

a dire la ricerca di universali linguistici. E di fatto, in modo più o meno esplicito, molte relazioni si sono affrontate con questo nodo centrale — come per es. quella di H. Andersen « Synchronic and Diachronic Tl. », di W. Dressler « Phonologische Prozesstypologie », di T. S. Sharadzenidze « Sprachtypologie und Sprachuniversalien » suggerendo possibili approcci a diversi livelli d'analisi.

Un punto importante emerso nel convegno è che Tl. non significa semplice classificazione di particolarità (come da alcuni è stato sostenuto anche a Praga) ma analisi di proprietà strutturalmente coordinate — il concetto di « tipo » della scuola praghese — e dei motivi del loro costituirsi in insiemi strutturati (A. Sgall, « Die Sprachtypen in der klassischen und der neueren Tl. »): a questo livello sono recuperabili i metodi d'analisi tanto dello strutturalismo tassonomico quanto quelli, più efficaci, del generativismo, vantaggiosamente liberati dai presupposti universalistici di tipo innatista che ne hanno spesso resa problematica l'accettabilità. Prima dello strato più profondo in cui si collocano le proprietà universali del linguaggio in quanto tale e dove il concetto di tipologia inteso come comparazione di strutture diverse non è operativo (R. Austerlitz, « The Frustrations of Linguistics Tl.: Limitations or Stimulants? »), vi sono altri livelli della struttura profonda dove si colgono categorie linguistiche generalizzate, le cui configurazioni sono caratteristiche dei diversi tipi linguistici esistenti (Birnbaum, già citato).

Sono questi solo alcuni dei punti emersi dai lavori delle dense giornate del simposio. Altre discussioni, come quella interessantissima dei rapporti tra tipologia e diacronia (K. H. Schmidt, « Histor. Sprachvergleichung und typologisches Kriterium ») o sull'ordine sintattico

come criterio tipologico (Z. Vennemann, « Language Type and Word Order »; V. Z. Panfilov, « Struktura predloženia v jazykach raznykh tipov ») possono qui per ragioni di spazio essere solo ricordate a testimoniare la vastità dei tempi affrontati e del posto centrale che lo studio tipologico occupa nella linguistica generale.

Malgrado assenze importanti fra coloro che avevano annunciato propria partecipazione (ancora una volta, in primo luogo, i sovietici, Andrejev e Melikišvili), si è trattato di un incontro quanto mai produttivo — contenuto nel giusto numero, sia di comunicazioni che di partecipanti — di cui bisogna dare pieno merito ai linguisti praghensi. [Ramat]

*5th International Conference on Computational Linguistics, (Pisa 27 agosto-1 settembre 1973).*

*Un'occasione per verificare lo stato dell'arte.*

La 5ª edizione dell'*International Conference on Computational Linguistics*, che dal 1965 riunisce ogni due anni sotto gli auspici dell'*International Committee on Computational Linguistics* gli studiosi delle più diverse discipline interessate alla elaborazione di testi o di dati linguistici con i calcolatori, si è svolta dal 27 agosto al 1º settembre 1973 presso l'Università di Pisa, con il patrocinio del CNR, del CNUCE e del Centro Scientifico IBM di Pisa, ed ha registrato più di quattrocento partecipanti da più di 40 paesi, circa il triplo delle precedenti edizioni, a testimonianza dello sviluppo di questi studi nel nostro paese e del prestigio internazionale a cui gode la Sezione Linguistica del CNUCE. Per motivi di spazio non è possibile elencare le 110 comunicazioni accettate dal *Program Co-*

a dire la ricerca di universali linguistici. E di fatto, in modo più o meno esplicito, molte relazioni si sono affrontate con questo nodo centrale — come per es. quella di H. Andersen « Synchronic and Diachronic Tl. », di W. Dressler « Phonologische Prozesstypologie », di T. S. Sharadzenidze « Sprachtypologie und Sprachuniversalien » suggerendo possibili approcci a diversi livelli d'analisi.

Un punto importante emerso nel convegno è che Tl. non significa semplice classificazione di particolarità (come da alcuni è stato sostenuto anche a Praga) ma analisi di proprietà strutturalmente coordinate — il concetto di « tipo » della scuola praghese — e dei motivi del loro costituirsi in insiemi strutturati (A. Sgall, « Die Sprachtypen in der klassischen und der neueren Tl. »): a questo livello sono recuperabili i metodi d'analisi tanto dello strutturalismo tassonomico quanto quelli, più efficaci, del generativismo, vantaggiosamente liberati dai presupposti universalistici di tipo innatista che ne hanno spesso resa problematica l'accettabilità. Prima dello strato più profondo in cui si collocano le proprietà universali del linguaggio in quanto tale e dove il concetto di tipologia inteso come comparazione di strutture diverse non è operativo (R. Austerlitz, « The Frustrations of Linguistics Tl.: Limitations or Stimulants? »), vi sono altri livelli della struttura profonda dove si colgono categorie linguistiche generalizzate, le cui configurazioni sono caratteristiche dei diversi tipi linguistici esistenti (Birnbaum, già citato).

Sono questi solo alcuni dei punti emersi dai lavori delle dense giornate del simposio. Altre discussioni, come quella interessantissima dei rapporti tra tipologia e diacronia (K. H. Schmidt, « Histor. Sprachvergleichung und typologisches Kriterium ») o sull'ordine sintattico

come criterio tipologico (Z. Vennemann, « Language Type and Word Order »; V. Z. Panfilov, « Struktura predloženiya v jazykach raznykh tipov ») possono qui per ragioni di spazio essere solo ricordate a testimoniare la vastità dei tempi affrontati e del posto centrale che lo studio tipologico occupa nella linguistica generale.

Malgrado assenze importanti di coloro che avevano annunciato propria partecipazione (ancora una volta, in primo luogo, i sovietici, Andrejev e Melikišvili), si è trattato di un incontro quanto mai produttivo — contenuto nel giusto numero, sia di comunicazioni che di partecipanti — di cui bisogna dare pieno merito ai linguisti praghensi. [Ramat]

*5th International Conference on Computational Linguistics, (Pisa 27 agosto-1 settembre 1973).*

*Un'occasione per verificare lo stato dell'arte.*

La 5ª edizione dell'*International Conference on Computational Linguistics*, che dal 1965 riunisce ogni due anni sotto gli auspici dell'*International Committee on Computational Linguistics* gli studiosi delle più diverse discipline interessate alla elaborazione di testi o di dati linguistici con i calcolatori, si è svolta dal 27 agosto al 1º settembre 1973 presso l'Università di Pisa, con il patrocinio del CNR, del CNUCE e del Centro Scientifico IBM di Pisa, ed ha registrato più di quattrocento partecipanti da più di 40 paesi, circa il triplo delle precedenti edizioni, a testimonianza dello sviluppo di questi studi nel nostro paese e del prestigio internazionale a cui gode la Sezione Linguistica del CNUCE. Per motivi di spazio non è possibile elencare le 110 comunicazioni accettate dal *Program Co-*

*mittee*, e distribuite in undici sezioni. Ci limiteremo pertanto a un breve commento sullo stato dell'arte nei corrispondenti settori così com'è emerso dal Convegno.

### *Semantical calculus.*

Il trattamento della semantica, uno dei temi dominanti del Convegno, è stato oggetto dei due *invited papers* (I. A. Mel'cuk, Mosca e W. Woods, Harvard), e ha dato luogo a accese e talora violente discussioni tra i rappresentanti delle diverse correnti linguistiche. Nel quadro di applicazioni come il *question-answering*, lo *speech understanding* e in genere i sistemi cosiddetti integrati i quali presuppongono che il calcolatore elabori gli enunciati come se li comprendesse, si cercano i mezzi per rappresentare il significato degli enunciati e per calcolare questa rappresentazione, di solito a partire dalla descrizione totale o parziale della loro struttura sintattica. Si sono delineati tre approcci principali. L'approccio logico (comunicazioni di S. Klein, E. V. Paduceva, S. R. Petrick), sotto l'influsso di logici e linguisti come Montague, Venneman, B. Hall-Partie, Keenan, vuole rappresentare il significato di una frase come una proposizione di logica simbolica, la quale viene considerata come un teorema il cui valore di verità viene calcolato per mezzo di tecniche di risoluzione rispetto a un insieme di assiomi che comprende la banca dei dati. La conversione degli enunciati in forme canoniche del calcolo dei predicati dovrebbe permettere di operare delle inferenze. Nell'approccio cognitivo (comunicazioni di D. G. Hays, R. C. Schank, Y. Wilks, ecc.), sotto l'influsso di linguisti come Fillmore, Quillian, Katz e Fodor, e nella convinzione che la struttura semantica sia strettamente radicata nel sistema cognitivo universale, il calcolo del significato avviene proiettando una rappresenta-

zione concettualizzata della frase (intesa come una descrizione di eventi in termini di relazioni del verbo con i suoi argomenti, sul tipo « casi profondi ») su una rete di concetti che dovrebbe riprodurre la struttura cognitiva del parlante. Nell'approccio procedurale (comunicazioni di W. Woods, J. A. Mel'cuk, B. Dostert, e F. Thompson, D. E. Walker, ecc.), sotto l'influsso di linguisti come Lamb, Halliday, Winograd, data una frase organizzata secondo una struttura nota, il suo significato è costituito dal valore restituito da una procedura che adopera gli elementi della frase per identificare, nella banca dei dati, un dato o un insieme di dati come referente della frase stessa. Il punto debole comune ai tre approcci sta nel fatto che sono stati applicati a sottoinsiemi lessicali molto ridotti. Inoltre si tende a far cadere la distinzione tra conoscenza linguistica e non linguistica del mondo, confine che, nonostante le affermazioni di principio, diventa sempre meno chiaro anche presso molti linguisti teorici.

### *Grammatical analysis.*

L'analisi morfologica automatica sembra aver raggiunto la maturità sia per la ottimizzazione degli algoritmi, sia per la sistemazione teorica, sia per la estensione del lessico trattato nelle varie lingue (comunicazioni di M. Kay, J. St. Bien, B. Sigurd, M. Quezel-Ambrunaz, Njobedkine, ecc.). Per quanto riguarda la sintassi, i sistemi di *parsing* risentono della situazione generale della grammatica trasformazionale. Come è noto, non esiste un modello generalmente accettato, e i trasformazionalisti non hanno mai tentato di specificare completamente una classe di grammatiche, condizione indispensabile per la computabilità. Essi si sono limitati a considerare insiemi limitati anche se altamente complessi di fenomeni, ciascuno proponendo modifiche alla

teoria sulla base di soluzioni specifiche escogitate di volta in volta. Lo sforzo maggiore è compiuto dalla équipe IBM (S. Petrick, W. Plath, P. Postal) che ha approntato un algoritmo di analisi secondo la teoria standard di *Aspects* potenziata con le aggiunte di J. Friedmann. Questa situazione è invece favorevole allo sviluppo di algoritmi di *produzione*, i quali generano automaticamente sequenze nel linguaggio di una data grammatica a partire da strutture iniziali date (Comunicazioni di I. Batori, S. Boisvert *et alii*, A. Riha e S. Machova). Questi programmi sono usati tra l'altro per verificare le capacità generative di una grammatica e per il *debugging* degli errori nella scrittura di un insieme di regole, compito di solito estremamente complesso.

#### *Study of formal properties.*

Le comunicazioni hanno cercato in genere di definire le relazioni tra lingue naturali e teorie che definiscono la nozione di computabilità. Sono stati esaminati quasi tutti i livelli linguistici: morfologico (W. Brecht) sintattico (J. P. Descles nel quadro della teoria di Z. S. Harris) semantico (S. K. Šaumian). Le comunicazioni di F. G. Pagan e di V. Borscev e M. Chomjakov hanno analizzato aspetti intrinseci del concetto di formalizzazione.

#### *Discovery procedures.*

Non si è fatto cenno alle ben note dispute sulle possibilità di meccanizzare in linguistica le procedure di scoperta. Le comunicazioni hanno fornito invece degli esempi concreti di come, dopo che sono state compiute alcune operazioni preliminari di analisi, è possibile rendere computabili delle procedure che portano alla identificazione di fatti ignoti. In questo settore è viva l'influenza di linguisti come H. A. Gleason jr., M. Svadesh, R. Oswald, S. M. Lamb, P. Garvin.

Vari i cambi di applicazione: ricostruzione di un protolinguaggio (J. Hewson), caratteristiche statistiche del linguaggio di malati mentali (P. Castrogiovanni), isofonie e moduli compositivi nella poesia popolare (A. M. Cirese), strutture percettive (R. M. Frumkina), identificazione delle sillabe (G. Gonenc), modelli della produttività lessicale (I. Prodanoff), scoperta dei componenti semantici su base statistica associativa (G. J. Wan Der Steen), inventario fonemico (S. Liebermann, V. Zivov), inventario di strutture sintattiche (A. Tretiakoff).

#### *Lexicology, text corpus editing.*

Il buon numero di comunicazioni relative alla struttura dei dizionari di macchina e alla loro applicazione (G. Ferrari, R. N. Smith, A. Stachowitz, F. Papp, J. Mathias, M. T. Weber, M. T. Wilton, ecc.) va inquadrato nello sforzo della lessicologia di chiarire le proprie premesse teoriche, con lo scopo dichiarato di trasformarsi da arte a tecnica in scienza applicata, e nella funzione sempre più centrale attribuita al lessico nella descrizione linguistica dalle più recenti correnti linguistiche. I dizionari di macchina sono visti come l'archivio adatto a registrare nel loro sviluppo dinamico le conoscenze prodotte dalle ricerche linguistiche. Per esempio a livello sintattico e semantico tali conoscenze vengono incorporate sotto forma di *features* sintattico-semantiche, rapporti di sinonimia, antonimia, codificazione di campi semantici, « strutture dei casi », costruzioni possibili, restrizioni e regole di concorrenza, ecc. Il dizionario macchina, considerato rappresentativo del lessico del sistema, è lo strumento che rende possibile quelle statistiche sul sistema che, a iniziare dai praghensi, molti linguisti (Martinet, Tagliavini, Dubois, Muljačić, Guiraud, Sapir, Greenberg, Hockett, Krámskí, Lyons) hanno considerato

indispensabile complemento alle statistiche sui testi. Inoltre dai testi, che le tecniche elettroniche permettono di sottoporre a spoglio in numero sempre maggiore, vengono registrate nel dizionario macchina le frequenze delle parole, delle loro accezioni, delle costruzioni grammaticali. Ciò permetterà da un lato di identificare sugli assi sincronico e diacronico sottoinsiemi di testi omogenei e di fondare così le basi teoriche della statistica linguistica, dall'altro di includere nel lessico i tratti sintattici e semantici non discreti, proposti da molti linguisti, quali Ross, Lakoff e Fillmore per rendere conto probabilisticamente del variare della accettabilità dei tratti sintattici e semantici in una data struttura. Sul versante della raccolta di vasti corpora di testi e della loro analisi sono da segnalare le comunicazioni di G. Buccellati e P. Saporetto (Banca dei dati assiro-babilonesi), R. Leonard (archivio lessicale dell'inglese), D'Arco Silvio Avalle (testi poetici italiani delle origini), O. Ikola e Y. Karialainen (archivio sintattico dei dialetti finlandesi), G. R. Wood (varietà regionali dell'inglese d'America), P. Baldacci *et alii* (V volume del *Corpus Inscriptionum Latinarum*).

#### *Quantitative description of language systems, text comparison.*

*Macromodelli:* Le prime analisi di grandi corpora di testi con il calcolatore hanno falsificato i modelli quantitativi del dopoguerra basati sull'equivalenza *langue : parole = popolazione : campione* (J. Ross, P. Guiraud, G. Herdan). La ricerca di modelli sostitutivi mira soprattutto a distinguere unità linguistiche la cui frequenza è stabile nel sistema, unità la cui frequenza è stabile in sottoinsiemi di testi omogenei, unità la cui frequenza dipende dallo stile, ecc. Comunicazioni: A. Zampolli, struttura statistica del lessico; N. Andreyev, probabilità dei tratti

distintivi nelle opposizioni; N. M. Arapov e H. M. Herz, relazione tra frequenze e età delle parole; R. D. Beebe, relazione tra frequenza dei tipi sintattici e genere letterario; V. N. Beloozerov e A. A. Polikarpov, modello sociolinguistico esplicativo delle correlazioni tra stati linguistici e sociali; A. Dzhubanov e B. Khasanov: modelli quantitativi delle lingue del Kazakistan. *Micro-modelli e stilometria:* indagini sulle caratteristiche quantitative che permettono di differenziare lo stile di un autore, di comparare stili diversi, di tentare attribuzione di testi adespoti, ecc. (Comunicazioni di G. Maloney, F. Meletti, S. Monsonogo e E. Kis, B. Follet, J. Raben, P. Fortier, M. Levison).

#### *Meaning extraction.*

A differenza della traduzione automatica, la documentazione automatica non richiede che il processo di analisi automatica del testo produca una rappresentazione adeguata della struttura semantica e sintattica degli enunciati. Tuttavia il rendimento di metodi semplici, quali il *Kwic-index*, l'associazione di parole, tabelle di frequenze del lessico tecnico, può essere aumentato dalla conoscenza delle strutture linguistiche dei testi, soprattutto tecnici, e dalla possibilità di formalizzare almeno in parte il loro contenuto (comunicazioni di B. Piotrowski e I. V. Palibina, D. Herault e A. Ljudskanov, J. Horecky, M. Murai, M. Gross, ecc.).

#### *Translation.*

La storia del precoce ottimismo e dei conseguenti insuccessi che caratterizzano la storia della traduzione automatica sono ben noti. Alla presa di coscienza che la possibilità di una traduzione automatica di alta qualità richiede una comprensione completa e corretta non solo delle strutture linguistiche, ma di tutto il processo della comunicazione verba-

le, le équipes che ancora resistono hanno risposto identificando diversi livelli di accuratezza, ciascuno dei quali incorpora il precedente, con una progressione che va dalla informazione lessicale, a quella sintattica, semantica, contestuale, pragmatica. I modelli attuali vanno raramente oltre il livello semantico, salvo quelli che, facendo proprio l'approccio della intelligenza artificiale, danno la prevalenza al livello pragmatico e contestuale, ma riducono drasticamente di conseguenza il sottinsieme delle strutture linguistiche accettate. (Comunicazioni di O. Kulagina, A. Melby, C. Boitet, T. D. Crawford, G. Potvin, A. A. Stachowitz, W. A. Slaby).

#### Conclusione.

Il Convegno ha avuto il merito di presentare, forse per la prima volta, un panorama completo delle tendenze, delle ricerche, delle realizzazioni attuali della *computational linguistics*. I limiti delle conoscenze linguistiche condizionano la *computational linguistics*, le cui esigenze di formalizzazione e le cui capacità di verificare modelli e regole determinano non di rado l'evoluzione delle teorie linguistiche. Per questi motivi riteniamo che saranno utili gli Atti del Convegno *Mathematical and Computational Linguistics*, due volumi a cura di A. Zampolli che appariranno presso l'editore Leo S. Olschki in Firenze rispettivamente nel marzo e nel giugno 1974. [A. Zampolli]

V. *Fachtagung für indogermanische und allgemeine Sprachwissenschaft* (Regensburg, 9-14 settembre 1973).

Si è tenuta nel settembre 1973 a Regensburg la quinta *Fachtagung* della Indogermanische Gesellschaft, l'attiva società, fondata nel 1953,

la quale, oltre al convegno generale ogni quattro anni, ha preso l'iniziativa di organizzare negli ultimi anni alcuni colloqui rivolti ad un ristretto pubblico di specialisti, che hanno suscitato un vivo interesse: nel 1971 a Bonn (*Anatolische Sprachen*) e a Colonia (*Methodendiskussion*), nel 1972 a Erlangen (*Probleme des indogerm. Ablauts*) e a Innsbruck (*Albanologisches Colloquium*).

Il tema proposto per il convegno del 1973 *Flexion und Wortbildung* ha suggerito un folto numero di trattazioni, che per comodità descrittiva possiamo ricondurre a due indirizzi metodologici. Da una parte sono stati presentati numerosi contributi alla ricostruzione di aspetti della flessione nominale e verbale dell'indeuropeo o di determinate lingue storiche nel quadro di principi teorici che definiremmo genericamente di strutturalismo tassonomico, dall'altra si è cercato di impostare — anche se in maniera non unitaria e comunque in un numero minore di comunicazioni — un discorso teorico in chiave di grammatica generativa.

È vero che su quest'ultimo filone sono state espresse alcune riserve (non riguardanti peraltro il merito dei contributi) e si è auspicata una più stretta aderenza ad argomenti tradizionali per i congressi di indeuropeistica (è stato anche proposto di considerare non pertinenti nei futuri convegni della Società le trattazioni di linguistica generale): a questa posizione, che pare indizio di un certo disagio dell'indeuropeistica di fronte al dilagare di teorie formalizzate che sembrano prescindere dal dato storico, si potrebbe rispondere con un'osservazione di Paul Kiparsky durante i lavori del convegno, che cioè il primo passo, il presupposto per costruire una storia linguistica è proprio fare della grammatica sistematica.

Dalla relazione di Kiparsky *Zum idg. Akzent- und Ablautsystem* si è