

ザールブリュッケンにおける機械翻訳国際会議報告

坂本義行
電子技術総合研究所

アメリカ、アジア、ヨーロッパにおける機械翻訳の現状に関する第一回国際会議と機械翻訳システム・フェアがザールブリュッケン市において、1986年8月20日から22日の3日間ザールランド大学応用情報学研究所の主催で行われた。この会議に出席し、Muプロジェクトについて発表し、デモを行う機会を得たので、会議の様子について、簡単に報告します。なお、この会議の目的は、アメリカ、アジアならびにヨーロッパにおける機械翻訳システムの現状を展望することである。発表されるシステムは、開発中または完成されたばかりのものである。会議では、完成された商用のシステムおよび開発中の新しい実用化システムについて広範囲なシステム紹介をおこなう。フェアでは、種々のFRONT-ENDシステム、種々のアプローチによる機械翻訳システムならびに機械支援型機械翻訳システムのデモを行うことである。

The Report on International Conference on State of The Art in Machine Translation in America, Asia and Europe and MT-System Fair

Yoshiyuki SAKAMOTO

Electrotechnical Laboratory

1-1-4, Umezono, Sakura-mura, Niihari-gun, Ibaraki-ken, JAPAN

The First International Conference on STATE OF THE ART IN MACHINE TRANSLATION in America, Asia and Europe and MT-SYSTEM FAIR was held at Saarbrücken during August 20th - 22nd 1986. We attended at this conference and read the paper on Mu-project, then I briefly introduce the outline of this conference.

The objective of the conference is to give an overview of the current state of MT-systems in America, Asia and Europe. The systems presented are under development, or recently finished. The conference presented a wide range of systems, well established commercial systems, new commercial systems under development as well as new and old research systems. The fair demonstrated the different possibilities of front-end systems, various types of MT-approaches and machine assisted systems.

1、概要

西ドイツの西南フランス国境近くザール川辺にあるザールブリュッケン市において機械翻訳システムに関する第一回の国際会議がひら開かれました。会議は、真夏の暑い盛りにおこなわれましたが、大陸性の気候で湿度が低く最適で朝晩肌寒い位の陽気でした。会議場は、山の上のザーランド大学に近いBürgerhausと呼ばれる公民館でおこなわれました。会場は1つで、議場の周囲をシステムフェア用のboothが取り囲んでような形をとっていました。このなかで、12のシステムのデモが行われました。

参加者は、12カ国から110名程になり、主催国である西独の74名を除くと日本が第2位で8名もの参加者があり、日本での活動を反映しているかのようなようでした。国別の参加を第1表に、デモに参加したシステムを第2表に示しました、

第1表 国別の参加者数		第2表 デモに参加したシステム	
西独	74	SYSTRAN (MT, commercial)	
日本	8	LOGOS (MT, commercial)	
ベルギー	5	SEMSYN (MT, research)	
オランダ	3	CAT of Ericsson (MAT, commercial)	
フランス	3	BAPCOMP (MT, research/commercial)	
ルクセンブルグ	3	Mu system (MT, research)	
デンマーク	2	EUROTRA-DK (MT-parser, research)	
米国	2	SPES (MT-analysis, research)	
スイス	1	PLAIN (MT, research)	
フィンランド	1	MARIS (MT-application, research)	
韓国	1	SUSY II (MT-parser, research)	
東独	1	ASCOF (MT-analysis, research)	
計	104名		

講演は、3日間で第3表にしめすような合計18件の発表が行われ、第1日目は、ドイツを中心としたMTシステムに関する発表、2日目は日本語を中心としたシステム、3日目は、EUROTRAに関する発表が中心をなしていた。このように、機械翻訳システムを一堂に会して、報告し、比較、討論する試みは、大いに興味があるところであり、お互いのシステムの発達に寄与するものであり、それがこの会の目的であるという話もあったが、対象とする言語対が異なり、その利用の仕方がことなるシステム比較することは困難で、今後課題を残した。ベンチマークテストのように、同じ文を翻訳させて、その結果を比較させるほどには機械翻訳システムの成熟度は高くない。最終日に、デモの担当者が、お互いに提出した文を翻訳させる試みがなされたが、結果を示したのはMuシステムだけだった。

全体の4割に当たる5件が日本語をソースまたはターゲット言語としているシステムの発表であった。このうち、2件がドイツと韓国の研究であり、どちらも日本との情報流通を研究の動機として挙げている。また、両方とも、日本のシステムをベースにしたもので、日本語の情報流通の要求の高まりとともに、機械翻訳システム研究における日本の位置の重要性を痛感した。

Muシステムは、全システムの中でも、一番完成度の高いものの1つであるといえる。また、他のシステムは、人間による前編集を十分に行うことによる翻訳の質を向上させる方針をとっている点からもいえる。

第3表 発表論文リスト

- 1) The SYSTRAN System
C.H. Stroll, SYSTRAN Institute GmbH, Luxemburg
- 2) The LOGOS Translation System
Peter Wgeeker LOGOS Corp., USA
- 3) The Machine Translation System ROSETTA
Lisette Appelo Philips Research Laboratories, NL
- 4) Transfer as a touchstone for Analysis
Jürgen Kunze Akademie der Wissenschaften, GDR
- 5) Extended Demonstration of SEMSYN
Martin Emele, Walter Kehl, Dietmar Rösner University of Stuttgart, FRG
- 6) Computer Aided Translation
Günther Stieglbauer Ericsson Information Systems, FRG
- 7) BAPCOMP's Universal Computer System
Burhan Bukhari FRG
- 8) Advances in the Japanese Governmental MT-Project
Makoto Nagao Department of Electrical Engineerings, Kyoto University, Japan
- 9) Unified Operation Control System for Mu Machine Translation
Yoshiyuki Sakamoto Electrotechnical Laboratory, Tsukuba, Japan
- 10) English-Japanese Machine Translation System LUTE -AID
Hirosato Nomura NTT Basic Research Laboratory, Tokyo, Japan
- 11) On the Korean-Japanese Machine Translation
Chang ho Park SERI/KAIST, Seoul, Korea
- 12) CALLIOPE - The French National Machine Translation Project
L. Guilbaud, GETA, Grenoble, France
- 13) Presentation of EUROTRA-DK's ECS component
Poul Andersen University of Copenhagen, DK
- 14) Linguistic Research and Software Test Systems for Machine Translation
Johann Halleer, Heinz-dieter Maas IAI, Saarbrücken, FRG
- 15) An LFG-based Approach to Machine Translation

Klaus Netter, Jürgen Wedekind University of Stuttgart

16) Program for Language Analysis and Inference (PLAIN)
Peter Hellwig University of Heidelberg, FRG

17) Multilingual Aspects of Reference Information Systems (MARIS)
Edith Kroupa Department of Information Science, University of Saarbrücken,
FRG

18) A CAP Parser Generator for German
Heinz-Dirk Luckhardt H.D Luckhardt, SFB 100/A2, University of
Saarbrücken,FRG

2. 各システムの紹介

2.1 SYSTRAN

商用化されたものから、開発が完了しているもの、さらに現在開発中のシステムを含むとヨーロッパ言語間の2カ国の言語対13組のMTシステム有しており、日英、英日についても開発中である。その特徴として、

辞書は、語幹辞書(STEM),熟語辞書、共起表現辞書の3種類が用いられている。

翻訳は、解析、変換、生成の3段階からなるtransfer方式である。

処理手順に特に新しものはないが、商用システムとしての豊富な経験からヨーロッパ言語に対する形態素処理には、多くの経験則が盛り込まれている。

辞書ひきに際して、ハイフンを含む語の処理、未知語の形態素情から意味情報を抽出する機能を持っている。たとえば、'PSEUDOPHYSKIST'にconcrete, countable, human, profession又、'MICRONEUROENTERLOGY'にabstract, mass, always singular, scientific discipline.

同形異義語の判別にフランス語では66、英語では、83の同形異義語分類表をもちいることべられているがその効果は不明である。

このシステムは、バッチ処理のみであり、システムが完成されたものであれば、非常に効率的なシステムと言えるが現在のところそれは不可能に近く、ユーザからも、interactiveな機能がない点、辞書の更新に関して不満だという声が聞かれた。

2.2. LOGOS

1985年から現在までの開発と利用状況について報告がなされた。英語からスペイン語への翻訳が非常に進んでいる。また、独仏の開発もすすめられている。

システムの納期について種々の言語対と共通ソフトを半年で出荷できる目処が立った。

このシステムは、menu-driven batch systemである。同時に、未知語、未知句の探索、ALEXと呼ぶ会話形辞書エディタ備えている。さらにALEXは意味を扱えるSEMANTHAと呼ぶ兄弟システムをもっている点が特徴である。翻訳に際しては、Subject Matter Code (SMC)を指定する。このコードは5個まで順位を付けて与えることができ、これが与えられた分野での用語選択の役を果たしている。これは、かなり有効な方法と考えられるが、このSMCをどの程度正確に付与できるかが問題である。

この論文では、利用者の経験にもとずいたデータが紹介されている。利用者が自分で翻訳した場合と比較して、post editingだけを行うことにより、4倍の量の翻訳が可能となった。

利用者の辞書作りに関して、利用者が辞書登録を6ヵ月行くと1分野当たり500から10000エントリーが作成される。その後は、新語が出現する割合が急速に減少し、新語の登録に悩まされることがなくなると報告されている。

2.3. ROSETTA (Philips 社)

Isomorphy Principle が最大の特徴である。これは、2つの言語間には対応するルール、表現がある。よって、各言語の文法を attune してパラレル同型の文法を書けばよいというものである。

しかし、2つの言語間の文法、表現、意味には、必ずしも、お互いに対応のとれない部分があり、これをどう扱うかが疑問である。結局、実用的なシステムを作る上では、この原則を崩さざるを得ない、これでは、この例外が、大変多くなり、先の原理を原理として保持しえないのではないか。

これから5年間の長いスパンをもった計画であるが、最終目標の中出でも、意味の扱いが単語の属性レベルのものにとどまっているのが問題である。

2.4. SEMSYN (Stuttgart 大)

ATLAS II の出力した Conceptual Structure からドイツ文を生成するシステムで、ソース言語の解析とターゲット言語の生成を全く別の機関で開発するという新し形の研究である。日本語から生成された Conceptual Structure がそのままドイツ文の生成に必要な情報を与えるかが興味あるところである。教、時制といった日本語では表面上表れない言語特性が Conceptual Structure に欠けているため、heuristic なルールを導入しているが常にこのような処理が施せるか疑問である。

2.5. EUROTRA-D (ザールブリュッケン大)

この発表では大きく2つのシステムが紹介された。

1つは、EUROTRA-D のプロトタイプで PROLOG を使った小さなシステムであり、もう1つは SUSY-II を EUROTRA のアイデアを一部使って拡大したもの (SPES)

プロトタイプは非常に小規模のものであるが、一方 SPES のほうは、SUSY を長年研究してきた暦もあり、辞書等も整っている。

2.6. PLAIN (Heidelberg 大)

知識ベースシステムと紹介された。翻訳システムや自然言語エキスパートシステムのための shell を提供するシステムとのことで、ターゲット言語の生成部がまだ作成されておらず、全体として、翻訳システムと成ってはいないが、知識ベースを全面に出したシステムとして興味をひいた、

知識ベースには inference rule が格納され、これらは、自動的につくられる。対象分野の文を入力し、解析、変換する。この過程でその文が inference rule にコンバートされる。自己学習能力として注目されるが、具体的な説明がなく、未だ構想の域に止まっている状態だとおもわれる。

2.7. MARIS

MARISはMultilingual Aspects of Reference Information Systemsの略出で、特定分野での情報翻訳サービスをするシステム開発のためのプロジェクト名である。

具体的には英独のターミノロジーデータバンクと、翻訳SUSYを使ったバイリンガルデータベース(タイトルと目録)の作成をする、

西ドイツの研究技術省(Federal Minister of Research and Technology)が資金を出しているプロジェクトで日本でいうとJICSTが英文文献データベースを作ろうとしているのに対応している。

現在8000タイトルの英独情報が蓄えられているそうだが、実際には、機械翻訳には使われておらず、来年から予定しているとのこと、その場合、post editが不可欠であり、中には翻訳できない文もでてくる。100,000件のタイトルを翻訳するとその分野でのvocabularyはほぼでつくすといっているが、何時の時点で人手より機械を使った方が速いという評価が出せるか興味あるところである。

MT システムの仕様比較表

システム名	SYSTRAN	LOGOS	ROSSETA	SEMSYN	MU	LUTE-AID	KANTU	EUROTRA-DK	EUROTRA-D(SPESET)	CALLOPE	LFG	PLAIN
status	商用	商用	研究	研究	研究	研究	研究	研究	研究	研究	研究	研究
システムのタイプ	transfer	transfer	transfer	實現表現からの抽出語の生成	transfer	transfer	transfer	transfer	transfer	transfer	transfer	transfer
言語	13のペア	5のペア	オランダ-スペイン	日-独	日英-日英	英-日	韓-日	オランダ parsingのみ	multilingual	仏-英 英-仏	独-仏	独-英
速度	150,000-300,000語/h IBM370/168		1,200-3600語/h		5,000語/h	4,000語/h			140,000(SPES)	30語/h		
辞書	英 60,000	独 90,000 英 50,000	5,000	独 3,800	70,000	10,000	2,000	4,000		12,000	200	
ルール		19,000(独英)	14,000lines		3,000	840	300	40		550(2,500lines)	独20 仏50	
開発言語	FORTRAN	FORTRAN 77	PASCAL	ZETA-LISP	UTILISP PLI		C,PLI	C-PROLOG	FORTRAN(SPE) PROLOG(EPT)		C-PROLOG	PLI
OS	IBM/MFV M IBM360/30 XX	CMS,MVS,WANGVS	VMS VAX11/780	SYMBOLIC S	FACOMOS/IVFAMSP	VMS(VAX) TOPS20(D EX2060)	KONIX SSM-16	ULTRIX VAX11/780	BS200,UNI X SIEMENS/SUN III	IBM4631	VMS4.3 VAX11/780	IBM/MVS SIEMENS BS2000
ハード												
その他	バッチ	バッチ ミニ電動 辞書エントリあり、ルールの英夏ユーザが可能	1991年までの計画	ATLAS IIに より出力された日本語 文の意味素を生成			ATLAS Iベースのシステム		SPEStは SUSY-IIをベースとしたシステム ETIは EUROTRAのフォロワー プロジェクト	ARIANE78をベースにしたもの		生成のインポートはまだなし
開発機関	SYSTRAN社	LOGOS社	ファイリップス社	Stuttgart大	日本の国家プロジェクト	NTT	KAIST		ザールブリュッケン リュネ大	フランス国家プロジェクト	Stuttgart大	Heidelberg大