

Verbmobilプロジェクト

— 内側から見て —

藤波 努 (fuji@jaist.ac.jp)
北陸先端科学技術大学院大学 知識科学研究科

Verbmobil プロジェクトは多言語間(独・英・日)の音声翻訳システムの研究・開発を目的としたドイツの国家プロジェクトである。本稿ではプロジェクトの現状を報告するとともに、これまでに得られた知見を述べる。プロジェクトの第一段階(1993-96)で開発されたプロトタイプシステムは74%の割合で正しく文を翻訳し、これは音声翻訳システムの性能としては良い方と思われる。成果が得られた理由のひとつとして、28もの研究機関(22大学と6企業)が幅広く協力し合ったことが挙げられる。この点を考慮した上で、音声処理技術の観点から見て重要なシステムの特徴とその実現過程を考察する。

Verbmobil Project

— An insider's view —

Tsutomu Fujinami
School of Knowledge Science
Japan Advanced Institute of Science and Technology, Hokuriku

Verbmobil is a long-term, German national project that aims to develop a speech-to-speech translation system applicable multi-lingually to German, English, and Japanese. We review the current status of the project to see what lessons we can learn. The prototype system developed in the first phase (1993-96) showed a good performance; 74% of input utterances are translated correctly. This is an outcome of wide ranges of collaboration between 28 institutions (22 universities and 6 industrial partners). We see some important features of the system in terms of speech technology and examine how they were realized.

1 はじめに

筆者は1995年10月より1998年3月までの2年半に渡って独 *Verbmobil* プロジェクトで働く機会を得た。本稿ではプロジェクトへの参加を通して得た知見を述べる。なお筆者は自然言語処理、特に意味論を守備範囲とするものであり、音声処理技術については利用者としてその片鱗に触れたに過ぎない。以下では、音声処理技術に

ついて素朴な感想を述べたり、あるいは誤った理解に基づいた私見を述べている恐れが多々あるが、部外者ゆえの誤ちとしてご容赦いただければ幸いである。

以下、第2章では第三者の立場で見たプロジェクトの概要を数字で説明し、第3章では音声処理研究という観点からプロジェクトで開発したシステムについて述べる。第4章では1997年から新たに始まったプロジェクト第2期について

解説し、第5章ではプロジェクトで働いた者として、感じたことをまとめる。最後に第6章ではプロジェクトの意義について考察する。

2 外から見た Verbmobil

Verbmobilプロジェクトは多言語間(独・英・日)の音声翻訳システムの研究・開発を目的としたドイツの国家プロジェクトである。音声翻訳システムの研究・開発が直接の目標であるが、プロジェクトの遂行を通して、ドイツのこの分野での技術レベルを世界一にすることを目指している。Verbmobilプロジェクトの概要を数字で見ると次のようなことがわかる。

期間 1993年より96年まで(第1期)、および
1997年より2000年まで(第2期)

予算 第1期 6500万マルク(約52億円)、第2期
5000万マルク(約40億円)を国が出資。
また諸企業からの出資をみると、第1期は
3100万マルク(約25億円)、第2期は2000
万マルク(約16億円)となっている。

参加機関 [第1期] 22の大学と6企業が参加した。
大学はほとんどが旧西ドイツにある総合大学である。
企業としては、ダイムラー・ベンツ、フィリップス、
シーメンス、IBMなどが挙げられる。機関としては
28であるが、細かくサイトごとに勘定すると100
近くになると言われていた。プロジェクトで働いて
いる人間は助手やアルバイトまで含めると200人
を越えていたという。

[第2期] 参加大学は16に減った。また企業からの
参加も4社となった。大きな変化はIBMが抜けた
ことである。このように規模は第2期では縮小
している。

システム 第1期で開発されたシステムは独英部
分について、70%以上の正確さで翻訳をお

こなう。システムは43のモジュールから成り、
辞書は約2,500語を収める。一発話を翻訳する
のに要する時間は平均0.7秒である。

以上がプロジェクトの概要であるが、かなり大規模なプロジェクトであることがおわかりいただけると思う。ヨーロッパの通貨統合を控えて政府が緊縮財政路線を押し進めるなか、単一の研究プロジェクトに数十億円単位で出資するのは異例ともいえる。

3 研究者から見た Verbmobil

3.1 システム概要

第1期において開発されたシステムの特徴を一つだけ挙げるなら、深い処理と浅い処理を組み合わせていることであろう。図1にシステム構成の一部を示す。入力された音声は音声認識モジュールと韻律解析モジュールに渡され、その後、形態素解析にかけられる。ここまでは深い処理、浅い処理とも同じであるが、浅い処理では統計的手法を使って翻訳文を合成する。一方、深い処理では構文と意味を解析し、得られた意味表現を変換し、変換された意味表現を元に文を生成するという違いがある。このように二つの異なったサブシステムを組み込むことにより、意味解析が必要な時は深い処理を行なうが、深い処理を必要としない簡単な発話(例えば挨拶や都合の良い日を尋ねる場合の定型的な言い回しなど)は浅い処理で済ませることができる。

システム稼働時の平均的な処理を分析した結果、各モジュールが全体の処理に占める割合は次の通りであることが判明した: 音声認識(38%)、韻律解析(16%)、構文・意味解析(25%)、談話分析と対話分析(14%)、意味翻訳(3%)、文生成(3%)。音声認識と韻律分析を合わせて50%以上の時間が音声認識処理に割かれていることがわかる。こ

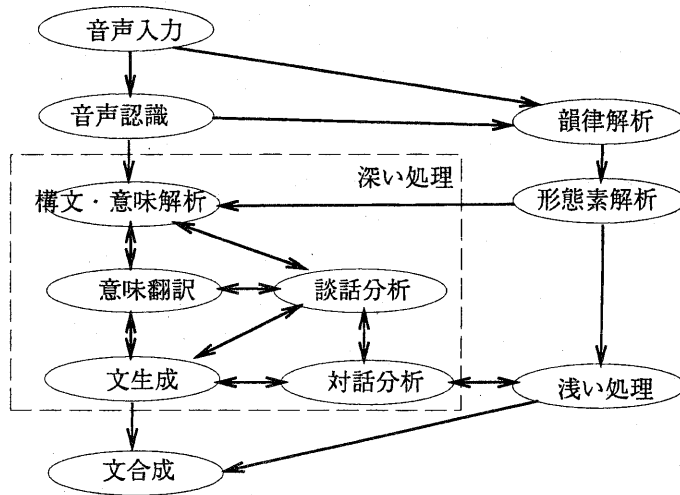


図 1: Verbmobilシステム構成 (部分)

の数字は様々な解釈の仕方が可能であるが、談話分析や対話分析、意味翻訳や文生成など、本来時間がかかる処理が全体に占める割合が少ないのは、これらのモジュールではまだルール数が少く、すぐに処理が終了するためと考えられている。

は、精度 92%で正しく特定、構文・意味解析への寄与については、候補数を 96%削減する(つまり 4%まで可能性を絞り込む)などの効果が得られている。

3.2 音声認識部分の性能と役割

音声認識部分を担当しているのは Weibel 教授率いる Karlsruhe 大学のグループである。認識率は公式には 73.3%とされている。筆者は音声認識の専門家ではないので、この数字をどのように判断してよいのかはわからないが、話者を限定せず、周りの騒音にもあまり左右されないことなどを考え合わせると良い性能を示しているのではないかと思われる。

また韻律解析については、Erlangen-Nuremberg 大学のチームが担当しており、文の切れ目を推定したり、構文・意味解析結果の候補を絞り込むために活用されている。文の切れ目について

3.3 韻律情報の利用

韻律情報の利用についてはまだ研究が始まったばかりであるが、ドイツ語を翻訳する際に有用な場合をいくつか説明する。

文ムードの特定 ドイツ語では疑問形と命令形が同じ構文になるという問題がある。話し言葉の場合は、疑問形は抑揚を最後に上げることにより、命令形の場合は下げることにより区別できるが、これは韻律情報なしには判定不可能である。例えば、

- (1) Machen Sie einen Vorschlag
make you a proposal

のような文は、もし最後に抑揚が上がってれば、“*Will you make a proposal?*”と疑問文に翻訳しなければならないし、逆に下がっているなら、“*Make a proposal!*”と命令形に翻訳しなければならない。

このような韻律による文ムードの違いは構文解析の一部と考えられる。開発されたシステムでは、実際に韻律情報を利用して、文ムードの特定に役立っている。

フォーカスの特定 ドイツ語にはストレスが置かれるか否かで意味が変わる語があり厄介である。以下に例を挙げる。

- (2) Können wir noch einen Termin
Could we a appointment
ausmachen?
arrange

問題となるのは、‘*noch*’にストレスを置くか否かである。もしここにストレスが置かれたとすると、英語の‘*another*’のような意味になるので、“*Could we arrange another appointment?*”のように訳さなければならない。一方、ストレスが置かれなければ、‘*noch*’は日本語でいう「あの一」のように口調を整えるために挿入された語と考えられるので、“*Could we arrange an appointment?*”のように訳すべきである。

この問題には談話構造の問題なども関係してくるため、まだ解決していない。とりあえずの処置として、ストレスが置かれた‘*noch*’と置かれていない‘*noch*’を別々の辞書項目として登録することにより対処しているが、あくまでも応急処置であり、理論上の問題がある。意味解析まで深く行う立場からは、フォーカスとその背景の情報を意味あるいは談話のレベルで記述しなければならないが、フォーカスについてはよくわからないことが多く、実用化できるほどの研究成果はまだ上っていないというのが現状であ

ろう。

同様の問題は日本語にも存在し、擬似名詞とされているものの扱いが難しい。例えば「など」の用法を見てみると、「来週の金曜日などいかがでしょうか」といった場合は、「など」にストレスが置かれ、「金曜日」に焦点を当てる特別な働きをしていると思われるが、「お茶やコーヒーなどが準備してあります」と言う場合にはそのような効果は観察できない。(ストレスが置かれることもない。) 少くとも独・英・日の三ヶ国語に共通する意味あるいは談話の構造を研究する必要があると考える。

4 Verbmobil第2期

Verbmobilプロジェクトが第2期へ進むのかどうか、関係者の重要な関心事であった。ドイツ国家としては、科学技術の研究費(さえ)も大幅削減という方針であり、噂では半数以上のプロジェクトは打ち切られるということだった。またIBMが第2期プロジェクトには参加しないことを明言したため、「産業育成のため」という大義名分も立て難くなった。多くの人が気を揉む中、第1期終了間際の1996年12月初旬になってようやく「続行」のサインが出た。これは何事も計画性を重んじるドイツ政府(あるいはドイツ人気質そのもの)にとって、相当の決断だったためと思われる。

第2期プロジェクトの特徴を要約すると、浅い処理の重視、大規模化、合わせ技の多用といった点が挙げられる。第1期ではIBMが「深い処理」アプローチを推進していたため、意味解析の部分に多くの労力が割かれていた。しかし第2期ではダイムラー社やシーメンス社のように、(良い意味で)言語処理に思い入れがない機関が主導権を握ったため、いかに統計的処理を利用して軽く早いシステムを構築するかに関心が移った。また政府からは実用化への圧力がか

かり、第1期で開発したシステムを実用可能なレベルまで洗練すること、より具体的には辞書を一万語レベルまで拡張するようにとの要請を受けた。また第1期で開発したシステムとの(見た目の)違いを出す意味もあって、マルチメディアと音声翻訳の統合や、多人数での会話を翻訳すること、Verbmobilサーバーを立ち上げて、インターネットやISDNを通して広くサービスを提供することなどが目標として挙げられるようになった。

第1期から第2期への変化を要約するなら、それは「産業への寄与」を重視している点であろう。すでに第2期プロジェクト発足にあたって政府からは、第2期終了時点で産業界からの参加が増えていない場合はプロジェクト打ち切り、と通告されている。また意味解析のように労多くして益が少ない研究よりは、音声認識のようにはっきり成果が確認できる分野へ資金を投入する形になっている。

5 内側から見た Verbmobil

Verbmobilプロジェクトは巨大であり、広大である。大学と企業合わせて200人もの人間が働いていることもあるが、参加サイトの位置から見てもドイツ中に広がっており、打ち合わせをするにも特急電車で片道数時間かけて旅しなければならない。大抵の場合は泊まりがけの会議となる。そのような打ち合わせが月に一度はあるので、研究費用に占める出張費用の割合が多過ぎるのではないかと底辺の研究者同士ではよく愚痴をこぼしあった。

仕事の進め方を説明すると、大きく半年単位で動いている。これは10月と3月の終わりに全体会議があり、そこで開発中のシステムをデモンストレーションしなければならないからである。全体会議はプロジェクトに参加している者が上から下まで半数以上集まり、広い会場で100人

近い人数で2日間に渡って討議する。各モジュールの開発責任者が30分くらいずつ自分たちの仕事を説明し、他部署から質問などを受ける。

このデモに向けて、デモ直前の2ヶ月くらい前から準備が始まる。2週間おき位に、まず音声認識モジュールが完成、それを受けて構文解析モジュールが完成、それを受けて意味解析モジュールが完成、以下、2週間の間を取って、意味翻訳モジュール、文生成モジュールが出来、結合して全体のテストとなる。ただし、これはあくまで理想で、一部では当然のことながら締切を過ぎても物が出来てこないことがあるし、後になって修正が入ったりする。したがって、常日頃から自分に関係する部署で働いている人間が何を考えているのかよく把握しておく必要がある。ほとんどの場合、仕様書は存在しないからである。

このやり方の一番の問題点は全体の進行を統轄する責任者がいないことである。一般にドイツ社会で仕事をする場合は、強烈な指導力の下、軍隊的な規律を以って目標に向けて一丸となり計画的に仕事を進めていく。ところが Verbmobil プロジェクトの場合、各サイトがそれぞれ権限を持っているため、「命令」することができない。言ってみればどの部署も「自由に」研究・開発できる「権利」があるのだ。したがって自分に関係するモジュールの開発に対しては「お願い」という形でしか意見を出せない。そのため実際に物を作り出すと、工程の前の方で独自に仕様を変更することもあり、現場は大混乱する。これでよく物が出来上がるものだと不思議だが、どういうわけか物は出来上がる。全体が何とかまとまるのは、研究資金を得るという大目標があるためであろう。プロジェクトが失敗して資金を引き上げられてしまっただけなら元も子もない。

いろいろと問題点や難しい点はあるものの、研究者相互の交流を促進するという点で、プロジェクトがドイツの科学技術に大きな益をもたらしたということでは、誰もが同意している。通常大学間、あるいは企業と大学との間では研究交流

といえるものは皆無である。しかし *Verbmobil* プロジェクトで何度も会議を繰り返すうちに、大学研究者にとっては、企業からの要求が何か、その背後にある事情などもわかってくるし、他大学の研究者の話聞くうちに周辺技術の動向などもわかるという利点がある。企業としてみれば、システム全体を構築するために必要な投資を軽減できるという利点の他に、自社研究者の教育という側面も重視しているようであった。

6 *Verbmobil*の意義

さて巨額の資金を投入してきた *Verbmobil* プロジェクトの意義は何であろうか？一言で言うならば、それは音声翻訳という難しい問題に対して様々なアプローチを試み、時間と労力を費して何らかの結果を出したということであろう。プロジェクトでの研究がすべて上手くいったとは正直考えられないが、現時点で考えられるアプローチは(人的・物的)資源が許す限りほとんど全て試してみたと言ってもそれほど間違いではなく、(語弊はあるが)失敗した研究でさえ、後からこの分野に入ってくる研究者には教訓として役立つのではないかと思われる。

また政治的な視点から見ると、*Verbmobil* プロジェクトはヨーロッパにおけるドイツの科学技術戦略の一部とも考えられる。ヨーロッパのもうひとつの雄といえばフランスであるが、音声・言語処理技術に限ればドイツの方が圧倒的に進んでいる。今後、ヨーロッパ各国の関係が通貨統合を経てますます緊密になっていくにつれて、自動翻訳技術は社会的基盤の一部として必要不可欠な技術になっていく筈であるが、その分野でフランスに大きく差をつけておくことが重要とドイツ政府は判断したのではないかと推測する。将来ヨーロッパ政府を樹立した時、ドイツの技術を使って情報インフラを整備することになれば、この分野で主導権を握ることができる

ためである。この点で、なぜ独仏翻訳ではなく、独・英・日という言語を選択したのかを考えると興味深い。第2期が始まる時、この三ヶ国語以外にも対象を広げようという議論があったが、案として出てきたのは韓国語や中国語であった。

参考資料 公開されている文書(論文も含めて)はほとんど全て電子化されており、インターネットを通じて入手可能である。

- *Verbmobil* ホームページ：
<http://www.dfki.uni-sb.de/verbmobil/>
- プロジェクト第1期概要：
<http://www.dfki.uni-sb.de/verbmobil/VM.English.Mail.30.10.96.html>
- プロジェクト第2期概要(ドイツ語のみ)：
<http://www.dfki.uni-sb.de/verbmobil/Vm.Info.Phase2.html>
- プロジェクト・レポート
<http://www.dfki.uni-sb.de/cgi-bin/verbmobil/htbin/doc-access.cgi>
- 音声翻訳における韻律情報の利用について
Christian Lieske, Johan Bos, Marin Emele, Björn Gambäck and C. J. Rupp. *Giving Prosody a Meaning*. In Proceedings of EuroSpeech '97. September 1997. 以下から参照可：
<http://www.ims.uni-stuttgart.de/projekte/verbmobil/Publications.html>

謝辞 本稿を執筆するにあたり、Martin Emele氏(IMS, Stuttgart 大学)からプロジェクトの第2期に関してご教示いただいた。また Reinhard Karger 氏(DFKI)より *Verbmobil* のビデオテープを無償で送って頂いた。両氏に感謝いたします。