

報 告

第2回 日米機械翻訳セミナー

—報告と座談会—

昭和40年6月19日 船舶振興ビル

報 告 和 田 弘

1. セミナー参加者

日本側

和田 弘 (成蹊大), 坂井 利之 (京大),
 田町 常夫 (九大), 野崎 昭弘 (東大),
 長尾 真 (京大), 平松 啓二 (電機大),
 山田 小枝 (立正大).

米国側

E.D. Pendergraft (U. Texas), V. Yngve (M.I.T.),
 D. Hays (RAND), H. Josselson (Wayne State),
 S. Kuno (Harvard), C. Fillmore (Ohio State),
 P. Garvin (Bunker Ramo), I. Sakai (Ohio State).

2. 見学旅行

5月5日

(i) Bunker Ramo. (Canoga Park)

P. Garvin から同所で計画している和文英訳のプログラムを聞いた。また訳文としての英文を CRT に出して post-edit する状況を見学した。

(ii) RAND Corp. (Santa Monica)

D. Hays の案内で JOSS という multi-access の計算機の実演を見学, M. Kay の漢字の Coding の研究が印象に残った。

Los Angeles 地区での MT 関連研究者の会合に出席して, 日本語および日本での MT の研究状況について約1時間話をした。

5月6日

(iii) U. of Calif., Los Angeles 分校

数学出身の H.P. Edmundson の説明, 図書館のオーディオメーション化の計画の話聞いた。また校内にあるアメリカの西部にある大学が共用している計算センターを見学した。

(iv) System Development Corp.

IBM の Q-32 計算機を見学, information retrieval の実験を見学した。

5月10日

(v) LRC, Univ. of Texas (Austin)

言語研究所は大学の campus から少し離れた独立したセンターで, その地下室に専用の 7044 が設置されていた。約50人の研究者がこのビルで研究に従事しており, 平松, 山田君の部屋も案内された。

Campus には別に計算センターがあって CDC 6600 があったが, これをさらに高性能のものに変える計画が進んでいる由。

5月12日

(vi) National Bureau of Standards (Washington D.C.)

Ida Rhodes が定年で退いてから, MT の研究はしていないが, いくつかの関連研究が行なわれていた。Sketchpad に似たもののほか, 計算機 PILOT も案内された。I. Rhodes も特に講演をしてくれた。National Science Foundation は Washington にあるので, 当日は Dr. See と Mr. Pronko も終日同行して, その活動内容を話した。

5月13日

(vii) Harvard Univ. (Cambridge)

久野さんに計算機研究所を案内していただいた。大形機1台で, とくに MT の研究をしているわけではないし, 一同は日本語が使えるので, あまり MT の話は出なかった。

(viii) M.I.T.

Dr. Yngve を訪れて, COMIT による実演を見学した。山田さん, 田町さんが日本語の文法を黒板に書くとき, 彼がそれをタイプに打込んで, 計算機が日本語を創作する実験を見せてもらった。

3. セミナー

5月17, 18日の両日, New York の Park-Sheraton Hotel で開催された. アメリカ側の出席者はさきに示した方々のほかに, N.S.F. の Dr. See, Mr. Pronko, RAND Corp. の Mr. Ziehe と Mr. M. Kay, Bunker Ramo の Mr. Penn. さらに通訳として日本人が1名加わりました.

われわれは論文を用意して行って発表しました.

すなわち

Input and Data Format of Japanese Texts:

和田 弘

Procedure for the Analysis of Japanese Text:

坂井 利之

Syntactic Description of Japanese Grammar:

田町 常夫

On the Dictionary Preparation: 野崎 昭弘

Japanese-English Translation regarded as Sentence Generation:

長尾 真

Syntactic analysis of Japanese, by Using the Property of Dependency: 田町常夫, 石原好宏の6論文*です.

アメリカ側は見学旅行の際に資料を提供してくれたため, セミナーではまとまった論文を中心に自由討論が大部分であったと申せます.

第2日目に N.S.F. と学振とを經由して日米科学協力委員会に提出する勧告書の草案を審議しました. 議論が活潑に出て, 原案が相当修正されました.

勧告書の訳を記します.

勧告

昨年のセミナーでの勧告を再確認するとともに, 次の5項目を一層に明確にする好機であるとの前置きにつづいて,

(1) 研究者の交換は日米協力の中で大切なものである. 現在の計画は1年間で成果をあげるには短かすぎるから, 延長するつもりで続けていこう. 別の計画も奨励しよう.

(2) 以下の勧告を実行するには日米の学会(情報処理学会と AMTCL)との支持が要る. 各学会は個人または委員会を任命して, それぞれのスポンサー機関と密接に連絡しよう.

(3) 文献などの交流をもっと促進しよう. 日本の文

献の題目をアメリカの The Finite String のリストに掲載しよう. アメリカのものを日本に配布する機関を決めよう.

(4) テキストの交換を便利にするために共通の code と format を採用できるように考えよう. 両国の学者はそれぞれの国語についての提案を出してもらいたい.

(5) 昨年と今回との2回のセミナーの結果, 専門家による小規模の会合が有効なことがわかった. 上記(3)および(4)の実体について seminar, planning meeting あるいは working group が時々必要だ. 少なくとも2年以内には次回の集合をしよう. 日米の両学会はテーマと会場所とを選定して, スポンサーに進言しよう.

4. 国際会議

5日19~21日の3日間, 同じ Park-Sheraton Hotel で International Conference on Computational Linguistics が開催されました. 日本はさきに6個の論文を送ったところ, 京大の坂井, 長尾両君による sentence generation の論文が発表されることになっていました. 論文は約40ぐらい参加者に配布されました. 2日目の夕, 晩饗会があって主催者たる AMTCL の会長 Prof. Lehman の講演がありました. そのあとで同氏の部屋に各国から1名ずつ代表が招かれまして, whisky をご馳走になりながら非公式に国際的な組織を作ることによって, こういう会議を開催することはどうだろうかと提案がありました.

ヨーロッパの人々はいずれも賛成しましたが, 私は反対しておきました. 本日ここにおられる久野さんも反対しておられたようです. しかし多数の意見によって推進することになり, その結果すでに Prof. Lehman から手紙が来ました.

それによると, Federation と Central Committee をつくることについて, 各国内の意見をまとめて7月中に返事が欲しいとのこと. もし大多数が賛成ならば, 1967年には会議を開くことに予定したい, との趣旨です. 返事を出さねばなりません.

質問 反対の理由は, 何か義務でもできますか.

和田 できます. できれば, 大した額ではないが費用を負担しなくてはなりません. さらに, 会合の都度誰か一人は出ねばならず, できれば同一人でありたいものですが, 無理でしょう.

国際的な会合はヨーロッパで開かれると思わねばなりません. 旅費が大変です. それにヨーロッパの諸国

* これらの論文はまとめたものが学会事務局にある. 希望者には有料で配布できる.

では、国語がいずれも違うのですからどうしてもその辺が議論の中心になって、日本語はやはりツンボ枝数におかれることになりましょう。

さらにこの動きを注目しますと、言語学者がその研究に計算機を利用しようといういわゆる *computational linguistics* の人々が中心になっています。わが国でいえば計量国語学会に相当するものです。理工科系の人はずである IFIP の 3 年ごとに開かれる *Congress* の中で *MT* のための場を持っています。

また仮に反対しても、設立されるでしょう。加入しなくとも 67 年の会議には参加することはもちろんできると思います。久野さんも理由を一つ、

久野 アメリカの連中は肩書を欲しがっている。またヨーロッパに行く理由を求めています。委員会ばかり沢山作っていて、意味がない。

座 談 会

出席者 和田 弘, 坂井利之, 久野 暉,
野崎昭弘, 長尾 真, 小笠原林樹,
伊藤正安, ベーツホッパー
司 会 中野道夫

1. M. I. T.

中野 はじめに、とくに印象に残ったことを伺いたいのですが

和田 アメリカでいま *MT* を一生懸命やっているところは、テキサス大学だけです。たいていのものはもうあきらめて、*Computational linguistics* の研究とあって、*N.S.F.* から金をもらって、好きなことをしていると見ました。ただ日本人がぼやぼやしていると、アメリカ人の方が日本語の機械翻訳を先にやってしまうのではないかという心配はあります。

長尾 *M.I.T.* は *Yngve* が彼の *depth hypothesis* が日本語でどうなっているかと興味をもって尋ねていた。彼の部屋には、マルチアクセスのステーションがあって、ここで彼がつくった *COMIT* を呼び出して日本語のジェネレーションの実験をしてくれた。田町さん、山田さん、彼はここで日本語をやったのは初めてだといっていたが、やはりプログラミングシステムは完備している。だから、言語学者は言語のことを専心して考えればよい。つまり計算機とかプログラミングを使う方に関しては、非常に自由に使える。そういう環境にあるということはやはり良いことですね。

中野 マルチアクセスというと、実際使う人ほど

うするのですか、*input* は何ですか？

野崎 簡単にいったら、オンラインのフレクソです。

和田 長尾さん、あなたもそれに似たようなことをおやりになっているようだけれども、能率についてくらべて頂けると……

長尾 それは比較にならないですね。彼等の場合だと、現に黒板に書いたものを、その場でポンポンとルールに打ちこめば、すぐ何百というようなセンテンスがパッパッと出てくるわけです。

小笠原 その出てくるセンテンスは立派なものですか。

長尾 それはグラマーの作りによります。結局グラマーどおりにプログラムが作り出してくるわけです。

小笠原 ゼネレーションをやっているのは、*MT* と関係があるのですか。

長尾 *MT* をやっているのではなくて、文法の検定ですよ。ゼネレートされた文章は全部われわれが見て、文章の意味を持っていなければまだ作られた文法が完全でないということです。

中野 ゼネレートしてみれば、文法のどこがおかしいか、すぐわかるわけですね。

花もあるきたい。みます。わたくしでも わたくしでも わたくしは花でも食べます。みます。あるきます。犬が花でも花をわたくしはみたい。わたくしがたべたい。たべたい。花は私でも犬もたべたい。犬も犬がわたくしでも犬はわたくしはわたくしはでもみます。わたくしでもあるきたい。犬がたべたい。ete

小笠原 *M.I.T.* はいわゆる *Chomsky* 言語学が中心になっていると思うのですが、彼の文法はやはりあすこで使われているか？それと、チョムスキーは *MT* に関心を持っているのですか。

和田 使っている人は多いでしょう。しかし、チョムスキー自身は、*MT* についての関心は、今は全然もっていない。興味はもっていたが、とっくにやめた。

2. 彼 我 の 差

中野 ハードウェアの面がうらやましいという声があったでしょう。

野崎 ハード、ソフト、データ伝送も含めて、要するに計算機については差が大きいですね。

和田 そうだ。彼等とはとにかく、機械翻訳をやりたい時はいつでも機械をつかっている。日本人は何か頭が良さそうな顔して、キョロキョロしているけれども、機械をいじりもしないのが多い。

坂井 MTの見学に行く前に、アメリカ帰りのMTの専門家でない人がいっていたことに、1950年台の初めに、アメリカで非常にさかんだったことを、今日日本でヤイヤイやっている。ところがアメリカ自体は、エンジニアが機械にかけて、これは面白いとやってみたものは、言語学的にみて何の翻訳にもなっていない。それで今度は言語学者がそれはいけないというので、非常にもの方へもどりすぎているのではないかと、すなわち、学問の研究というのは確かにオッシレーションはあるけれども、少しその反動が大きすぎるのではないかというような印象をもったわけです。もう一つは、アメリカでは機械とソフトウェアが進んでいて、確かに環境は十分なだけけれども、とくに新しいものが出たというのではなしに、日本でやっても別に今からでも遅くはない。わたくし自身はそういう反省のもとで、初めから機械を使ってやっているけれども、機械を使ったMTも、日本ではやっつけられる余地は十分あると感じて帰ったわけです。わたくしの感じでは、ランドにしろ、テキサスにしろ、辞書なんかだとやはり何万という単語を入れている。久野さんの文法にしろ、われわれが知っているようなものより、相当に密な内容をもったものをもっている。それでいろいろ議論をしているわけで、だからいろんな面もわかってきている。抽象的なことだけでないということはあるですね。もちろん周囲環境は日本にくらべて恵まれている。考えようによっては、ああいう環境があれば、われわれでもある程度おもしろいことができるのではないかという気がする。そういうことです。

野崎 少し長い目でみた話だと、アメリカの方が計算機そのものばかりでなく、計算機についての教育が普及していることの差を感じました。ほんとうによいMTを長期的な見透しで実現しようとしたら、だいたい大学における教育から考えないと結局進まないんじゃないかな。

中野 皆さん、だいたい量的な人員、設備、環境、それは確かに良いが、質的には日本と比べてどうということはないと。

坂井 パーソナリティーがないという感じはわたくしはした。

和田 それをいうと危ないんだよ。一面の真理だが、坂井さんがさつきいわれたことですが、アメリカは「アツモノにこりてナマスを吹く」という感じがする。要するにMTをしていないのだから、MTがそんなに進歩しない。アメリカがCLをやっているから、日本もすべきだとは思わない。やはりMTを頑張れば良いと思います。

田町 小生の印象ではアメリカは計算機は進んでいる。Vocabularyもある。しかしMTやMTに関係のあるCLの面では特にこれはと思うものにお目にかからなかったこと。卒直にいうと、すぐれたアイデアでとられたデータがない。あるいはデータはあってもどうしてよいかわからないというのが現状で、特に意味に関しては役に立ちそうな具体論もデータもないようだということ…などです。これは日本に比べて劣っているというわけではありませんが、その点役に立つデータをとることが必要で、数の上だけで“日本はよほど頑張らなければ”というのは当たらないと思います。

3. セミナーの印象

中野 今度はセミナーの印象をひとつおきかせください。

和田 先に申したとおりで、セミナーでアメリカ人は良い話しをしませんでした。自分の方が経験者のような顔をしていた。日本人がどのくらいのレベルの話しをするか、まあ聞いてみてやろうということだった。日本は1年間にたいして進歩してないが、彼等が思っていたよりはやっていたという印象を受けたのではないかな。

坂井 わたくしはレゾリューション・コミティにいて、項目としてどういうことをうたうかを協議したのですけれども、その初めの日からハッキリしたことは、俺のところにも人物交流のチャンスを与えろということで、これが文章にも強く出たことです。ところで日本人をどの程度欲しているかということ、自分らが支払うものに対して有利であればよいということで、よい言葉でいえば、cooperationでやりたいことが、ハッキリでいた。もう一つはテキサスの話してペンダクラフトが非常に大きな組織をつかってやっていると、それに対してガーヴィンが非常に違った反対の意見を出している。アメリカの中にもいわゆるフランス流とアメリカ流の考え方が分れていた。

中野 カンファレンスの方では、日本の講演に対し

て、良い質問は出ましたか。

長尾 あの時は30分のもち時間をほとんど使ったので、その時は出なかった。しかし、たとえばランドのハーバーという人などは、自分のと非常によく似ているとっていました。ドイツのボンの MT の研究所のホップが“アメリカ人は意味のことについてほとんど関心をもっていないと、あれはけしからんことである”とっていました。彼等も私達と非常によく似た考えをもってやっているのだそうです。

4. トランスフォーメーションナルグラマーの評価

小笠原 久野さんはトランスフォーメーションナルグラマーのその位置というのはどういうふうにお考えですか。

久野 やはりフレーズ・ストラクチャ・グラマーの時代は過ぎて、どのグループもそれではもう十分でないということがわかって来た。型はいろいろ違うが、なんらかのかたちで Chomsky 流のトランスフォーメーションを組み入れている。

和田 Chomsky のグラマーは、日本語に直接適用できないから、つまらない。チョムスキーの本の附録は英語についてだ。

野崎 あの附録を例に挙げるのは拙いでしょう。

和田 それは知っている。シンタクテックストラクチャの始めの方に書いてあるマセマティカルな所はいいと思う。あそこなら日本語にアプライできる。

長尾 トランスフォーメーションナルグラマーは非常に有益ですネ。いろんなことを説明するのに、だけれどその……どういうふう一般的なレベルで利用できるかということになると非常にむかずかしいと思う。

皆はトランスフォーメーショングラマーの云々とかいろいろなことを意識していつてるけれども実際上それを有益に使った人は今だにないのではないか、否定をとるかというくらいのことですネ。一見よくて皆とびついたらけれども、にっちもさっちもいなくなつたという危険性をいつも考えるべきだといいたい。

久野 別に Chomsky がいいだしたから、トランスフォーメーションをいうのではなくて、要するに言語にそういう面があるので、それを使わないと、いろいろな面を描き出せない、結局説明が不足になるわけですよ。

5. IFIP での MT

長尾 IFIP の MT が少ないようですけれども、あれが何故そんなつまらんのが集まったか不思議ですネ。

中野 それは日本だけですか、全体ですか？

長尾 全体ですネ。

小笠原 それは、その前の MTCL の方にそちらの人達が主力をおいたということですか？

久野 MT の良いペーパーがないのです。

和田 あまりやってないのに、どうしてあるんだ。

中野 それが世界的傾向ですか？

野崎 MT という意味をせまくとれば、良いペーパーはそうはないんじゃないですか？

久野 一度フレームワークを出したら、それから1年後に文法がこれだけよくなったって書いてもしょうがないでしょう。

中野 おしまいにしてしまおう。ありがとうございました。