

Euralex の参加報告、INFOTERM, GMD, IINTE 訪問報告
(ヨーロッパの言語活動の報告)

田中康仁(姫路短大)、横山晶一(電総研)、荻野孝野
山階正樹(NTT研究所)、林 達也(富士通研究所)

Euralex(ヨーロッパ辞書会議)に参加(研究発表等をした内容ならびにヨーロッパの各研究所を訪問した内容を発表する。ヨーロッパの研究所はウインのINFOTERMとドイツのボンにあるGMD、ポーランドのIINTEを訪問した。これらの研究機関での活動状況を調べた内容を報告する。

Report on Euralex'88 Congress and Visit to
INFOTERM, GMD, and IINTE

Yasuhito Tanaka(Himeji college), Shoichi Yokoyama(ETL)
Takano Ogino , Masaki Yamashina(NTT), Tatuya Hayashi(Fujitsu)

The 3rd Euralex (European Lexicography) International Congress was held by Akadémiai Kiadó at TOT (Mezőgazdasági Szövetkezetek Háza) in Budapest, Hungary, 4-9 September, 1988. This congress is not so popular in Japan, and in this report we describe the history of this Congress. And then, the content of the program is surveyed.

Some of the authors visited GMD in West-Germany, INFOTERM in Austria, and IINTE in Poland before the congress. They will introduce the status of computational linguistics in middle and eastern Europe which is rarely known in Japan.

I Euralex (ヨーロッパ辞書学会) 参加報告

1) 開催日, 場所, 日本からの参加者

開催日: 1988年9月4日~9月9日

場所: ハンガリーの首都 ブタペスト

T・O・Tホテル

日本からの参加者: 横山晶一(電総研), 荻野孝野, 山階正樹(NTT研究所), 田中康仁(姫路短大), ペーレント・メリー(東工大), 平井さん(ペーレント・メリーさんの秘書) 以上6名

日本人としてはこのほかドイツのチュービンゲン大学の江沢建之助教授が参加した。

2) 参加者の国別人数, 主要組織

国名	人数	国名	人数
ソビエト	42	チェコ	4
イギリス	37	スウェーデン	3
ハンガリー	25	オーストリア	3
西ドイツ	23	オーストラリア	3
アメリカ	20	デンマーク	3
イタリア	9	ベルギー	2
オランダ	9	スイス	2
イスラエル	9	マレーシア	2
フランス	8	ユーゴスラビア	2
ブルガリア	7	イラン	1
アイスランド	6	アルバニア	1
フィンランド	6	中国	1
日本	6	ノルウェー	1
ポーランド	5	南アメリカ	1
東ドイツ	5	ブラジル	1
スペイン	5	31カ国	252名

参加人数上位国における組織

ソビエト	Institut russkogo jazyka	AN USSR	15
イギリス	Oxford University 出版関係		13
イギリス	Collins・Cobuild 関係		6
イタリア	Instituto Di Linguistica Computazionale		5
ハンガリー	MTA Nyelvtudományi Intezet		4
アメリカ	IBM TJ Watson Research		4
ソビエト	Institute jazykoznanija	AN USSR	4
西ドイツ	Universität Erlangen-Nürnberg		3
西ドイツ	Ernst Klett Verlag		3

参加国の人数は、政治的な近さ(共産圏の参加)、距離的な近さ、辞書作りの伝統の深さなどを物語っているようである。

距離的な近さでいえば、ポーランド、チェコ、ソビエト、西ドイツ、オーストリア、イタリア、ユーゴスラビアなど、隣接諸国からの参加が多く、我々が知り合った何人かは、飛行機ではなく国際列車を利用しての参加であった。同じ陸続きで地方に行く気安さで、国際交流ができるヨーロッパ事情を感じた。

政治的な近さ、参加のしやすさのせい、表でわかるように、ソビエト、ブルガリアなどの参加が目立つ。ソビエトが一番参加人数が多いことに気付いたが、後で考えてみると、全体的なシンポジウムや講演でもロシア語での話しはほとんどなかった。全体講演で、プログラムで in Russian と書かれていたものも、当日は英語で行われた。それに較べて、参加人数が1けたにもかかわらず、シンポジウムではフランス人は必ずフランス語になったのは印象的である。ヨーロッパにおけるフランス語の位置は大きい。

辞書作りの伝統という意味では、イギリスの参加と発言が大きい。実際的人数よりもっと多く感じさせられた。これは、発表件数の多さ、発表内容の重み、EURALEX におけるイギリスの地位 (Subcommittee 4名のうち、3名がイギリス人) などからくる総合的な印象によるものである。イギリスからの参加者37名の所属の内訳は、出版関係者27名、教育研究関係者5名、その他5名で、具体的に辞書作りに参加している人たちが圧倒的に多い。出版関係者の中でも、特に、数の上でOxford 出版関係者13名、Collins・Cobuild 関係者6名が目立つ。発表関係でもCollins・Cobuild 関係者の若い人や女性の発表など、内容的にも印象深いものが多かった。

参加上位国を組織別にみると、それぞれの国によって、参加者の所属する組織や専攻に特徴があった。イギリスは辞書出版関係者、イタリアはZampolliの所属機関からの参加者、アメリカは、cognitive science, computational linguistics, psychologyを専攻する大学、民間組織の研究者(特にIBM Watson 研究所から4人、なお、IBMについては、デンマーク、フランスをいれると6人)、西ドイツは、ジューメンスの一人を除いてほとんど大学の研究者という具合であった。日本についていえば、31参加国の11番目の参加人

数で、参加者の組織に偏りはなかったが、どちらかというところ、計算機関係が中心であった。

3) 発表内容

一般セッションは、10のサブセッションに分かれ、同時に4つのセッションが進行する形で進められた。10のサブセッションの内訳は次の通りである。

PL Prelexicographical Theory-General (辞書に関する一般的な理論) ……………	20
BL Prelexicographical Theory-Bilingual (二カ国語辞書に関する一般的な理論) ……	10
GL General Lexicographical Topics (辞書学に関する一般的なトピック) ……	8
CS Case Studies (ケーススタディ) ……	15
CO Computational Lexicography (計算辞書学) ……………	17
PR Project Reports (プロジェクト報告) ……	3
DU Dictionary Use (辞書の使用) ……	7
TT Terminology and Translation (専門用語と翻訳) ……………	12
LK Lesser Known Languages (あまり知られていない言語) ……………	12
PH Phraseology (語法) ……………	3

計 107

このうち、CO(Computational Lexicography)に限って、発表者を国別にみると、次のようになる。

アメリカ	4 (うち、IBM関係者3名)
オランダ、西ドイツ	3
ソビエト、日本	2
ブルガリア、チェコ	} 1
ハンガリー、イタリア	

イギリスのCollins・Co-build関係はかなりきちんとした計算機活用がされているが、あくまでも辞書作成のための計算機活用ということで、主流が辞書記述者、編集者になるためか、発表は辞書記述の内容そのものに関する部分で、COでの発表はなかった。

4) ヨーロッパ辞書会議の概要

辞書学(lexicography)は、ヨーロッパにおいては、単一言語、二国間言語、語源などのさまざまな辞書を作成する際に必要な知識、語彙、語義などに関する基礎的な学問としてかなり古くから盛んであり、この学会も、それらの成果を発表する場として設けられているものである。

この学会の設立は1977年で、1983年に第1

回の会議をExeter(イギリス)で開催し、1986年に第2回の会議をZürich(スイス)で開催した。これらはそれぞれ開催地の名前をとって、LEXETER ZÜRILLEXと呼ばれ、予稿集も出版されている。第2回以降はCOLINGと同じように、偶数年の隔年開催であるが、開催地がCOLINGと重なったのは偶然の一致である。ちなみに、辞書学会の次回開催地は1990年にPisa(イタリア)の予定で、PISALEXという名称に決定しているが、COLINGは同じ年に、Helsinki(フィンランド)で開かれる予定である。

今回の会議はBudaの丘の上にあるTOT(Mezőgazdasági Szovetkezeti Háza)において開かれた。これは農業国であるハンガリーの農業協同組合によって建設されたホテル兼会議場で、参加者の多くは、ここで会議に参加するとともに宿泊も行き、交流を深めることが出来るようになっている。今回の参加者は、開催地が東欧圏だけあって、東西の幅広い国々にまたがっており、地元ハンガリーはもとより、西ヨーロッパの多くの国々(イギリス、ベルギー、フランス、オランダ、西ドイツ、デンマーク、イタリア、スペイン、フィンランド、アイスランド、オーストリアなど)、東ヨーロッパの多くの国々(東ドイツ、ブルガリア、ユーゴスラビア、チェコスロバキア、ポーランド、アジア圏を含むソ連など)、さらにアメリカ、日本、イラン、マレーシアなど多岐にわたった。

辞書学という学問の性質上、女性の参加者も非常に多く、男性と女性の割合は6対4くらいであった。また、参加者の専門も、伝統的な辞書編集者、コンピュータを用いた辞書編集に携わるもの、人工知能研究者、出版関係者、大学教授(実践的な言語学者が多い)など幅広いが、会議全体としては、文科系の色彩が濃いものであった。従って、Proceedingsも、ごく簡単なドラフト程度のもので、正式なものも追って刊行の予定である。

プログラムは、開会直後に行われた、ハンガリーアカデミーのF.Kieferによる基調講演、「言語学的、概念的、百科事典的知識」(Linguistic, conceptual and encyclopedic knowledge)と2つの合同セッション、2つのシンポジウム、パネル討論のほか、一般セッションが、4つの並行セッションとして行われた。この会議の公用語は、英語、ドイツ語、フランス語、ロシア語であり、それぞれの言語による発表、討論が活発に行われた。

4.1) 会議の内容

4.1.1) 基調講演

基調講演は、ハンガリーアカデミーのF. Kieferによる“Linguistic, conceptual and encyclopedic knowledge: some implications for lexicography”(「言語学的、概念的、百科事典的知識：辞書学との関係」)と題するもので、この学会参加者すべてが共通して抱えている問題点を具現化し、要約した基調講演にふさわしいものであった。

Kieferの立場は辞書編集者のもので、新しい語彙項目が生成され、それが辞書に取り入れられた後に言語学的(または語彙项目的)、概念的、百科事典的に生じる「ずれ」の例を上げ、それらにどう対処していくべきかを(解決策は示さなかったが)論じたものである。最初の、言語学的な知識の「ずれ」の例ではたとえば、“cup”の例がある。“cup”の典型的な意味は、「ガラスでできた円錐台形のコップ」という意味であるが、形状、サイズ、材料などの異なるものが作られるようになっていく(たとえば「プラスチックのコップ」)。これによって典型的な意味からの「ずれ」が生じる。また、そのほかの単語でも、比喩的な意味の場合には意味のずれが生じている。

次に、概念的なずれがある。これは、ある単語が参照している内容にずれが生じる場合である。たとえば、“opera house”は、米語では人間まで含んだ概念であるのに対して、英語では建物しか表わさないという例がある。“school, university”なども建物なのか、地名なのか、中身(学問、講義など)なのかといった問題がある。つまり、ある単語の表わす概念の範囲といったことが辞書を作る上で問題になる。

百科事典的な知識のずれという問題では、「足のあるヘビ」という例が挙げられた。つまり、これはファンタジーの世界のことであるか、あるいは科学上の新発見であるかどうかであって、後者の場合には、(もし辞書が百科事典的な記述をしている場合には)内容の書き換えなどが必要と思われる。

このように、この基調講演は、辞書の編集者が抱える問題点を特に記述のずれまたは変遷に絞って論じたものであるが、計算機上での知識表現などとも共通の問題点を含んでおり、興味深いものであった。

4.1.2) シンポジウム「辞書とコンピュータ」ワーキンググループ

これは、前回のZürichにおける第2回大会において設置された「辞書とコンピュータ」ワーキンググループのこれまでの活動報告と今後の取り組みについて討議するものであった。内容は、このワーキンググループ設置の趣旨、プロジェクト“computational lexicology”(計算語義学)の提案と参加国募集、用語についての情報などである。計算語義学のプロジェクトについては、一応英語を中心として、フランス語、ドイツ語、イタリア語、ロシア語、スペイン語の計6カ国語について研究を行うことになった。そのほかの言語については、無視するというのではなく、このワーキンググループとしては、一応これだけの範囲で、種々の問題点について検討していこうという趣旨のようである。そのために、用語についての情報網を完備するために「辞書ワークステーション」を作って各国に配布するなどという提案もなされたが、まだ具体化までには時間がかかりそうである。また、それに関連して、各国の電子化辞書プロジェクトについてのアンケートも配布された。

このアンケートに基づいて、上記ワーキンググループの整備・発展を行う心積もりのようである。

4.1.3) パネル討論

パネル討論のテーマは、「辞書の記述に文法事項を含むべきか否か」というものであった。文法事項を含むべきでないとする立場からイギリスのR. W. BurchfieldとフランスのA. Reyとが発言し、文法事項も含めるべきであるとする立場からアメリカのA. MakkaiとイギリスのR. Ilsonが発言した。前者の立場は、辞書の限られたスペースの中にどの程度の新語を盛り込むかを取捨選択することに苦心しなければならない現状では、文法的な事項は文法書または文法事典といったものに任せて辞書は語義の記述に専念すべきであるというものであった。それに対して後者の立場は、語義記述のみに限っても、そこには必ず文法情報を書き込む必要があり、しかも文法情報の記述なしには説明が難しい辞書記述もあるというものであった。

4人のパネリストの主張のあと、会場の参加者からも発言を募り、幾人かの参加者から発言があったが、会場の大勢は、辞書記述には文法事項を含めるべきであるという立場をとっていた。最後に、約150人の参加者について、どちらの立場をとるか、挙手による

投票が行われたが、文法事項を辞書記述から排除すべきであるという立場のものはわずか24名で、残りはすべて文法事項を含めるべきであるという立場であった。

4.1.4) 一般セッション

一般セッションは、4つの並行セッションが同時進行していたために、すべてについて述べることは不可能である。ここでは、主としてCO(Computational Lexicography, 計算辞書学)のセッションの概要を述べる。

COのセッションは、その名の通り、辞書学でカバーしている学問の範囲内に何らかの形でコンピュータを使用したものであるが、著者の立場によって、その内容は大きく3つに分けられる。

第一は、自然言語処理や知識表現など、計算機科学分野の研究者で、人工知能関係のシステム用の辞書を指向する立場にある者である。従って、機械指向型の電子化辞書という面で、EDRの目指す方向に最も近いと考えられる。

第二は、辞書編集のためにコンピュータを用いるという立場の研究者である。文科系の人々がコンピュータを使いこなすのは容易ではないし、そのための環境も整っているとは言い難いが、これに関連した発表も少数ながら何件あった。

第三の立場はその中間ともいえるもので、辞書編集のためのコンピュータによる支援環境を構築しようという立場から、さまざまなシステム作りを試みている。以下では、各々の論文の内容に立ち入って、やや詳しく解説する。

まず第一の立場に属するものとしては、COの最初に書いてあるKitchovitch(ブルガリア)の辞書管理システムについての発表が挙げられる。

Kitchovitchは、従来のデータベース管理システムでは十分な辞書を作れないとして、独自のデータモデルに基づき、IBM-PC上に、木状の辞書を構築している。この辞書は英語、フランス語、ロシア語について(混合か別々かは不明)作られている。辞書の内部は3層から成る木状で、まず、単語の語幹(メモリ節約のため語幹のみを格納してある)の下に品詞と同義語・類義語のエントリが来る。品詞の部分には名詞動詞などの品詞名が書かれ、さらに品詞名の下に、たとえば名詞なら、性・数などの情報が書かれる。同義

語・類義語の部分には、同義語のエントリ、その格などが書かれる。現在のところまだ、形態素や構文解析レベルにとどまっているが、将来的には意味解析まで拡張する意向のようである。辞書を木状にするという類似の考え方は、Viks(エストニア)の発表にもみられる。Viksは、いくつかの伝統的な辞書を電子化して、それを辞書データベースとして利用する構想もっているが、そのためには辞書記述用の形式言語の開発が必要であるとし、それらによって作られた辞書は、Kitchovitchのものとはやや形が異なるが、やはり木状になると述べている。

データベースの側面からみると、上のViksの論文もそうであるが、言語学者や、辞書作成者を支援するための、いわば第三の立場にある発表もいくつかあった。たとえば、van der Woudenの発表は、オランダ政府の援助のもとに設立された辞書情報センター(Center for Lexical Information (CELEX))について述べたものである。

CELEXは、主として純粋および応用言語学の研究者を対象とした、多言語、多機能で、種々の理論を網羅した辞書データベースをつくるもので、まだそれほど大規模なものは完成していないが、1989年の初頭には稼動し始めるとのことである。

データベースでは、別のセッション(TT)になるが、情報処理関係の専門用語辞書に関するチェコ語のデータベースを作り、その中で類義語の処理について論じたMachova(チェコスロバキア)の発表があった。しかしながら、これはシステムとしてはそれほど大規模なものではなく、内容も、まだ単語を羅列したレベルにとどまっているように思われた。

また、機械翻訳や人間の翻訳のための辞書データベースを作る動きもある。Calzolari(イタリア)の発表した論文は、英語とイタリア語それぞれの機械可読辞書を融合して、二国間の翻訳用データベースを構築するものであった。これは、上記のvan der Woudenのシステムと同じように、言語学や辞書学、さらには言語教育用データベースまで意図したものである。

機械翻訳用の辞書という立場からは、フランス語とドイツ語、チェコ語とロシア語という、どちらも系統的には近い言語のための機械翻訳用辞書の発表があった。まず、Heid(西ドイツ)は、LFG(Lexical Functional Grammar)(語彙機能文法)の記述を用いた、独仏、仏独機械翻訳用辞書について発表

した。LFGの特徴や、言語の系統が近いことなどから、その記述は比較的浅いレベルにとどまっていたが、ドイツ語とフランス語の語順の違いや、曖昧性の解消をねらった辞書記述法などを（小規模ながら）試みていた。Oliva（チェコスロバキア）によるチェコ語とロシア語との間の機械翻訳システム用の辞書も、やはり系統的に近い言語であることも手伝って、文法的な記述をなるべく辞書に依存した格好に書き直して処理するのが目的であるように思われた。

一方第二の立場からの発表はそれほど多くない。ここでは、その中から一つだけ例を挙げる。Spears（アメリカ）の発表である。Spearsは、熟語辞書を作る場合の辞書の配列の問題について論じた。熟語辞書は、前置詞や冠詞を含んでいる場合に、どのエントリのところに配列するかは工夫を要する。また、固有名詞の取り扱い、さらに、ハイフンやコンマを含んだ熟語の配列をどうするかという問題もある。

Spearsは、一つの試みとして、語間のスペースやコンマを圧縮してそれを一つの単語と考え、全体をソーティングするというやり方をマイコンレベルで試みている。辞書の配列の問題は、辞書編集者にとっては大きな問題の一つであり、類似の問題が他のセッションでも論じられていたようである。

第三の立場をとるものの一部はすでに上に述べたが、辞書編集者を支援するためのシステムを作るという立場をはっきり打ち出していたのがIBMによる2件の発表であった。まずNeff（アメリカ）とCantor（デンマーク）による発表は、辞書編集のためのワークステーションの仕様を述べるようなもので、IBMで開発したWordSmith, Lexical Database, Terminology File System (TFS) といった辞書編集者用のツールを、ワークステーションのウィンドウシステムとして実現できるような、辞書編集者用エキスパートワークステーションの紹介とでもいえるような内容であった。もう一つの発表は、Ravinらによるもので、こちらは、電子化したシソーラス（The New Collins Thesaurus, 1984. 16, 800エントリに対して287,000語を収容）を、辞書編集者が改定するに際しての問題点を述べるものであった。Ravinらの着眼点は、シソーラス中に埋もれている情報をどのように明確化するかということと、シソーラスに含まれている不整合性をどうやって（半）自動的に抽出するかということにあった。ま

た、異なるエントリの間で類似の記述がある場合にはそのインタセクションをとることによって、メモリを軽減する手法も示した。IBMでは、これらのシステムを将来的には自然言語処理などと結び付ける意向も持っているようである。

まだIBMのシステムほど大規模ではないが、辞書作成のためのPCレベルの発表としては、Vollnhals（西ドイツ）によるSiemensのものがあつた。Vollnhalsのものは、どちらかといえばDesktop Publishingで辞書を作ることを意図したもので、たとえば英語以外のヨーロッパの言語にあらわれる種々の文字（ウムラウト、アクセント）を画面上にフォントとして出力し、プリンタにも出せるようにするといったシステムをウィンドウマネージメントを含めて論じていた。

そのほか、ニューメディア関連では、CD-ROMについての発表もあった（Wekker & Lemmens（オランダ））が、これはCD-ROMに関する単なる解説で、研究要素として特に目新しいものはなかった。

最後に、日本からの発表について少し述べる。日本からは、田中と横山が1件ずつ発表したが、プログラムの都合で、mixed sessionでの発表となり、発表時間もやや短めで、残念ながら十分に意をつくした発表とはならなかった。田中は、多義語の曖昧性を知識データを用いて解消する手法を示した。また、横山は、電子化された冊子体辞書から日本語のアクセント情報を抽出し、それを統計的に処理した結果について、特に外来語のアクセントを中心に述べた。

4.1.5) Euralex の感想

(1) EDRおよび日本への関心について

今回の学会参加者の間では、コンピュータ関係のものは別として、日本EDRの設立と活動がそれほど知られているとはいえない。しかしながら11月の「電子化辞書国際会議」のパンフレットが全部捌けたように、この活動が知られれば、この会議の参加者の関心を引くことは疑いない。

今回の会議では、もう一つ、日本からの参加が参加者のひそかな関心と呼んでいたようである（第1回Exeterでの会議に三省堂の上西俊雄氏、第2回Zürichでの会議にチュービンゲン大学の江沢建之助准教授が参加）。日本人は、髪の色や肌の色で、このような学会では目立ちやすいが、遠い日本から何をしにきたかということが注目を集めたようである。

(2) 国際会議の運営について

今回の会場は、宿泊と会議場とが（多くの参加者にとって）同じ場所であったので、多くの参加者の間に親しい雰囲気が生まれた。また途中の遠足（王宮美術館の見学）、コンサート、また度々の共同ディナー、レセプションなども皆の協調を深めるのに役立った。ランチ、コーヒーなどはあらかじめチケットが渡されて飲食できるシステムであり、農業国ハンガリーにふさわしく、豪華なものであった。係員の対応なども優れていた。

(3) 参加者および発表者について

最初にも述べたように、参加者は大きく分けて、文科系からは伝統的な辞書編集者、出版社、大学や研究所における言語学研究者で、理科系は主として計算機メーカや人工知能関係の研究所からきた者が多かった。会議の運営事態はいわゆる「長老」たちが行っていたが、発表は積極的に若手にやらせ、それを適当に補佐するという姿勢が目立った。辞書編集に携わる者は女性が多いことを反映して、この会議も女性の参加が非常に多かったが、発表や発言などもこれらの女性が活発に行っていた。特に、辞書

編集に実際に参加している若い女性が生き生きと発表・発言している姿は日本の学会でも（一部を除いて）あまり見られないことであった。

言語学の研究者は、この学会が辞書学会であるために、実際のデータに基づいた発表が多く、少数の例外的な事項を掘り出してそれを理論的に論ずるといふ雰囲気よりは、実際のデータに即して悩むという、応用言語学的な色彩の強い参加者・発表が多かったようである。しかしながら、並行セッションのため、このような発表をそれほど多く聞けなかったのは残念である。

(4) 言語について

何度か述べたように、公用語が英語、フランス語、ドイツ語、ロシア語であり、それぞれの言語での発表が行われ、質疑応答も（発表言語とは必ずしも一致しない）これらの言語で行われたので、日本からの参加者は、皆言語のハンディを感じさせられた。特にシンポジウムによっては、英語とフランス語とで丁々発止とやりあい、また会場参加者も当然のようにそれを受け入れているのを見聞して、語学的なハンディが予想外に大きいことを痛感した。

II INFOTERM(オーストリア、ウイン)

(1) 訪問日、場所、訪問者

訪問日：1988年8月29日

訪問場所：INFOTERM

Affiliated to ÖN
(Austrian Standards
Institute)

Heinestraze38 Wien2
Austria(A-1021)

訪問者：山階正樹(NTT)、林達也(富士通)
田中康仁(姫路短大)

相手側受入者：Dr. Christian Galinski

(2) 訪問目的

- ① INFOTERMの活動を知ること
- ② ウイン学派の活躍した、ウインを知ること。

(3) INFOTERM訪問内容

INFOTERMはUNESCOの下部機関として設立され、国際的なターミノロジー活動の機関である。

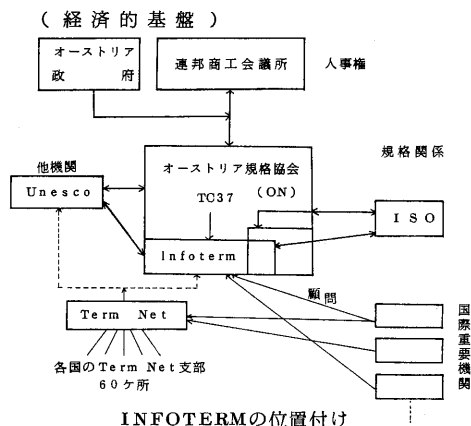
INFOTERMの活動はターミノロジー学の教育、研究、活動、ドキュメンテーションのための情報センターとして、また情報伝達および情報分析センターとしての役割を果たしている。

例えば、各国の、また国際レベルでのターミノロジー活動の顧問となり、必要な助言を与えることを行っている。またターミノロジーに関する各種の大会、国際会議の開催等を行っている。

INFOTERMはISOとも協体制にあり、ターミノロジー活動の国際ネットワーク(Termnet)センターである。

しかし、経済的基盤はオーストリア連邦経済省、連邦商工会議所、オーストリア規格協会に依存している。人事権はオーストリア連邦商工会議所にある。メンバーはガリンスキー所長以外に10名程度である。各国からの研究員も受入れており、日本からも平山(アイ・エヌエス株式会社)(女)、が参加して研修を受けていた。

INFOTERMの位置付けを図示すると、次のようになる。



ターミノロジーは我々は単純に用語のことであると思いがちである。これも一つの側面であるが、ターミノロジーは概念に付けられた名前であるとも考えることもできる。

INFOTERMでのターミノロジーは「概念や用語の体系化の理論や活動」と考えている。この考え方はオーストラリア人オイゲン・ヴェスター(Eugen Wuster 1898~1977)によって、ターミノロジー学として創始された。この考え方はDr.フェルバー氏によって発展され、現在Dr.クリスチャン・ガリンスキー氏に引き継がれている。

最近では科学・技術の進歩がめまぐるしく、新しい概念が次々と発生し、又古い概念が新しく変化することもしばしば起っている。このため膨大な数の新概念、新しい用語が次々と作り出されている。又、既存の用語が変化するという事も発生している。このため概念の明確化、用語の調整が必要になってきている。

また一方では機械翻訳、文章理解、人工知能の研究の中から用語や概念の取り扱い等が重要な問題となってきた。用語や概念の考え方を利用して人工知能システムの知識データとして辞書や用語の項目をもちいることが考えられている。

INFOTERMではターミノロジー活動の中で標準化、規格化、シソーラス、クラシフィケーション等にも力を入れている。

世界各国との情報交換を行い標準化・規格化を行っている。この組織としてTermnetが作られている。

これは通信回線を使ったものではなく、各国との連絡網としての活動である。

我々は INFOTERM の訪問によりターミノロジーについての基礎的知識を得た。

しかし、INFOTERM は国際機関でもあるため具体的作業をコンピュータ処理を用いて行っているというものではなかった。常にトップダウン的なアプローチでの考え方であった。

日本ではボトム・アップ的なアプローチが強い面があるので少しの足りない側面があった。これは立場の違いかもしれない。

INFOTERM は映画「第3の男」で有名になった大観覧車の近くにある。

INFOTERM はターミノロジー理論の研究や各国のターミノロジー活動の助言、指導、調整、標準化等を目的としている。これ迄に2万~2万5千語のターミノロジーデータバンクが彼等の手法に基づいて構築されて居り、そのデモを見学した。概念や用語の体系づけや規格化の方法論について参考になるものと思われる。】
林の報告書より

(1) 見学した Austrian Standards Terminology Data Bank には約25,000語の用語についての概念的な定義、概念関係が収録されており、それらの用語は定義、概念関係によって検索可能である。ここに収録されている定義等は数人の合議で決定するとのことであり、非常に時間のかかる地味な方法であるが、精度の高い知識として計算言語学の分野でも利用可能と考えられる。

(2) Dr. Galinski は最初に手間がかっても正確な情報を体系化しておくことが重要であり、Filing System においても同様に最初から規格化した情報を蓄積していくことが最終的には有利であると述べている。

(3) そのため、自然言語処理技術によって生のドキュメントを規格化された表現(機械処理可能な表現)に変換して蓄積していくメリットは大きく、自動索引等の情報加工技術は上記の観点からも非常に重要な技術と考えられる。】

山階の訪問内容

田中は林、山階とは別に同じフロア内で開かれたワークショップに参加した。ワークショップでは概念や知識体系を計算機のために作ってゆくにはどのようなことをすればよいかという議論がなされた。

議論内容の項目は次のとおりである。

- ① 抽象化の方法(タームバンク、ファクトバンク)
- ② 表現方法
- ③ 獲得方法
- ④ 体系化認知方法
- ⑤ 歴史と科学
- ⑥ 応用分野
- ⑦ 各国の活動と組織

これらの項目をあげ、さらに詳細化はどのようにするかと話し合われた。

アプローチは常にトップダウンであり、日本人のアプローチとは異なる面があり大変参考になった。このワークショップは今後も続けられ、さらに詳しい内容に発展させるようである。

(4) ウインの町について

ウインは音楽の都、バロックの都、文化の都、建築の都、森の都、歴史の都、である。

ハプスブルグ家が神聖ローマ帝国の皇帝としてオーストラリア・ハンガリー帝国時代までの640年間、多民族をも統一してきた。

ウインは数多くの観光名所がある。

特にシェンブルグ宮殿、ベルベデーレ宮の見学、国立オペラ座でのオペラ観劇、市立公園のヨハン・シュトラウス像やモーツァルト像、ゲーテ像、などの見学等楽しいものが多い。

G. M. D本部(ドイツ国立情報処理研究所)

(Gesellschaft für Mathematik und Datenverarbeitung)

(1) 訪問日、場所、訪問者

訪問日: 1988年8月31日

訪問場所: G. M. D 本部応用情報工学研究所
Dept. for Applied Information Technology

場所: Schloz Birlinghoven
D-5205 Sankt Augustin
in West Germany

ボンから車で30分位 東に行った所にある。

訪問者: 山階正樹(NTT)、林達也(富士通)
田中康仁(姫路短大)

相手側受入者: Dr. M Agi 国際部日本担当
Ms. Dr. Maria Theresia

Rolland (グループリーダー)

(2) 訪問目的

① GMD で作られたシソーラス・データベースを知るため。

(3) 訪問内容

GMDは1968年に設立された国立機関で、現在の人員は1300名、そのうち700名が本部に属している。必要経費は、全体の約90%を連邦政府が負担し、残り10%を機関の所在地である二つの州(Nord Rhein West Falen, Hessen)が負担している。このGMDでは、最近EC(欧州共同体)の委託を受けて、3年間の年月をかけてEC圏内のシソーラスデータバンクを開発した。データバンクは500種類のシソーラス情報から成り、10分野51領域にまたがっている(一般45, 情報25, 電気機械40, 金属科学29, 宇宙気象36, 農林64, 医学28, 土木建設61, 社会科学145, 文化芸術24)言語別では、英語222, 仏語92, 独語83, オランダ語5, イタリア語2, デンマーク語1, マルチ語92となっている。現在このオンライン検索システムが構築されて稼働して居るのでそのデモを見学した。このように多数のシソーラス情報を集約したシステムは他に例がなく、我々にとっても有益な情報源である。林報告内容より

① Thesaurus Data baseは、欧米における約500種類のシソーラスの書誌情報をデータベース化したものであり、シソーラスに関する種々の情報をオンラインで検索できるシステムである。このデータベースはECの要請により構築されたものであり、そこでのECの目的は①情報システム間のデータ交換の容易化、②情報検索のためのシソーラスの開発と利用、③多言語シソーラス開発のベースである。上記の目的に向けて、書誌情報の徹底的な収集から開始するというドイツ的で堅実な方法で研究が進められている。

② 上記のシソーラスのうち、約80種類はMultilingual thesaurusであり、一つの概念体系について、複数の言語をMappingする作業が実際に進められている。

③ このデータベースの内容は“Thesaurus Guide”として刊行されており、本資料により欧米における種々の分野におけるシソーラスの内容を知ることができる。

山階報告内容

Dr. M. T. Rollandさんは美人で優秀な人であった。オーバヘッド・プロジェクターと端末を使い、シソーラス・ビブリオグラフィ・データベースの説明を行った。オペレータを配置し、説明を行った。シソーラスが各国でどのように作られているか、分野別、言語別、マルチリンガルか、モノリンガルかもわかるようになり、出版元までデータ・ベースに入っている。幾つかの表により内容を説明する。

Number of the thesaurus descriptions

Thesaurus descriptions:
-the detailed descriptions in the main part
-the concise descriptions in the annexes

Main groups	Main part Annexes			
	Multi-lingual	Mono-lingual		
A General	7	38	45	40
B Information Area	5	20	25	28
C Mathematics and Physico-Technical Sciences	7	31	40	52
D Physico-Chemical Technology	7	22	29	50
E Astronomy and Geosciences	2	34	36	48
F Agriculture and Nutrition	12	52	64	40
G Biomedical Sciences	6	22	28	31
H Regional and Environmental Sciences	13	48	61	67
J Social Sciences	30	115	145	119
K Culture and the Arts	3	21	24	36
	92	406	497	511

Number of thesauri per language

total no of thes.	main groups	languages						
		mu	en	fr	de	nl	it da	
45	A General	(+1) 6	20	9	6	4		
25	B Information Area		5	13	4	3		
40	C Math. and Physico-Technical Sciences	(+2) 5	16	6	11			
29	D Physico-Chemical Technology		7	13	3	6		
36	E Astronomy and Geosciences	(+1) 1	22	11	1			
64	F Agriculture and Nutrition		12	22	22	7	1	
28	G Biomedical Sciences		6	14	3	5		
61	H Regional and Environmental Sciences	(+3) 10	29	6	12		1	
145	J Social Sciences	(+4) 26	61	23	28		2	
24	K Culture and the Arts		3	12	5	4		
497		(+11) 81	222	92	83	5	2	1

Frequency of languages in the multilingual thesauri

No of ML groups	main groups	19 different languages																	
		en	fr	de	it	al	es	pt	gr	da	no	sv	ru	ar	in	ja	ch	pe	br
6	A	5	5	1	1	1													
5	B	5	3	2			3												
5	C	5	3	5	1									1					
7	D	6	4	7										1					
1	E	1	1	1	1														
12	F	12	7	6	3	1	6	1											
6	G	6	5	4	1	1							1						
10	H	10	10	5	1	1	4	1	1	1				1	1			1	
26	J	26	20	10	2	2	8	2	1	1	1			1	2	1		1	1
3	K	3	3	1															

81 Optimum:
groups : 8 3 3

個々のシソーラスの内容ばかりでなく、分野、言語、Mono, Multiシソーラスの区別等も指定できる。そして各種統計資料も出力できるようになっていた。

研究に先だつてまずデータ・ベースを作り、それから方針を立て体系だてて分析を行ってゆくという方針はさすがドイツ人という感じがした。我々の研究にも参考になる点が多々あった。

マルチリンガル・シソーラスまで完全にできているのではないかと思ったがそこまでは完成していないようであった。

(4) GMD (ドイツ国立情報処理研究所) 訪問にあつて

GMD 訪問にあつては東京にある GMD 東京事務所から何回となく Telex や手紙を送ってもらい、我々の訪問を円滑に進めることができた。東京事務所の高久かおりに深く感謝する。

Dr. M. Agi 氏とはボン駅で 1988. 8. 31. 10 時 38 分着の急行列車で、目印は日本の国旗ということであつた。しかし、ボン駅は小さな駅なので 3 人別々の車両に乗っていたがすぐ会えることができた。

GMD 東京事務所 : 東京都港区赤坂 7 丁目 5 番
56 号 ドイツ文化会館 1 F
03-586-7104
FAX 03-586-7187

ポーランドの研究所

(1) 訪問日、場所、訪問者

訪問日 : 1988 年 9 月 2 日

訪問場所 : Instytut Informacji
Naukowej Technicznej
Ekonomicznej (Institute
for Scientific, Technical
and Economic Information)
I INTE, Ul. Jasna 14/16,
00-041 Warsaw Poland

訪問者 : 田中康仁 (姫路短大)

相手側受入者 : Dr. E. Scibor and Director
Dr. hab. J. Bankowski

(2) 訪問目的

- ① シソーラスを体系的に収集している研究所だと聞いたためと、シソーラスの収集状況の見学のため
- ② I INTE での講演
講演題目は「Solution of Multivocal Words by utilizing Knowledge Data」(Euralex の内容と同じ)
- ③ ワルシャワを見学し、見聞を広めるため

(3) 訪問内容

1968 年～1981 年まではユネスコからシソーラスの研究資金が出ていたので、シソーラスを収集する活動を行っていたが、現在は資金がなく活動を行っていないとのことであつた。

1968 年～1981 年まではシソーラスの目録のようなもの(本の名前、出版社、内容等について)を出版していたが、それは行っていないとのことであつた。しかし今までに集めたシソーラスの本については内容を見ることができた。これは大変参考になった。

訪問目的 2 の講演は Euralex の発表内容をここで講演した。約 20 名ばかりの参加者があつた。内容に興味を示し、質問も幾つかあつた。

(4) ポーランドについて

ポーランドは第 2 次世界大戦で、ドイツ(ヒトラー)の侵略とソ連からの侵入とに会い、ワルシャワの町は 95% まで破壊されてしまった。しかしポーランド国民の努力により、レンガの 1 つ 1 つまで吟味し復元していった。しかし、民族問題、宗教問題、対外債務をかかえポーランドは困っている。しかし、国民の生活はそれほど困っているようでもなかつた。週末にはホテルではダンスを楽しむ人々であふれていた。

但し、外貨を必要とする輸入物資は高価である。

ポーランドは外国人に対して1日US\$15ドルをポーランド通貨に両替し、使用しなければならないという規則があったが、これは通行税と思いホテル代その他で使用した。外貨獲得と外国人の安全のための通行税のためであろう。

ポーランドはDr. Zenon Grabarczyk (大学教授)がワルシャワの町を案内してくれ、文化宮殿、ショパンやコペルニクスのモニュメントを見物し、キューリ夫人の博物館に立寄り記帳した。3日間のポーランドの旅であったが彼の案内は私に大変有益であった。

おわりに

ヨーロッパ旅行は日立製作所基礎研究所の新田義彦氏の提案で始まったが、彼の参加中止により内容を少し変えて行った。

Euralexへの参加は電総研横山の提案により、多くの人々によびかけ田中、荻野、山階の参加になった。所属や旅行の目的も少しづつ異なるため、目的地で合流し、解散し、宿泊その他は別という行動であった。全員元気で、無事目的を実行できたことは幸いである。また、各国の研究所の人々、Euralexの人々に感謝したい。特にEuralexでは江沢建之助教授(チュービンゲン大学)に大変感謝する。

この成果をもとに計算言語の研究を発展させたい。

田中のポーランド訪問については、有益な助言を下された中村幸雄氏(情報科学技術協会)に感謝する。

参考文献

- (1) K. R. K. ハートマン編、木原研三・加藤知己翻訳
監修：辞書学 その原理と実際、三省堂(1984)
- (2) シドニー・ランドウ著、小島義郎・増田秀夫・高野嘉明訳：辞書学のすべて、研究社出版(1988)
- (3) M. Snell-Hornby(ed.): ZÜRILEX
Proceedings '86, A. Francke Verlag,
Tübingen(1988)
- (4) R. R. K. Hartmann(ed.): LEXeter
Proceeding '83, 出版社不明、Tübingen
(1984)
- (5) GID For the Commission of the
European Communities THESAURUS
GUIDE 1985 North-Holl and ISBN
0-444-87736-3
- (6) 尾関周二、クリスティアン・ガリンスキ編著
ターミノロジー学(ウエスターの言語哲学とその応用)
文理閣 1987.6
- (7) 岡谷 大、ターミノロジーの方法：ドキュメンテーションへの応用とAIとの関係
情報の科学と技術 37巻9号 pp405~412
情報科学技術協会 1987.9
- (8) 岡谷 大、タームバンクの概要とタームネットにおける役割
情報管理 VOL 30 №7
日本科学技術情報センター 1987.10
- (9) Felber, H Terminology manual
Paris UNESCO 1984
- (10) Terminology and knowledge engineering:
proceedings/International Congress on
Terminology and Knowledge Engineering,
29 Sept.-1 Oct. 1987. Univ. of Trier, Fed.
Republic of Germany, Hans Czup; Christian
Galinski, ed. Organized by the International
Information Centre for Terminology
(INFOTERM) and the Association for
Terminology and Knowledge Transfer. -
Frankfurt/M: INDEKS Verlag, 1987.
ISBN 3-88672-202-3
- (11) J. M. Sinclair (ed.): Looking Up-- An Account
of the COBUILD Project in Lexical Computing.
Collins (1987).
- (12) 小島 義郎: 英語辞書学入門、三省堂選書110
(1984)