこれは可逆性技術の使用対象であると思います。いまあなたはフィードバックを生起させるレベルについてはあまりおっしゃいませんでしたが。

答　　それはいい質問です。それについてはあまり言いませんでした。極端な場合における総合的なアンティシペーションフィードバックループは十分に発達した発話を構成することだと申し上げました。これは語形変化とともに、例えば実に表層的なストリングとともに、それからこれをまず形態学等にフィードバックすることで、単に代名詞があいまいであるかを見つけようとすることです。これは正しい考えではないでしょう。さきに申しましたように、フィードバックする正しいレベルを見つけなければなりません。現在はこの考えを問題に応じて異なるレベルで使用しています。

これが意味するのはたとえば代名詞生成のため、これを表層意味論的構造と呼ばれるものに対して行なっているのです。これは状況スキーマタに似ています。そのような表層に適応づけされた意味論的適応を持つ提案がないとすればどうでしょう。それがフィードバックを行なう正しいレベルであると思われます。この問題に対して、十分に語形変化した形式にまでレベルを下げるのは適当ではありません。ですから、あなたは正しいのです。このフィードバックはいろんなレベルでそれを行うということを意味しています。省略については行なうべきでないと思います。そこでは生成においてもっと下のレベルにまで行かなければなりません。たとえば順序変形を行なわなければならないでしょう。省略の再構成には語順が重要なのですが、最終語形変化は重要ではないのです。それは問題のレベルに依存するのです。

問　　あなたの仕事と私の話している仕事には何か類似性があるようです。あなたの話の中で、生成器が、あいまい性に対してまさに話そうとしていたことをチェックし、見直すという方法が示されました。最初のスライドで、これはソース入力の場合にも当てはまると暗示されましたが、最後には、あなたはそれを行なわなかったとおっしゃいました。これは正しいですか。

答　　正しいです。

問　　それならば、同様に入力をチェックするようなシステムを考えておられますか。

答　　はい。入力に対してそれは本当に対称的なことだと思います。あなたが入力するので、私が言ったように、このアブダクションの過程の中で、話し手はなぜ正確にこのようなことを言ったのだろうかということを理解しなければなりません。完全な対話翻訳システムのためにはこれが重要であると思いますが、私は現在対話翻訳について仕事をしているわけではないので言わなければなりません。私たちのプロジェクトは対話についてですので、まだこれをやっていないのです。しかし原則的には違いがあるとは思いませんし、

現在の仕事のほとんどは間違っていると思います。アブダクションプロセスとしての理解としてのみ、計算処理可能な言語学の中に現在の仕事があるのです。たとえば、ＳＲＩでのマーク・スティックルおよびジェラルド・ホッブスの仕事です。これは指摘したようにここの誘引という考えです。一方こっちには演繹があります。これは完全に適合すると思います。しかし、申しましたようにこれは非常に野心的な目標です。ですから高品質の翻訳と申し上げたのです。

話し手の意図がなんであったか、また聞き手の意図はなんであるかを本当に理解するのは多くの翻訳でも、人間の通訳でさえそれほどうまくいっていないと思います。これは本当に大変なことです。ですから、革新的なアプローチであると申し上げたのです。長期的にはこれに取り組むべきであると思います。それは構造モデルをもたらすでしょう。それでいいと思いますが、これにはまだ何年もかかると思います。このモデルで試作システムを構築するのは現実的でないでしょうが、研究課題として、取り組むべきだと思います。そしてこのアンティシペーションフィードバックループを、さきに言いましたように、いろんなレベルで行えますし、私たちが省略生成、代名詞で行なったように、それはいま現在実時間で働くと思います。そして充分に成熟したモデルは将来のためのものです。

中村孝（大阪大学）

Parsing Utterance Mechanism Based on Black board Model

私たちの研究所の方では音声言語のプロジェクトの一環ということで、特に音声認識の終わったあと、ダイアローグ・アンダースタンディング・システムというんですか、自然言語ベースの対話の理解のシステムというのを実際つくっておりまして、現在プロトタイプ・システムというのを構築しています。

今回の発表では特にダイアローグ・アンダースタンディング・システムの中から個々の発話のパージングのメカニズムということについてお話したいと思います。

先ほど申しましたように音声認識システムが前段にありまして、その結果音声認識された結果というのが入ってきました後、それを実際に解析しまして一体どういうような対話になっているのかということをやるようなシステムということで考えております。

一応全体のダイアローグ・アンダースタンディング・システムというのは３つに分けて考えていまして、個々の発話をパージングするシステム、それからダイアローグ全体の構造、ダイアローグペアーとかそういうような構造というのを解析するシステム、それから全体としてマンマシンのダイアローグというのを扱いますので、そのために必要なダイアローグのマネージメント、実際どういうふうにダイアローグを誘導していくか、次に出てくる発話というのは一体どういうものかという予測ですとか、そういうふうなことのマネージメントのシステムというふうに分けて考えております。そして私どもの、特に私のグループでやっておりますのが個々の発話の解析のシステムということで、特に対話、スポークンダイアローグに出てきますグラマティカリーイルフォームドな個々の発話というのをどういうふうに解析するかということについて研究を行ってきました。

バックグラウンドに当たるものなんですけれども、個々の発話の解析ということなんですけれども、先ほど申したようにグラマティカリー・イル・フォームドなものを実際に解析してしないといけないわけなんですけれども、これを解析するための手法といいますか、そういうバックグラウンドなアプローチとしまして、大きく３つのものを考えて行っております。

プリファレンスセマンティクなアプローチ、それからインクリメンタル・ディスアンビギエーションアプローチという大きな２つのアプローチを使いまして、その実際にやる段階におきましてセマンティックロールパターンマッチングというメソッドを使いまして、最終的にセマンティックドリブンなパージングメカニズムを構築していこうということになっております。

プリファレンスセマンティックなアプローチといいますのは、これはおそらく皆さん賛成していただけると思いますけれども、特にグラマティカル・イル・フォームドなアトランスを解析しようと思いましたら、ただ一つの正しい解が出てくるということはまずあり得ませんので、幾つかの可能な解釈というのを生成しまして、その中からなるべく確からしいものというのを取り出してやって、それを最終的な解析結果とすることになります。

どちらかというとこちらの方に重点を置いているんですけれども、インクリメンタルにディスアンビギュエーションするアプローチとなります。人間の個々の文の対話理解のシステムといいましたら、おそらく割とインクルメンタリーというんですか、文の初めから普通大体解釈していきまして、一つの文が終わった段階でその文の意味が確定するということになります。文の最後に来るまで解釈を中断していて、終わった段階にガサッと解釈するというのではなしに、だんだんとわかっていくということになります。それからガーデンパス文なんていうのもあるんですけれども、おそらくそういう感じであまりバックト

ラッキングというのは少ないだろうというふうに考えます。

そういうふうなこと、人間のアナリシスプロセスというのをなるべく取り入れてやろうというわけで、個々の発話の中の単語の役割ですね、意味ですとか、構文上の役割というものを徐々に明らかにしていく、インクルメンタルに明らかにしていく、そのときに実際に解析の途中で得られますようなコンストレインツを利用してだんだん明らかにしていくという形で解析を行おうと。ですから個々のロールオブワーズに対してハイポセシス、仮説をどんどんつくっては次の単語、どんどん入ってきたら、その仮説がだんだん減っていって、最終的に文の最後までいったら個々の仮説というのがある程度一つの解釈としてできると、定まるというふうに考えます。

こういうふうなことを実際にできましたらレフト・ツー・ライトでなるべくバックトラックが少ないようなパージングというのが可能になるのじゃないかというふうな形で、このアプローチというのをなるべく取り入れようというふうに考えてやっております。

もう一つバックグラウンドの最後の３つ目なんですけれども、セマンティックロールパターンのマッチングというのを使おうと。先ほど言いましたように個々の単語の役割ということで、セマンティックロールというものを取り上げまして、これについての仮説候補ですね、個々の名詞句に対するセマンティックロールの候補というのを出します。個々の名詞句ですから、文の全体としてはそういうセマンティックロールのパターンというのが一応できます。それと、動詞の方のレキシコンの方にありますそれぞれのセマンティックロールのパターンのマッチングをとりまして、その動詞のどういうふうな意味がとられているかというのを決定します。それに基づきましてそこのナウンフレーズのケースフレームですとかそういうのを決定してやろうということになります。

セマンティックロールというものについてもう少し説明いたしますと、ここでは一応私たちがセマンティックロールと言っておりますのが、日本語の動詞のレキシコン、ＩＰＡで開発しました動詞のレキシコンであるＩＰＡＬと呼んでいるものがあるんですけども、これで取り入れられているもので、日本語でいいますと述語素というふうに呼んでおります。一応それをセマンティックロールというふうに呼んでおります。ここのナウンフレーズと動詞の関係、ケースフレームみたいな感じですとか、それからナウンフレーズとナウンフレーズの関係、ディペンデンシーグラマー、かかりうけに類するものというような形のパターンが一応載っております。これを使うと。

それからもう一つ、ＩＰＡＬの方の動詞というのは、一つの動詞について幾つかのサブエントリーというものからなっています。個々のサブエントリーというのは、意味的ですとか、構文的な差、一つの動詞でも意味的、構文的な差があるものをそれぞれグループという形にして、個々のサブエントリーに関して意味ですとか、セマンティックロールのパターンというのが一応書いてあります。これに基づいてサブエントリーを決定してやれば動詞の意味　―　一つの動詞の中でもそれはどういう意味かというのが決定されるということになります。

そのＩＰＡＬのエントリーの例なんです。これ日本語のものなので　―。たとえばレッスンの例なんですけれども、サブエントリーがそれぞれこうあって、こうあってそして後ずらっと続いていると。それでここの０２インＡと書いてあるのがたとえばセマンティックロールのパターンということになります。

それからセマンティックロールというふうに言いましたのが大体こういう感じでありまして、これも大体ケースフレームですとかそういうのに対応することになります。

こういうのを使いまして、セマンティックロールパターンのマッチングという手法を使ってなるべくインクルメンタルな解析を行おうということになるわけなんですけれども、大体の流れというのはこうなります。

大きな流れでいいましたら最初に出てきたセンテンスにつきましてまずＮＰごとに分割しまして、それぞれのＮＰについてセマンティックロールのキャンディデートを出して、出てきたやつ、セマンティックロールのキャンディデートでいうと組み合わせとか、それぞれのコンストレイントを考えまして、フィルタリングを行いまして、最終的にバーブのレキシコンの方とのマッチングをとって、バーブのレキシコンのサブエントリーが決定されますから、その段階でＮＰのケースロールとかサブエントリーとか、その辺のものが明らかになると。特にここのキャンディデートでやると、それのフィルタリングのとこにつきましてインクルメンタルに仮説というものをつくっては絞り込みというものをやっていこうと。

それで具体的にはまず最初のＮＰが入ってきましたらナウンフレーズが入ってきしましたら、それについての候補というのが割とどっと出てくると。２つ目のが入ってきたらそれについてのがどっと出てくると。その両方の組み合わせというものを考えましたら、組み合わせについてコンストレイントでありますからちょっと減ると。３つ目のＮＰというが入ってきましたらそれについてのがまた単独だとかなり出るんですけども、ワン、ツー、スリーで３つの組み合わせだとまた減っていくということになりますので、こういう感

じでなるべくフィルタリングを行いまして、マッチングの手間を省いてやろうということになります。

その3つのスティップについて簡単に説明しますと、最初にセマンティックロールのキャンディデートというのをつくり出すんですけれども、それは個々のナウンフレーズについて、ナウンフレーズについていますセマンティックフィーチャーですとか、助詞になりますとサフィックスですとかその辺の情報からキャンディデートを出してやると。そのために必要な知識といいますのが、それぞれルールの形で書いておくということになります。最終的にこのやりましたら個々の名詞句についてそれのセマンティックロールのキャンディデートというのがリストの形で出てくる。

いまセマンティックフィチャー・意味素性と言ったんですけれども、じゃＩＰＡＬの方で取り上げられていますセマンティックフィーチャーというのはこの程度のものでして、あまり細かくなくて、どうもこれじゃ不十分じゃないかという面もあるんですけれども、一応ＩＰＡＬの中にこういう記述されていますので、これを実際に使って行っていくと。

それから次にフィルタリングの方の話なんですけれども、フィルタリングしてやるのにそのセマンティックロールの組み合わせの情報というのが必要になりますので、セマンティックロールのコオカレンスのルールというのを一応つくりまして、それでありそうにない組み合わせというものをとっていくという形で、実際のフィルタリングというのを行っています。普通は大体たとえば2項の組み合わせのことでしたら、Ａというセマンティックロールとｂというセマンティックロールというのが1つの文だと普通は出てこないというふうな感じで書いてあります。

このセマンティックロールのコオカレンスのルールなんですけれども、一応ＩＰＡＬの辞書のセマンティックロールのパターンを全部抽出しまして、これちょっと見にくいですが、たとえばこれ2項でたとえばこういうセマンティックロールと、こういうセマンティックロールの間のやつがどれぐあるかというのを全部一応出してみまして、たとえば空白のやつはそういう存在は一応ないということになっています。この辺のデータを使いまして、ないものとか、少ないものというのはなるべく全然やめてしまうか、順位というのを下げてしまう。大きいものだったらその順位を上げてやる。一応いまのところ2項と3項のことに関してこういうデータがとってありまして、後は大きくなればその組み合わせということでやっていくことになります。

そういう形でフィルタリングができました後フィルタリングしたキャンディデートの方とそれから動詞の方のセマンティックロールのパターンというもののマッチングをとりまして、サブエントリーがどういうふうになっているかということを決めてやると。この場合にもどういうふうなマッチングがあるかというのが問題であります。そのサブエントリーが決まりましたら、何遍も言っておりますように名詞句のケースロールというのはそのパターンとのあれで決まりますし、そしてサブエントリーが決定しますから、そのサブエントリーロールに書いてある個々の動詞の性質がわかってくるというふうになっています。

いまのところシステム全体の拡張性ですとか柔軟性なんかを考えまして全体ブラックボードモデルみたいなものを使ってやろうということでつくっておりまして、ノレッジソースですとか、それからブラックボードのハイアラーキーというのは一応こういうふうな感じでやっております。

一応説明はそこまでなんですけれども、ごくプロトタイプなシステムが一応いまできたところでして、まだまだ将来の問題点というのはかなり残っています。ここにずらっと並べたようなものがありますし、先ほど出しましたようなルールですとか、そのようなものというのもまだまだやらないといけないところです。たとえば統計的なというんですか、辞書の項目の中からセマンティックなパターンというのをとっておりますので、実際のコーパスからそのセマンティックなパターンとかいうのを統計的にとっているわけじゃないんですね。ですから実際の文ですとか、実際の発話ですとかそういうのを直接反映しているというわけではないというのもあります。それからフィルタリングの絞り込みのルールなんていうのはもうちょっと強力にした方がええのか、それともも少し緩めにした方がええのか、実際いろんな面でやっていかなければならないことがいろいろ残っております。

もう終わりでして、そういうことで一応パージングメカニズムということで、イルフォームのパージングメカニズムというのを一応つくってみたと。プロトタイプのものなんですけれども、おそらくこのインクルメンタルなジェネレーション・セレクション・オブ・ハイポセシスというこういうことというのはおそらく今後もずっと残っていくことになると思います。以上です。

質疑応答

問　　マッチングにはどういうメジャーを使っているのでしょうか。たとえば確率的なメジャーを使っているのか、論理的なものだけでやっているのかお伺いしたいと思います。

答　いまのところマッチングの方は、キャンディデート候補で出しました順位づけの方がありますので、それの上位のものというのを一応優先します。お聞きのとおり実はこのところにマッチングのとり方が将来の課題になっていまして、それの方も実はどういうふうにマッチングメジャーをとるのか。それからたとえばマッチングといいましても、セマンティックロールパターンの方の数とそれから実際出てきたキャンディデートとの項目の数が違うときですとか、そういうときに実際どういうふうにしてどうやるのか。それから2つのパターンをやったときにどっちを優先するかというのは実はあまり詰めてませんので、いまところ将来の課題ということで残っています。

Session 8

Future Direction of Language Processing for Automatic Telephone Interpretation (Panel Discussion)

長尾（京都大学）　外国からの方はたった１人でございますので、できましたらまたフロアーの方からも積極的な御発言を期待したいと、こういうように思います。

それからこのセッションは最後のセッションですからできましたらＡＴＲの今後の研究の進め方についてもいろんな期待とか御忠告とかがございましたら積極的にしていただくのもありがたいかというふうに思いますのでよろしくお願い申し上げます。

言い忘れましたけれども私は京都大学の電気工学科におります長尾と申します。どうぞよろしくお願い申し上げます。

きょうは司会をさせていただくということであまり意見は言わないことにしたいと思いますので、早速時間もありませんから第１のスピーカーからお話を伺います。第１のスピーカーは東芝の天野さんでございまして機械翻訳の研究を随分してこられた方でございます。よろしくお願いします。

天野（東芝総研究情報システム研）　東芝の天野です。私どもの研究室では２通りの仕方でダイアローグにかかわるような研究をしておりまして、１つはドメインを非常に限ってそのかわり対象知識をもって推論機構をもって対話をすると。１つは機械翻訳システムなんですが、前の方のドメインを非常に限って対象知識、推論機構を持つというダイアローグシステムと言いますのは基礎研究のパイロットモデルのようなものは皆様も御経験があるかもしれませんが、設計者が入力した文にはうまく答えられますけども、一般にスポンティニアスな発話を入力しますとなかなかうまく働かないというようなもので、言語現象を広くどんな言語現象がダイアローグで起こるのかということを広く集めようとしますとなかなか難しいということがございます。

そうこうしているときに機械翻訳システムは一方非常に深く分析はできませんが、広い分野を非常にタフに扱うことができるというような特性があります。その機械翻訳システムを使いましてダイアローグと言いますか、会話システムを構成してみました。これでいろいろな対話にかかわるようなデータをとってみたわけです。きょうはそのことについてお話してみたいと思います。

このシステムを使いまして去年、一昨年と２年間にわたりまして一昨年はここにありますようにテレコム８７の会場へ１台持ち込み日本との間でビジターと我々とが会話をしたわけです。去年はドイツでハーノーバーメッセに持ち込んで実験をしたわけです。ここで行いましたのはこれはもう言うまでもなく機械翻訳システムとキーボードコンバセーションファンクションというものを結合しましてオンラインのトランスレーティングダイアローグシステムを構成しましてどこまで一体浅い処理で機械翻訳は対象知識という意味では非常に浅いですから、浅い知識でできるかと、会話が。というようなことをやってみたわけです。

このときに当然ダイアローグというのは非常に難しいものですので、いろいろな問題が起こることは予想しておりますが、どうそれに対処するかということなんですが、モデルとしましてはコンバーサーがおり、真ん中に機械翻訳システムをつくったマシンーインタープリーターがいると。コンバーサーの役割としてはアタランスフォースとコンテキスト・アンダースタンダーであると。マシンインタープリーターはシンタクテックとセマンティックなアナリス、トランスファーとターゲットランゲレッジジェネレーション、ここで当然いろいろなノイズが起こります。レキシカルなエラー、シンタクティックなエラー、セマンティックなエラー、トランスファーのところではレキシカルなエラーが起こりましょうし、アナリシスと、ジェネレーションのところではシンタクティック及びセマンティックなエラーが起こります。こういうものがノイズになって誤訳を引き起こします。

それに対するカウンターメジャーというのはもうシステムの方ではつくることができなくって、対処できないわけで、それに対しては人間のコンバーサーの知識といいますか、コンバーサーに期待したわけですね、一つはコンバーサーがコンテキストを使って理解してくれるであろうと。もう一つはもう聞き直してほしいと、くださいというこの単純なものでやったわけです。ちょっとイメージを出すためにお見せいたしますと、これが英語側の画面で、これが自分の会話、発話、これが相手の発話、これが翻訳されたもの、これは相手の発話と、現時点の一字一字打っている発話です。

日本語側では当然日本語が出てまいります。当然考えられますことはメタダイアローグが非常に多く出るだろうと、聞き返しが起こりますのでわからないところはメタダイアローグが非常に出るだろうということで、たとえばその例ですが、Ｅというのは英語側、Ｊというのは日本語側でＴというのはその翻訳という意味ですが、英語側からこういうことを言ってきたわけですね。そのときにこのクラッシュコースというのは辞書になかったわ

けです。辞書にない場合には品詞は推定いたしますので大体こういうものは冠詞がありますから名詞でしょうし、品詞を推定いたしまして原語のまま表示します。そうすると日本語側ではここの「Crash Course」だけが翻訳されないで出てきますので、何だかわからないわけで聞き返したわけですね。そういうような聞き返しがどんどん起こるわけです。この場合ですと、この場合は２月前の「月前」が処理できなくってこのまま相手方へ英語側へ出てしまったわけですね。相手方から「ここのところ何だ」と、こう聞いてきたわけです。「６月だと」こちらから日本語側から言い直してあげたと、相手はわかりましたと、こういう会話が起こったということです。

きのういろいろ先生方のお話を聞いておりまして感じましたことは非常に音声とキーボードダイアローグシステムで似たような現象が非常に起こっていると。直接同じではないですが、対応できるようなものが非常に起こっているという感じがいたしました。　一つには音声では必ず認識間違いが起こりますが、当然キーボードではタイピングエラーが起こります。そのほか大文字でないといけないものが小文字のまま打たれてしまうと、音声は大文字、小文字は区別できませんので同じような現象ですね。これはいまのところ機械ではもちろんスペリングチェッカーというものがありますけれど、スペリングチェッカーは複数の候補を出すこともありましょうし、時間もかかりますからうまく処理できない。

それからフィルドスペースという話がきのうも出てきましたが、これは逆の問題でスペースがとれてしまった場合ですね、こういう場合も何だかわからなくなりますので会話がメタダイアローグになってしまうと。

それからある言語で話しているところへ別の言語が入ってきた場合、そのスイッチングをどうするかというお話がありましたが、あの問題は多分人間でも解けなくて今私が多分ここで知っていたとしてスワヒリ語を急にしゃべったら同時通訳の方が困ってしまうと思いますが、同じようなことが起こりまして、この部分が日本語なわけですよ。ところがこの日本語はＧＯＭＥＮはたまたま英語にあるわけですね。ところが文法的には不定詞がこの後ろへ来れませんので文法的には間違いなわけです。そうしますとシステムはもう混乱してしまいまして完璧なワード・フォー・ワード・トランスレーションになってしまいます。ここはワード・フォー・ワードのは切れ目なんですが、ハウがどのように、アバウトについてゴーが行くと、ナサイは英語にもないですからこのままと。こういうようなことが起こってくるわけです。

これもやはり別のフランス語が英語の中に混じってしまった場合ですが、先ほど申し上げましたように、辞書にないものはそのまま出てきますので、日本語側では何ですか、フランス語で、たまたまフランス語知っておりまして、ですからこう聞いたわけなんですね。こういうようなことが起こるわけです。

それから全く知らない語というか、ニューワードというきのう言葉がほかで出てきていたと思いますが、突然こういう語を入れられたわけですね。これが大文字で、そうすると当然これは辞書にありませんし、文法的にも訳がわかりませんから、そのまま日本語で出てくるわけです。この場合、ＬとＥが大文字であれば固有名詞だということがわかるんですが、小文字のままですから日本側で受けてる日本人は何のことかわからなかったわけです。英語でない全然別の言語かと思って聞いたわけですね。「What is lass cngval?」すか、こう聞いたわけです。怒られたわけですね、私の名前だと。九州などこの場合だと固有名詞だということがわかるわけですが、大文字ですから。ただ固有名詞も地名か人名かはわからないわけで、日本人にはわかりますが、ヨーロッパの方には、当然わかりませんで人だと思ったわけです。ここで日本人が言い直してこれを九州の出身ですということでこうなりますので、どうも地名であろうということが理解していただけたというような、こういう現象です。

それからこれはもうコンベンション、ランゲージコンベンションにかかわるかと思いますが、我々日本人は温かいコーヒーを飲みたいと普通に言います。温かいはこれ辞書で普通に引きますとワームです。ホットとまあ載ってないと思います。そのままリテラルに訳しますとワームコーヒーと。そうするとどうも英語の語感ではこれおかしいんだと思います。本当にそれでいいのかと聞いてきたわけですね。ホットのものじゃないか、こういうようなものというのはどう対処するか、コンベンションにかかわりますから、いわゆる辞書の問題なんですが、普通の意味での辞書の問題とはちょっと変わっているということですね。

それからきのう田窪さんから出たと思いますが、「ああ」とかいまの「ええ」とかですね、「あのう」とかいうようなダイアローグマーカーでしたっけ。ああいうものは対話、会話の中で、しゃべり言葉の中では非常に対応されて困るであろうと、ですがそれは実は我々文書翻訳もやっているんですが、文書翻訳の中でも同じようなのが出てくるわけです。フレーズストラクチャーのバイオレーションと言っておりますけれども、たとえばこういうような副詞句のようなものはかなり完全ではないですが、かなり自由度を持って文の中へポコンと入ってきたり、こういうような分詞構文ですともうどこへ入ってくるかわか

らないというようなものが起きまして、こういう文法を書こうとしますと非常にやっかいなものになるわけですね。話としましてはこうきてここへ飛んでつないでいってる方が素直なわけですから、そのほかにたとえば括弧で注釈がどこへでも入り込みます。こういうようなものというのは「ああ」とか「ええ」とか言うのと同じようなレベルのものだということです。

ちょっと全体はどんなぐらいの規模で実験をしたかということをまとめてみますと、8日間時差の関係で我々徹夜はなかなかできませんので、2時間、日に2時間です。78人の方とやりまして大体 1,429の発話をやりました。翻訳率は90%で、いまここで示しましたような例は大体10%ぐらいこういうような例が起こってきたわけです。文書翻訳の方ではほとんどあらわれないような問題というのが会話では非常に頻繁にあらわれるということも非常に難しい問題なんです。

一つには文章がフルセンテンスではなくって、フラグメントで、部分で来ると。これきのうも藤崎先生からお話がありましたけれども、たえとば日本人というのはコンベンションにかかわって何々ですからとかいうようなところで切ってしまって結論を言わないというようなことがありますので、こういうようなものを完全に扱えるようなパーサーにしておかないと複文だけで終わってしまう場合があるわけですね。そういうものを処理できないといけない。コンベンションとか言いましたが、そういうようなものの全体だと思います。簡単な方のコンベンションというのはたとえばイエス、ノーがはいと、いいえと一応リテラルにはなるんですが、わかりませんかという日本語がこう英語になるんですが日本人としてははい、わかりません。はいというのはリテラルにはイエスなんですが、これは英語で言えばノーでないといけないはずなんです。

それからテンスの方でもはい、わかりました、アイシーとそれからアイアンダースタンドと普通軽く言いますが、日本語では大体わかりましたと過去形、日本語にはテンスがないという言語学者ももちろんいらっしゃいますので問題がありますが、一応形式的には過去形ということ、わかりますとは言いますけれど、普通にはわかりましたと過去形で言ってしまうと。それからさっきのワームとホットの問題ですとか、このコンベンションの問題というのは非常に広範でかつシステマティックではないから難しい問題がございます。メタダイアローグを減らすには当然コンベンションをいかに扱うかということですね。スピーチアクトなども多分日本語のスピーチアクトと英語のスピーチアクトでは言語現象としては違ってあらわれてくるでしょうからそういうものをきちんと集めておかないといけないというわけで、我々文書の翻訳を長くやっているわけですけれども、そこではセマンティックのあたりまではかなりやってるんですが、やはりこのインタープリテーションと言いますか、インテンションレベルのアプローチを、こういうものを扱うためにはこれからかなりやっていかないとうまいダイアローグのシステムというのはできないと思います。以上です。

長尾（京都大学）　どうもありがとうございました。いろいろ御質問があるかと思いますが、皆さんのお話が終わってからいろいろ御質問を伺う方が時間的にはいいんじゃないかと思いますのでそうさせていただきます。

それでは次は田中穂積先生、東京工業大学の先生ですけれども、非常に有名な方でございまして自然言語処理全般について非常におもしろいことをやっておられます。

田中穂積（東京工業大学）

A Consideration on Parallel Parsing Algorthms

本日は少し言語処理に関するちょっと話でパラレルパージングアルゴリズムについてちょっと考えてみたことをお話してみたいと思います。

パラレルパージングなんですけれども、この青棒で示しましたものが文章だとしますと 1,000個なら 1,000個のプロセッサーが与えられまして1台1台全部違う文章をパーズすれば、これもパラレルパージングでありますし、それからプロセッサーの使用率、同時に走る全部同時に動かすわけですから非常に早い。これはあまりパラレルパージングについて考察をするといっても当たり前の話であまりおもしろいことはないわけです。いま私がここで言おうとしている、あるいは皆さんがパラレルパージングというようなことでお話するのは数個のプロセッサーがあってそれが1つのセンテンスを寄ってたかってパーズするという、それが終わったら次へいく。そういう意味ですからこちらの方で考えるアルゴリズムというのはもしかするとこちらよりも効率は悪くなる可能性はあるわけですが、オンラインでテキストがどんどん入ってくるようなもの、たとえばＡＴＲでやっておられるようなそういうスピーチがどんどん入ってきて、それを同時通訳しなくちゃいけないというような場面の場合には1つの文章を寄ってたかって、大勢でパラレルにパーズするということが必要になってくると思います。

その際パラレルパージングのアルゴリズムをどういうように評価するかということがいろいろあるわけですが、普通いろいろシュミレーションなんかやりましてどれぐらいのプロセスが同時に走るか、たとえば6台なら6台のＣＰＵが与えられてそのうちの実際は3台ぐらいが同時に入ってくるのであればこれはもう50%ぐらいのパラレルしか出てないからアルゴリズムとしてはあまりよくないんじゃないかとかいう、そういう議論があるんですけれども、それは本当だろうかということをちょっとここで考えてみたいと思います。

それはどういうことかと言いますと、結局1つの文章を大勢で寄ってたかってやるというようなことになりますと、たしかに同時になるべく多くのプロセッサーがみんなビジーで走れば一見いいように思うんですけれども、実はアルゴリズム、パラレルのアルゴリズムによって必ずしも非常にビジー状態が続いているんであるけれども、効率的な解析が中で行われるかどうかということについては疑問があるわけです。端的に言ってしまえば非常に無駄をするプロセスをどんどんどんどんつくり出して、それらのほとんどは死んでいくようなプロセスであるとしますと、どんどんどんどんそういう無駄なプロセスをつくり出してやってプロセッサーに割り当てて走らせればたしかに見かけ上は非常にビジー状態になるんだけれども、そういう無駄をつくり出すようなアルゴリズム、無駄なプロセスをつくり出すようなアルゴリズムというのは必ずしもパラレルパージングの観点からするといいアルゴリズムとは言えないのではないか。

それではどういうふうな指標でパラレルパージングのアルゴリズムを評価したらいいかというようなこともちょっと考えてみたんでそれについてお話したいと思います。

我々がちょっと実験してみたのは実はＣＭＵの富田さんがつくられたシェネライズドＬＲパージングアルゴリズムをベースにしていろいろ考察を進めたわけですけれども、これは富田さんの文献からとった文法なんですが、これは有名なＰＰアタッチメントのためのあいまい性が生じる、そういう文法で、この文法を使いますとＬＲパージングテーブルに非常にコンフリクトなエントリーにコンフリクトが起こってというようなそういう文法になってるわけです。

富田さんのアルゴリズムを富田法と呼ばしていただきますと、富田法ではコンフリクトが起こったときにそこをブレッドファーストにサーチを進めるということで、そこの部分に並列性があるということで、並列のアルゴリズムとしても非常に相性がいいものであろうというふうに考えられるわけで、それで実験に使ったわけです。富田法とはちょっと違うかもしれませんけれども、ＬＲ表の中にコンフリクトが起こったときに、コンフリクトごとにそれぞれプロセスを発生させてその先のパーズをやるというふうなやり方が考えられます。富田さんのアルゴリズムはそういう横型探索でＬＲ表にコンフリクトが生じたときに複数の可能性ですね、アクションの可能性が生じたときにナイーブなブレッドファーストのストラテジーをやっていくと、非常にその先同じことを何度もやるというようなことが起こり得るので、富田さんのパーザーはスタックの構造をある時点でマージするわけです。これがマージがない場合にはここでコンフリクトが起こるとそれぞれパッと進むと、それぞれバラに進むと。一見非常にパラレルこのあたり動いているように見えるわけですね。富田さんのアルゴリズムはここでコンフリクトが起こると、シフトレデュースコンフリクト、いまの場合起こっているわけですが、シフトレデュースコンフリクトが起こると、レデュースを優先させてシフトは待つというようなことをやります。シフトというのは一語一語処理すればシフトがシフト、シフト、シフトで処理がだんだんだんだん後ろに進むという、そういうもののわけですけれども、そのためにここで待ちが生ずるわけですね。

そういうことがあるわけですが、ちょっと次のスライドで御説明しますが。これも同じもんでんですが、先ほどのただ単にブレッドスファーストに開いたというやり方ですと行った先でこの赤丸で書いてあるところ、ここここうなってるこの赤でくくってあるここは全く同じことを両方の処理の過程でやってるわけですね。つまりここの部分は再計算になっている。それに対してこの赤で書いたところは富田さんのやつだとここに相当してるんですけれども、ここのプロセスというのは1回だけやればいいということで、そのためここで待ち合わせをして同じ計算二度やらないようにしてるという、そういうアルゴリズムなわけです。これとこれとどっちがいいかという、まあそういうことですね。

それからもう一つはいまうちでちょっと考えてる話をちょっとさしていただきますと、ここで待ちがあるのはちょっともったいない。可能な場合にはこの2つをどんどんバラに進めるようなアルゴリズムはないかということで、そういうアルゴリズムを我々のところでちょっといまやり始めてますので、それをやりますともうちょっとこれがこうタイム、時間ですね、時間もうちょっと短くて済むという、そういう具合になっていますが、とにかくこのアルゴリズムとこのアルゴリズムの違いは一見ＣＰＵがどれぐらいビジーかということで見るとこちらの方がビジーなように見えるんだけれども、同じ計算を何遍もやっているということで、必ずしもいいアルゴリズムとは言えない。そういうことになります。

それでじゃどういうメジャーで測るかということですが、1つは今先ほど示しましたス

ライドで丸で書いてあるところがプロセスがつくられているところだと考えてください。そうしますと最後の問題解決にとって有効に働いたプロセスをＰＶとします。しかしながら先ほど赤丸で書きましたように問題解決にはちゃんとコントリビュートはしてるんだけども、一応ＰＶとデュープリケートしている。そういうものをＰＤＶという。

それからいまの例ではそういうことは起こらなかったんですが、プロセスがつくられたんだけれども死んでしまうと、途中でエラーか何かでデリートされてしまうと、そういうのをＰＩＶとしますと、トータルのプロセスナンバーはこういったものの総和になると。それでこういったメジャーでパラレルなアルゴリズムを測ってやればいいんじゃないかということなんです。このＴというのはトータルの時間ですね。解析する。あるアルゴリズムを適用したときに問題が解けるまでかかったトータルの時間、それ掛ける本当に有効であったプロセス、ＰＶですね、ＰＶを全体のＰを全体のプロセスをこのＰＶの総和で割ったようなもの。それを掛け算してやる。無駄なことをやるアルゴリズムというのはこの数が非常に大きくなるんですね。この比率が小さくなるということで、こういったメジャーを使ってやるといいアルゴリズム、パラレルのいいアルゴリズム、悪いアルゴリズムということがもうちょっと評価できるんじゃないかということで、ここではそれをちょっと実験してみますと、これＰＰアタッチメントの１つのやつなんですが、ＰＰアタッチメントが１つですとそんなにファクター変わらないんですが、ＰＰアタッチメントが増えてくるに従ってこのアルファーの値が断然こちらの方が悪くなっていくということでアルファーが小さければ小さい程良いということが言えるんじゃないか。我々の場合はもう少し待ち合わせをやらないようにしてますので、これよりもアルファーのファクターがよくなるということで、こういったものでパラレルのアルゴリズムを測りながらいいパラレルプロセッシングのアプローチを探る必要があるんじゃないかという、そういうことでお話を終わらせていただきます。

司会　それでは次のＵＭＩＳＴのソマーズ先生にお願いします。

Harold L. Somers（ＵＭＩＳＴ）

前に言いましたように、私は２度話することを頼まれるというかなり幸運な立場にあります。さて、要約集をごらんになれば、私がこの２番目の機会で何を言おうとしていたかわかるでしょう。そして私が何を言おうと計画しているかはすぐにここにいる皆様にわかると思います。それは、音声による翻訳とテキストによる翻訳とはかなり違うということです。多くの意見の相違はありますが、実際は、この２日間でみてきたように、類似するところもあります。ですから、私は以前の話を続けます。

ここに実物大模型があります。模型が何を意味するか理解されていることを望みますが、これは私が今朝説明していたシステムとの活動を想像した模型です。これをあなたたちとともに調べる目的は、なぜ私たちが思考するのかについての考え方を提供することです。それは、そこに付随している特徴についての考え方であるかもしれませんが。それでこれにざっと目を通してみましょう。それは、３ページか４ページですので、それほど長くはかからないでしょう。

そこで、ちょうどＡＴＲでの筋書きのように、私たちにはユーザーがいて、会議の記録についての情報を欲しがっているとします。そしてそのユーザーはシステムを理解し、情報を欲しいと言います。最初の入力にタイプミスがあるとします。そのミスは明白なものなので、訂正も明白なので、システムは、簡単に「情報のことですね」と言うでしょう。その言葉が何を意味していたのか、システムには簡単に推測できます。そして、システムは「それにはいくらかかるか、支払い方法について知りたいですか」というようなことを言うでしょう。それはここで処理されるメニューのようなものです。ユーザーは「Ａ」と答えます。ここでは相当にロバストでなければならないことに注目して下さい。ユーザーはいろんな答を出すでしょう。大文字Ａ、小文字ａ、括弧のあるなし。「最初の物」「いくら」等あらゆる種類のものです。

それからシステムは、ユーザーの言葉で、依頼を生成するでしょう。その依頼は送信される予定のものですが、その確認を求めるでしょう。そして、この質問を送信するでしょうが、それは読まないでおきます。

それから、この答が返ってきて、それはまた翻訳されます。そしてそれはこんな場合であるかも知れません。一方の端には良く似たシステムがあり、同様のインタラクションを行ないます。答は、「一人当り35,000円かかりますが、来月は応募に40,000円かかります。」

この対話で納得する人もいるでしょうが、ユーザーが英国人かなにがしかで、「ポンドではいくらになるますか」と言うとします。するとシステムが、「これは予期しなかった質問だ。それを変換してみることはできる」と言います。ここでの考え方は、ユーザーが、想定されていた質問を逸脱したということです。そして、システムは「うまく変換できるかどうかわからないが、もしお望みならやってみよう」と言います。そして「会議の料金はポンドでいくらになるか尋ねてみよう」と言います。これは非常に賢いシステムであると思います。代名詞の照応を解決し、ユーザーに対し質問を明確にするからです。
それはとても大がかりなシステムです。ですから、「オーケー」ユーザーはやってくれと言います。ユーザーの答がいつも「オーケー」であるとはかぎりません。「はい」という言い方にはたくさんのいろんな言い方があります。

システムは質問を送信し、それから私は、これが日本人にとってたいてい自然な答であると思います。なぜなら、私たちはまだ対話モデルからはずれているからです。そのうえ答は、伝統的な機械翻訳方式で直接翻訳されなければなりませんでした。その翻訳はそんなに良い翻訳ではないと思います。確かにそれは、ユーザーが理解するためにはそんなに良くありません。

ここでの問題点は彼らが、日本のお金と言ったことです。そこでシステム、そこでユーザーは、それは何の意味だというでしょう。今度はユーザーが質問しているのです。さあ、彼はシステムまたはユーザーに質問して、明確化しようとしているのでしょうか。まあそんなことは問題ではないでしょう。システムはそれが何を意味するのか知りません。なぜなら、それがシステムのモデルの中に存在しないからです。それでシステムは、オーケー、それを質問として発してみようと言います。「お願いします。」

そしてここではもう一方の端のユーザーも、この対話を受け取ったときには協力しなければなりません。理解しなければなりませんが、一度言ったことをそのまま繰り返すのは、ばかばかしいのでやめにしましょう。もっと役に立つように言い換えてみましょう。このときにはこのユーザーは日本のお金で料金を払ってくれと言います。ここでこのユーザーは、質問の意味を取り違えたに違いないと言います。この人は料金をポンドで払いたいのではなかった。こっちの人は彼がそのつもりだったと思ったのですね。彼は、ただ35,000円がどのくらいか知りたかっただけです。高い会議かどうかを知りたかっただけです。それでシステムは、それを翻訳するけれど、やってみましょうかと言います。しかし、このユーザーはノーと言います。彼は困難な状況になったこと、すでに対話モデルからはずれてしまったことをさとります。彼はひょっとすると35,000円がどのくらいか知る方法を見つけるかも知れません。

それでこのユーザーが発する次の質問は、料金には何が含まれているのかということです。ところで、こう言うべきかも知れません。つまり、たぶんシステムはその時にはモデル対話に戻ったことを知り、何をするべきかに自信を持つのです。システムは質問をパラフレーズするでしょう。質問にあいまい性がなかったとしても、システムにとっては質問をパラフレーズすることが必要なのです。システムが一言一句そのまま繰り返すことは愚かでしょう。言われた言葉をそのまま繰りかえすのであれば、システムが本当に言われたことを理解したかどうかわからないからです。ですからパラフレーズすることが必要なのです。システムはこれを送信します。標準的な答があり、それは参加料金には議事録と歓迎会も含まれているというふうに翻訳されます。

ここでユーザーはこの用語の変化にちょっととまどうかもしれません。「参加料金は会議料金と同じですか。」あきらかにこれはシステムへの質問です。もう一方のユーザーに対してではありません。そしてシステムはこれをメタ対話として扱います。つまり、これを自身の辞書に対するあるシステムとして扱います。「それでは、私は情報処理学会の会員であるが、会員割引はありますか。」システムは疑問を出して、ちょっと質問をリフレーズします。

会話はとても調子良くすすみます。この時点では会話です。「このときには割引がありません。」「わかりました。支払いはどうすればよいですか。」そしてこの下の方にありますが、システムは、「わかりました」というようなものでさえ、彼が言ったのはこれで正しいのだろうかと確認するためにパラフレーズします。「会議料金をどのように払えばいいのでしょうか。」ユーザーは「はいどうぞ、送信してください」といいます。それでシステムはその質問を送信します。答が翻訳で返ってきます。ミスプリントがあります。すみません。それは私のミスでこのシステムのせいではありません。「支払いは銀行送金でお願いします。案内に記載してある我々の銀行口座に振り込んで下さい。期限は今年の末です。」

さてここは私の気に入っているところです。ユーザーは「案内をなくしました」と言います。そしてここで、ちょっといい会話が交わされます。そうでしょう。私たちはユーザーがこれで何のことを言っているのかわかります。ところが、システムは残念ながら理解していません。システムはそれほどには良くないからです。しかし、システムの対話モデ

ルの中で「私は案内をなくした」に非常に近いものがあります。それは、「私は案内を読みました」です。たとえば、多分システムは、「案内を送って欲しいですか。」「案内を読みました。」というようなことについて考えるでしょう。そしてそれがどこで得るのか想像できるでしょうが、これはまた別のサブダイアローグではないでしょうか。そうです、システムは「案内をなくしました」をパーシングしようとしているのです。システムに「なくした」という単語はありませんが、それに非常に近い単語を捜し当てたということにしましょう。しかしそれは送信するべきものか確認します。それは、送信するべきものではありません。

それで、それでユーザーは、「いいえ銀行について詳細を尋ねて下さい」と言います。ここでとても興味深いと思えるのは、まあもちろんこの対話を作ったのは私ですが、私はこれを生の状況で作りました。これがどのように見えるべきかということについてはあまり考えませんでした。そして、この時点でこのユーザーが、この終端のユーザーとしゃべるよりも、システムと話始める方が、自然であると思えたのです。彼が、「銀行について詳細を尋ねて下さい」と言うことに注意して下さい。

さあそれで、システムはその依頼、その質問を送信します。そして答を得ます。こういうことが続き、最後にシステムは「他に聞きたいことがありますか」と言います。ユーザーは「そうは思いません」と答えます。それは、「いいえ」ということです。なにか終わりの言葉を送信しましょうか。ここで私たちはなにか最終の対話計画、どうやって対話をすっかり終わらせるかというところにいます。しかしまだここにすこしインタラクションが残っています。なぜなら、システムはどのように、そしてどんな最終の言葉を言うべきか知りたがるからです。ていねいにしようか、それほどていねいではないようにしようかといったようなことです。私たちの作ろうとしているのは、こういう考え方のものなのです。

長尾（京都大学）　どうもありがとうございました。非常におもしろいお話だったんですけども、それでは最後にＡＴＲに飯田さんにお願いしたいと思います。

ＡＴＲでのこれから考えておられることじゃないかと思いますけれども。

飯田仁（ＡＴＲ）

Advanced Dialogue Translation Techniques: Plan-Based, Memory-Based and Parallel Approaches.

ＡＴＲの飯田です。先ほど午前中にＡＴＲの方の音声言語の翻訳システムの言語処理部分についてお話しましたけれども、あまり時間がなくて詳細にわたった御説明ができなくて申しわけなかったんですが、どうもこれからお話する内容も少し量が多くて十分にお話できなくなるんではないかとちょっと心配しておりますが、これからの私どもが考えている機械翻訳のアドバンストテクニックというんですか、そういうものについて少し紹介していきたいと思います。

それで一応これからの機械翻訳の目指すものとして３つ課題を考えておりまして、１つはハイクオリティーのトランスレーション、これは当然どんなものにでもまた要求されることになると思いますが、それから音声認識を統合したところでのトランスレーションという意味でロバストなトランスレーション、それからこのロバストなトランスレーションというのは次のハイスピードなトランスレーションというのにもまたかかわってくるわけですが、計算コストを非常に削減してやろうという狙いもあります。

この３つについてそれぞれどういうアプローチをとっていくかというお話をしていきたいと思います。それで１つはまず１番目についてはダイアローグを扱うということでありまして、これはあくまでも音声と言語が統合化したシステムという意味で対話ということになるわけですが、その点でやはり対話のためのプラグマティク、それを何らかの形で扱っていかなきゃいけないだろうと。当然プラグマティクと言われるものはダイアローグを展開していくための共調的な展開の方法、そういうものが何らかの形で対話の中に知識として潜んでるはずで、そういうものをうまく使っていかなければいけない。それから聞き話し手の間での待遇関係とかですね、待遇表現、そういうものを着目してゼロ代名詞の同定とか、そういうことをやっていこうと。それからダイアローグのアンダースタンディング全体にわたってはコンテキストプロセスティングという意味でも何らの形でそのプランベーストのインフィレンスを行っていく必要があるだろうというふうに考えています。

それからロバストなトランスレーションに関しては大変コンテキストに依存した表現が多いとか、それからイディオマティクな表現が多いというようなことで、それらをどのような意味的な内容を持ってるかということをきちんと計算して翻訳するということをやっていますと大変計算コストはかかります。そういう点でかなりの用例を集めてそのような

用例からそういうこのような問題についての翻訳の解を探索していこうというアプローチ、それがもう少し発展させればケースベイストな事例、事例に基づいた翻訳をしていこうというような話になっていくかと思います。

それからハイスピードなトランスレーションに関しましてはこれは当然のことながらいろいろな形で計算コストを下げていかなきゃいけないわけですが、物理的に並列的に早くさせようという意味で並列化処理を行っていこうと、そういう点で先ほどの田中先生からの御報告にありましたパラレルプロセッシングの手法がどう取り入れていこうと考えているんです。それであまり時間ありませんが、これらについてもう少しだけ中身についてお話します。

まず第1点目のプラグマティクスをある程度扱っていかなきゃならないだろうという点に関しましてはプランベーストな手法で発話の意図、そういうものを解釈していくというアンダースタンディングの方法を基本的にはとりますが、その中でここで4つのプランを上げております。ダイアローグプラン、それからドメインプラン、省略しておりますけれどもコミュニケーションプラン、それからインターラクションプランというものを上げておりますが、このダイアローグとコミュニケーションとインターラクションというのがここではプラグマティクスにかかるノリッジでありまして、まず簡単なところからお話しますとインターラクションプランというのが基本的にはクエッションがあるとそれに対してアンサーをする、さらにはそれにコンファーメーションの応答をするというような非常にローカルなところでの対話を構成するため基本的なノリッジというか、会話を実行するためのノリッジというものになっていく。それからコミュニケーションプランというのは会話を展開するためにたとえばある情報がほしいときにはどういう形の発話を、つまり相手に何か情報提供を求めるような依頼の表現をするというような一つはアブストラクトなレベルでの会話を展開するための知識であります。

それからダイアローグプランというのは対話全体にわたっての構成についてどのようにこの相手とコミュニケーションすればいいかというところの知識でありまして、当然のことながら電話会話であれば初めに「もしもし」というようなことから、「私はだれですが」というようなあいさつが始まって、そういう紹介ですね、そういうところから始まってそれから主題の導入があって、それからどうもありがとうございました。「失礼します」というような対話を終了するためのやりとりがあります。というものがこのダイアローグプランになります。

それでこのような枠を設定しておいて、まずはとりあえず非常に共調的なあるタスクを、あるゴールを達成しようとするダイアローグについて取り扱っていこうということです。ちょっと説明細かくはできませんが、それにはクエスチョンがありますと、それに対してのアンサーからそこから次へのアンサーというふうに続いていきますが、こういうインターラクション、クエスチョンに対してアンサーとか、そういうものが組み合わさっていく中で、ここのこのドメインプランというそれぞれのドメイン、領域に応じた目的達成のための行為の手順ですね、そういうものとの関係づけが行われて対話全体の構成がつくり上げられるということを考えております。

それでこの辺もあしたオープンハウスのときに幾らかのデモンストレーションができます。それから多分時間があまりないのでひとつ飛ばしますが、午前中の話にも少し触れたのですけど、ゼロ代名詞の同定等に関しましてはコンストレイントというように、と言われる条件をですね、うまく使いながら穴埋めをしていくというようなことをやっておりまして、一つの発話の中での待遇表現、そういうものに基づいてどちらが話し手と聞き手がどちらが目上であるかとか、そういう情報を使いながらエとかyouとかそういうものを補っていくという方法をとっております。それでこの方法はさらに発話と発話の間にもわたって情報を伝播させて処理するということもできるようになっています。

それから先ほどの2番目のロバストネスというところでお話しました点につきましてそのイグザンプルベースな機械翻訳ということで少し実験を始めてますのでそれについて少しお話しておきます。

まずダイアローグでなくてもかなりいろいろなコンテクスチュアルな表現というのは当然のことながら出てくるわけですが、ここでは一例としまして「の」表現について、名詞の名詞ですね、さらに「の」については「での」とか、「からの」とかそういうもう少し発展形はありますけれども、ここでは例に挙げましたのは京都での会議というものをきちんと意味解析をして翻訳するという過程ではなくて、用例に基づいてそれの翻訳結果、ここに最終的に結果出ておりますが、The Conference in Kyotoというふうにinという前置詞でそれら2つの名詞を関係づけるという翻訳結果を出す過程についてちょっとお話します。

ここでたまたまいろいろ用例を挙げたのが東京での滞在、香港での滞在、それから大阪の会議とか東京の会議、こういうものが用例として挙げられるわけですが、たまたま結果として英語の方は全部inになってるんですけど、ですからこの用例だけを見ると全部英訳inにすればいいかと思うんですが、決してそういうわけではなくて、たとえばVISA Cardで

の支払い、Registration feeの支払いなんかで、VISA Cardでの支払いなんていうのがあるわけで、このときの「での」なんかでありますと多分この英語正しいと思いますけれども、a payment with VISA card とか、そういうようにwithという前置詞をとるのが適切であるということで、いろいろな訳しわけが考えられるわけです。

それでここではこのようなインプットに対して用例との近さを測ってその近いものを最適な候補として翻訳結果として上げます。この場合は東京での滞在とか香港での滞在というものと非常に近いよという結果が出ておりまして、その距離の設定の仕方につきましては、一つはこのようなシソーラスと言われるような意味的な概念関係のハイアラキーを示したノリッジベースをもとにしまして、たとえば先ほどの東京での会議という会議というものとそれから東京とかそういう場所ですね、そういうものとの関係がどの程度離れているか、いまの場合ですと先ほどのは京都での会議というのが問題であったわけで、京都での会議、それから用例の方が東京での滞在というようなのがあったわけで、会議と滞在のかかわりがどの程度離れているかということを測るわけです。

それでこのようなシソーラスという意味概念のハイアラーキーを用意しておきまして、それぞれの意味的な関係の距離の近さ遠さをですね、ここですとたとえば3分の1づつに設定しておいて、会議とこの中に上がってこないものとしてはもう距離が1離れてて、それから相談というところで合って会議とか何かもう一つこの相談の下にミーティングとかがあった場合に、それは近さが3分の1であるとかというふうに設定しておきます。そのような計算をしましてここでは距離としてはそれぞれの概念ハイアラーキー上の距離とそれからそれぞれのウエートのサンメーションをとるという形で計算をします。

詳しいところはちょっと省略いたします。

それでもう1つだけちょっとお話をしないといけないんですが、パラレルプロセッシングについて解析手法をですね、少しどのくらい高速にできるのかということを実験し始めております。ここの部分です。

これは私どもが素性ベーストな構造に基づいた解析手法をとっているということで、その枠組みの中で考えたパラレリズムであります。それで基本的な考え方はまずレベルを一応3つに分けておきまして、この中では、解析の中で基本的な操作につきましてユニフィケーションということだったんですが、そのユニフィケーションという操作を並列化することによって高速にするという話、それからその前に当然シンタクティックなプロセッシングを並列化で高速にするという話があるわけですが、まだそこの部分、レベル2については実験をしておりませんけれども、レベル1のパーザー及びグラマーにですね、ディペンドした部分について並列化をまずやるということを行ってきています。

それでその実験結果をちょっと示しますが、パラレリズムの方法としましてはここに素性構造のユニフィケーションの様子を示しましたが、このように素性構造は2つあった場合にですね、それぞれユニファイしています。この場合デコンポジション、ＦＳ１、ＦＳ２というデコンポジション2つに可能性を分けてやっていかないといけない。こちらの部分でもということでこのユニファイ1のところでは4通りのユニフィケーションの操作を行っていくということになるわけですけど、それをここの部分ではこのようなユニフィケーションに対していまＦＳ１からを全部展開したものをですね、最終的にはデコンポジション3に関するような、そういう展開したような状態をつくっておきまして、ツリーとしましては全部同じになるんですけれども、それぞれの素性構造がそれぞれのツリーのノードになるというようなものをつくっておいてこれらを同時に計算するということをやらせます。

結果としまして私どものところではIntelのＩＰＳＣという並列マシン、Hypercubeでありますけれども、それを使っておりまして、メモリーが16メガバイト、ＣＰＵ持ってるというもので、大体ルールが23、それからレキシコンの数が53なんですが、そういうものについてセンテンスを少しやってみまして、大体32プロセッサーぐらいであるものによってはだから 200秒ぐらいかかっているのが、20秒ぐらいというか、1桁ぐらいのところまで落ちるというぐらいのところまで実験結果が得られております。

それでは以上でいまの話はそれぞれの技術についてお話しましたけれども、また後ほど時間がありましたら、全体をまとめてお話をしたいと思います。

長尾（京都大学）　どうもありがとうございました。天野さんから始まりましていろんな言語表現のとんでもない表現がたくさん出てくるというような話、それから田中さんのパラレルアナリシスの問題とか、それからSomersさんのシステムを入れた場合にどういうふうになるかというような将来のことであるとか、飯田さんがいまＡＴＲで今後どういうことを考えられていこうとしているかというようなお話いろいろあったわけでございますが、とにかくこの対話の翻訳という問題というのは非常に難しい問題を山ほど含んでいるわけでして、どういうふうな立場で見ていくかというのはなかなか難しいと思います。たとえばこの理論言語学的にきちっと研究していかなければならないテーマ山ほどありますし、

また工学的立場からどういうシステム化を考えていくかということも非常にたくさんの解決すべき問題があるわけです。それからきのうからきょうにかけてもいろいろなお話があったわけですけれども、言語的な研究をしている人から見ると音声における言語処理というのはまだまだ音声の中にいるんではないかと。それから多分音声の方々から見ると言語の方でやっている言語処理というのは音声の方ではなかなか使えないと、つまり音声と言語とはまだまだ離れている、融合がうまくいってないんじゃないかというような感じを抱いておられた方もおられたんじゃないかと思うわけであります。

そういった意味でいろんな観点からいろんな意見があり得ると思うんですけども、ここでフロアの皆様方にこの４人の先生方のお話についての御質問、御意見、あるいはもっとそれを離れて今後のこのテーマに関する方法についての積極的な御提案といったようなことをお伺いしたいと思います。

５時15分まであと25分間なんですけれども、ひとつよろしくお願いいたします。何か皆さん方御意見ございませんでしょうか。いかがでしょうか、何でも結構ですけれども。どうぞ。

問　　さっきの天野さんの翻訳の例なんかでパーズしそこなうとか間違えるところなんかで、随分メタ言語的な要素ですね、名前を聞いたり名前の部分とそうでない部分を切り間違えたりというふうなことがあるんですけれども、たとえば日本語なり英語なんかでそのメタ言語的なことを言うというような表現ですね、そんなところをもう少し研究すると一部分はその切り間違いとか誤解の部分が直るんじゃないかと思うんですけれども。

答　　そうですね、メタダイアローグのところはお見せしましたのは一部なんです。たとえば幾つかの例の中で先ほどの「Do you know kyushu」みたいな九州を地名でなく人名と間違えたなどというものはこれはどうしようもなく起こるものでしょうし、たしかにシステムの方で努力すれば直るものもあります。そこのあたりは我々もはっきり言ってまだこういうことはやったばかりで体系的に研究をしていません。大体体系的にできるのかどうかがわからないんですね。むしろ我々の工学にかかわるものから言語学者の皆さんに御意見とか御助言をいただきたいと思うんですけれども、こういうことに関しまして。どうでしょうか。

司会　こういう問題については例えば音声という立場から見るとおそらく何らかのイントネーションか何かのところでの特徴があって、メタ言語的なところは少し音声的な何かのキーがあるんではないかというような気もしますけれども、音声の研究でそこまできちっとまだまだ特徴が取り出せてないんではないかというような気もしますけれども、たとえばもっともっとこの音声の中でそういう言語的な特徴というのが潜んでいるのを逃しているんじゃないかいうようなことについていかがでしょう。

杉藤（大阪樟蔭女子大学）　大阪樟陰女子大学杉藤でございますが、私の調べたところでは「あの」とか「その」とか、いま長尾先生もその途中でメタ言語的なことをお入れになりましたけれども、そういうときのイントネーションに特徴がありまして、声を下げるとか上げるとかというはっきりしたインフォメーションが入らないで、皆中途半端な高さでそこでポーズが短く入る。そういうようなかなり特徴的なものが私の方では抽出されております。

長尾（京都大学）　どうもありがとうございます。

杉藤（大阪樟蔭女子大学）　そのお話でもうちょっと進むようでしたら次のクエスチョンはまた後で言います。
もう一つ問いがございます。

長尾（京都大学）　じゃどうぞ続けて。

杉藤（大阪樟蔭女子大学）　よろしゅうございますか。日本語のあいまいな表現と言いますか、文末が「ですけれども」とか、先ほども話題に出ましたけれども、こういうあいまいな表現というものを考えに入れるとき、たとえば英語とドイツ語とか、フランス語との関係とかというそういう翻訳談話的な研究で出てくる問題点とここへ日本語を扱うときに出てくる問題とで、何か大変やりにくい異質的なものがあるのでしょうか、そこら辺をついでにちょっと教えていただきたいなと思っておりますので、ついでに言わせていただきました。

司会　どなたかこの問題について御意見とかお答えいただける方おられませんでしょうか。どうぞ。

○　先ほどちょっと東芝の天野さんがおっしゃいましてすぐに杉藤先生が反論なさったんですけれども、きのうから皆さんに私が個人的にちょっとお話してた問題なんですが、そのあいまいという言葉はちょっと避けていただきたい。物言い方で我々の日本語の方がはるかに完結しない文章でとめることが社会慣習的に多い。それは完全な文章あるいはこういうこと自体が日本人の間のいまコミュニケーションとして少しエキスプリシットすぎるというかね、それでもって私どもの中でもそうですけれども、日本語で言うときはただ理由を言って結論を言わないで止めるというのが普通のあるいは礼儀の正しい言い方です。

ところが英語で言うときにはそういうことは許されない。慣習上許されないからほとんど必ず、もちろん英語での場合にも理由だけ上げて結論を言わないというようなそういうような社会的慣習の違いを難し過ぎると言えば難し過ぎますけども、どこまでというか、会員の申し込みあたりまでならいいかと思いますけれど、たとえば3日までに来なきゃならないんですけれども、それから先言わないのが普通なんです。したがって3日のホテルは予約してくださいなんていうことを一々言わないですね。そういうことの社会的慣習の違いを乗り越えて本当にハイクオリティートランスレーションというときに、それはじゃターゲットランゲレージのときには何かシテスムはよく考えて補わなきゃならん。逆に言うとそのターゲットランゲージが日本語の場合にはシステムはよく考えてドロップさせなきゃならないというようなことがあると思うんです。そういうことを言い出しますとＡＴＲは15年じゃ研究は終わらないと言われるかもしれませんけれども、実はですね、クオリティーのいいトランスレーションを目指すとすればそういうことのコンベンションというかね、それをもう少しやらなきゃならないではないかということを実は個人的に天野さんや杉藤さんに私はちょっと申し上げていたことかあるんですが、そういうところを事例毎にだんだん詰めていくとだんだん明らかになるのが多分工学的なやり方だと思いますけれど、そういうところでちょっとあるいはＡＴＲや天野さんというか、日本でそういう問題を一番扱っていらっしゃる方々の御意見をちょっと伺いたいように思います。

長尾（京都大学）　どうもありがとうございました。いろいろ意見があるでしょうと思いますのでこれはこの辺でとめますが、機械翻訳システムでは結構そういうなんは知らん顔をして落としてる分はもう既にたくさんありまして、それじゃ何か御意見ありますか。

問　その「crash course（集中訓練）」についての例は非常に面白かったのですが、それについての決定的な点は、あなたが crash course という言葉を知らなかったと認めるだろうということです。そしてこれはコンピュータのようなひどい質問かも知れませんが、まず crash course が一つの単語であって二つの単語ではないということにどうやって気づくかと言うことです。つぎには、 crash（衝突）がなにか自動車に関係したことであって、courseが何か競争コースを意味するものであるということを知っていて、どうやって crash course は合成的なものではないということを知るのでしょうか。

答　簡単に答えますと、「crash courseハ、ナンデスカ」と言ったのはもう一方にいる人間であったのです。「crash courseハ、ナンデスカ」と言ったのはシステムではありませんでした。談話を交わせるシステムと、人間が談話できるように援助するシステムとを、区別することはとても大切だと思います。個々に解決しなければならないいろんな違った問題があるのです。

○　メタ対話に戻りたいのですが、最初の仕事は、メタ対話を少なくとも3つの型に分類することが必要だということです。一番目はメタ対話というよりは主題の明確化のようなものです。たとえば、所得税とはなんのことですか。これは引用ではありません。所得税とはなんのことですか。しかし、「なんですか」とは「所得税の事例には何が該当するでしょうか」または「所得税についてどうすることが必要ですか」ということです。そして、2番目は言語学的な、意味論的な明確化です。この場合、入力から、つまりもとの文章から何かが引用されます。「これこれのことはどういうことですか」という場合がそうです。3番目は特に難しいのですが、人が対話能力を問題にする場合です。たとえば、公演で見たようなものをシステムにタイプ入力する場合などです。もし「私は何を尋ねることができますか」、または「なんについてのシステムですか」、または「どういう主題領域の質問に答えられますか」を自動システムにタイプ入力すればどうなるでしょうか。これはとても難しい問題です。なぜなら、システムを外側から見なければならないからです。最初の問題は比較的優しい問題です。2番目の問題は、メタ対話で私たちがだいたい意味するものです。3番目は何か機械自体の外部からの調査のようなものを意味します。と

ころで、まず問題になるのはこれらを区別しなければならないことです。なぜなら、それらにはまったく異種のメカニズムが必要だからです。

○　ここで非常に大切なことは、対話のシナリオを収集したり作ったりすることです。私たちは手近にある仕事を忘れ安いのです。そしてここでの想像的な模式対話はとてもわかりやすいものでした。そこで提案したいのですが、将来、話言葉の対話の、タスクをねらいとした構造の機械通訳における研究開発では、人間の通訳を通して二人の参加者が会話を交わす実験が、価値あるであろうということです。たとえば、一人はとても上手な人間の通訳で、日本語について深い知識があるが、英語についてはそれほどでもない。そしてまたその反対とがいて、良い通訳を一つの回線に、あまり良くない通訳をもう一方の回線に配置して、何が起こるか、そして、何が必要であるかを見るのです。なぜなら、そこには多くのことがありすぎてすべてを担当できないからです。それで、参加者のスタートに間違いがなかったか、想像していた通りの現象が起こっているかどうか、また最初想像だにしなかったような現象が起きているかどうかを見ることができます。

問　田中教授に質問します。私は教授のパラレリズムが、正確にはどの様に働くのかを理解できたかどうか心もとないのですが、教授の計算には重複がありますので、つまり、二つのプロセスが同じ計算を２回行なっていますので、この処理の流れのどこかで、コンピュータ処理の道で、分配を失っていることを意味します。ここで起き得る問題は、コンピュータ処理の道での分配が、パーシング演算で生成されているデータ構造と非常に強く結びついているということです。コンピュータ処理の分配を失うと、構造分配をも失います。そして、構造分配を失うということは、パーシングの次の、より意味論的な計算を繰り返さなければならないということです。処理の開始点でこの分配を失うことは、処理の後の部分で多くの分配と追加的な計算を余儀なくされます。このようにして、総体的な労力が増すという結果になり、それが問題になると思います。どうすればこれを処理できるかについて、なにか考えをお持ちですか。私が尋ねているのは……

田中　いまの御質問はプロセッサーの方のシェアリングとそれからデータ構造のシェアリングというのは必ずしも片方を立てようとすると片方は立たないというふうなことがあるという点が御質問だったかというふうに思うんですけれども、私が、オーバーヘッドプロジェクトで説明した例もまさにそうなってるんですね。富田法ですとグラフストラクチャスタックということでデータ構造のシェアリングをやるわけですね。それをやることによって同じ計算を避けるということをやるわけですね。それを使わない最初の 1と書いた方のやり方ですとデータ構造のシェアリングをやらないかわりにプロセッサーの方がシェアーというか同じこと二度やるようなことにあの場合にはなってるわけですけども、そういうことでなかなかパラレルを考えるときにどっちをどうやったらいいかという、たしかにトレードオフがあると思います。

私がちょっと提案したメジャーというのはそのあたりが両方考えられるようなメジャーになっているので、そういうメジャーの大小をそれぞれのアルゴリズムでやるということでコンプリマイズというか、トータルとしてどちらのアルゴリズムがいいかということはわかるんじゃないかいうことを言ったわけで、実際アルゴリズムとしてどういうものがいいかというのはもうちょっと御質問の方のおっしゃっているようなポイントをもうちょっと深く考えながらもうちょっと研究する必要もあるんじゃないかと思うんですけれども。

長尾（京都大学）　よろしいか。どうぞ。

荻野（筑波大学）　筑波大学の荻野です。きょうの自動通訳といいますか、のためには音声に関する研究というのが非常に大事だと思うんですけれども、たとえばきょうのお話でも何回か出てきましたが、音声認識をしてたとえば文字の列に直してそれから先に言語処理を行うとした段階で非常にたくさんの情報が落ちてしまうわけですね。たとえばポーズがどこに置いてあったかということによってある種の情報を我々は使いますし、イントネーションがどうであったかとか、あるいは場合によってはヘジテーションのことを伝えたりですね、あるいは場合によっては非常にどこをはっきり言いたいか、トピックがどうだとかですね、いろんな情報を我々伝える。もちろんしゃべることによっていまだれがしゃべっているとか、どういう気持ちでしゃべっているとかいろんな情報が伝わる。それを落とす必要ないわけですね。それをもちろん十分使って解釈しなくちゃいけないと思うんですが、いまの音声認識というのはそういう情報を取り出すことがどの辺まで可能なのかということと、それがどの辺まで利用されるのかですね、現状とそれから将来というか、これからの方向性あたりをできればこれは飯田さんあたりなんでしょうか、にお聞きしたいと思うんですが。

飯田（ＡＴＲ）　そうですね。私よりも音声の専門家の人の方がいいと思いますが、一言だけ言っておきますとプロソディーと言われるような情報を使ったいろいろ言語的なアンビギュイティーを解消する問題というのが随分いろいろなところで話はされてきてて、特にジェネレーションなんかにおいて、ジェネレーションというか、音声合成というようなところではそういう情報がかなりいろいろ使われてきてると思うんですけど、実際に音声認識から言語処理へそういう情報を渡していって使うというところまではまだあまり十分行われてないんじゃないかと思います。

それでその点について私の方からも音声処理の専門家の人にその辺の現状を聞きたいと思います。

藤崎（東京大学）　私はこれらの質問に答えるのにもっともふさわしい場所にいるわけではありませんが、たぶん、音声研究者の代表として、何かを言うべきだと思います。
まず、統語論上の、音響上の明示、または相関的な音声上の明示、統語論的な情報や談話情報の音声上の相関というのは、まだ形成の端緒についたばかりです。数年しかたっていません。そしてまだ研究段階です。非常にうまく、また注意深く発音されたデータについて初めてなされたというところなのです。そして、飯田博士がおっしゃったように音声合成の分野で即座に応用されています。しかし、それが音声理解ということになると、実際の話し手はずっと多様性に富んでいます。語調を整えるために中断したり、肯定的な語調で始めたのを、途中で気を変えて、質問の語調にしたり、いろいろです。私は話し手の実際の行為、つまり韻律がまったくあてにならないことを知っています。韻律の使用に関してはあまり強調しないでおきましょう。なかには役に立つものもありますが、ときどき、誤解を招くこともあります。統語論上のあいまい性を無くすことに関する限り、日本語でも英語と同じように研究があり、韻律の助けで統語論上のあいまい性が解決されることはありますが、構造的なあいまい性をなくすことはできません。ですから、あいまい性を無くすために韻律に過剰な期待をよせることはできないのです。もちろんある特徴を役立てることは可能でしょうが、大きな期待は持てないのです。

長尾（京都大学）　どうもありがとうございます。はい、どうぞ。

田中（姫路短期大学）　姫路短期大学の田中といいますけれども、きのうは音声認識に関して、きょうは言語に関してあったわけなんですけれども、いろいろな提案とか問題点とかそういうものがなされたわけですね。それのキーワード的なものをピックアップしてですね、そしてアプローチがどういうふうにあったかというふうな解決策とかやろうとしている、そういうサマライズしたようなものをですね、つくっていただく。これは長尾先生からあそこにいらっしゃる榊松社長さんにお願いできればいいことかもしれませんし、そういうものをつくっていただくと今後の研究がですね、非常にうまくいくんじゃないかと。それから我々研究者というのはリソースが限られているわけですから、これがどのぐらいまでにリサーチのレベルでは達してるとか、エンジニアリングのレベルではどうかとかですね、そういうものもあると。あまり方法についていい悪いはですね、言わない方がこれは将来決まることですから、そんなものができればいいんじゃないかと思いますけれども、いかがでしょうか。長尾先生。

長尾（京都大学）　どうもありがとうございました。たしかにおっしゃるとおりだと思います。どうぞ。

Church ここにいるのは田中博士です。あなたはふたつのパラレリズムモデルから始められ、直接迫るモデルを早く簡単に放棄され、難しいモデルについて続けられました。それで、あなたの議論は実時間を得るためのものだったと理解しております。さて、実時間問題を放棄しましょう。つまらないものが、実際は重要なものだとわかるのではないかと思います。事実、それは私の取り上げた、動詞－目的語の組に使うための物です。私たちは、何千万という言葉をパースしなければなりませんが、私たちのパーサーは一日に百万の言葉しかパースすることができません。それで、 Sunのようなワークステーションをたくさん探して、各ワークステーションに百万の言葉を送信するというのが、便利ではないかと気づきました。それで一週間に百万の言葉を収集して、一台の Sunを見つけ、百万の言葉を送信し、そしてまた百万の言葉というふうに続けていきます。それは直接な方法でありますが。

長尾（京都大学）　どうもありがとうございます。ほかにいかがでございましょうか。

それじゃどうぞ。

嵯峨山　私は音声の専門家なのであまり今回の話専門家ではないのでそのまた立場からお聞きしたいと思います。

１人の英語をきょうはさんざん聞いている立場としましては、英語を我々は学習してきてつまり外国語の獲得ということをみんなやってきてるわけなんですが、そういったプロセスというものを分析するとか、あるいは子供が言語を獲得していくような過程とかいうようなことなどはちまたではよく話題になるわけなんですが、今回そういうことが全く話題に出なかった。それは全然問題にならないのか、それとも忘れた方がいいことなのか、あるいは関係づけることはできないのか、できるのか、あるいは逆に言うといまこれから目標にするものは何歳程度の知能というものを実現しようという具合な言い方ができるのか、それとも知能としてはある面に対しては10歳だけれども、ある面に関しては20歳という具合にもう全くつぎはぎのようなものになるのか、その辺について人間との対比というのは今回全く話題にならなかったので、そういうことで何でもよろしいですからちょっとお聞かせいただきたいと思います。

長尾（京都大学）　非常にとんでもないおもしろい質問をいただきましてありがとうございます。飯田さんいかがでしょう。

飯田（ＡＴＲ）　私が用例主導な機械翻訳を一部導入していくと、ローカルなストラクチャーについてそういう話をしましたけれども、そのイクザンプルベーストのそもそものＭＴへの適用というのは長尾先生が先ほど言うの忘れましたけれども、始められている話であって、認識の違いはもしかするとあるかもわからないですが、そのときはまた長尾先生の方にコメントいただければいいと思いますけれども、私の方はそういうかなりイディオマティクとかコンテクチュアルなものを習慣的に身につけていくという過程を部分的にはやはりどうしても入れざるを得ないだろうと、それと言語的には構成的なものですね、そういうものも同時にそこで加味しなければいけないと。さらに一部言葉だけちょっと述べたんですが、事例ベーストな事例主導リーゾニングと言われるようなものに基づいた機械翻訳の方向というのも考えていく必要があろうと。それはどういうことかというと、こういう状況においてはこういうふうにトランスレーションをした方がいいとかですね、その状況、状況に応じたときのトランスレーションの方法ですね、そういうものを事例として蓄えていって、そういう知識を使いながらの翻訳を一部やっぱり目指すべきだろうということです。

長尾（京都大学）　どうもありがとうございました。時間がもう過ぎましたので、ここでやめたいと思いますが、このパネルをきちっとした結論を出さずに皆さんの御自由な意見を述べていただきましたのもこの問題が非常に難しすぎて簡単にまとめができるようなものではないということだからでございます。この２日間の会議を通じましていろんなことが提案されて非常に進歩があったかと思いますけれども、むしろ難しい問題点の方がどんどんたくさん浮き彫りにされてきたと、つまりやればやるほど難しい問題がどんどん増えてきていると、もっともっと人とお金と国際協力でもってやらなければいけないと、そういことがひしひしと感じられたんじゃないかと思います。それじゃ皆さんどうもありがとうございました。このパネルを終わります。

閉会の言葉

樽松

ＡＴＲのシンポジウムに参加された皆様に心よりお礼を申し上げます。そして講演された皆様には、素晴らしい内容をどうもありがとうございました。海外から、そして国内のいろいろな団体から、多くの方々が参加されました。こういった会議がまた将来もどこかで持てることを望んでおります。皆様ありがとうございました。