

コーパスを利用した談話・対話研究のための基礎的研究

(Towards the corpus-base researches on discourse and dialog)

土屋俊

(Syun Tutiya)

千葉大学文学部教授

(Professor, Faculty of Letters, Chiba University)

Abstract:

This study aims to set up an objective and reliable standards for the creation and utilization of spoken dialog corpora in order for those researchers with different backgrounds who share interests in human spoken dialog and thereby to explore the universal nature of human communication so that artificial intelligence research benefits from such interdisciplinary efforts and results. In the course of the study, the spoken dialog corpus has been defined to a reasonable extent and some interesting phenomena in human spoken dialogs, including overlaps, backchannels, turn taking and common grounding have been empirically explored with innovating observations and insights.

1. 研究目的

人間の言語的コミュニケーションの最も基本的な形態である音声対話の構造と機能について、言語学、心理学、認知科学、社会学、人類学、人工知能、音声処理などの各分野の成果をもとに学際的に総合的にアプローチすることによって、人間が相互に理解しあい共同して作業する際のメカニズムを解明することを目的とする。このために、電子化された先端的な音声対話・談話コーパスの利用によって客観性と厳密性を実現することが重要な目的である。また、この目的のために、国際的な標準化を推進することによって、言語の差異を越えたコミュニケーションの普遍性について考察することを第二の目的とする。

この研究から得られることが予想される成果は、音声対話コーパスを実際に作成し、それを異なる分野の研究者が活用できる環境を作成することであり、また、その活用によって、音声対話におけるさまざまな現象について新たな知見を得ることである。

2. 研究経過

1998年5月に千葉大学およびかずさアカデミアホールで国際ワークショップを実施して、この分野で先駆的な業績をあげている中堅、若手の国際的研究者約30名に呼びかけて音声対話コーパスの作成方法、その表現方法、その共有方法についての議論を行った。あわせて、この目的の重要性を一般に啓蒙することを目的として、公開講演会を開催した。このワークショップにおいて、討議した発話意図の観点からの発話の重層的なタグ付け、共通基礎付けの概念による話題共有化プロセスのタグ付けの基本的理解をもとにして、人工知能学会談話タグワーキンググループを中心として、日本語による音声対話(とくに課題志向対話)のタグ付けの方式を詳細にわたって規定し、それに基づいて実際の音声対話収録を

行い、タグを付したコーパスを作成した。このコーパスは、実際の音声対話研究に活用され、発表論文リストにみられるような研究を生んだ。以上の第一年度の研究成果を踏まえ、第二年度には、コーパスを活用した研究をさらに推進するとともに、人工知能研究における音声言語研究の位置付けを、言語学、心理学、認知科学、社会学、人類学、音声処理との関係で明らかにするべく検討を加えた。これらの点はまだ十分に公表可能な業績となっていないが、当初の予定をさらに越えて、理論的観点からの考察にも踏み込んだことは大きな成果であると考えられる

3. 研究成果

3.1 音声対話コーパスの標準化

資源の共有化には、さまざまな理由が考えられるが、例えば、音声対話コーパスについては、少なくとも2つの理由によって共有することが必要であると考えられる。

1. 自由対話という課題を含めて)さまざまな課題の遂行の際の対話を収録して、そのデータに基づいてさまざまな理論化、システム開発を行なうが、その理論やシステムのより妥当な検証・評価のためには異なる課題、設定、状況における対話のデータを利用することが必要となる。このような利用が可能となるためには、コーパスを共有することが必要である。
2. 対話に関する一般的理論化のためには、(自由対話という課題を含めて)さまざまな課題の遂行の際の対話を比較して検討することが必要である。音声対話コーパス作成の技術的、時間的、人的コストの膨大さを考慮するならば、質的に多様で量的に大規模な対話コーパスをそれぞれの研究者、研究グループが作成することは困難であり、また、非効率的である。したがって、各研究グループが作成したコーパスを研究者

コミュニティが共有する研究資源とする必要がある。

音声対話に関する情報として共有すべきものは、基本的には「音声」に関する情報と状況に関する情報である。対話の状況の映像による記録は、現段階においては状況の記録の一部として考えるべきであろう。なぜならば、ビデオカメラなどによる記録は、つねにひとつのアスペクトからの姿の記録であり、状況を完全に再現するものではないからである。同様に、音声を転記したとされるテキストも音声に関する情報の一部であり、すべてではない。あるいはむしろ、そこに非音声の情報(動作など)も記述されるとするならば、状況に関する情報も含んでいるので、転記テキストは複数の観点からのタグ付け(アノテーション)として特徴づけることができるかもしれない。

音声対話コーパスは音声ファイル、転記テキストファイル、談話タグファイルなど論理的には異なる形式を持つ複数の文書から構成されることになる。たとえば、音声ファイルは典型的にはストリーム・オブジェクトであるし、転記テキストファイルはおそらく階層的構造を持つタグ付テキストファイルとなるであろう。したがって、コーパスの表現方式を考える場合には、単一の文書ではなく、複合的な文書を交換する必要があることに注意しなければならない。典型的と考えられる音声対話コーパスが含むと考えられる文書には、さらに高次のタグ(たとえば、談話セグメント、発話意図、共参照など)を収容するものがある。これらさまざまな文書を統合的かつ整合的に複合し交換する方式を開発することは、個々のタグセットを考案することとともに十分な準備を行なう必要がある。

音声対話コーパス共有のためには、対話コーパスの記述に必要な情報をすべて表現することができる記述方式を確定し、その仕様を共有することが第一に必要である。その仕様は少なくとも、

1. 単位を設定し、単位を基礎として階層的に情報を組み上げることができること、さらに、それらに対し、特徴(タグ)を付与できること
2. 単位または単位を基礎として組み上げられた情報に、開始時間、終了時間、発話者といった属性を設定できること
3. 単位または単位を基礎として組み上げられた情報とは、直交する別の単位を許容できること
4. 音声以外の言語的、非言語的現象の発現が、音声発話との関係で表現できること
5. 話し手およびその他の収録の条件が持つ属性

を表現できること

という5つの条件を満たすものでなければならない。

第1の条件における単位は、音声対話の場合、通常、発話とされる。しかし、発話という用語についてはほとんど疑義がなくなっているのに対して、操作的な定義は依然として多様なものが並存しているという状況である。この点が明確に表現されているならば、研究者集団相互でコーパスを共有できる可能性が生まれる。発話を基礎として、韻律、形態素、発話の機能などの特徴的現象、談話構造、相互参照のような文脈的現象を特徴付けることができる。注意しなければならないのは、単位の定義を適宜変更する必要性である。たとえば、相手の質問に肯定的に答えて、さらに話しを続けるときには「はい、それでは」と表記され得る1発話を行なうが、この場合「はい」に対しては応諾というラベルを付与し、「それでは」という働きかけの表現に対しては、それ以降の発話機能の一部と考えたいとする場合が十分に納得できる。この場合には、発話という単位を細分化したタグ付けを行なうことが必要である。また、ひとつの依頼機能が、表面上複数の発話によって実現されるということはよくある。

研究をすすめるにあたって、音声データを直接利用するのではなく、音声を文字化したテキストを利用することも多い。音声データを基本とすると、文字化も一種の特徴付けと考えられる。文字化に関しては、どのような表記を利用するか(ひらがな、漢字かな混じり、ローマ字、IPA)、どのように文字を配列するか(付録A参照)について、選択肢が存在する。原理的には、表現と内容の問題は独立に扱うことが望ましいが、現実には、それぞれの研究目的に応じて、表現が選択されているのが現状である。

第2の条件は、第1の条件を実現するための必要条件であるとともに、対話が少なくとも2人の参加者によって遂行されることを考えるならば、それぞれの参加者による発話の開始と終了が独立に記述されなければならないことがわかるこの独立性による要請から、音声は話者ごとに別々のチャンネルに保存されることが望ましい。また、話者ごとに音声ファイルが1つであれば、ファイル先頭からの時間情報がそのアイデンティファイアとして利用するため便利である。

第3の条件は、最初に決めた単位とは別の単位を扱うための必要条件となる。たとえば、発話の重なりを記述する場合、誰のどの発話が他の参加者のどの発話と、発話中のどの時点からどの時点まで重な

っているかを記述できる必要がある。これにより、最初に決定した単位を、研究の途中で変更したい旨に対処できることになる。

第4、第5の条件は(大規模)コーパスの交換と共有が、従来のバランストコーパスによるアプローチの目指す側面を補完するために必要なものであり、言語学的、社会言語学的、心理言語学的に重要な情報を保持する必要があることを強調している。通常、対話収録は単一の対話の収録にとどまることなく、たとえば課題志向対話であれば課題ごとに、自由対話であっても時期ごとになど、複数の対話をまとめて行ない、それをまとめて記録する。これは、たとえば、安定した基準で転記することができる転記担当者を確保したり、機材の有効活用を図るなどの点で自然な判断であるが、このような傾向は、対話コーパスがたんなる集積ではなく、どのような状況で、どのような意図のもとになされた収録の記録であるかということを示明することの必要性を示唆している。同時に、収録の条件(環境、機材、課題など)と発話者の属性(性別、年齢、出身地、各対話におけるそれぞれの発話者の役割)、音声ファイル情報(サンプリング周波数、量子化ビット数など)をすくなくとも表現することが必要である。

3.2 音声対話について学ぶこと

今回の研究は、人間の音声言語によるコミュニケーションのすべての特徴を解明することを目的とするものではないが、人間的なコミュニケーションがもつ重要な特徴である「共同的行為」「協調行動」という側面を、経験的な裏づけをもって指摘することができるようになったと思われる。

人間的な会話には現象的に異なる二つの重要な局面が観察できる。すなわち、2人以上の会話参加者が、あたかも順番が決まっているかのように交代で言葉を発するという局面と、相手が話している途中であるにもかかわらず自分が話し始めてしまい結局相手に割り込むことになってしまうという局面である。われわれが観察してきた会話は、すでに指摘したように「課題志向対話」(task oriented dialog)と呼ばれる種類の対話であり、ごく気楽でカジュアルな会話とか、教室や法廷におけるように発言の手順があらかじめ決まっているような会話と区別されるものである。これらの課題志向対話(教示、協議、助言など)では、多くの場合、情報のやり取りが課題の達成にとって重要な役割を果たしているため、情報の正確な伝達を実現しやすいように会話参加者が行動し、発言することが期待される。正確な伝達という条件から連想されると思われる会話の様態は、

一方の人が喋っているときには他方が黙り、その他方が話す順番のときは最初に話していた人は黙って聞くというようなものであろう。しかし、実際にはそうでない場合がきわめて多いのである。

このことは、人工知能の機能という観点からどのような意味をもつであろうか。たとえば、電話でホテルの予約をするとき、ホテル側の対応を人工知能に任せることにしてみよう。予約を取りたい人は、ホテル側の人工知能に向かって、名前とか滞在期間とかを告げることになるであろうが、その際、人工知能がなんの相槌もうたず、すべての情報を伝え終わってからはじめて次の質問をするということであるとすると、なにか不自然な感じ、雰囲気か醸し出されるということになるのではないだろうか。相手が喋っていても、たとえば、相槌やそれに相当する言葉(たとえば、相手の言ったことを短くそのまますぐに繰り返すというような発言)を適切なタイミングで発するほうが会話の進行はスムーズになることが予想される。

問題は、何をどのタイミングで差し挟むかということである。今回の研究からは、まだそれに対して確定的なことを述べることはできないが、しかし、良質の音声対話コーパスが、本研究の直接間接の成果として作成されつつあるので、この点の解明はそれほど遠いことではない。実際すでに、相手の割り込みを誘発するような音の出し方については、有力な仮説が提案されている。すなわち単語を終わる前に、低い音からいったん高い音に移ってまた低い音に戻るといったような変化をつけると、相手はそれで話しやめるといった兆候がなくても割り込んでくるという法則性があることが知られている。

また、割り込みについての考察で明らかになったことは、「割り込み」という言葉遣いから想像される印象とは反対に、割り込み割り込まれる対話参加者同士の関係はけっして敵対的のものとはいえないということである。相槌についてはとくに、(たしかに個人差の問題もあるにせよ)親しい者同士のほうが頻繁に相槌を打ち合うということが知られている。また、相手の言っていることを理解していることを伝える表現も、共通した課題を解決しようとしているときには、相手が話し終わらないうちに発せられることが頻繁になる。このことは、共同で協調的に課題を遂行する場面での会話の特徴として指摘することが可能である。

このようにして人間のように会話することができる人工知能を作るためには、内容の理解だけではなく、会話の仕方を仕込んでやる必要があるのでは

る。仕込むべきことのなかには、どういふタイミングで話し始めるかについて自然な振る舞いができるようにするというようなことが含まれるのである。もちろん、これ以外にも会話する人工知能ができなければならないことは多い。しかし、この報告で述べたようなコーパスを利用する方法によって人間の言語についての知見を増大させ、そのような人工知能へ少しずつ接近しているということが出来る。

4. 今後の課題と発展

2000年の段階においては、すでに述べたように音声対話コーパスを作成、利用するための基礎が確立され、実際の研究活動が実質的な公表成果を伴いつつ生み出されつつあることが確認される。今後においては、このような成果を以下をもとにして、より一層、人工知能研究における応用を追求する必要がある。同時にこのような研究成果が、言語学、認知科学に対して及ぼす理論的影響について一層厳密に考察する必要があるであろう。前者については、すでに、科学技術振興調整費「話し言葉工学の構築」プロジェクトや文部省新プロジェクト準備作業に反映されているが、後者についてはさらに研究をすすめることによって、国際的にも独自の貢献が可能となるであろう。

5. 発表論文リスト

[1]市川, 荒木, 市川, 石崎, 板橋, 伊藤, 柏岡, 加藤, 熊谷, 樽松, 小磯, 田本, 土屋, 中里, 堀内, 前川, 村上, 山下: 談話タグ標準化の現状, 人工知能学会言語音声理解と対話研究会, SIG-SLUD-9703, pp. 41-48 (1998).

[2]岩渕淳子・榎本美香・大谷京子・嶋野健・土屋俊「日本語地図課題対話における相手発話中の発話開始現象」信学技報 (SP98-70) (1998)

[3]Ichikawa, A., M. Araki, M. Ishizaki, S. Itahashi, T. Itoh, H. Kashioka, K. Kato, H. Kikuchi, T. Kumagai, A. Kurematsu, H. Koiso, M. Tamoto, S. Tutiya, S. Nakazato, Y. Horiuchi, K. Maekawa and Y. Yamashita: "Standardising Annotation Schemes for Japanese Discourse", *Proceedings of the 1st International Conference on Language Resources & Evaluation*, pp. 731-738 (1998). (<http://www.slp.cs.ritsumei.ac.jp/dtag/>).

[4]Yasuo Horiuchi, Akira Ichikawa, "Prosodic Structure in Japanese Spontaneous Speech", *Proceedings of International Conference on Spoken Language Processing*, pp. 591-594, Sydney, December, 1998.

[5]土屋俊、堀内靖雄、石崎雅人、前川喜久雄「音

声対話コーパスの共有化へ向けて」人工知能学会誌 14 卷 2 号, pp. 261~272 (1999).

[6]堀内靖雄, 中野有紀子, 小磯花絵, 石崎雅人, 鈴木浩之, 岡田美智男, 仲真紀子, 土屋俊, 市川薫「日本語地図課題対話コーパスの設計と特徴」人工知能学会誌 14 卷 2 号, pp. 261~272 (1999).

[7]山下洋一, 小磯花絵, 堀内靖雄「音声対話に対する談話セグメントのタグ方式の検討」人工知能学会誌 14 卷 2 号 pp. 282~289 (1999).

[8]Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa and Yasuharu, Den, "An Analysis of Turn-Taking and Backchannels based on Prosodic and Syntactic Features in Japanese Map Task Dialogues", *Speech and Language*, Vol. 41 No. 3-4, 1999.

[9]Mark Core, Masato Ishizaki, Johanna Moore, Christine Nakatani, Norbert Reithinger, David Traum and Syun Tutiya, *The Report of The Third Workshop of the Discourse Resource Initiative, Chiba University and Kazusa Academia Hall, Department of Cognitive and Information Sciences, Chiba University, 1999.*

[10]Akira Ichikawa, Masahiro Araki, Yasuo Horiuchi, Masato Ishizaki, Shuichi Itabashi, Toshihiko Ito, Hideki Kashioka, Keiji Kato, Hideaki Kikuchi, Hanae Koiso, Tomoko Kumagai, Akira Kurematsu, Kikuo Maekawa, Shu Nakazato, Masafumi Tamoto, Syun Tutiya, Yoichi Yamashita, Takashi Yoshimura, "Evaluation of Annotation Schemes for Japanese Discourse", *Proceedings of ACL Towards Standards and Tools for Discourse Tagging*, pp. 26-34, June 1999.

[11]榎本美香・土屋俊「日本語地図課題対話における相手話者発話中の発話開始現象(2)」(1999) 人工知能学会 言語・音声理解と対話処理研究会 SIG-SLUD-9902-7, (1999)

[12]土屋俊「言語行為論の終焉から語用論の将来へ」日本語用論学会シンポジウム, 京都 (1999).

[13]Syun Tutiya, "Philosophy of language in light of spoken dialog corpora" (in Proceedings of Symposium for Spontaneous Speech Engineering), Tokyo, February 2000.

[14]Mika Enomoto and Syun Tutiya, "Overlaps and Interruptions: Towards a hearer's model of turn taking", *The Proceedings of the 4th Workshop on Semantics and Pragmatics of Discourse and Dialog*, Gotenburg, June, 2000.