

日本における情報検索の現状と将来*

出席者 関根 智明⁽¹⁾(司会) 小沢 大二⁽²⁾
 菊池 敏典⁽³⁾ 小林 胖⁽⁴⁾
 水谷 静夫⁽⁵⁾ 渡辺 竜雄⁽⁶⁾

関根 今度、『情報処理』で特集号を出すのです。それぞれのテーマについては原稿ができることになっているんですが、それらでは触れられない点とか、生の面白い話を座談会でやりましょうということです。

ところで面白いことに、情報処理学会の会員で IR を専門にやっている方はほとんどいないわけです。少なくとも私が IR の専門家でないことは事実です。

そこで皆さんにきょうお集まりいただいたわけですが、いま自分がどんなことで IR に関係しているかということをお話ししていただきたいのですが……。

菊池 われわれ、あまり IR といういい方はしないんです。ドキュメンテーションといういい方をするほうが多い。IR とドキュメンテーションを区別する必要があるのかないのかもはっきりわからないのですが、要するに、情報センターが何をやっているかという科学技術に関するという一つの枠がありますけれども、その中で特に文献情報処理、IR といってもファクト・リトリバルの問題もあります。いまのところ情報センターではドキュメント・リトリバルの機関であると考えていただいていいと思います。

そういう範囲でできるだけ IR を手広くやろうという構想を持っているわけです。私が数年前からやっておりますのは、いわゆる化合物の機械検索です。これはそれとして面白い問題を持っているのですが、同時に非常に簡単に言語のモデルであるとか、そういう考え方が成り立つわけで、その辺で言葉との結びつきという多少面白い問題がこれから出てくるんじゃないかと思っております。

水谷 関根さんが IR の専門家じゃないというなら、私のほうはもっと専門家じゃない。ただ私は割合に物好きなものですから、何にでもフラフラッと興味を持ちやすいので、一応世間的に専攻は国語学ということになっていて、言葉の問題を扱うのが専門ということになりますが、国語学に限らず言語学でもそうですが、実際にいろいろな言語の情報処理をするとき必要になるような知識の蓄積が言語研究、あるいは国語研究にあるかと申しますと、学問はそれぞれお家の事情があって、だいぶ目的が違うわけですね。

私自身の言葉の問題を考えますときに普通言語学でも国語学でも記号としての言語を扱うわけです。それはいわば物としての言語なんですが、言語という言葉は二とおりのニュアンスがあって、一つは普通に誰でも考えるように、記号の列としてのストリング言語ですが、そのほかにストリングの使い手のワザ、行為ということが問題になるわけです。その行為のほうに普通の言語学者より関心が強かったものですから、必要な情報を引出すなんていうのも言語を媒介にする場合には言語行為の一種になるということで関心を持っているわけです。とくに、言語学者が皆いやがる意味論に対しての関心が割合に強いものですから、意味をフォーマルな形で扱おうという希望を持っておりまして、それが幾分ファクト・リトリバルの問題に触れてくるんじゃないかと思えます。

もう一つ、形が非常に違いますけれども、実は国語学は IR に当たることを昔から国語学の範囲でやっているのです。と申しますのは、だいたい国語学とか言語学といわれているものは最近まで言語の歴史の研究がおもだったわけです。そこで、いろんな言語の変遷やなんかをたどるときに、昔の言葉ですから今から話し手に確かめるということができませんので、僅かに切れっぱしが残っていたところから言葉拾ってこなければならぬ。それから、まとまった索引は、たとえばホトトギスという言葉がいつの例が一番古く、どんなような使い方をされているかというような

* Information Retrieval at Present and in Future by Tomoharu Sekine (Keio University), Daiji Ozawa (the Ministry of Foreign Affairs), Toshinori Kikuchi (the Japan Information Center of Science and Technology), Yutaka Kobayashi (the Japan Information Center of Science and Technology), Shizuo Mizutani (the Tokyo Women's Christian Colledge) and Tatsuo Watanabe (the Ministry of International Trade and Industry)

(1) 慶応義塾大学 (2) 外務省 (3), (4) 日本科学技術情報センター (5) 東京女子大学 (6) 通商産業省

知識を持っていないと成り立たない学問なものですから、そういう索引ができていくわけです。

この索引ということが意味が IR の場合と違います。キー・ワード的なものじゃなくてキー・ワードは一覧してどこに辞典があるかということを書いた本ですけれども、そういうことをつくるということが基礎的な研究の一つになっておりますので、私が大学にありましたときには古代語の研究をしておりまして、自分でも古事記の総索引をつくったという経験もありますので、たとえば、ドキュメント・リトリバルでこんなことだといわれると、あっそうか、文献の代わりにたとえばアメとか、タカマガハラというのがくるのだなど、そして、何という雑誌の何巻何号にあるという代わりに、上巻の第何丁の表何丁にくるんだなということがすぐわかって、それで一応興味を持ったところなんです。

小林 私は IR というよりは長いあいだ図書館のほうをやっておったわけです。きょうしゃべりたいと思ったのは IR のうちの I とは何ぞやということで、情報の受け渡しの状況をつかまないと何をやっているのかわからなくなりはしないか、という問題に興味をもっているのです。

渡辺 私のところは IR かどうかちょっと問題があるんですが、行政事務の機械化のシステム問題を考え電子計算機の適用をはかるという仕事をやらされることになったのが七年前です。その頃の事務処理のテクニックというのは何でもかんでもコードづけを丹念にしないとできないらしい、そんなことをしないで済む方法はないのかという程度でした。そこで、このようなことも機械に押しつけてしまおうと考えて、そのあとは腕力でやってきた、というような次第です。それは IR だという人がいるので、IR とはそういうものかという感じで今まで進んできたわけです。

ですから、現在やっている仕事はいろんな業務を機械のシステムにのせていく、あるいはのせるシステム自体を考えるということです。経済統計からはじまって、文章で書いたものにポツポツ入ってきていますが要するに、形式化の問題、あるいはプログラムの問題はくるところまでできており、問題は文章の扱いということになっております。去年の夏から、ソースラックをつくってみようということで今悪戦苦闘をしております。実は、なるべくよそ様の話を聞いてということで慎重にかまえていたのですが、外務省で作業を行なったのを知って勇気づけられたわけです。あちこちで随

分議論される割合には実際に行なっているケースは極めて少ないのが実態でしょう。

IR の問題といいますが、むしろシステム・デザインの問題が本業で、その必要上そういうことをやっておりますので、IR の話ということになってもよくわからない。

小沢 私はそれこそ真正銘文科系の出身ですから技術的なこと、もしくは電子計算なんかにについては一番わからない立場です。ただ、ご出席されている方の中でおそらく一番電子計算機で苦勞しているほうに入ると思います。なにしろ仕事の範囲が職員の給料計算までやっておりますから……。

外務省は 38 年に機械を入れまして、IR の旗を訳もわからず振り回したわけです。外務省は戦前から非常に情報の管理というのはやかましいところなんです。電子計算機が現在対象としている、いわゆる IR の対象は生の情報です。したがって、古いものとか、あるいは物体になったもの、たとえば図書などは除かれています。

現在のところ、省内に IR についての PR をして IR というのはとても便利で何でも出せるぞといったため IR について過大評価をされてしまった上、IR というのは非常に容易なものだという印象をあたえてしまった。これはもう省内だけじゃありませんで、多くの人がそういうふう考えている。私共のところは民間の方がよくくるのですが、IR については情報を機械に入れると何でも出てくるというような簡単な考え方を持っている方が多いようです。

当面の問題としては、一応実用化ということで踏み切っております。現在 120 の在外公館から情報が送られてくるわけです。それをコンピュータに入れて検索サービスしろということになったのですが、IR を実際にやってみてインプットとアウトプットということに非常に問題があるようです。言いかえれば、Storage, Retrieval, Service ということですが、システムを効果的なものにするのは実は大変です。

いま 20,000 語程度のソースラックを実際に使っているわけですが、とにかく 20,000 語に入らない言葉がワンサカ出てきまして、メンテナンスが非常に大変なんです。金もかかるし、時間もかかって、割にあわないので IR というのは WR の間違いで割にあわないんだということでボヤいているわけです(笑)。

ただ方向としてはやはり IR というのは今後日本でも重視していかなければならない分野であって、とく

に情報だけではなく人事管理とかあらゆる分野に IR が使われるわけです。そういった点でも考え直したい面があると思います。

さし当たっての計画としては、とにかくインプットを解決するという意味で国際通信を計算機に入れること、要するに在外公館から入ってくる電信を電波の形でコンピュータに入れ、自動的にストアする方向を考えなければならないということ。それから言語関係の勉強、とくにソソーラスの問題をかなりやらなければならないということ。それと、いろんな検索方式をテストしてみましたが、あまり公式を複雑にしてみると、現場の人がインデックスをつけられなくなってしまふ。これに関係する問題重点的には自動索引の問題をやりたいということです。

さし当たって少なくとも OR と IR を結ぶ糸口として Content Analysis の問題に取りかかろうと思っています。

関根 次に、自分の関係しているところでどんな具合にして IR ということが問題になってきたかということをお伺いしましょうか。

水谷 IR という言葉が日本に入ったのはいつごろですか。どこからはじまったのですか。

渡辺 私のところは昭和 34 年頃、国産の電子計算機が本当に使えるのかどうか経済分析計算の簡単な問題から入って見たわけです。

関根 そのときに IR がありましたか。

渡辺 なかったんです。統計データをあらかじめ磁気テープに入れておかないと計算のたびにパンチをやらなければならないので、統計データを入れて、キイをつくろうと思ったところ、あらゆるビュー・ポイントで引けるキイを考えますとコードが、何十桁にもなってしまいます。それでナチュラル・ランゲージで処理した方がよいということになりましたが、それにも問題がでてきました。

ちょうどそのころ電気試験所の木沢さんが情報検索機を発表したので、情報検索機の検索ボタンで、ナチュラル・ランゲージの検索テストをしようということで、はじまったわけです。それで 2、3 年やっているうちに、皆がお前のところは IR やっているんだというから、そういうものかなと思って、それが 37 年ごろですね。

関根 自然言語でやっていたというのは、どんなことをやっていたのですか。

渡辺 まず統計の名前をそのまま入れる。たとえば

日本というアジアにあって、独立国というふうに日本という国の属性を全部コードで引けるようにしようと思うと十何桁になります。それに統計の内容などを全部入れると、一つのデータをインプットするのに何十桁ものコードをつけなければいけない、そのためナチュラル・ランゲージで入れて計算機の中で何とかさせようかということになりました。

関根 そのときのナチュラル・ランゲージにはどんなことまでやるつもりだったわけですか。

渡辺 最初はディスクリプタだけを扱うつもりだったのが、けっきょく文章を皆入れたって同じじゃないかというようなことになりました。

関根 だけど、文法を入れて、たとえば論理的な関係だとかということも……。

渡辺 最初は、論理シンタックでも考えればいいのだということでした。木沢さんのところの検索機は、論理シンタックス・パッケージになっているわけです。

関根 その頃、身のまわりでも全然知らなかったわけですね。

渡辺 私の知っている限りでは、科学技術センターの情報検索でした。あらかじめ文献をきちんと整理して、キイ・ワーズにしてコードに近い格好でしたが。

関根 それはいつ頃できたのですか。

菊池 '61 年からだったかな。

関根 36 年ですね。それは知っていたんですか。

渡辺 知っていました。それから 38 年になって小沢さんのところが出てきて……、気遣いは俺だけじゃないらしいと（笑）。

関根 小沢さんのところはいつから？

小沢 IR そのものをやろうということで調査がはじまったのは 35 年なんです。

関根 その動機は何ですか。

小沢 私共、電信を扱っているものですから、そういう生の情報は何とかして早く処理したい、ということが一つと、それから、情報を整理しても、それぞれの課とか人間によって整理の方法が違うものですから、人が休むと出てこないわけです。それじゃ困るので何とかしろということになったのです。

直接のきっかけはやはりキューバ事件ですね。あれで電報がワンサときたわけです。これは体系的に電報を整理していくと各国の態度といったものがわかるんじゃないかと、ただ、お役人というのとはにかくどっちへ転んでもいいような報告が多いんですよ。ただ、

キューバ事件のようにたとえば船という事実がありますと、船が引き返すか、進むか、イエスかノーか、どっちかでしょう、二進法ですよ。とどまるという見解は成立しませんよね。ですから進むか、退くかどっちかの報告になるわけで、これは面白いというわけで研究に着手しました。

私が IR という言葉を初めて知ったのは、たしかアメリカから送られてきた資料でウェスタン・リザーブの先生が書いているタイトルの一部に入っていたというのが一つと、もう一つはアメリカン・ドキュメンテーションとかいう雑誌に当時 IR という言葉を使ってどなたかが発表なさっておったようです。それが私共が知っている範囲では最初なんです。

実は国内のことについてはよく知らなかったのです。ただ科学技術情報センターと通産省のことは割合知っていたのですけれどもね。アメリカから調査資料を随分集めたのですが、アメリカというのはすごいことをやっているわい、という程度ですね。

菊池 われわれの場合は、さっきも言いましたように、とにかくいわゆる科学技術情報活動をやろうと、その中に狭い意味の IR、計算機に情報蓄積して、そして検索するというのも一部として含まれるわけです。

関根 それは、発足のときからそういうふうになっていたんですか。

菊池 そうです、予算ののってました。

小林 これは確かウェスタン・リザーブ大学の文献か何かを読んで、のせたんです。

関根 そうすると、日本の中で動き出した頃のきっかけというのは自然発生的という面もあるけれども、ウェスタン・リザーブ大学の影響が強かったということですか。

菊池 ええ。けっきょく金属のコードがウェスタン・リザーブのコードだったわけです。ですから、センターはウェスタン・リザーブ直輸入みたいな形で出発したわけです。

関根 それはいつ頃ですか。

菊池 それがセンターができたのが 1957 年ですから、その頃から予算としてはのっていたわけです。

小林 計画としては少なくとも前年に予算要求をしていますから……。

関根 そうすると 1955 年ぐらいには知っていたということになりますね。

そのころ知っていたのはそこだけだったのかしら。

渡辺 そうでしょうね。私のところなんかそれから 5 年間……、まあもっとも科学技術情報センターと私達が扱う内容があまりにも違っていて、自然科学系統の文献は整理がつくものであるという世間的な考えを持っていたのですよ。

小林 新聞に出たのは機械翻訳のほうが早かったですね。

菊池 あれはヤマトまできたときじゃないですか。

水谷 それは新聞の普通の読みものにはなっていなかったでしょう。

関根 ヤマトは '57 年ぐらいですか。

水谷 設計にかかったのが '57 年ぐらいじゃないですか。

関根 それから、ヤマトと限らず、関心は割合言語屋さんのほうではもっと早くから出ましたね。

水谷 いや、言語屋のほうではそんなに関心はなかったんです。むしろプログラミングのほうをやっていた人達です。

関根 日本じゃ？

水谷 えっ、アメリカでもそうでしょう、言語学者が動員されていますが言語学者の数はそんなにない。

渡辺 そうですね、私が、34、35 年頃計算機は実際に使えるか、使えないかという調子で実務ベースの経済予測のモデル計算をかけてみたりしておったときにはすでに電気試験所の蓼沼さんの機械翻訳のプログラムの研究も進んでおりましたから。

そのころ木沢さんの検索機もでき上がっていたのです。

関根 けっきょく急激に IR、IR と言いつつ出たのはいつごろからですか。

渡辺 それは小沢さんのところの看板が上がってからですよ。

関根 世の中はどうでしたか。木沢さんのところの機械なんてかなり関心を持った？

いわゆる汎用の計算機をやっているほうの側はどうでしたか。

渡辺 検索機のパッケージを汎用機のプログラムにしてテストをはじめていたのですが反応はみられず、39 年になって小沢さんのところが動き出してこういことができるということになったら、汎用機のメーカーがうるさくなってきました。外務省で効果的なシステムを作ったということが、世間の注目を集めたと思います。

関根 三菱なんかはいつごろですか。

渡辺 三菱はもっと前でしょうね、だいたい 35～36 年じゃないかと思います。

関根 東レなんかも早かったんじゃないですか。やる気構えになって初めたのは。

渡辺 そうですね。

関根 通産省なんかと同じぐらいだと思っていいですかね。

渡辺 通産省で実用化が早かったのは特許のプログラムだったのです。これは 37 年に合金組成とステロイド化合物の部門の審査のための文献検索を電子試験所からもらってきた汎用機でテストしてみたらできてしまったのです。

小沢 そうですね、私が 38 年に行ったときには特許庁の人が Patent Office に研究員として駐在していましたね、どんどん情報を送っていました。

うちはどっちかという通産省を利用したわけですよ。通産がやっているからということでくっつけたんです。たしか私は 37 年の春にうちのお偉方をつれて渡辺さんのところへ見学に行っているんです。そのときにもう検索機は見せていただきました。

関根 そうすると、37 年ごろからポチポチ意識してきたんですかね。

渡辺 要するに化学構造式とか、金属の元素記号とか、特殊なものしかやれないのではないかという感じでおりました。

関根 情報センターの金属はいつごろですか。機械が入ったのは。

菊池 '61 年だったですか。

関根 それはわかっていただけですね。

小沢 いや、36 年に TOSBAC が入っているというのは知っていましたけど。

菊池 一生懸命ためこんでいたときじゃないですか。

小林 外から見ると両者が一つのきっかけじゃないですか、外への。

小沢 うちが興味を持たれたのは、むしろ漢字の処理をやったということじゃないですか。

関根 それはあとでしょう。

小沢 いや、最初からです。

渡辺 私のところで本当に全面的にやってみようという気になったのは、逆に小沢さんのところのを見てからなんですよ。

小沢 適当に利用し合った形ですね。

関根 情報センターなんかも利用したわけですか。

小林 うちをよく言えばオーソドックスにやる主義だから、外に対する PR という点で弱かったですね。

関根 よその国でやっているからやろうというライターはありましたか。

渡辺 それは特許の場合にありました。社会科学関係のほうはなかったですね。

菊池 Luhn が KWIC を発表したのは '58 年でしたか。

渡辺 影響しなかったとみていいのでしょうか……。

関根 そのころはセンターでどんなふうな反応を示しましたか。反応はあったか、なかったか。

小林 うちには直接ないですね。むしろアメリカがケミカルタイトルが出たということですね。世界に実用性を示したわけですから。

菊池 だから、けっきょく当時の考え方として、ウエスタン・リザーブ流のこつこつためこんでやるのと、KWIC みたいに索引誌を出すのと、二つの方向が並行していたようですね。アメリカでね、センターはウエスタン・リザーブ流の行き方をしていたので KWIC のほうは慌てることはない。

関根 何故ですか、それは。

小林 一つは機械がないからですね。

関根 それから、いままで機械と接触していないから。

菊池 そういうことですね。それで勘みたいなのがなかったわけですね。

小沢 IBM が一番先にメーカで旗をあげていると思うんです。あれはいつごろですか。

関根 なにしる Luhn のが最初でしょう。IBM のジャーナルで論文をみたのが一番最初なんで、あれは '58 年ですね。

小林 '58 年にワシントンで国際会議をやったときに、あの論文が出たわけですよ。

関根 なるほど、そのときに機械のことを知っていればやっぱり反応の仕方は違ったと思いますか。

菊池 もう一つは KWIC に対する評価ですがね。やっぱりタイトルからキーワードを抽出してもつまらないと、それは強かったですよ。

関根 もう少し本当に偉いと、実用的になるという意味と物事の面白さと並行して反応するはずなんだけれども。

菊池 欠点だけ目についているわけですね。オーソドックスないろんな知識のもとに。

関根 何か知らんけれども、そんなぐらいで出てきたのでしょう。

たとえば Luhn の論文は、IBM ジャーナルでみたわけです。IBM ジャーナルを何故見たかという、計算機のはりの雑誌だから見たんです。ドキュメンテーションの本だから見たわけじゃない。

関根 ぼくは通産省でやっていたということは聞いたけれども、それに対して興味はなかったですね。

菊池 先生が Luhn の論文を読んだときは興味なかったですか (笑)。

関根 いや、それは 704 か何かでやっているという話なんですよね、で、そのころの日本の機械というのは、2203 の頃ですよ。あんまり自由に使えないし、あっそうかなということ、あれならできんだらうねということ。

菊池 計算機が自由に使えなかったという要素もありますね、われわれの場合には、どこへ持って行って、誰に頼んだらやってくれるのかわからない。

関根 お宅に入った機械だって当時としては最新鋭の武器だったんですよ、木沢さんの機械だって今となってはんでチャチでしょう。

小林 うちで戦略を誤ったのは、あれを実験機械と規定して使えば使い道はあったと思いますよ。ところが、実用化の機械として使っちゃったものだから。

渡辺 そう、その点はお宅のお蔭で私のところは助かりました。

関根 そのころは機械使っていたわけね。

渡辺 経済分析計算は大いにやっていたのですよ。さらに木沢さんの機械をもらってくればノルマはかからない。それで検索問題をのびのびとテストできるようにしたいということで、お宅のケースは私には参考になったのです。

関根 だから計算機屋さんのほうから言ってもそうじゃないかと思うんですけど、'58 年は何があったか知らないけれども、おそらく日本でアセンブラをつくるとか、コンパイラをつくるとか。

いわゆる汎用の言語をつくるほうに圧倒的な興味があった。アプリケーションのほうで、IR というようなことを考える余地はなかったですね。お話としては、それはできるだろうという可能性はあったけれども。

渡辺 35 年ごろまでは殆んど数値計算の方のサブルーチンをつくるのが精一杯でしたね。

関根 だから、世の中一般もそうだったんじゃない

かな。

小沢 面白い記録があるんですよ。37 年の予算要求のときに導入機種決定をやったわけです。そのときに各メーカーうちの IR の構想を出しました。そのときのレポートがまだあるんですよ (笑)。いま各メーカーさんはそれを突きつけられると参っちゃうやつが随分ありますね。でも一応 IR ということについてはわかってたかどうか知りませんが、書いてはきていますよね、少なくとも 1 ページか 2 ページは。

小林 うちのあれでも日本に出先のメーカーさんは興味はなかったらしいですね。機械を入れて事務用にやれば儲かっているじゃないかと。

渡辺 私はやっぱり一つの転機は外務省が汎用機ではじめたということだと思います。

関根 僕らは積極的な反応を示さなかったですね。

小沢 そうですね。IR をやるときにうちのほうで大学の先生方にいろいろ相談を持ちかけているのですけれども、IR というのは何だというようなことで、主に工学部の先生ですけれども、よくおわかりになっていなくて、どうしようもなかったですね。

関根 それから、ドキュメンテーション協会なんてあるわけですね、いわゆるドキュメント屋さんというと怒られるかも知れないけれども、そういう人がいまもいるわけですね、そういう人達の反応というのはどうなんですか。

小林 それは現在宇宙飛行ぐらいの感じじゃないですか。

関根 そうですか、現在というんじゃないで、たとえば外務省でこんなことをやり出したとか、通産省でこういうことをやっているという話があるわけですね。それに技術センターで何かやっているということがありましたね、そういうところにドキュメント屋さんというか、図書館屋さんというか……、ところでこの二つは同じだと思っていいいわけですか。

渡辺 違うでしょうね。

菊池 ちょっと違うんです、ドキュメンテーション協会というのは UDC 協会が発展的になったのです。

関根 電子計算機の世の中になっても、まだ趣旨は変わらないのですか。

小林 いまは変わっています、明らかに。

小沢 シンポジウムが箱根なんかでやられているのをみましたけど、随分変わっています。

小林 別の講習会屋さんでしょう。

渡辺 ドキュメント屋さんが私のところにアプロー

ちしたのは昭和 38 年の春です。これは国立国会図書館が主になって、支部図書館なんかと共催して 200 人ぐらいきました。それがドキュメント屋さんとの接触の最初なんです。

関根 その後どうですか、積極的にそんな方向に向かっていくのですか、それともそれほどでもないのですか。

小林 そこでやっぱり分かれるのでしょうか、旧東海道線と新幹線と（笑）。

関根 両方とも国鉄経営ですか、ドキュメント協会の中に両方あり得るわけですか。

小林 旧東海道が多いかな。

関根 そうすると、ドキュメント協会というのはあるけれども、IR 協会というのではないわけね。IR 協会というのはどこでやるのかな、やはり情報処理学会がやるのかしら。

渡辺 そうでもないですね、結局 IR そのものが独立していないのではないかと思うのです。経営管理か何かやる立場でシステムのことを考えてくると、その中に必要なテクニックとしてあるというようなことですね。OR 屋が IR をやったり、在庫管理とか人事管理とか研究管理、などの管理関係者が IR をやったり、要するに、IR というのはそういうものらしいですね。

関根 現状はそんなふうに言えるということでしょうね。しかし、行き当たりばったりにやらないで、整然とやるところがあっていいでしょう。

渡辺 私は昭和 39 年頃から渡辺茂先生と関根先生がやるものだと思っていたのですよ（笑）。それで小沢さんと二人で、一緒に研究会に入ってみたのが、ついこの前という次第です。

小沢 組織化されたのは新しいけれども。

渡辺 ええ、先生と渡辺先生がそういう方面のリーダーだと思っているのですが。

関根 何となくつき合いがはじまった。しかし、そういう点から考えて図書館学科というのが学校にありますが、あれはどうなんですか。

渡辺 藤川さんが機械とのアプローチというのか、ドキュメント屋さんとの接点の問題を考えておられるようです。ほかに、一昨年からはじまった情報センター主催のドキュメンテーション研究集会在が大きく評価されますね。

小林 慶応大学の図書館学科というのは来年から修士課程を設けるといっていますね。

関根 IR はちゃんとやることになっているわけですか。

菊池 結局、IR をもう少し正確に定義づけなければいけないんじゃないですか。

水谷 特許庁が合金の IR をなさるときに、私が国語研究所にいたときですけれども、何人かお見えになってお話しをしたわけです。そのときそういうものは簡単にできるのでいいでしょうが、そうでないものについてはどういうふうにお考えですかと聞いたら、何も成算がないというわけです。アメリカではナチュラル・ランゲージはキー・ワードとして使っているというのですが、キー・ワードにナチュラル・ランゲージを使うだけだったら別に言語学の問題でもない。名詞だけだったら文法じゃないですから、文法が入ってきて初めて言語の問題になるわけですね。ただそれは考えようによってはこんなふうにできるんじゃないかと、私そのときにホラをふいたきりきちんとやっていないので恥しいのですけれども、ナチュラル・ランゲージがそのままいいものだとしなくても、荒っぽい意味で、二値論理に対応させることができると、とくに技術関係の文献ならできる。そしたら完全な形で二値論理の論理意識にならなくても、コンピュータで荒こなしするぐらいのことはできるだろう。そういうプログラムはかける可能性があるんじゃないかと言ったんですけれども、特許庁の方に十分わかっていただけなかったと思います。

結局、ナチュラル・ランゲージを使ってもキー・ワードの段階でそれ以上にナチュラル・ランゲージを使うという気持はなかったんじゃないか、ことにドキュメント・リトリバルが主だったわけでしょうから、ことさら得体の知れないナチュラル・ランゲージを使っていることはないというふうにご考えておられたんじゃないか。

ところがわれわれのほうは情報という言葉では言いませぬけれども、われわれの側で言う言い方に近く言うという意味内容というものはそういう技術屋さんがお考えになるような情報とはちょっと違うわけですね、そういうところに食い違いがあって、たまたま何かの拍子にせっかくそういう出会いがあってもそのままになってしまった。

渡辺 水谷先生がおっしゃった意味で考えなければならぬという風潮を与えたのは中井浩さんの「思考工学」からですね。その年の秋に、日科技連の数学計画シンポジウムでもとりあげております。

小沢 システムを開発するとき、何が一番参考になったかということから見ていきますと、それは今でも使っていますが、JICST ですよ。この論文は創刊号からずっと借りて使っているわけです。それから水谷先生がやっているので計量国語学というのがあるんです。これが言語関係で非常に参考になりまして、この二つはフルに使っています。

実は 38 年頃ですか、情報処理学会に加盟しようというので「情報処理」という雑誌が出てますね、これは IR だというので飛びついて雑誌を見た途端にがっかりした(笑)ことがあったんですよ。だから、こういう座談会とか特集をやるというのは 4~5 年遅いわけですよ。

渡辺 そういう感じはしますね。もう一つはプロフィット・ベースに結びつきにくい内容なんです。

関根 それではここでそういう話に移りましょう。というのは、たとえばセンサーというのがあったほうがいい。しかしセンサーをつくらうと思うとこれは膨大な仕事ですね。じゃ誰がつくってくれるのか、誰が金を出すのかという問題が出てくる。

渡辺 それは誰かやってくれるだろうと思ってなるべく手をつけなくていいわけですよ。

関根 現在のところ SMART というのが一番完備した IR システムに見えますね。SMART みたいなシステムをじゃどこがつくるのか、どのくらい金を使って誰がどういうふうにつくるのが順当なのだろうか。

小林 本格的にやれば二桁ぐらいですね。

菊池 SMART の予算はたしか年間 7 万 5,000 ドルぐらいでしょう。

今年で 3 年目か 4 年目ぐらいじゃないですか。

関根 あれは SMART になる前からやっていたわけでしょう。

小林 そうです。

関根 その前の投資があるわけですよ、言語関係のシステムとしての。

小林 シンソーラスでも何万ドルとかかっているでしょう。

関根 たとえばウエスタン・リザーブのああいうのをつくるんだって……、大変ですね。だから日本でないのは当然だという気がする。つまり、人力も時間も金もない。

小林 そこまでいくと親方日の丸主義になりますけれども、政府投資がなすすぎるんですよ。民間でやる

のはちょっと無理でしょう。

それで大型計算機にもそういうシステムがないとすれば、機械はあいかわらず使わないということになるんですよ。

渡辺 しかし、政府でやるのも無理みたいな感じがしてきました。これは官民合同でやらないといかんという感じですね。

関根 たとえばシンソーラスだって全般的なものは国語研究所で、機械関係だとか、各産業別とか、それぞれ専門分野に対しては、それぞれになるでしょう。

水谷 制度上、無理ですね、設置の趣旨が違いますからね。

関根 だから結局ないわけですね。

水谷 そう、ないんです。国語問題を解決していくためのそれが、いままではいわゆる国語問題だとか、運動屋の意見に左右されたわけですね、そうじゃなくて、一応それがいいのならそれを裏付けるし、それが悪いんだったら悪いということが示せるような科学的研究をすること、こういうことを言うともた怒られますけれども(笑)。そういうふうに読める条文です。

渡辺 電気試験所はやっぱりハードを追っかけることが精一杯の時代があって、気がついたら少し遅かった?……。

関根 だけどまた電気試験所がやるのがふさわしいかと言われると、別にそうふさわしいとも思いませんよ。

渡辺 このような問題を電気試験所が整理して指摘しておいてもよかったのかも知れませんね。

関根 一番しょっぱなだったから。

渡辺 そう、ハードを重点に行なって、いわゆるコンピュータの機能を完成された形で想定してみるということが欠けたような感じはしますね。

小林 うちで青写真をつくったんだろうと思うんですよ。研究所をつくるという路線だけ入れまして、来年はさし当たり研究費ぐらい出そうと、それと理事長の考えでもあって、大学でもそれに見合う研究をやってほしいと、あるいは学界に持っていくとちょうど合うわけですね。そしたら実用化を含む研究になりますし。

関根 あるいはアメリカみたいにいろんな学界がやってもいいかも知れない。学界関係の IR をやるということを中心にして、それからシンソーラスをやってもいいわけですね。

小沢 そうすると金のかかり方が違いますね。

菊池 学界がやるべきだと思いますね。論文審査に IR がないと各学界ともまずいんじゃないかと思えますね。

小林 特許庁だってそうですよ。あそこは何を根拠にして審査しているか私は非常に疑問に思います。世界のあらゆる分野の特許を審査できないわけですね。ドイツの特許法じゃ過去百年の文献と書いてありますよ。

関根 やっぱアメリカが一番盛んで、いろいろありますね。

しかも実的にやっているでしょう。メドラスにしてもそうだし。金儲けでやっているのは、もとはやっぱり政府の金で研究したやつでしょう。

菊池 そうです。たしか何年間か政府から金を貰って気のきいた研究をやってその結果を……。

関根 だから結構、何か知らんけれども、金を使ってやっていますよね。FBI だったか人間の……。

小林 ええ、海軍の情宣面とか、いろんなところでやっていますね。

水谷 で、割合にアメリカは無駄金使っているようで使っていませんね。

小林 目標としてアメリカ程度でやるとすれば日本はスピードアップするためには無駄金使っても止むをえないと思えますね。

関根 しかも、いまの場合無駄金使わないでかなりできるわけですね。かなりわかっているから。

小沢 少なくとも技術的にはそんなに遅れていないと思います。

関根 考え方としては遅れていないということですね。実際にインプリメントするときには何もないと。

菊池 ただ彼らの場合はやっぱり膨大ないままでの蓄積があると思うのです。マニュアルでやったという、その蓄積はすごいんですよ。それが日本にはないんです。

小林 そのマニュアルの IR 処理技術はケミカルアブストラクトだってそうですね。あれすらいまの日本では機械はできないですよ。

関根 SMART はやっぱり典型的のといっていいでしょうね。ケミカルアブストラクトというのは一つの典型ですね。それからウエスタン・リザーブも一つの典型ですね。ほかにどんなのがありますか。

菊池 メドラスがありますね。あと国防省のドキュメンテーションセンター（もとの ASTIA）などもそうですね。

それから IBM のインフォメーション・リトリバル・センター。

関根 やり方としては KWIC は原始的だというけれども、KWIC 索引を皆に配っても十分役立つでしょう。KWIC はアメリカならほとんどのところで使っているね。日本では三菱原子力がこのあいだ自分のところの文献をやっていたけど、サイテーションとか、KWIC とかいうのは機械的にできることだからどんどんやればいいですね。

日本でも少なくとも KWIC ならプログラムはできているからやろうと思えばできるわけです。SMART をつくろうと思ったら言語翻訳を何とかしてくれないと困るわけだから、機械翻訳をやる必要がありますね。メドラスみたいなのだったらこれはやればできることは明らかだからサッサと金を出してそういうシステムをつくれということですね。

菊池 もう一つ、日本が見落しているのに SDI があると思うんですよ。

関根 それも原理的には問題ない。やればできますね。

渡辺 結局 IR のために金をかけるということよりも社会資本のかけ方というような感じですね。メドラスの例も、IR に金を出すという出し方ではありません。

関根 もうそれだけ金をかけて、人間をかければすぐにもできるというふうなものがあり得る。たとえば KWIC をつくるとか、SDI をつくるということではできる。やはり翻訳みたいなことを根本的にやろうとしたら研究活動をしなくちゃならない。

もう一つはソーラスづくりだとかたとえばソーラスをどんな具合につくろうか。あるいはコーディング・システムをどうしようかというようなことでしょうね。

小沢 われわれ、いろいろやってみた経験からいいますと、最初非常に安易に考えて、アメリカとかそのほかの開発されている手法を外務省がそのままただけば IR はいくんだという考えがあったわけです。実際やってみると自分のところのことをやるのには自分のところに合ったシステムをつくらなければならない。そうすると、システムにしる技術にしる、やっぱりかなり負担がかかる。

関根 だけど、外務省でやっているようなのは一つの典型的なパターンですよ。そしたら典型的なパターンとして別に外務省がやらなくてもほかにそういう

ところがあつたらやってもいいことですね。たまたまやるところがないから外務省がみずからやらなくちゃならなかったということが言えるし、それは研究としても登場しうるんじゃないですか。

小沢 ええ。ただパターンが割合にないですよ。私共がそういうパターンを使おうと思つても。

関根 たとえば通産省で外務省と同じようなことをやろうとしたら、そのパターンをそのまま使つたらいいでしょう。

渡辺 使えないです。

小沢 一つは IR も利用の仕方によって全然違ふんです。

関根 その利用の仕方が通産省にないというのは嘘だと思ふんですよ。いま現在は違ふのかも知れんけれども、通産省でもって外務省でやっているような発想法の IR が用がないというようなことは絶対なくて、やろうと思えばあると思ふですよ。

渡辺 それはそうですけれども、ただ問題は、IR という何かある意味で抽象的なシステムというのは考えられないわけです。で、結局通産省が通産省というオーガニゼーションにもっともピッタリした機能としてのシステムが最初考えられて、そこに必要なテクニックとして IR が採り入れられてくるということになるわけです。

関根 ところで、現在いろんな経験をしていますでしょう。それに基づいて、たとえば、いま考える理想形というのはこんなことであつてそれをやるためにはどのくらい金があるかというどんな具合になりますか。たとえば菊池さんのところは？

菊池 一つの考え方として、外国の雑誌をたくさん取り寄せていますから、それから価値のある論文を機械が自動的に翻訳して、日本語の抄録をつくつて、その抄録から自動的に KWIC のような簡単な索引誌ができて、一方においてそれを磁気テープのようなものにたくわえてリアルタイムのユーザ・サービスができて、その材料を使つて SDI、そういう形のもので……。

小林 公式の青写真は4年後に30億という数字です。

関根 それはてんで安すぎますね。

小林 ちょっとまだ弱いと思ふますね。

関根 たとえばそれでタイムシェアリングがあつてどうのこうのといつたら30億なんて、機械の費用だけで足りないんじゃないですか。そうすると、それに

対するシンソーラスをつくるとか、情報を整理するとか、システムをつくるというのは何にもないんだからそれをつくることを考えるわけですよ。

小林 平屋の長屋に建て増ししているわけです。同じ坪数だからだめなんですよ(笑)。とつばらつて小さくてもいいものをつくるというんでないと本当は改善にならないと思ふんです。

関根 言語翻訳をきちんとやろうとしたらどうでしょう。何がきても一応やるというようなのをやろうとしたら。

水谷 ちょっと金が見積れないんじゃないですか。というのは、私は悲観的に見るんですけども、これはかなり実用文としていい翻訳結果を出そうというふうに考えますとそういうことを研究する人間の育成からはじまらなければならぬわけです。実は大学院ではもう遅いんですね。だからそこから考えたら大変な金になるでしょう。

小林 うちも本格的にやればそういうことになりまますよ。

渡辺 私のところでは量的に本当に徹底しますと、世界中の経済情報を、特許技術情報も含めて洩れなくインプットできるようにするというのが夢です。

もう一つは、それと同時に現在個人の行政官の勉強によってさばっている問題が随分あるのですが、個人の持っている情報を個人の所有にしないで、全体で使えるようにすることです。そのためには、行政イメージに対応した検索の内容をもつシステムをつくる必要がある。それは単なる IR というようなことではなくて、いろんな OR 的な処理をも伴うことになりまます。

小林 それで、デジジョン・メーカーをやつていたらやれなくなつてしまうでしょう。

渡辺 デジジョン・メーカーではありません。デジジョンにつながるための質のよい情報供給機能としてのメカナイゼーションを考えているということです。

関根 さし当たつて、いまはどの辺ですか。

渡辺 そうですね、百年かかるスケジュールの第1日目か2日目くらいじゃないですか(笑)。夢物語でもなく、公式的でもない形でいいますと、昭和47年までに建物は約5,000平方メートルくらい、機械はレンタルベースで年間十億円ないし20億円の見通して検討を進めております。建物は、既に今年の夏から工事にかかりました。

関根 だいたいどのぐらいの容量なんですか、機械は。

渡辺 機械は目下、白紙の立場で研究中です。

関根 現在どのぐらい使っているのですか。

渡辺 本省、特許合計で1億8,000万です。ただしIR 独自の子算はありません。IR について、事前処理の費用が今後問題になります。

関根 そちらのほうはどうですか。どのぐらいの金がかかりますか。

渡辺 何倍か、かかるんじゃないでしょうかね。

小沢 ぼくのほうは割合に皆さんと比べると小じんまりとしているほうでやりやすいのです。でも大変には大変です。うちのほうでいま一番大きな問題は海外との連絡網を計算機を中心に編成することです。これはべらぼうに金がかかります。で、着々とやろうということなのですが、とにかくその国の事情によってだいぶ事情が違います。たとえばアフリカの奥地から緊急を要する情報がはたしてどのぐらいあるかという問題もあるでしょう。

関根 逆からいうと、そんなところで何か緊急なことが起ったときにはものすごい緊急なことにきまっていますから(笑)……。

小沢 それもありますね。それがはたして日本の安全にどの程度ひびくかということがありますが、そうすると経済問題に入るのでね。うちの場合、情報の一件当たりのストレージのコストが高いんです。

関根 何ですか。

小沢 通信料を考えなくてはなりませんからね、情報の価値を金に換算した場合、一体どれだけの金がかかるか、それがコストとの関係でどうなるかということを実際に考えなければいけないわけです。とりあえず、日本の安全とか、海外に行っている日本人の安全を含めてシステムを整備することが大切です、頭が痛いんです。

関根 だいたいどのぐらいですか。

小沢 そう大したことはないですが、42年度から37年計画を立てて、電位情報の Online Realtime System を編成する予定で大型機2台を入れる構想をたてております。これは年間十億ぐらいですむ程度のもなんです。

そういうことからいうと、大したことはないんです。問題は人ですね、これはことに現在 IR をやろうということになると私共の場合痛感しているのですが、とにかく優秀な人がいないんですよ。たとえば論理数学

とか言語学とか、心理学とか、そういった人達で、しかも IR をやろうということに限定されますとまずいませぬね。そういう人の養成ということから考えていかなければならないので、先程関根先生がいわれたように、協会とか大学の養成機関があればこの点がはぶけるのです。

金としてはそういうことで小じんまりとしてはいるのですが、通信網を考えるとべらぼうになる。これはとても 200 億とか 300 億ではすまない。

関根 その 10 億ぐらいの金で間に合うというのは 5 年ぐらいのあいだにそのぐらいの規模にしようということですね。

小沢 3年計画でそのぐらいの規模にしよう。

関根 いま考えて、だいたい理想的だと思えるようなものをしようと思うとそれの何層倍ぐらいになるのですか(笑)。

小沢 100 倍はかかるでしょうね。たとえば人から何から集めたら。ただ日本の産業規模とか、国の事情というか、台所、そういった諸々のことを考えますと、おのずからせいぜいここにいえるのは菊池さんのほうから出た 30 億ぐらいというのが常識的な数字でしょうね。

関根 テクニカルにそういうことでやらなければならないことはそれぞれのところでありませんか。

小林 昨日の会議で出たのは、アメリカでできているのはそのまま線路をつなげばいいんじゃないかという意見がありました。それに対して愛国心だけではいけないと思うので、そのへんでどういう点を押えれば日本でやる必要がはっきり出るかというようなことですね。

関根 それは僕はそういう点非常に好き勝手なことをやるから、たとえばメドラスみたいなものでやっていいならどんどんやたらいいんじゃないかと思えますね。でも日本語の研究というのはやっぱり日本人がやるでしょう。日本に金がないからアメリカへ行って日本語の研究をやるということになるわけね。そうするとやっぱりアメリカがつくってくれちゃうのかな。つくるのは日本人であっても金を出してくれるのはアメリカだから。

水谷 ええ。

関根 そうなるんですかね。そうなくてもソースをつくるというのは日本でやらなければいけないね。それも向うでやってくれるかな。

小林 共同事業でできると思っていますね。うまく

やれば。

関根 いずれにしても徹底的に機械翻訳をやらなくちゃならないことは事実ですね。それからコードのつくり方なんて目標に入れて引っ張り方をどうのこうのということでもってコードのつけ方がなんていうのがありますね。

それから、ファクト・リトリバルということになると、これは言語翻訳ができて IR がシステムがやっていくようになって、それが完全ならおのずからできると思いますか。

水谷 それはできないと思います。それは別に、たとえばそうしますと今の状態ではいわゆる世間様でいっている IR じゃなくて、証明論に近いでしょう。幾何を証明させるとか、ある効率に矛盾がないかチェックするとか、性格はそういうことのほうにかなり近いんじゃないですか。

関根 何に比べてですか。

水谷 いま世間で普通に考えている IR というものよりは初めの段階ですよ。

関根 いま普通いわれている IR は文献検索……。

水谷 ええ、キーワードで必要なものを引っ張ってこようなんていうことは本質的に違うんです。

関根 それも証明問題というか、もっと違うんじゃないですか。

水谷 それはもちろん違いますよ。だけど問題全部のタイプとしてはそういうものに近いんじゃないですか。何か一応広い意味でも論理の系があって、その中でいろいろに変形して、それでパターンになっているものを合わせるということになるわけでしょう。

関根 そういう意味ですか。

水谷 ええ。

関根 だけど、人間はきまったことをやっているわけではないからね。

水谷 だからそれは全然違うわけですよ。

小林 去年その会議に出たんですが、科学技術の場合にまず数値データをおさえるほうが本質じゃないかという気がするんですよ。必ずしも文章構造そのものの決定的解明を待たないでもそれをやるべきであるし。

水谷 もちろんそうです。

関根 やるという意味だってもうどんどん実用的で、たとえば翻訳にしたって曖昧な翻訳したってかまわないですね。人間が読んでワアやっているな、機械また間違っただっていいじゃないから、それはど

んどんやるべきですよ。

小林 結局は論文の最後はデータなんですよ。文章はそれを説明しているだけのことなんです。

関根 説明が論として必要なんであって、データとしては結果にすぎないとすればどっちが重要かというのは又問題になりますよね。

少なくともファクト・リトリバルというのはそんなことなんですかね。

水谷 それから、やっぱり自動抄録というようなこともかなり大きな問題だと思うんですよ。機械翻訳するたって、初めからやみくもに翻訳したって、いくら機械が早くなったってこれはかなわんでしょう。したら一応抄録をつかって、英文なら英文の抄録をつくらせて読めるやつに読ませて、これ訳す価値があるかどうか聞いてからやったほうがかえって得じゃないかと思えますね、抄録の研究というのは今減っていますね。

関根 本当の抄録というのは翻訳ができた上での話になって、セマンティックスに入っているんだから単純に直すという話じゃないわけですね。

水谷 論文の価値があるかどうかということは別として、極端なことを言えば、その論文の中でこの部分は価値があるかどうかということは入るわけですよ。翻訳がとにかく横のものを縦にするか、縦のものを横にすればいいわけですからね。

関根 そういう抄録もあるし、そうでない抄録もあるということになっているね。

関根 ほくも中を判断してということができるとしたら全部コンデンスされた形でできると思うんだ。論文なんか全部書き直してくれて、一定水準のものに直して出るんだということが可能だから、それは随分問題だと思います。それは違い将来の問題ですね。それからあとどんなのがありますか。

渡辺 情報の評価の問題は、ファクトリトリバルの問題につながり、今後多くの実験が必要と考えられます。何かそういうあとから生きる無駄、それが今一番必要ではないかと思えますね。

関根 それは IR というものをもっと豊かにするためにはやらなければいけないことですね。もっとも基礎的な実験だと思いますよ。

小沢 まだ無駄を無駄として認めてくれませんかからね。

関根 そういうことと、もう一つ今すぐにもできるようなシステムをどどんつくることね。クイック

でも何でもいいから。

小沢 そういう面もありますね。私なんかどっちかという、いいハードがなければだめだということですね。

関根 それはもうそういうことを実際にいろいろわかってどうこうしたときに実現するためにはカナモノとこうくる……。

小沢 そういふ点が大きいですね。

水谷 やっぱ最後はヒューマン・コンピュータのいいものがないとだめでしょう。それを養成することについて大学であまり真面目に取り上げていないんじゃないですか。

関根 少なくともいまいえることは、IR なんてい

うのを教えてる学校はないはずだね。

渡辺 役所の一つの行政管理なり、会社の経営管理なり、物の考え方をかえていくわけでしょう。そうすると今の大学教育の中にテクニシャンの養成というよりは、物の考え方のリーダーを育てることにより大きな期待をもちたいと思います。

関根 大学というのはそういうところだと思うのですけれども……。それは、IR に関してだけいえるんじゃないくて、物事全てがそうだから。まあ、IR というのはいままで何にもないものだから、とくにそういう感じが強いわけですね。それではこのへんで……。

(41年7月8日)