

NIEUWE WEGEN IN DE LEXICOLOGIE

Automation is a useful technique when intelligently applied and wisely introduced. It is, however, neither a magic wand nor a universal panacea,

J. HARRINGTON

Die Automation schenkt dem Menschen durchaus nicht das Paradies auf Erden, im Gegenteil: sie vermehrt seine Arbeit noch und erzwingt ihn zu einem intensiveren Leben,

R. BUSA S.J.

VERHANDELINGEN DER KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN, AFD. LETTERKUNDE
NIEUWE REEKS – DEEL LXX – No. 1

NIEUWE WEGEN IN DE LEXICOLOGIE

WITH A SUMMARY IN ENGLISH

F. DE TOLLENAERE

N.V. NOORD-HOLLANDSCHE UITGEVERS MAATSCHAPPIJ
AMSTERDAM – 1963

uxori
piaeque Christinae memoriae

INHOUD

VOORBERICHT	7
INLEIDING	9
HOOFDSTUK I <i>Het „Centre d'Étude du Vocabulaire Français” te Besançon</i>	19
§ 1. Ontstaan, organisatie, taak	19
§ 2. Het werk op het „Laboratoire d'Analyse Lexicologique”	20
§ 3. Wat reeds in het CEVF werd verwezenlijkt	31
HOOFDSTUK II <i>Het „Centro per l'Automazione dell'Analisi Letteraria” te Gallarate</i>	37
§ 1. Geschiedenis	37
§ 2. Organisatie	41
§ 3. Tegenwoordige werkwijze	44
§ 4. Werken	60
§ 5. Toekomstplannen	66
§ 6. Slotbeschouwingen	71
HOOFDSTUK III <i>Elektromechanische of Elektronische Apparatuur?</i>	73
§ 1. De elektronische rekenmachine	73
§ 2. Huidige stand der elektronische lexicologie	75
Verenigde Staten	75
Engeland	88
Italië	92
Sovjetunie	93
Tsjecho-Slovakije	93
België	94
Nederland	95
Euratom	98
§ 3. De elektronische woordindex	98
§ 4. De toekomst	100
HOOFDSTUK IV <i>Een centrum voor Lexicologie in Nederland?</i>	104
§ 1. Woordenboek en Thesaurus	104
§ 2. Moderne Lexicografie	108
§ 3. Gemechaniseerde Lexicologie	110
§ 4. Een centrum voor Lexicologie te Leiden?	112
§ 5. Het LLC a) Taak en doel	115
b) Excerperen of ficheren?	117
c) Enkele of dubbele kaartenverzameling?	119
d) Woordstatistiek	119
e) Organisatie	121
BESLUIT	123
ENGLISH SUMMARY	128
REGISTER VAN PERSOONSNAMEN	141
LIJST VAN AANGEHAALDE LITERATUUR	143

VOORBERICHT

De kiem van dit werk ligt besloten in een gesprek met Professor Dr. L. Kukenheim te Leiden, waarbij toevallig ook het „Centre d'Étude du Vocabulaire Français” te Besançon ter sprake kwam.

Toen ik kort daarop de aandacht der Rijkscommissie van Bijstand voor het „Woordenboek der Nederlandsche Taal” vestigde op de lexicologische vernieuwing die in Frankrijk aan de gang was, en op het mogelijke belang ervan voor het WNT, ontving ik de opdracht, een studiereis naar het CEVF te ondernemen. Professor B. Quemada stelde me in 1960 in de gelegenheid om, met behulp van de machines van het „Laboratoire d'Analyse Lexicologique”, het tweetalige *Vocabulaire* van Noel van Berlainmont [1536]¹⁾ machinaal te ficheren.

In 1961 bezocht ik, eveneens in opdracht van de R. v. B., het „Centro per l'Automazione dell'Analisi Letteraria” te Gallarate bij Milaan, waar Professor Dr. R. Busa s.j. me de gelegenheid bood werkwijze en techniek van het Italiaanse centrum voor gemechaniseerde lexicologie te bestuderen. Ik nam eveneens deel aan het tussen 6 en 10 juni 1961 te Besançon gehouden „Colloque international sur la mécanisation des recherches lexicologiques”.

Reeds in 1960 kwam ik in contact met de „Stichting Studiecentrum voor Administratieve Automatisering te Amsterdam”, en wel met de heren Dr. M. Euwe en A. B. Frielink. In november van datzelfde jaar was ik, op verzoek van de Stichting, aanwezig op het „Kolloquium über maschinelle Methoden der literarischen Analyse und der Lexikographie” te Tübingen. Op 8 maart 1961 organiseerde de SSAA, in samenwerking met het WNT, een colloquium over „Automatisering en Taalkunde”. Ook mocht ik de door de Stichting tussen 10 februari en 24 maart 1961, aan de Technische Hogeschool te Delft georganiseerde technische cursus bijwonen.

Dank zij een introductie me verschaft door Professor Dr. S. Wouthuysen bij zijn collega Professor Dr. Ir. A. van Wijngaarden, directeur van het Mathematisch Centrum te Amsterdam, bleek deze bereid, me in 1961 en 1962 kennis te laten maken met de grote, door de elektronische rekenmachine geboden mogelijkheden op het gebied van de verwerking der linguïstische informatie. Hierbij heb ik op een bijzonder prettige wijze samengewerkt met de heer J. A. Th. M. van Berckel, verbonden aan het Centrum, met wie ik menig detail in dit boek heb besproken.

Zo is dan, zonder vooropgezette bedoeling, het hier thans voorgelegde boek gegroeid. Zonder het begrip van de Voorzitter, Dr. L. G. Kortenhorst,

¹⁾ Zie B. Quemada, *L'Inventaire des dictionnaires bilingues* [1961] en *L'Exploration mécanique des dictionnaires bilingues* [1961].

en de leden der Commissie van Bijstand voor het WNT, die me de nodige vrijheid lieten bij het zoeken van deze nieuwe wegen, zou het ongeschreven zijn gebleven. Op de belangstellende welwillendheid van de Secretaris der R. v. B., de heer J. Platteel, heb ik nooit tevergeefs een beroep gedaan.

Het illustratiemateriaal voor de hoofdstukken I en II werd me door de firma's BULL en IBM welwillend ter beschikking gesteld. Voor verschillende wenken naar aanleiding van sommige details ben ik de heer H. de Klerk, van BULL-Nederland, en de heer C. J. Brandt, van IBM-Nederland, bijzonder erkentelijk. Professor Dr. W. Z. Shetter is zo vriendelijk geweest het Engels van mijn „Summary” te kuisen.

Ten slotte past nog een woord van dank aan mijn vriend Dr. A. J. Persijn, die me bij de correctie der proeven heeft bijgestaan.

januari 1963.

INLEIDING

Het woord *lexicografie* wordt vrij algemeen gedefinieerd als „het schrijven van woordenboeken”¹⁾ of als „de wetenschap van de woordenboeken der taal”. Beter is de omschrijving „kunst”²⁾, resp. techniek³⁾ van het samenstellen van woordenboeken”. Toch is het gebruik van dit woord niet altijd even ondubbelzinnig⁴⁾.

1) „The making of dictionaries”, *Problems in Lexicography* v [1962].

2) „Arte de componer diccionarios” (J. Casares, *Introducción a la Lexicografía moderna* 11 [1950]); „l'arte di compilare i dizionari” (B. Migliorini, *Che cos' è un vocabolario?* 1 [1951]).

3) S. Ullmann, *The Principles of Semantics* 64 [1951]: „the writing or compilation of dictionaries, which is a special technique rather than a branch of linguistics”.

De definitie van U. Weinreich in *Problems in Lexicography* 26 [1962]: „The semantic description of individual terms (in such contexts as may be appropriate)” past beter bij de semasiologisch-alfabetische dan bij de onomasiologische of ideologische lexicografie.

4) J. Marouzeau in zijn *Lexique de la terminologie linguistique*¹⁻² [1933-1943] b.v. beschouwt beide woorden, *lexicografie* en *lexicologie*, als synoniemen; in *La linguistique ou science du langage*² [1944] heeft zijn tweede hoofdstuk „Les Mots”, waarin de morfologie en de semantiek der woorden wordt behandeld, als ondertitel *Lexicographie*. De derde editie van zijn *Lexique*, „augmentée et mise à jour” [1951], bevat echter de volgende toevoeging: „On réserve quelquefois le nom de *lexicologie* pour l'étude du vocabulaire et celui de *lexicographie* pour la science des dictionnaires”. Naar aanleiding van het dubbelzinnige gebruik van de term *lexicographie* door Marouzeau, schrijft S. Ullmann, *Précis de sémantique française* 33, noot 2) [1952]: „il est préférable de réserver celui-ci (t.w. *lexicographie*) à la compilation de dictionnaires”. Bij J. B. Carroll, *The Study of Language* [1953] heeft *lexicography*, waar de „morphology” niet toe behoort (zie blz. 24), de betekenis semantiek. Op blz. 236 schrijft hij trouwens zelf: „I have deliberately substituted the term *lexicography* for what Bloomfield (1933, p. 138) describes as semantics”. Het maken van woordenboeken behoort volgens Carroll niet tot de linguïstiek, doch tot de „philological studies” (65), bepaaldelijk tot de „linguistic philology” (66). Op blz. 92 echter schijnt de „lexicography” toch weer tot de linguïstiek te behoren.

In het voorbericht (gedateerd 7 maart 1959) van Schönfelds *Historische Grammatica*, spreekt A. van Loey over „aanvullingen op lexicografisch gebied”, waar duidelijk geen aanvullingen van lexicografische, doch van lexicologische aard zijn bedoeld. Op blz. 55 van A. D. Booth, L. Brandwood en J. P. Cleave, *Mechanical Resolution of Linguistic Problems* [1958] wordt *lexicography* gebruikt in een verband waar ik de voorkeur zou geven aan *lexicology*.

Omgekeerd heeft P. Paardekooper, *Syntaxis, Spraakkunst en Taalkunde* 183 volg. [1955] een hoofdstuk „Lexicologie”, waarin hij duidelijk het maken van woordenboeken op de korrel neemt. J. J. Mak spreekt in de inleiding van zijn *Rhetoricaal Glossarium* [1959] van de „eisen der moderne lexicologie” (blz. x) en van „de werkkamer van een moderne lexicoloog” (blz. xii), waar ik liever *lexicografie*, resp. *lexicograaf* zou lezen.

Ook W. von Wartburg gebruikt de woorden *lexicografie* en *lexicologie* blijkbaar zonder onderscheid door elkaar, waar hij, in zijn artikel *Betrachtungen* [1939], het de ene keer heeft over „wissenschaftliche Lexikographie” (blz. 10), terwijl hij op

Kost een goede omschrijving van het woord *lexicografie* betrekkelijk weinig moeite, anders evenwel is het gesteld met de term *lexicologie*; over de betekenis van dit woord lopen de meningen nogal uiteen. Ik laat hier enkele definities de revue passeren, die van dit woord zijn gegeven.

In een artikel getiteld „Words, Logic and Grammar” gebruikt H. Sweet in 1876 *lexicology* in een betekenis die zowat gelijkstaat met „semantiek”, doch tevens het schrijven van woordenboeken omvat¹⁾. In het vlak van de semantiek liggen eveneens de omschrijvingen van U. Weinreich: „The study of recurrent patterns of semantic relationship, and of any formal (phonological and grammatical) devices by which they may be rendered”²⁾ en D. S. Worth „(the) description of semantic fields”³⁾. Volgens Littré⁴⁾ is lexicologie een „partie de la grammaire qui s'occupe spécialement des mots considérés par rapport à leur valeur, à leur étymologie, à tout ce qui est nécessaire de savoir pour composer un lexique”⁵⁾. Marouzeau ziet de lexicologie — die hij, zoals gezegd, niet onderscheidt van de lexicografie — als de „étude du vocabulaire enregistré dans les lexiques, c'est-à-dire des mots et locutions..., considérés soit dans leur

blz. 12 schrijft: „Diese beiden grossen Sprachen werden (na de voltooiing van de „Oxford Dictionary” en het „Wörterbuch” van Grimm) kaum in absehbarer Zeit erneut einer umfassenden lexikologischen Behandlung teilhaftig werden”.

1) Sweet schrijft immers ook: „The real distinction (*tussen grammatica en woordenboek*) is that grammar deals with the general facts of language, lexicology with the special facts” (*Collected Papers of Henry Sweet* 31 [1931]). Deze veel geciteerde, later door Jespersen overgenomen definitie is een beetje ongelukkig geformuleerd; grammatica en woordenboeken behandelen dezelfde stof, maar vanuit een verschillend gezichtspunt, t.w. het algemene, d.w.z. systematische, resp. het bijzondere, d.w.z. dat van het behandelde woord; zie in dit verband B. Migliorini, *Che cos' è un vocabolario?* 1 [1951] en F. F. Odendal, *Leksikografiese Probleme* III, 51 [1961-'62]. De Engelse linguïst M. A. K. Halliday heeft onlangs het verschil tussen grammatica en lexicologie als volgt omschreven: „la lexicologie, c'est l'étude des relations ouvertes; la grammaire, c'est l'étude des relations fermées. Autrement dit, là où il y a un choix parmi un nombre fixe de possibilités, nous sommes dans la grammaire” (in *Études de linguistique appliquée* 1, 9 [1962]). Wanneer hij op blz. 20 schrijft: „Dans la plupart des dictionnaires on trouve, bien entendu, des morceaux supplémentaires: classe de mot (de la grammaire), étymologie (de la linguistique historique), prononciation (de la phonétique ou de la phonologie, selon le dictionnaire), mais ceux-ci sont des corps étrangers, et ne font pas partie de la lexicologie proprement dite)”, dan heeft hij gelijk. Het domein van het woordenboek omvat dan ook meer dan het strikt lexicologische. Daarom kan men in de door Halliday gegeven omschrijving van grammatica en lexicologie, dit laatste woord niet zo maar vervangen door woordenboek.

2) In *Problems in Lexicography* 26 [1962].

3) In *a.w.* 82 [1962].

4) Afl. 16 waarin het woord voorkomt, verscheen in 1867; zie P. J. Wexler, in *Cah. de Lex.* 1, 99 [1959].

5) Hiermee stemmen geheel overeen de definities „science de l'histoire et de la signification des mots relativement à leur étymologie” (*Dictionnaire usuel Quillet Flammarion* [1959]) en „science, connaissance raisonnée des mots sous le rapport de l'étymologie, des acceptions, etc.” (*Petit Larousse* [1961]).

signification (sémantique) soit dans leur forme (morphologie)" [1933]. F. de Saussure, *Cours de linguistique générale*³ 185 [1931] geeft de omschrijving „science des mots”; blijkens de context bevat de lexicologie echter niet de morfologie. Casares ziet de lexicologie als de studie van vorm en betekenis der woorden vanuit een algemeen wetenschappelijk gezichtspunt¹⁾, terwijl Migliorini schrijft: „lessicologia è lo studio del sistema lexicale di una lingua”²⁾. De lexicologie, volgens Migliorini, verwaarloost de vorm niet, en valt dus geenszins samen met de semantiek; verder ziet hij de lexicologie als een zuiver synchronische, beschrijvende en catalogiserende discipline, de semantiek daarentegen, die zich bezighoudt met de verklaring van de betekenisontwikkeling der woorden, als een diachronische³⁾. S. Ullmann geeft de volgende definitie: „the science of the word”, „the branch of linguistics devoted to (the) study (of the words)”⁴⁾; deze omvat de morfologie en de semantiek, of liever, wat hij noemt de „lexical morphology” en de „lexical semantics”⁵⁾.

Zag Littré de lexicologie nog als de combinatie van semantiek en etymologie, thans zien vele taalkundigen (Marouzeau, Migliorini, Ullmann) ze als de vormleer en de semantiek omvattend.

Daarnaast heeft men het woord *lexicologie* ook in zeer beperkte betekenissen gebruikt. Zo spreekt E. Tappolet in zijn boek *Die romanischen Verwandtschaftsnamen* 3 [1895] van „vergleichende Lexicologie”, waarmee hij bedoelt wat thans vrij algemeen onomasiologie heet, de tegenhanger dus van de semasiologie of de semantiek. Een zeer speciale betekenis heeft het woord ook bij de Deense taalkundige K. Møller. Lexicologie is volgens hem de wetenschappelijke behandeling van woordgroepen; zo zou de studie van synoniemen en homoniemen, systeem- of begripsgroepen, concrete en abstracte woordgroepen en van de sociale

1) *Introducción a la Lexicografía moderna* 11 [1950].

2) *Che cos' è un vocabolario?* 1 [1951]. Op zichzelf zegt deze definitie niets, aangezien het woord waar het om gaat door een afleiding van zichzelf omschreven wordt. Op dezelfde bladzijde schrijft Migliorini echter: „Chiamiamo lessico il complesso delle parole e delle locuzioni della lingua, considerate specialmente nei loro significati”.

3) Met deze nauwe begrenzing van lexicologie en semantiek zal niet iedereen het eens zijn; in een noot op blz. 2 schrijft Migliorini trouwens dat een descriptieve, synchronische semantiek identiek zou zijn met de lexicologie.

4) *The Principles of Semantics* 26, resp. 30 [1951].

5) *a.w.* 33 [1951]; zie ook zijn *Précis de sémantique française* 33 [1952]. In zijn boek *Semantics* 29 [1962] schrijft hij: „Lexicology will deal not only with words but with all types of morphemes entering into the composition of words”.

Wanneer K. Møller, *Leksikologi og Leksikografi* 12 [1959] schrijft, dat men in de Engelse en Amerikaanse linguïstiek naast *lexicography* zelden het woord *lexicologie* vindt, en dat men in plaats daarvan liever *lexicography* gebruikt in een zeer ruime betekenis, is dit in zijn algemeenheid niet juist, zoals uit de hierboven geciteerde voorbeelden blijkt. Het is echter wel waar, dat in het boek van J. B. Carroll, *The study of Linguistics* [1953] het woord *lexicologie* niet in de index voorkomt, wel echter *lexicography*.

stratificatie van de woordvoorraad er wèl toe behoren, doch de betekenisleer en de etymologie niet ¹⁾).

In zijn werk *La méthode en Lexicologie*, met de ondertitel *domaine français*, geeft de Franse taalkundige G. Matoré een definitie van de woorden *lexicologie* en *lexicografie*, die, naar hij zegt, is geïnspireerd door de woordparen *ethnologie* en *ethnografie*. De lexicologie is een „discipline de caractère synthétique se proposant l'étude des faits de civilisation". De lexicografie daarentegen is de „étude analytique des faits de vocabulaire, discipline linguistique" ²⁾. Hieruit meen ik te mogen opmaken dat volgens Matoré de lexicologie geen „discipline linguistique" wil zijn, of zich te goed acht om dat te zijn. Volgens hem berust de lexicologie weliswaar op de gedetailleerde analyse van de feiten van de woordenschat, ze gaat uit van „des groupes de mots considérés statiquement du point de vue notionnel" ³⁾, maar de lexicologen moeten tot elke prijs verder gaan dan dit stadium, ze moeten, zo zegt hij, komen tot „l'intellectualisation des faits étudiés" ²⁾. Volgens Matoré heeft de lexicologie tot taak „une explication des faits de société" ⁴⁾; op de basis van „l'étude des mots" moet ze komen tot het doen begrijpen van „le processus des évolutions sociales" ⁵⁾. Matoré wenst m.a.w. de lexicologie uitsluitend te zien als een sociologische, niet als een linguïstische discipline ⁶⁾.

We laten nu ook enkele Nederlandse stemmen horen. Het woordenboek van Koenen, t.w. de 25ste druk, bewerkt door J. Naarding, en verschenen in 1960, definieert *lexicologie* als „kennis v. de betekenis en afleiding der woorden". Van Haeringen in *Kramers' Nederlands Woordenboek* ⁷ [1956] omschrijft het woord als „leer van de woordenschat en de betekenis der woorden". Evenals bij Littré is in het woordenboek van Koenen de lexicologie een samenvattende term voor semantiek en etymologie. Volgens de Utrechtse hoogleraar omvat de lexicologie eveneens de semantiek („betekenis der woorden"); het eerste deel van zijn definitie echter heeft blijkbaar zulk een algemene betekenis, dat het voegwoord *en* hier een beetje vreemd lijkt. In een artikel, getiteld „Van Glossarium tot Thesaurus. Lexicologische Theorien en lexicographische Realiteiten" sluit L. Kukenheim zich aan bij een tweede definitie die Littré van het

¹⁾ *Leksikologi og Leksikografi* 9 en 10 [1959].

²⁾ *La méthode en Lexicologie* 88 [1953].

³⁾ *a.w.* 13 [1953].

⁴⁾ *a.w.* 5 [1953].

⁵⁾ *a.w.* 6 [1953]. De definities van *lexicografie* en *lexicologie* in de nieuwe bewerking, onder leiding van Gérard Venzac, van de *Dictionnaire de la langue française* door Littré en Beaujean [1959] zijn geheel door het werk van Matoré geïnspireerd. Dit is m.i. rijkelijk voorbarig.

⁶⁾ Men kan betwijfelen of dit in wezen antilinguïstische standpunt de instemming van veel taalkundigen zal vinden; zie b.v. de streng afwijzende kritiek van H. L. Scheel, *Neuere Arbeiten zur Lexikologie* 261-262 [1955]. S. Ullmann drukt zich wel bijzonder gematigd uit waar hij schrijft „some linguists will feel that Professor Matoré has gone too far", *Semantics* 252 [1962].

woord *lexicologie* gegeven heeft, en die luidt „*recherches sur les dictionnaires*”. De Leidse romanist ziet de lexicologie als de theorie van de samenstelling van woordenboeken; ter onderscheiding van de lexicografie verduidelijkt hij de term lexicologie nog als „theoretische lexicologie”¹⁾. Dit komt, afgezien van de toevoeging theoretisch, overeen met de onderscheiding die de Zweedse taalkundige A. Noreen reeds in 1903 heeft gemaakt²⁾.

Als ik de geciteerde omschrijvingen kritisch beschouw, wil het me allereerst voorkomen dat men er, afgezien van het feitelijke woordgebruik, principieel goed aan zal doen de woorden *lexicografie* en *lexicologie* niet door elkaar, min of meer als synoniemen dus, te gebruiken. *Lexicografie* zou ik daarom, de Nederlandse woordenboeken zijn het hierover trouwens roerend eens, alleen maar willen gebruiken ter aanduiding van het schrijven van woordenboeken, of, zo men wil, als term voor het handwerk bestaande in het schrijven van woordenboeken. Voor de theorie, de wetenschappelijke grondslag van dit handwerk, zou ik, in tegenstelling tot Littré en Kukenheim, niet van *lexicologie* willen spreken, doch liever daarvoor de term *theoretische lexicografie* reserveren, dit naar het voorbeeld van een verbinding als *theoretische geschiedenis*; die, op haar beurt, door J. Romein, is gevormd naar het voorbeeld van *theoretische natuurkunde*³⁾.

Tot zover is het gemakkelijk, maar nu de lexicologie. Ik heb dit woord wel eens omschreven als „leer omtrent de ordening van een taalschat volgens wetenschappelijke beginselen”⁴⁾. Welke ordening me daarbij voor de geest heeft gestaan, weet ik niet meer; ik weet echter wel dat ik deze definitie nu niet meer zou onderschrijven, om de eenvoudige reden dat ze nog te veel zegt. Immers met „leer van de woordenschat”, zoals het eerste deel van de definitie van Van Haeringen luidt, of met „wetenschap van de (in het woordenboek voorkomende) woordenschat”⁵⁾, zoals ik

¹⁾ *Lev. Talen* 1959, 15. Zie eveneens L. Kukenheim, *Esquisse historique de la linguistique française et de ses rapports avec la linguistique générale* 151 [1962].

Verg. *Winkler Prins' Encyclop.* 12, 722 [1951]: „Onder lexicologie wordt ook wel eens verstaan: de leer omtrent de samenstelling van lexica, leer omtrent de beginselen der lexicografie”.

²⁾ „Läran om ordböckers utarbetande heter lexikologi, själfva författareverksamheten däremot kallas lexikografi”, *Vårt Språk* 1, 50 [1903].

³⁾ Zie J. M. Romein, *Het vergruisde Beeld* 19 [1939]: „de nadelen van de specialisering kunnen alleen overwonnen worden . . . door het zuiver stellen der betreffende problemen. Wil men voor dit nieuwe vak een nieuw woord: men noeme het *theoretische geschiedenis*”. Dat de verbinding *theoretische natuurkunde* het voorbeeld is geweest, werd me in een brief van Romein van 25-11-1961 bevestigd. Vergelijk echter b.v. ook *theoretische biologie*, *theoretische statistiek* enz.

Elders trof ik aan „the theory of lexicography”, J. Sledd, in *Problems in Lexicography* 147 [1962].

⁴⁾ *Van Dale's Nieuw Groot Woordenboek der Nederlandse Taal* [1950].

⁵⁾ *Winkler Prins' Encyclop.* 12, 721 [1951].

Vergelijk J. Casares, *Introducción a la Lexicografía moderna* 11 [1950], die de

het woord naderhand eens heb omschreven, is m.i. alles gezegd. Dat deze definitie niet etymologisch is, dat het *lexicon* (van het woord *lexicologie*), het woordenboek erin is vervangen door de woordenschat, moet men maar op de koop toe nemen. Het verdient, naar het me voorkomt, aanbeveling de term *lexicologie* niet verder te preciseren. Als gevolg van de recente ontwikkeling op het gebied der taalkundige documentatie bestaat er namelijk behoefte aan een term, die ruim genoeg is om zowel het studie-object van de lexicograaf als dat van de grammaticus te omvatten. We hebben een woord nodig dat, al naar het uitkomt, als vlag kan fungeren ter dekking van een zeer disparate lading: voor de lexicografie en de etymologie, de semasiologie en de vormleer, de leer van de woordsoorten en de woordvorming, de syntaxis, de stilistiek en het onderzoek der woordfrequentie ¹⁾. Waarom we zulk een veelkleurige, of als men wil, witte vlag nodig hebben, zal, naar ik hoop, in de loop van mijn uiteenzetting voldoende duidelijk worden.

De nieuwe wegen in deze opzettelijk vaag gehouden lexicologie zijn in de laatste tien jaar, aanvankelijk vrij langzaam en moeizaam, gebaad door enkele pioniers in de Verenigde Staten en in Europa. Dezen stonden min of meer toevallig voor de taak een grote hoeveelheid taalkundige informatie te moeten verwerken. Het besef dat hun taak met de traditionele methoden van het lexicologische handwerk niet kon worden verwezenlijkt, deed hen uitzien naar nieuwe, meer eigentijdse. Dat ze daarbij probeerden de machine als bondgenoot in te schakelen, lag voor de hand. Deze voortrekkers waren de mening toegedaan dat, ook op taalkundig gebied, een beroep zou kunnen worden gedaan op methoden die gebruik maken van zelfwerkende systemen, machines of instrumenten. Ze onderzochten of de automering of automatie (eng. *automation*), zoals de moderne automatisering met behulp van elektronische apparaten wel eens wordt genoemd (in tegenstelling tot de vroegere automatisering met behulp van de traditionele mechanische middelen) ²⁾, uitkomst zou kunnen brengen. Uit tactische overwegingen, om de vakgenoten niet al te zeer te laten schrikken, of uit principiële overwegingen, omdat de term automatisering de indruk kan wekken dat de mens er niet meer aan te pas komt, sprak men daarbij meestal van mechanisering, een woord waarmee men reeds meer dan een eeuw was vertrouwd. Hiermee was dan een ontwikkeling ingezet die, in tegenstelling tot de mechanisering van de negentiende eeuw, het traditionele lexicologische handwerk niet alleen grondig zou beïnvloeden, doch zelfs volkomen revolutioneren. De grote onafwend- en

lexicografie beschouwt als een kunst en de lexicologie als een wetenschap: „la lexicología . . . estudia estas materias (*t.w.* el origen, la forma y el significado de las palabras) desde un punto de vista general y científico”.

¹⁾ De term *linguïstiek* zou in dit verband te ruim zijn; algemene taalwetenschap b.v. valt niet onder het begrip lexicologie. Ook in het artikel *Neuere Arbeiten zur Lexikologie* [1955] van Scheel heeft de term *lexicologie* een zeer ruime betekenis.

²⁾ *Vaktermen* [1959].

onafwijsbare „revolutie” van deze eeuw, de automatisering, die, via ponskaartmachines en elektronische machines, elke dag meer en meer om zich heen grijpt, en waarvan de invloed in toenemende mate het uitzicht der dingen verandert, had ook de taalkunde bereikt. Dat dit relatief laat is gebeurd, valt te begrijpen. Immers de meeste linguïsten hebben weinig of geen belangstelling voor de exacte wetenschappen; de methoden ervan zijn hun vreemd.

De machine en de zgn. geesteswetenschappen of humanistische wetenschappen zijn vanouds nu eenmaal geen onafscheidelijke bondgenoten. De mystiek der zgn. „elektronische hersens”, die via de vulgarisering in de krant tot ons komt, is bovendien niet bevorderlijk voor onze houding tegenover de moderne uitvindingen der ingenieurs. Toen de voormalige bondsprezident Heuss eens op bezoek was in een groot rekencentrum, moet hij gezegd hebben: „Da sind die Maschinen die zwar denken können aber keine Phantasie haben”. Fantasie hebben deze reusachtige machines stellig niet, doch echt denken kunnen ze al evenmin¹⁾. Immers het zijn de ingenieurs die deze machines hebben geconstrueerd, die gedacht hebben en een vernuftig schema in elkaar hebben gezet, waardoor enige tientallen op zichzelf eenvoudige, elementaire machinale verrichtingen in de juiste volgorde herhaald, tot een sluitend geheel worden gecombineerd. De door de machine genomen beslissingen zijn alleen mogelijk, wanneer de voorwaarden ervoor eerst door de mens in de machine worden vastgelegd. Zonder programmering, d.w.z. zonder het geven van opdrachten door de mens, kunnen de machtigste machines niets; als ze defect zijn, vermogen ze zichzelf niet te helpen. Wat uit de machine komt moet er m.a.w. in een of andere vorm van tevoren worden ingebracht; wat men haar geheugen noemt, is slechts een enorme voorraadschuur waarin informatie kan worden bewaard. De machine schrijft en leest weliswaar duizenden malen vlugger dan de mens het zou kunnen, doch heeft geen verstand. Men heeft de elektronische rekenmachine dan ook wel eens „ein Vollidiot mit spezieller Begabung” genoemd. De machine is dom; ze heeft geen hersenen, doch is een serviel werktuig, een slaaf van de menselijke geest. Dat de automatisering en de mechanisering de mens overbodig zouden maken, is slechts bedrieglijke schijn. In werkelijkheid blijft de mens belangrijk, ja wordt juist belangrijker, omdat de machines hem in staat stellen steeds grotere taken op zich te nemen en tot een goed einde te brengen. De automatisering heeft tot gevolg een niet slechts verstandige, doch bij uitstek humane arbeidsverdeling: de machine krijgt het slavenwerk, de mens het mensenwerk toegewezen.

Dat als gevolg van de automatisering de humanistische wetenschappen hun „humaan” aspect zouden verliezen²⁾, is een angstgevoel dat ik niet

¹⁾ Voor een goede uiteenzetting over het zgn. denken van de informatieverwerkende machine, zie H. Zemanek, *Automaten und Denkprozesse* 3, 18, 56 en 57 [1962].

²⁾ „Every good humanist feels ambivalent about the intrusion of technology into his domain”, S. M. Parrish, *Computer Concordances* 1 [1962].

alleen als „overdreven”, doch als volkomen ongegrond zou willen betitelen. Op het eind van de 15de eeuw heeft de boekdrukkunst een revolutie in de lexicografie ingeluid; de manuscripten met glossenverzamelingen, de glossaria, aangelegd om ontsiering van codices te vermijden, of om als doorlopende verklaring bij een tekst te dienen, werden vervangen door gedrukte vocabularia, die later, in de 16de eeuw, zouden leiden tot beroemde dictionaria. Is de lexicografie door de boekdrukkunst, deze mechanisering der renaissance, „gedeshumaniseerd”? Of is het soms niet juist de boekdrukkunst geweest, die de ontplooiing van een nieuwe lexicografie mogelijk heeft gemaakt? Het is ook mijn overtuiging dat de automatisering van onze eeuw voor de toekomst van de lexicologie op zijn minst even belangrijk zal zijn als de boekdrukkunst dat vier eeuwen geleden is geweest ¹⁾.

Waarom zouden wij lexicologen achterblijven, wanneer we zien hoe op een ander gebied van de linguïstische automatisering met zoveel energie wordt gewerkt? Ik bedoel de automatische vertaling, die zolang met gevoelens van meewarigheid, skeptisisme en medelijden door taalkundigen en vertalers is gadeslagen. Wanneer aan de automatische vertaling meer geld, tijd en energie wordt ten koste gelegd dan aan de automatische lexicologie, is dit natuurlijk te wijten aan het feit dat, in deze tak van linguïstisch-technisch onderzoek, door de grote mogendheden grif belangrijke bedragen worden geïnvesteerd. Dit is niet te verwonderen als men beseft welk een enorm praktisch belang met de automatische vertaling is gemoeid. De vooruitgang der exacte wetenschappen, die op zijn beurt de militaire kracht en de wedloop naar de ruimte conditioneert, kan op den duur (voor een gedeelte) afhankelijk worden van de vertaal-machine ²⁾.

Het is de taak van de moderne lexicologie met een zeer grote hoeveelheid linguïstische informatie te werken; hoe groter deze hoeveelheid is, des te betrouwbaarder worden de resultaten van het onderzoek. „En linguistique, comme en toute science”, schrijft A. Meillet, „la solution de beaucoup de problèmes tient à un degré de précision de plus dans la détermination des faits; les solutions qui semblent acquises deviennent incertaines quand on serre de près les dépouillements sur lesquels elles reposent” ³⁾. Maar hoe groter de hoeveelheid lexicologische informatie is, des te bezwaarlijker ook wordt het opbergen, classificeren en terugvinden ervan. In zoverre krijgt de lexicoloog een taak, die tot op zekere hoogte verwant is aan die van de documentalist die belast is met het opbergen

¹⁾ L. Kukenheim, *Van Glossarium tot Thesaurus* 33 [1960].

²⁾ Zelfs al zouden op dit ogenblik alle problemen der mechanische vertaling geheel bevredigend zijn opgelost, dan nog zou de praktische toepassing der mechanische vertaling ten zeerste worden belemmerd door het feit dat het automatisch lezen, althans het alfabetisch automatisch lezen nog in de kinderschoenen staat.

³⁾ *Linguistique historique et linguistique générale* 5 [1921].

en terugvinden van informatie¹⁾. De gewenste precisie, exactheid en volledigheid in de lexicologie zijn alleen mogelijk door toepassing van kwantitatieve methodes. Lexicologen hebben in het verleden met hun sigarenkistjes grote en eerbiedwaardige resultaten bereikt; hun materiaal was echter maar gering. Een werk als het MNW berust praktisch alleen maar op literaire bronnen. Wat weten wij thans over de Nederlandse terminologie van de jacht in de middeleeuwen²⁾, over de woordenschat van de oude landbouw, van het oude handwerk? Niets, of maar bitter weinig. Het inventariseren van de totale woordenschat gaat de krachten en de tijd van de enkeling of van een klein groepje van specialisten te boven; het is alleen te bereiken door een radicale wijziging in de totnogtoe gebruikte methoden³⁾. Een dergelijke ommekeer in de gevolgde werkmethoden moet onvermijdelijk gepaard gaan met een even radicale wijziging van de geest waarin het lexicologisch werk tot op heden werd ondernomen⁴⁾.

De Franse linguïst A.-J. Greimas, thans hoogleraar te Ankara, promoveerde in 1948 te Parijs op een dissertatie over *La Mode en 1830. Essai de description du vocabulaire vestimentaire d'après les journaux de mode de l'époque*. Het verzamelen van het materiaal uit de Franse modejournalen alleen al had hem vijf jaar gekost. In 1960, tijdens een colloquium van lexicologen te Besançon, zei hij, niet zonder bitterheid, dat dit een absurd werk was geweest, omdat de machines het veel vlugger en beter hadden kunnen doen. Gelukkiger dan Greimas was een Amerikaans romanist die een index aan het maken was op Ronsard. Toen hij al met het werk bezig was, ontdekte hij het bestaan van een gemechaniseerd lexicologisch centrum; het gevolg was dat hij jaren van zijn leven voor werkelijk wetenschappelijk onderzoek had gewonnen.

In verschillende landen, ten oosten, maar vooral ook ten westen van het IJzeren Gordijn, werken linguïsten, in nauwe samenwerking met ingenieurs en technici, met mechanografische en elektronische middelen, aan het systematisch toegankelijk maken van lexicologische informatie, aan de automatisering der lexicologie. Deze mechanisering, die in een dringende behoefte voorziet, zal zich ongetwijfeld doorzetten. Ze vormt

1) Zie *Inf.* Nr. 16, november 1961, blz. 10-13.

2) Men vergelijk hiermee de rijke oogst op het gebied der Oudfranse jachtterminologie, die de Zweedse romanist Gunnar Tilander, nog geheel met de oude methoden, heeft binnengehaald.

3) E. R. Vincent, *Mechanical Aids* 57 [1955] betitelt de methoden die tot op heden bij het maken van concordanties in de meeste gevallen werden gevolgd als „stone-age methods”.

4) B. Quemada, *Mécanisation* 40 [1959]. Vergelijk *John Diebold Answers Twenty Questions* (4): „Truly fruitful results from automatic data processing systems require a fundamental change in approach, an understanding that the best applications are not the mechanization or streamlining of existing procedures, but a willingness to rethink the problems of an entire business in terms of ultimate goal and final product” [1959].

een onderdeel van de nog maar kortgeleden begonnen toepassing van wiskundige methodes in de taalkunde¹⁾.

De volgende bladzijden zijn bedoeld als een inleiding tot, een kennis-making met dit werk, op grond van ervaringen die schr. in twee der voornaamste centra ervan, t.w. dat van Besançon en dat van Gallarate, heeft opgedaan. Wat de nieuwe werkwijzen ons bieden, is te belangrijk, dan dat de lexicologen er nog langer aan mogen voorbijgaan.

¹⁾ „The use of mathematical methods and computing machines in the most diverse fields of science and culture is typical of our age”, F. Shirokov, *Computer deciphers Maya hieroglyphs* 32 [1962].

HOOFDSTUK I

HET „CENTRE D'ÉTUDE DU VOCABULAIRE FRANÇAIS” TE BESANÇON

On ne fait pas de bonne lexicographie sans bonne lexicologie, P. Imbs, in *Cah. de Lex.* 2, 17 [1961].

§ 1. ONTSTAAN, ORGANISATIE, TAAK

In 1957 werd te Straatsburg, in het kader van het centrum voor Romaanse filologie der Universiteit, het „Colloque de Lexicologie et de Lexicographie Françaises et Romanes” gehouden. De bedoeling ervan was om, in het licht der huidige linguïstiek, de voorwaarden te onderzoeken waarin een algemene thesaurus der Franse taal tot stand zou kunnen komen.

Tijdens dit colloquium hield B. Quemada, toen nog „chargé d'enseignement” aan de universiteit van Besançon, een lezing over „La technique des inventaires mécanographiques”. Deze voordracht, waarin spreker de tegenwoordige en toekomstige mogelijkheden der lexicologie met behulp van machines behandelde, was bijzonder sensationeel¹⁾. Hij wees op de grote bezwaren die aan het traditionele lexicologische handwerk zijn verbonden, t.w. het moeizame sorteren van een apparaat bestaande uit miljoenen fiches en de moeilijkheid om een dergelijk apparaat voor velen op verschillende plaatsen toegankelijk te maken. Vervolgens wees hij op de wenselijkheid om een woordapparaat te bezitten dat niet uitsluitend de lexicografie ten goede zou kunnen komen, en op het gebruik van mechanografische apparatuur in de Verenigde Staten, in Sovjet-Rusland en Italië, niet alleen bij het samenstellen van bibliotheekcatalogi, maar ook van indices op teksten. Quemada vermeldde dat hij in contact was met pater Busa, die in Gallarate met mechanografische middelen de *Index Thomisticus* bewerkte, en kondigde aan dat in de toekomst het langdurige en kostbare vervaardigen van fiches met de hand, geautomatiseerd zou worden, terwijl eveneens het mechanografisch sorteren en reproduceren ervan mogelijk zou zijn. Bij dezelfde gelegenheid deelde hij mee, dat er te Besançon een „Centre d'Étude du Vocabulaire Français” „in statu nascendi” was²⁾.

Dit „Centre”, in november 1957 door B. Quemada opgericht als een instituut van de „Faculté des Lettres et Sciences humaines” der universi-

1) Zie K. Baldinger, *Automation und Lexikologie* 541 [1959].

2) Zie B. Quemada, *Technique des inventaires* 56 [1960].

teit van Besançon, stelt zich ten doel het onderzoek van de Franse woordenschat te stimuleren en te coördineren, alsmede de verbinding en de samenwerking tussen Franse en buitenlandse navorsers en specialisten op dit gebied te verzekeren. Het „Centre” is dus geen onafhankelijk centrum van onderzoek, doch een onderdeel van de universiteit. Van het Ministerie van Onderwijs te Parijs heeft het formeel de opdracht gekregen studie te maken van de woordenschat der klassieke Franse tragedie.

Het eerste werk bestond in het inventariseren van oude(re) bewijsplaatsen met het oog op de historische studie van de Franse woordenschat. Een andere taak die het CEVF zich had gesteld, was de oprichting van een „Laboratoire d'analyse lexicologique”, ten einde een algemene en volledige lexicologische inventaris van de Franse woordenschat tot stand te brengen. Begin 1959 werd dit „laboratoire” een feit. De inventaris die men zich ten doel stelt, moet zijn: volledig, cumulatief, d.w.z. hergroepeerbaar, gemakkelijk toegankelijk, verspreidbaar, d.i. gemakkelijk te reproduceren, en ten slotte financieel realiseerbaar, d.w.z. niet duurder dan het traditionele handwerk. Een inventaris waaraan dergelijke eisen worden gesteld, is alleen te verwezenlijken met machines, met mechanografische en elektronische apparaten. Deze werken automatisch vanaf de invoer der gegevens tot de uitvoer der resultaten.

Het lexicologisch apparaat zoals dat op het „laboratoire” tot stand komt, is uiteraard nauw verbonden met de lexicografie. Voor elke lexicografische onderneming op het gebied der Franse taal is het immers te gebruiken. Het is echter bepaald niet zo dat dit apparaat aan zuiver lexicografische doeleinden ondergeschikt zou zijn gemaakt, m.a.w. uitsluitend bedoeld zou zijn als grondslag voor een woordenboek. „J'ai fait le Centre pour avoir un inventaire de la langue française dans toutes les directions”, zei Professor Quemada me.

§ 2. HET WERK OP HET „LABORATOIRE D'ANALYSE LEXICOLOGIQUE”

Uitrusting

Het „Laboratoire d'Analyse Lexicologique” werkt met een machinepark geleverd door de Franse maatschappij BULL te Parijs. Het park dat 50 miljoen A.Fr., d.i. 500.000 N.Fr. heeft gekost, bestond in 1960 uit de volgende apparaten:

- 1 (thans 2) handponsmachine P 80
- 2 (thans 4) bandponsapparaten verbonden met elektrische schrijfmachines¹⁾
- 2 ponsbandleesapparaten verbonden met
- 2 (thans 3) alfanumerieke ponsmachines PELEROD MC

¹⁾ De bandponsapparaten en de elektrische schrijfmachines zijn echter niet afkomstig van de firma BULL.

- 1 SCR (*Sélectionneuse, contrôleuse, retourneuse*)
- 1 photo-lecteur
- 1 (thans 2) vertolker
- 1 reproducerende ponsmachine PRD
- 1 tabelleermachine BS ¹⁾ verbonden met
- 1 totaalkaartenponsmachine PC
- 1 reporteuse-vertolker
- 1 (thans 2) sorteermachine D 3
- 1 tussen- en uitsorteermachine (*interclasseuse*)
- 2 alfanumerieke controleponsmachines VINOD ²⁾.

De klassieke kantoormachines waren echter niet zonder meer voor lexicologische doeleinden te gebruiken; er was een aanpassing, vaak van ingewikkelde aard, noodzakelijk, die nog nooit eerder was verwezenlijkt. Deze machines, die normaal slechts 36 alfanumerieke tekens verwerken, dienden geschikt te worden gemaakt voor de behandeling van 59 tekens, d.w.z. 23 meer. Een nieuwe codificatie van de grafische tekens van het Frans was noodzakelijk; het normale aantal van twee ponsingen per kolom moest hiervoor worden opgevoerd tot het absolute minimum van drie. Verder dienden ook veranderingen te worden aangebracht aan de sorteer- en tabelleermachine, de vertolker en de interclasseuse. De verschillende machines die te Besançon in gebruik zijn, komen verderop nog ter sprake bij het behandelen van de op het „Laboratoire” gevolgde werkwijze.

De kaart

De mechanografische kartotheek bestaat uit een verzameling fiches, d.w.z. kaarten van het standaardformaat 187,3 mm bij 82,54 mm. De kaart is vervaardigd uit karton dat bestand is tegen mechanische vervorming. In deze kaarten worden gaatjes geponst. Het aanbrengen van een cijfer op de kaart gebeurt door een ponsing in een der kolommen en wel ter hoogte van de cijfers (zwarte vakjes) 0 tot 9, of van een der bovenste posities 12 en 11. Het aanbrengen van een letter of van een interpunctieteken maakt een tot drie ponsingen in dezelfde kolom noodzakelijk; zo heeft de *s* de ponsing 11, de *a* de ponsingen 7 en 11, de *w* de ponsingen 7, 9 en 12 ³⁾. De mechanografische inventarisering in de kolommen 37-80 berust op het ponsen van cijfers in deze kolommen; zo krijgt b.v. een rijmwoord de ponsing 9 in kolom 71. Op deze kaarten is ook nog plaats voor maximaal 19 regelen druks, resp. 10 of 12 (als men de bovenste posities 12 en 11 meetelt) regels, als men afziet van de ruimte bestemd voor de ponsingen. De kaarten kunnen bovendien als

¹⁾ Thans bovendien nog een tabelleermachine TAS.

²⁾ Men zie verder de afbeeldingen achter in het boek.

³⁾ De klassieke mechanografie kent in de regel slechts twee ponsingen per kolom.

gewone, traditionele fiches beschreven worden. Verder kunnen de ponskaarten nog worden gebruikt om er een microfilm in aan te brengen; dit is een interessante mogelijkheid voor het geval dat men een oude inventaris snel mechanografisch toegankelijk wil maken. De ponsgaten beantwoorden aan een alfabetische of numerieke code, die vertolkt kan worden in gegevens welke op de bovenrand van de kaart worden gedrukt, t.w. door een machine die men vertolker noemt. Verscheidene van deze regels die op verschillende kaarten staan, kunnen machinaal op één andere kaart worden afgedrukt, namelijk door de reporteuse (die ook als vertolker kan worden gebruikt), tot een maximum van 19 regels, t.w. op een „witte” kaart, waarin geen ponsingen zullen worden aangebracht. Elke kaart heeft 80 kolommen voor het aanbrengen van code-tokens en hun vertolking. Men heeft acht verschillende kaarttypen en bovendien een volledig witte kaart. Het kaarttype waarop alleen cijfers voorkomen (continu genummerde regels en cijfers der kolommen), de zgn. standaardkaart, komt in vier kleuren voor, en wordt voor verschillende doeleinden gebruikt; de kleur wordt bij het begin van elk werk afgesproken.

De voornaamste, t.w. de woordkaart welke dient voor de thesaurus, verdient een nadere bespreking. Ze is ingedeeld in de volgende kaartvelden, ook wel ponsvelden genoemd, of rubrieken:

- kolom 1: code van de kaartsoort: dient voor de hergroepering van elke kaartverzameling en de controle der diverse machinale operaties ¹⁾).
- kolom 2-16: code van de auteur, code van de tekst en verschillende referenties: deel, bladzijde, vers of regel, nummer van het woord in het vers of in de regel.
- kolom 17-36: woord in de spelling van de tekst.
- kolom 37-40: aanduidingen van grammaticale aard, b.v. genus en getal, of transitief, intransitief enz.
- kolom 41-49: codificatie der betekenissen volgens zgn. begripvelden of gebruikssferen.
- kolom 50-52: datering van de tekst.
- kolom 53-56: gegevens over de auteur (leeftijd, generatie, beroep, godsdienst enz.).
- kolom 57-61: gegevens over het werk (proza, poëzie, literaire tekst, techniek enz.).
- kolom 62-66: gegevens over de context (variant, correctie, conjectuur, type van de spreker enz.).
- kolom 67-71: gegevens betreffende het woord (cursief, hoofdletter, tussen

¹⁾ Elk type kaart heeft een ponsing in kolom 1; de woordkaarten krijgen de ponsing 3, de tekstkaarten 4, de „carte mot second élément de mot composé” krijgt 8.

aanhalingstekens, plaats van het woord in de perikoop, rijm).

kolom 72-80: numerieke code van het woord (stamboeknummer van het woord in het algemene repertorium der vormen, wat de mechanische hergroepering der spellingvarianten en der buigingsvormen mogelijk maakt).

Uit het geringe aantal kolommen dat voor elke rubriek werd gereserveerd, mag niet worden geconcludeerd dat het aantal mogelijke aanduidingen in code eveneens uiterst klein zou zijn. Zo betekenen vier kolommen voor gegevens van grammaticale aard niet, dat de ponskaart hier slechts $4 \times 10 = 40$ mogelijkheden zou hebben. Uitgaande van 10 regels (posities)¹⁾ zijn er bij één ponsing weliswaar niet meer mogelijkheden dan hokjes (ponsposities), t.w. veertig, doch bij twee ponsingen wordt het aantal perforatiepatronen groter en bij drie nog groter. Tachtig kolommen van tien regels betekenen bij een enkele ponsing per kolom, althans theoretisch, 800 mogelijkheden van codering en classificatie²⁾.

Dit grote aantal mogelijkheden der mechanische inventarisering wordt in de praktijk slechts gaandeweg ook daadwerkelijk toegepast. Niemand kan trouwens van tevoren zeggen of de aangebrachte „zonage”, d.i. indeling in ponsvelden, ideaal is, of voor sommige onderdelen niet te veel en voor andere te weinig plaats werd ingeruimd. Gegevens die nu reeds op de kaart voorkomen, resp. daarop gemakkelijk aangebracht kunnen worden, zijn die betreffende de datum, de auteur en het werk; deze aanduidingen worden achteraf in serie op de verschillende kaarten van een tekst aangebracht. De numerieke code van het woord, waarbij de alfabetische volgorde door een numerieke wordt vervangen, vergt echter een langdurig en moeilijk voorbereidend werk, dat vermoedelijk pas op lange termijn verwezenlijkt zal kunnen worden. De problemen die de praktische toepassing van de inventarisering meebrengt, zijn nog lang niet alle theoretisch, laat staan praktisch opgelost; de mogelijkheden echter zijn er.

Neemt men als uitgangspunt het maken van een alfabetische woorden-

¹⁾ Er zijn eigenlijk 12 posities of regels; in de twee bovenste posities worden in de mechanografie vaak besturingsponsingen aangebracht.

²⁾ Ponsst men, i.p.v. 1 hokje per kolom, 2 of 3 hokjes (steeds in dezelfde kolom), dan wordt het aantal mogelijkheden enorm groot: twee ponsingen geven 4, drie ponsingen 8 mogelijkheden enz. Bij 4 kolommen van tien posities zijn de mogelijk-

heden als volgt: 1 ponsing: $\frac{40}{1} = 40$ mogelijkheden; 2 ponsingen: $\frac{40 \times 39}{1 \times 2} = 780$

mogelijkheden; 3 ponsingen: $\frac{40 \times 39 \times 38}{1 \times 2 \times 3} = 9.880$ mogelijkheden; 4 ponsingen:

$\frac{40 \times 39 \times 38 \times 37}{1 \times 2 \times 3 \times 4} = 91.390$ mogelijkheden.

RÉFÉRENCES																	FORME									GRAM.				CODE SENS									DATE			AUTEUR				ŒUVRE					CONTEXTE					MOT					CODE NUM.																																																															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9																																													

CENTRE D'ÉTUDE DU VOCABULAIRE FRANÇAIS - 2

Afb. 1.

Woordkaart in gebruik bij het CEVF.

lijst, dan kan men daarbij de volgende stappen onderscheiden:

- prepareren van de tekst
- ponsen van de band
- aanmaken van ponskaarten
 1. tekstkaarten
 2. woordkaarten
- controle en correctie der
 1. tekstkaarten
 2. woordkaarten
- „compostage” (ponsen van het rangnummer van het woord in de regel)
- alfabetisch sorteren (voorafgegaan door het sorteren der woorden naar hun lengte)
- drukken van een voorlopige woordindex
- controle en correctie van de woordindex
- afdrukken van de definitieve woordindex
- inbinden van de woordindex.

We zullen nu achtereenvolgens de verschillende, hierboven slechts schematisch aangegeven werkzaamheden bespreken.

Prepareren

Vóór de gegevens in de machine kunnen worden ingevoerd, d.w.z. vóór een tekst machinaal geëxcerpeerd of in zijn geheel geficheerd kan worden, is dikwijls een zekere technische voorbereiding door de taalkundige specialist vereist.

Dit prepareren van de tekst verschilt al naar gelang van het doel dat men zich stelt en van het soort van teksten waarmee men te maken krijgt. Elke tekst brengt dus in zekere zin altijd zijn eigen machinale problemen mee. Het essentiële bij het prepareren is, dat de machine een tekst krijgt ingevoerd die verdeeld is in perikopen waarvan de lengte een zeker aantal tekens (70), beantwoordend aan het aantal beschikbare kolommen op de standaardkaart, niet te boven gaat. Teksten in verzen zijn al vanzelf in dergelijke perikopen ingedeeld, zodat het probleem zich praktisch alleen maar voordoet bij prozateksten. Bij prozateksten die geen voorbereiding door de taalkundige specialist vereisen, is het mogelijk ze vrij „mechanisch” door de ponstypiste te laten verdelen in perikopen ¹⁾. Tijdrovend wordt het prepareren van een tekst, wanneer men het nodig vindt ook de resultaten van de tekstkritiek, b.v. emendaties of varianten, erin aan te brengen. De tijd die voor het prepareren vereist wordt, is dus zeer variabel; hij wisselt echter niet alleen naar de externe aard (poëzie/proza) van de tekst, doch ook naar de eisen en verlangens van de taalkundige specialist die hem prepareert, wat op zijn beurt weer afhankelijk is van de interne aard van een gegeven tekst.

¹⁾ Zie B. Quemada, *Mécanisation* 23 [1959].

Het mechanografische werk

De geprepareerde tekst wordt door een ponstypiste, met behulp van een ponsende schrijfmachine met 59 tekens, geponst in een papieren bandje met 8 sporen, waarvan 7 voor informatie en 1 voor controle (er is ook een positie- of dienstspoor voor transport van de band). Dit laatste gebeurt, volgens een bepaalde code, automatisch door het aanslaan van letter- en cijfertekens op een elektrische schrijfmachine van het gewone type. Hierbij wordt het bandje afgewikkeld op een bandponsapparaat dat met de schrijfmachine verbonden is. De door de schrijfmachine geleverde informatie wordt elektromechanisch omgezet in de door het ponsapparaat vereiste vorm. Samen met de tekst worden door de typiste ook bepaalde tekens, dienstcodetekens of bewerkingscodetekens, in de vorm van besturingsponsingen aangebracht, die naderhand de machine automatisch bepaalde opdrachten zullen laten uitvoeren. De ponsband wordt gebruikt als tijdelijke informatiedrager; de erin vastgelegde gegevens worden in een later stadium overgebracht in ponskaarten¹⁾. De onmiddellijke verificatie van de geponste tekst, met mogelijkheid tot directe correctie, kan geschieden met behulp van het blad papier in de schrijfmachine, dat de getikte (en geponste) tekst met de dienstcodetekens reproduceert. Een grondiger verificatie kan naderhand gebeuren door het lezen van de ponsband met een leesapparaat dat automatisch de geponste tekst door middel van een ontvangende schrijfmachine reproduceert. Als deze tekst klaar is, kan men hem vergelijken met de oorspronkelijke tekst van het boek en de fouten aanstrepen. Daarna kan men de eerste ponsband automatisch dupliceren, waarbij men foutieve regels of grotere tekstgedeelten kan verbeteren²⁾. Het is ook mogelijk de eerste band te verbeteren zonder hem te dupliceren³⁾.

Is de definitieve ponsband eenmaal klaar, dan gaat hij naar een ponsbandleesapparaat, dat de, via een verwisselbaar schakelbord, ermee gekoppelde ponsmachine automatisch, kolom na kolom, met de snelheid van ongeveer 12 kolommen per seconde, alle bewerkingen laat uitvoeren die nodig zijn voor het tot stand komen van de ponskaart. Constante gegevens worden eenmaal uit de band gelezen en opgenomen in een geheugen om te worden geponst in alle kaarten waarvoor ze gelden. Bij de kaarten die op grond van de band geponst worden, onderscheidt men, naar gelang van het doel en de aard van het werk, twee types, t.w. woordkaarten en tekstkaarten. Voor het samenstellen van een

1) Bij sommig lexicologisch werk is de band als uitgangspunt praktischer dan de ponskaart; het ponsen van deze laatste gaat langzamer, terwijl bovendien nog een tweede ponsing, de controleponsing, nodig is.

2) Te Besançon werd in 1960 de ponsband nog niet systematisch gecorrigeerd, met als gevolg een veel te groot aantal foute ponskaarten.

3) Men knipt dan de foutieve gedeelten uit de band en plakt er, met behulp van speciale apparatuur, nieuwe, foutloze stukjes in. Zie ook blz. 91.

woordindex b.v. heeft men alleen maar woordkaarten nodig die in de bovenrand, boven de eerste regel van de ponsposities dus, de nodige referenties (kolom 2-16) en het woord (kolom 17-36) bevatten. Voor concordanties of voor een lexicologische inventaris daarentegen heeft men bovendien nog de tekstkaart nodig, die in de bovenrand de referenties (kolom 2-16) en een regel tekst (vers of perikoop: kolom 14-80) bevat.

De tot stand gekomen ponskaarten, die elk een hoeveelheid informatie in de vorm van ponsgaten bevatten, worden daarna door de reporteuse, die voor het lexicologische werk als vertolker gebruikt wordt, vertolkt, d.w.z. de referenties en het woord, of de referenties met de regel tekst, die tot dan toe op de ponskaart alleen maar in de vorm van ponsgaten voorkwamen, worden nu automatisch gedrukt in de ponsvrije bovenrand, desgewenst ook in een van de verschillende interlinies van de kaart (snelheid 3.900 kaarten per uur).

In dit stadium vindt te Besançon een eerste verificatie van de kaarten plaats. De ponsstypiste vergelijkt de tekst op de ponskaarten met die in het boek en vervangt foutieve kaarten door nieuwe, die ze met de hand op de ponsmachine ponsst en daarna machinaal laat vertolken door de reporteuse. Bij het ponsen behoeft de typiste geenszins de foutieve kaart geheel opnieuw te ponsen; immers door een foute kaart gedeeltelijk als moeder- of stamponskaart te gebruiken, kan men ze gedeeltelijk, t.w. het niet-foutieve gedeelte, automatisch kopiëren. Tekstkaarten en woordkaarten moeten apart worden gecorrigeerd; immers ze komen weliswaar tot stand op grond van een en dezelfde ponsband, doch volgens verschillende pons- en vertolkingsbewerkingen.

Een tweede verificatie vindt thans plaats, door de tekst van de verschillende kaarten op een doorlopende lijst in de volgorde van het oorspronkelijke werk af te drukken met een tabelleermachine. Ook deze werkt volgens een door de gebruiker bepaald programma, dat wordt vastgelegd in een schakelbord. De tabelleermachine krijgt de kaarten in de gewenste volgorde toegevoerd en drukt ze onder elkaar af. De machine die thans te Besançon in gebruik is, heeft een maximum snelheid van 150 kaarten (=regels) per minuut. Hierbij kunnen echter, evenals bij de hierboven genoemde reporteuse, slechts 45 verschillende tekens worden afgedrukt¹⁾, zodat sommige tekens die wel op de schrijfmachine voorkomen, dus ook in de band en op de kaarten worden geponsst, niet tot hun recht komen, zo b.v. niet het onderscheid tussen *v* en *w*. Het gevolg hiervan is dat men niet kan controleren, of de ponsstypistes wat dat betreft fouten hebben gemaakt. Een nieuwer model (BULL, imprimante 300 T.I.) kan echter 59 tekens drukken met een grotere snelheid. Bij de

¹⁾ Dit is mogelijk dank zij een zonestelsel in de afdrukposities; in sommige zones worden standaard alfanumerieke wielen met 36 tekens, in andere echter speciale wielen zonder cijfers, doch met 10 extra alfabetische tekens gemonteerd.

verificatie en de correctie van de ponskaarten zijn natuurlijk allerlei variaties mogelijk. Bij het tot nog toe te Besançon gevolgde systeem echter, waarbij dus de ponsband niet wordt gecorrigeerd, zijn twee verificaties van de kaarten (1°. kaarten met het boek; 2°. tabelleerlijst met het boek) wel het absolute minimum.

Nu volgt de zgn. „compostage”, waarbij de woorden op de woordkaarten voorzien worden van het volgnummer in het vers of in de perikoop. Deze operatie, waarbij de kaarten weliswaar met grote snelheid, doch uiteraard een voor een bewerkt worden, wordt uitgevoerd op de reproducerende ponsmachine, PRD (*Perforatrice, Reproductrice, Duplicatrice*). Deze operatie is mogelijk doordat de typiste, bij het ponsen van de band, het laatste woord van het vers of van de perikoop voorziet van een speciaal codeteken, het zgn. rijmcodeteken. Dit laatste, dat op de woordkaart in kolom 71 op regel 9 wordt geponst, dient niet uitsluitend om er een register van rijmwoorden mee te kunnen aanleggen, doch ook om de woordorde in de regel te kunnen aangeven; deze woordorde immers maakt het mogelijk automatisch de woordkaarten opnieuw te groeperen in de oorspronkelijke volgorde van de tekst, indien de kaarten b.v. tussendoor alfabetisch gerangschikt werden.

Zijn woord- en tekstkaarten eenmaal behoorlijk verbeterd en de woordkaarten voorzien van een nummering die de volgorde in het vers of in de perikoop aangeeft, dan worden deze laatste, de woordkaarten dus, gealfabetiseerd in een elektronische sorteermachine. Deze laatste sorteert de kaarten in twee bewerkingen (eerst *A-N*, vervolgens *O-Z*) met een snelheid van 700 kaarten per minuut. Per letter waaruit de woorden bestaan, zijn, zelfs bij 3 ponsingen, slechts 2 doorgangen per kolom nodig. Als gesorteerd moet worden op b.v. maximaal 10 letters, worden de kaarten $20 (= 10 \times 2)$ maal door de machine gevoerd, zodat dan per minuut $700 : 20 = 35$ kaarten worden gerangschikt.

De alfabetische orde van de sorteermachine is natuurlijk zuiver automatisch. Voor een moderne tekst die in een strikt uniforme spelling is gesteld, is dit uiteraard geen bezwaar, integendeel. Maar vele teksten waarmee de lexicoloog te maken krijgt, vertonen deze uniforme spelling niet, zodat het nodig is deze machinale, strikt formele, alfabetische volgorde met de hand bij te werken. Hiertoe wordt op de tabelleermachine het alfabetische resultaat van de sorteermachine afgedrukt; aan de hand van dit alfabetisch overzicht is het mogelijk snel bij elkaar te plaatsen wat bij elkaar behoort (b.v. *ÿs* en *ys* die bij het mechanisch sorteren in verschillende vakken terechtkomen, spellingvarianten en buigingsvormen die thans onder de hoofdvorm gerangschikt worden), en kaarten met verwijzingen toe te voegen.

Liggen de woordkaarten eenmaal in de gewenste volgorde, dan gaat men over tot het afdrukken van de definitieve woordindex, desgewenst op een speciaal voor Besançon geconstrueerd drukwerk. Dit bestaat uit een VIN (*Vérificateur, Intégral, Numérique*) of controleponsmachine,

die ponskaarten kan lezen, verbonden met een ontvangende elektrische schrijfmachine ¹⁾). Dit kleine drukwerk heeft niet alleen het voordeel dat het lijsten vervaardigt in gewoon machineschrift, doch ook dat het in staat is de 59 tekens weer te geven van het toetsenbord waarmee de ponsband tot stand komt. Deze in machineschrift vervaardigde tabellen kunnen, na fotografering op het gewenste formaat, in offset gedrukt worden.

Duur van de bewerkingen

Bij het maken van een woordindex op een klassieke tragedie van 1900 verzen, of ongeveer 18.000 woorden is de tijdsduur van de verschillende machinale bewerkingen als volgt:

– totstandkoming van de ponsband (typen met eventuele machinale verificatie)	18 uren
– ponsing van de kaarten	3 „
– vertolking van de kaarten	4½ „
– controledruk	2 „
– alfabetische classificatie	7 „
– definitieve afdruk van de index	2 „

Totaal: 36½ uren

Dit eindcijfer is machinaal gesproken niet geflatteerd. Met nieuwere en snellere machines zouden deze 36 uren waarschijnlijk wel tot 25 of 20 te reduceren zijn. Toch duurt de totstandkoming van de machinale index op een tragedie van 18.000 woorden te Besançon veel langer dan 36 uren. Immers in dit overzicht is de verificatie met correctie alleen maar vertegenwoordigd met de hierboven pro memorie vermelde verificatie van de band, die echter, zoals gezegd, in 1960 nog niet plaatsvond. In elk geval, bij het te Besançon gevolgde procédé duurt de verificatie en de correctie minstens even lang als de totale machinale bewerking.

Hoe nuttig indices in de lexicologie ook zijn, voor lexicografische doeleinden zijn ze toch uiteraard niet meer dan een voorstadium. Immers wat de lexicograaf nodig heeft zijn fiches, kaarten met een woord gevolgd door een context. Een kaart die de lexicografische zeer nabijkomt, is de concordantiekaart. Wanneer men weet dat de omvang van een concordantie, waarbij elk woord slechts een minimum aan context van twee verzen meekrijgt, meer dan twintigmaal groter is dan de tekst zelf, dan is het duidelijk dat op dit speciale gebied, dat der concordanties, de mechanisering wel de oplossing, ja zelfs de enige mogelijkheid is.

¹⁾ Dit drukwerk is eigenlijk een variatie van een „convertisseur carte-bande” of kaart naar band omzetapparaat, d.i. een combinatie van een controleponsmachine VIN met een bandponsapparaat; de combinatie leest de ponskaarten en reproduceert de afgetaste gegevens in ponsband.

Heeft een mechanische concordantiekaart een context die uitvoerig genoeg is, dan is ze tevens bruikbaar voor het zuiver lexicografische werk. Aangezien men uitgaande van de ponsband niet alleen woordkaarten doch ook tekstkaarten kan maken, kunnen deze laatste, wanneer ze eenmaal vertolkt en geverifieerd zijn, gebruikt worden om op de woordkaarten een context aan te brengen, waarvan de lengte door de bewerker te bepalen is. Het aanbrengen van deze contexten van wisselende lengte geschiedt met de reeds in een ander verband genoemde reporteuse, die met een snelheid werkt van 3900 kaarten per uur. Het mechanografisch aanbrengen van de context is een bewerking die voorlopig nog vrij veel tijd kost, minstens zoveel als het ponsen van de tekst op de ponsband. De wisselende lengte van de gewenste context vormt een probleem op zichzelf ¹⁾).

Beweeglijkheid van de kaartverzameling

Bij het bespreken van de mechanografische ponskaart heb ik reeds gewezen op de vele dimensies die een dergelijke kaart heeft; deze dimensies hangen af van de gegevens die de kaart bevat. Gegevens die eenmaal in ponskaarten werden vastgelegd, kunnen herhaaldelijk snel naar verschillende gezichtspunten worden gegroepeerd. Uit een lexicologische inventaris op ponskaarten kan men machinaal, in één enkele bewerking met behulp van de sorteermachine, de kaarten halen die beantwoorden aan één criterium (b.v. alle kaarten uit de 16de eeuw tussen 1500 en 1520) of aan een combinatie van criteria (b.v. de namen der kleuren, in prozateksten tussen 1600 en 1630). Een mechanografische lexicologische kartotheek is stellig niet duurder dan een die tot stand komt langs de weg van het traditionele handwerk. Maar zelfs als de kosten gelijk te stellen zouden zijn ²⁾, dan staat tegenover een grotere aanvankelijke investering aan de basis een enorm aantal classificatie- en dus ook exploitatiemogelijkheden. In vergelijking hiermee zou men de traditionele kartotheek bijna steriel kunnen noemen.

Kaartenproductie en prijs

De-mogelijkheden der produktie aan kaarten hangen uiteraard af van de grootte en de bezetting van het machinepark, alsmede van de tijd dat de machines in werking zijn. Het was in 1960 nog niet zo, dat alle machines van het „Laboratoire” te Besançon constant op volle toeren draaiden. Als dat wel het geval was, zou men tien miljoen kaarten per jaar kunnen vervaardigen. Uitgangspunt der produktie is de hoeveelheid tekst die op een ponsband wordt geponst. Een goede ponstypiste ponst ongeveer

¹⁾ Zie blz. 69.

²⁾ Het archief van het „Goethewörterbuch” te Tübingen bezit circa anderhalf miljoen traditionele fiches, die per stuk 1 DM hebben gekost. Voor een bedrag van 1.500.000 DM zou men 3 à 4 lexicologisch-mechanografische centra kunnen uitrusten.

10.000 woorden per dag. Eind januari 1959 werden de machines in werking gesteld: in september 1960 bevatte de kaartverzameling van het „Laboratoire” ongeveer 1.600.000 kaarten, opgeborgen in stalen kaartenbakken, die elk ongeveer 2500 kaarten bevatten; eind 1960 waren er zowat 2.000.000 en in 1962 4.000.000 kaarten ¹⁾).

De prijs der blanco kaarten bedraagt, volgens een opgave van BULL-Nederland f 6,25 per duizend (manilla) en f 8.— per duizend (wit), d.w.z. resp. 0,6 en 0,8 cent per stuk. Een machinale woordkaart kost uiteindelijk 6 A.Fr. of 0,06 N.Fr. zonder context, en 12 A.Fr. of 0,12 N.Fr. met context, als deze laatste niet uitvoeriger is dan vier regels.

Personeel

In september 1960 waren bij het „Laboratoire” werkzaam: 2 (pons)-typistes, die de banden ponsten en incidenteel ook wel de machines bedienden, en 2 machinebedienden, belast met de werkzaamheden vanaf de ponsband. Dit personeel dat tot de universiteit behoorde, was bij het „Laboratoire” gedetacheerd. Verder had men nog de beschikking over twee gepensioneerde onderwijzers, die gedurende halve dagen werkten; hun taak bestond in het vergelijken van ponskaarten met de originele tekst.

§ 3. WAT REEDS IN HET CEVF WERD VERWEZENLIJKT

1) *Indices en Concordanties*

Officieel bestaat de taak van het CEVF in het opmaken van de woordinventaris van Franse poëtische teksten uit de 16de en 17de eeuw; ook 19de-eeuwse teksten werden echter reeds bewerkt ²⁾). Men beperkt zich daarbij niet tot het op ponskaarten brengen van de woorden, doch legt, op grond van de voorhanden woordkaarten, tevens indices en concordanties op de geïnventariseerde teksten aan. Tal van indices en concordanties (b.v. op Racine, *Bérénice*) zijn reeds klaar en worden eerstdaags in offset-druk uitgegeven.

Rijmwoordenindex

Bij het ponsen van de band wordt, zoals gezegd, elk rijmwoord van een speciaal codeteken, t.w. een ponsing op regel 9 in kolom 71, voorzien. Op grond hiervan is het mogelijk, mechanisch een index van de rijmen in een bepaald werk aan te leggen. Deze rijmen kunnen daarbij op verschillende manieren worden geclassificeerd, hetzij alfabetisch, hetzij in de volgorde van de tekst, zodat men de verschillende rijmparen verkrijgt. Met behulp van de reproducerende ponsmachine PRD is het eveneens mogelijk een zelfstandige rijmkaartenverzameling aan te leggen, waarbij

¹⁾ Zie *Cah. de Lex.* 3, 60 [1962].

²⁾ Voor een lijst van deze teksten zie *Cah. de Lex.* 3, 60 [1962].

op de kaarten, naast elkaar, de twee rijmwoorden met hun referenties voorkomen. Dit maakt dan een hergroeperen en drukken van de corresponderende rijmen mogelijk. Als gevolg van de verschillende mogelijkheden van het alfabetisch sorteren, kunnen ook rijmindices worden gedrukt volgens de alfabetische volgorde der beginklanken of der eindklanken.

Frequentie-index

Een totaalkaartenponsmachine (ook de PRD kan als zodanig fungeren) gekoppeld aan de tabelleermachine, maakt het mogelijk, terwijl de lijst der ponskaarten wordt gedrukt, frequentiekaarten te maken, die per woord aangeven het aantal keren dat dit voorkomt. Deze verzameling frequentiekaarten wordt dan machinaal gerangschikt volgens de afnemende orde van het gebruik, waarna er een lijst van wordt gedrukt. Het CEVF heeft reeds eenmaal een dergelijke frequentielijst gemaakt, echter niet als onderdeel van het eigen werkprogramma. In opdracht van het „Centre du Français élémentaire pour la Biologie” namelijk werd een frequentielijst aangelegd van de woordenschat der plant- en dierkunde.

2) *Repertorium der Franse woordvormen in omgekeerde volgorde*

Wanneer in een woordenboek de lemmata in omgekeerde volgorde worden gerangschikt ¹⁾, d.w.z. volgens de alfabetische volgorde niet der begin- doch der eindletters, verkrijgt men een lijst die zeer belangrijk is voor vormleer en metriek ²⁾. Wat de vormleer betreft, weet men in een oogopslag, waar, d.w.z. bij welke woorden, een bepaald suffix voorkomt. Dit betekent dat men bij het machinaal exciperen of ficheren van een tekst de suffixen en de tweede leden der samenstelling kan verwaarlozen, voorwaar geen geringe besparing van tijd, energie en kosten. De grondslag voor een dergelijk repertorium werd reeds gelegd, namelijk door de „Littré à rebours”, een uitvoerige kaartverzameling, die daarna op de tabelleermachine in de vorm van een lijst werd gedrukt, en waarin de lemmata van het woordenboek van Littré alfabetisch werden gerangschikt op de laatste, voorlaatste letter enz. ³⁾.

3) *„Trésor des dictionnaires français”*

Met het maken van een inventaris van de woordenschat der Franse woordenboeken werd een begin gemaakt (Nicot [1606], Richelet [1680], Furetière [1690], Académie [1694]). Deze inventaris moet het mogelijk maken, in één oogopslag te zien of en wanneer een woord in de woorden-

¹⁾ Hieraan gaat een rangschikking volgens de woordlengte vooraf.

²⁾ Over het nut van een retrograde index, zie J. Štindlová, *Classement inverse* [1961]; E. Mater, *Herstellung eines rückläufigen Wörterbuches* [1962]. Zie verder nog blz. 56.

³⁾ De omgekeerde Littré zal eerstdaags verschijnen; zie *Bull. d'Inform.* 4, 55 [1961].

boeken voorkomt. Naderhand kunnen op grond van deze inventaris ook automatisch speciale woordenlijsten worden samengesteld, b.v. een lijst van alle woorden die door de woordenboeken worden gekarakteriseerd als behorend tot de volkstaal.

4) „*Répertoire historique du vocabulaire français*”

Met een voor de etymologie of woordgeschiedenis van de Franse woordenschat zeer belangrijk werk werd eveneens een aanvang gemaakt, t.w. met een historisch repertorium van de Franse woordenschat. De vele oude vindplaatsen die in de laatste halve eeuw werden gepubliceerd, zijn nog nooit systematisch verzameld. Het CEVF nu is hiermee begonnen. De basis van het werk is een woordponaskaart waarop in de kolommen 2-5 een lijst van de voornaamste historische woordenboeken (FEW, Bloch-Wartburg, Dauzat, Dictionnaire Général, Littré, Godefroy, Huguet, Tobler-Lommatsch, *Matériaux pour l'Histoire du Vocabulaire Français*) werd gedrukt. Op deze woordkaart komt het woord voor met dien verstande, dat men voor elke betekenis ervan telkens een aparte woordkaart aanlegt. In de interlinie, naast elk van deze historische werken, worden de gegevens die deze over het woord verschaffen vermeld, b.v. θ , d.i. het woord ontbreekt, of θd , d.w.z. het woord wordt wel vermeld doch zonder datum enz. Wanneer al deze gegevens bij elkaar zijn gebracht, wordt de oudste vindplaats in de kaart geponst. Op grond van dit kaartstelsel is het mogelijk een staat van de bekende dateringen te drukken en uit te geven.

Deze grondslag moet nu verder worden aangevuld met verwijzingen naar dateringen en vindplaatsen die in verschillende taalkundige tijdschriften en publikaties zijn verschenen. Bovendien moet de op deze wijze tot stand gekomen compilatie regelmatig bijgehouden en aangevuld worden. Oudere dateringen en vindplaatsen die door de onderzoekers en medewerkers van het CEVF gesignaleerd worden, worden eveneens op kaarten gebracht.

5) „*Datations et Documents lexicographiques*”

Een publikatie die een direct gevolg is van de werkzaamheden in verband met het „*Répertoire historique*” is de reeks *Datations et Documents lexicographiques*, waarvan, als uitgave van het CEVF, in de reeks *Matériaux pour l'Histoire du Vocabulaire Français*, een eerste deel van 265 bladzijden (*abacot-azyme*) in 1959, en een tweede deel van 213 bladzijden (*baba-byzantinisme*) in 1960 is verschenen.

6) *Oprachten voor instituten of particulieren*

Het CEVF aanvaardt en krijgt werkopdrachten van particulieren en van romanistische centra in binnen- en buitenland. Franse instellingen als de „*Faculté des Lettres*” te Parijs, te Lyon en te Poitiers, het „*Centre de Philologie romane de Strasbourg*”, het „*Centre de Recherches et*

d'Études pour la Diffusion du Français" van de „ENS de St Cloud" ¹⁾, het „Comité d'Action Scientifique de Défense Nationale" ²⁾ en het „Centre National de la Recherche Scientifique" lieten reeds machinaal lexicologisch werk door het „Laboratoire d'Analyse Lexicologique" uitvoeren. Uit Zweden kwam een opdracht binnen voor het machinaal ficheren van een uitvoerige Oudfranse tekst, uit de U.S.A. een voor het maken van een index op de werken van Ronsard. In 1962 was het „Centre" zelfs zogoed als geheel in beslag genomen door een statistische inventaris van het moderne Duits, zulks met het oog op het Grunddeutsch-project van Professor Dr. J. Alan Pfeffer ³⁾ van de universiteit van Buffalo (thans Pittsburgh). Uit Marokko kreeg men zelfs een aanvraag voor het machinaal op kaarten brengen van Arabische teksten, iets waar de machines van het CEVF uiteraard niet op berekend zijn. Op het „Laboratoire" werd bovendien ook werk uitgevoerd met actieve medewerking van buitenlandse navorsers verbonden aan de universiteiten van Manchester, Leiden, Poona en van de Akademie van Wetenschappen te Praag.

Als gevolg van deze opdrachten, beschikt het „Laboratoire" over meer geld dan de 30.000 N.Fr. van het jaarlijkse krediet, dat uitsluitend bestemd is voor de „frais de fonctionnement". Toch weet men op het CEVF veel te realiseren met naar verhouding weinig geldmiddelen.

Prof. B. Quemada

Aanvankelijk door de Franse collega's skeptisch bejegend, heeft de directeur van het CEVF reeds na korte tijd het nut, de levensvatbaarheid en de noodzaak van zijn mechanografisch experiment overtuigend weten te bewijzen. Tot op zekere hoogte is in het buitenland de belangstelling voor het CEVF groter dan in Frankrijk zelf. Behalve uit de reeds genoemde landen waren te Besançon reeds buitenlandse lexicologen uit India ⁴⁾, Polen ⁵⁾, Tsjecho-Slowakije (Dr. Jitka Štindlová) ⁶⁾, Oost-Duitsland, t.w. van het *Wörterbuch der deutschen Gegenwartssprache* (Dr. Ruth Klappenbach) en West-Duitsland (Prof. Dr. K. Baldinger uit Heidelberg met studenten), op bezoek. De toenemende belangstelling voor zijn werk is

¹⁾ Zie *Bull. d'Inform.* 4, 51-52 [1961] en G. Gougenheim et P. Rivenc, *La préparation du vocabulaire scientifique général* [1962].

²⁾ Zie *Bull. d'Inform.* 4, 51 [1961]. Uit dit contact is voortgekomen de studiegroep voor taalstatistiek, onder leiding van R. Moreau, van het „Laboratoire d'Analyse Lexicologique"; zie *Bull. d'Inform.* 6, II [1962].

³⁾ Zie *Bull. d'Inform.* 4, 51 [1961]; J. Alan Pfeffer, *Grunddeutsch: Werden und Wesen* [1962] en *La mécanisation dans la préparation statistique de l'allemand fondamental* [1962].

⁴⁾ Professor Sohoni, die het plan heeft opgevat aan de universiteit van Poona een gemechaniseerd centrum op te richten voor de normalisering der talen van India.

⁵⁾ Zie Halina Lewicka, *Metody mechanograficzne w leksykologii* [1959].

⁶⁾ Zie thans ook Jiřina Smrčková, *Mechanografické metody ve francouzské jazykovědě* [1961].

voor de directeur van het CEVF een steun en een aansporing om verder te gaan op de ingeslagen weg.

Hoewel Quemada zich niet als eerste heeft beziggehouden met de mechanisering der lexicologie, is het echter zijn verdienste, dat hij heeft gezien dat de langs mechanische weg tot stand gekomen theologische indices en concordanties konden leiden tot een volledige vernieuwing van de Franse lexicologie. Hij heeft niet alleen de enorme mogelijkheden gezien die de ponskaarten voor deze tak der linguïstiek zouden kunnen bieden, doch bovendien de moed en het doorzettingsvermogen gehad om het experiment te wagen. Dank zij zijn visie is een radicale ommekeer in de Franse lexicologie mogelijk geworden, zodat de totstandkoming van een grote Franse thesaurus concrete vormen kan gaan aannemen¹⁾. Het CEVF is nu reeds een onmisbaar werktuig van de Franse lexicologie geworden. Het mogen dan de ingenieurs van BULL zijn die de traditionele ponskaartmachines voor lexicologische doeleinden bruikbaar hebben gemaakt, de richtlijnen voor deze ombouw zijn afkomstig van Quemada, bij wie taalkundige eruditie en belangstelling voor en inzicht in machines op een even zeldzame als harmonische wijze samengaan. Hij is trouwens afkomstig uit een familie waarin de beoefening der exacte wetenschappen traditie is. Bij al het spectaculaire dat het laboratorium van het CEVF de belangstellende bezoeker biedt, doet men er toch goed aan, niet de grootscheepse conceptie te vergeten die achter dit experiment schuilgaat. Hier heeft de lexicologie, sinds de uitvinding der boekdrukkunst tot op onze dagen nog als zuiver traditioneel handwerk beoefend, het „industriële” stadium bereikt.

Als orgaan van het CEVF dient het tijdschrift *Cahiers de Lexicologie*, waarvan in 1959 het eerste, in 1960 het tweede en in 1962 het derde deel is verschenen. Uit het door Prof. Quemada geschreven hoofdartikel over de mechanisering in de lexicologie, dat daarin is opgenomen, heb ik voor mijn uiteenzetting rijkelijk kunnen putten. Daarin verschijnen, naast theoretische bijdragen op het gebied der lexicologische mechanisatie, ook bibliografische overzichten, waarin methodologische bijdragen van lexicologische en lexicografische aard worden gesignaleerd en waarin men ook systematisch verzameld vindt wat in de afgelopen jaren aan studies over de Franse woordenschat is verschenen.

Naast de *Cahiers* verschijnt ook een voorlopig nog gestencilde publicatie, namelijk het *Bulletin d'Information du Laboratoire d'Analyse Lexicologique*, waarvan reeds zes nummers (1960: I en II; 1961: III, IV en V; 1962: VI) zijn verschenen; deze *Bulletins* zijn meer bedoeld als praktisch werk- en verbindingsorgaan tussen de medewerkers aan en de gebruikers van het „Laboratoire”. Ik heb reeds vermeld dat de meeste rubrieken van de gemechaniseerde lexicologische kaart van Besançon voorlopig nog oningevuld blijven. De mogelijkheden tot exploitatie zijn

¹⁾ Zie hierover blz. 109.

afhankelijk van een voorafgaande normalisatie van de methode der analyse. Het *Bulletin* nu dient o.m. als forum van discussie om te komen tot definitieve criteria op het gebied der mechanische codificatie. De resultaten der colloquia die met dit doel regelmatig te Besançon worden gehouden, vindt men in het *Bulletin* opgenomen.

Als eerste taak stelt men zich, te komen tot collectief aanvaarde criteria voor de mechanografische exploitatie der rubriek „grammaire” (kolommen 37–40) op de ponskaart. Als inleiding tot de discussie over de mogelijkheden ervan verscheen in het eerste *Bulletin* een concept van de hand van H. Mitterrand, assistent aan de Faculteit der Letteren, over de normering van de mechanografische beschrijving der grammaticale vormen. Het tweede *Bulletin* bevat, behalve aanvullingen van de schrijver zelf, nieuwe voorstellen van de hand der heren Bonnard en Greimas. Het voorstel van Greimas heeft als grondslag gediend van de colloquia¹⁾ die in augustus 1960 te Besançon over dit onderwerp zijn gehouden, en die geleid hebben tot een „projet de codage morphologique” van de hand van Klaus Heger, van de universiteit van Heidelberg. Dit „projet” werd, samen met het verslag van het in augustus 1960 gehouden colloquium, gepubliceerd in het derde *Bulletin* (gedateerd 1960, verschenen voorjaar 1961), dat verder ook nog artikelen van H. Bonnard en G. Gougenheim over de codering en classificatie der grammaticale gegevens bevat.

In 1961 (30 juli–26 augustus) werd aan het CEVF, in aansluiting aan de colloquia van 1960, een „séminaire de lexicologie” georganiseerd, dat aan de taalkundige statistiek was gewijd²⁾.

In de nabije toekomst zullen de *Cahiers de Lexicologie* een taak krijgen die ver uitgaat boven de oorspronkelijke, nationale bedoeling. Een der conclusies van het internationale colloquium van Tübingen³⁾ in november 1960 was, dat de *Cahiers* van Besançon zouden fungeren als tijdschrift van het internationale informatiecentrum voor gemechaniseerde lexicologie. Dat de officiële wens geen officiële werkelijkheid is geworden – dit bleek duidelijk op het in 1961 te Besançon gehouden „Colloque international”⁴⁾ –, is wellicht minder belangrijk dan het feit, dat de *Cahiers*, sinds de verschijning van de tweede jaargang, de spreekbuis zijn geworden van een internationaal gericht forum. In deze omstandigheden kan men zich afvragen, of het niet gewenst moet worden geacht het speciaal romanistische, niet direct mechanografische gedeelte van de *Cahiers*, b.v. de bibliografie van het Franse woordonderzoek, onder te brengen in het *Bulletin d'Information*.

1) Zie ook K. Heger, *Automation in der Sprachwissenschaft* 537 volg. [1960].

2) Zie het programma in *Bull. d'Inform.* 4, 50 [1961].

3) Zie over dit colloquium de verslagen van G. Natalis [1960], F. de Tollenaere [1961] en E. Mater [1961].

4) Zie over dit congres F. de Tollenaere, *Het „Colloque international sur la mécanisation des recherches lexicologiques”* [1961]; J. Štindlová, *Uplatňování metod mechanizace a automatizace v lexicologické práci v zahraničí* [1962].

HOOFDSTUK II

HET „CENTRO PER L'AUTOMAZIONE DELL'ANALISI LETTERARIA" TE GALLARATE

E forse l'unica ricompensa terrena che egli possa attendersi è che si riconosca a che prezzo sia stata pagata la satisfazione di aver lavorato affinché altri possano più agevolmente lavorare,

R. Busa S.J., *Varia Specimina* 13 [1951].

§ 1. GESCHIEDENIS

De Italiaanse jezuïet Roberto Busa, die reeds in 1941 als docent verbonden was aan de filosofische faculteit van het Aloisianum te Gallarate bij Milaan¹⁾, promoveerde in 1942 aan de Gregoriaanse universiteit te Rome op de rationele conceptie van de „presenza" bij Thomas van Aquino²⁾.

Enkele jaren later werd hij gewoon hoogleraar in de wijsbegeerte aan de filosofische faculteit van het Aloisianum, waar men reeds bijna anderhalve eeuw studie maakt van de wijsbegeerte van Thomas van Aquino. Hij stelde in 1946 zijn oversten en andere deskundigen voor, een kartotheek aan te leggen van alle woorden die voorkomen in de werken van Thomas³⁾.

De redenen die een dergelijke *thesaurus verborum* noodzakelijk maakten, zette Busa uiteen op het internationale congres voor wijsbegeerte, dat in oktober 1948 te Barcelona werd gehouden⁴⁾. Aan het slot van zijn mededeling *L'interiorità secondo la scolastica*, wees hij op de noodzaak het onderzoek van het scholastieke denken een ruimer filologisch-grammaticale basis te verschaffen. Hem stond voor de geest:

a) een kaartenverzameling die naar schatting ongeveer 13 miljoen fiches zou omvatten. Als voorbeeld voor deze kartotheek zag hij die welke ten grondslag ligt aan de *Thesaurus Linguae Latinae* (TLL) te München⁵⁾. Deze berust namelijk op een door vijf Duitse universiteiten

¹⁾ Zie P. Tasman, *Lit. Data Proc.* 249 b, noot *) [1957].

²⁾ R. Busa, *La Terminologia Tomistica dell' Interiorità* 6 [1949].

³⁾ R. Busa, *Elektronentechnik* 20 b [1957]. In E. en K. Delavenay, *Bibliography of Mechanical Translation* 28 volg. wordt deze publikatie van Busa, met nog andere, ten onrechte onder de rubriek „mechanical abstracting and analysis" gerangschikt, i.p.v. onder de rubriek „indexing and concordance".

⁴⁾ Busa, *Complete Index* 424 [1950].

⁵⁾ Busa, *Complete Index* 425 [1950]; R. Busa, *Elektronentechnik* 21 a [1957].

aan het eind van de vorige eeuw bij elkaar gebracht „heerleger” van tien miljoen kaarten, elk met een tekst van ongeveer tachtig woorden¹⁾; zo zijn er alleen voor het woordje *et* veertig bakken met elk ongeveer 1.200 kaarten.

b) indices en concordanties die met behulp van de kaartverzameling gemaakt zouden worden²⁾.

Voor het financieren van dit werk nam Busa zich voor, een beroep te doen op Italiaanse industriëlen.

Aanvankelijk vormde het ontbreken van een kritische uitgave van de werken van Thomas een handicap. Aangezien de door de dominicanen te bezorgen kritische editie echter nog lang op zich zal laten wachten, werd besloten toch maar aan het werk te gaan. Een volledige index zou immers in ieder geval een onwaardeerlijke hulp zijn voor de tekstkritiek³⁾; bovendien zou de kartotheek naderhand, met behulp van de computer, in overeenstemming kunnen worden gebracht met een nieuwe kritische uitgave.

Een uitvoeriger uiteenzetting van de redenen waarom een *thesaurus verborum* op Thomas tot stand moet komen, vindt men in een onderzoek van Busa op het gebied van de terminologie bij Thomas van Aquino⁴⁾, dat in 1949 verscheen. In de voorrede van dit werk constateert de auteur, dat de moeilijkheden die we ondervinden bij de codificatie van het denken van de grootmeester der scholastiek, voor een groot gedeelte te wijten zijn aan het feit dat we de terminologie van de auteur, als het ware het geraamte van zijn filosofie, niet voldoende beheersen. Een terminologie die formeel gelijk blijft, kan toch in de loop der tijden van inhoud veranderen. Wil men verder komen dan het opprikken van dorre omhulsels van dode insecten, schrijft pater Busa, dan zal men zich van de inhoud der terminologie van Thomas zo objectief mogelijk rekenschap dienen te geven. Een terminologisch onderzoek dient methodologisch vooraf te gaan aan het onderzoek van het filosofische denken van de auteur. Als onontbeerlijke voorwaarde hiertoe zag Busa het tot stand brengen van een index, resp. een concordantie op de 13.000 foliobladzijden⁵⁾ der

1) R. Busa, *Varia Specimina* 14 en 34 [1951].

2) Van de *Summa Theologiae* van Thomas bestaat een, toen nog ongedrukte, concordantie door Prof. Roy J. Deferrari, van de R.-K. universiteit van Washington, en Zuster M. I. Barry. De bedoeling was de concordantie uit te geven, hetgeen financieel echter niet mogelijk bleek. Ten slotte kon in 1956 tenminste een gedeeltelijke index verschijnen: *A complete Index of The Summa Theologica of St. Thomas Aquinas*. Een, eveneens onvolledige, index door Professor Peter O'Reilly op de *Summa contra Gentes* was in voorbereiding. Zie R. Busa, *Complete Index* 425 [1950]. Over de werkwijze van Deferrari en P. O'Reilly, zie R. Busa, *Varia Specimina* 20-22 [1951].

3) R. Busa, *Complete Index* 425 [1950].

4) R. Busa, *La Terminologia Tomistica dell' Interiorità* [1949].

5) R. Busa, *Varia Specimina* 34 [1951].

werken van de „doctor angelicus”, welk werk tevens het materiaal zou kunnen leveren voor een toekomstig woordenboek op Thomas¹⁾. Hij kwam echter spoedig tot het besef, dat het samenstellen van een index op Thomas met de traditionele middelen uitgesloten moest worden geacht.

Het verzamelen van de terminologie der „interiorità”, d.w.z. *in* met zijn samenstellingen, synoniemen, antoniemen en dergelijke²⁾, kostte de schrijver meer dan vier jaar tijd. Bij het verzamelen van de bewijspplaatsen werd hij er zich van bewust, dat men de garantie moet hebben, dat de bouwstoffen integraal verzameld en geanalyseerd worden. Het gevolg hiervan was dat zijn materiaal weliswaar compleet was, doch slechts betrekking had op een gedeelte van het werk van Thomas³⁾.

In 1949 reisde Professor Busa, als „tutor” van een jongeman, door Noord-Amerika. Op deze reis, die hem zowel in Canada als in de Verenigde Staten bracht, sprak hij aan verschillende universiteiten over zijn plan tot het samenstellen van een woordregister op Thomas. Via de „Library of Congress” kwam hij, door bemiddeling van het „Massachusetts Institute of Technology”⁴⁾, in contact met de Amerikaanse firma „International Business Machines” (IBM) in New York. In november 1949 begon Busa zijn proeven met het ponskaartensysteem van de IBM; men meende toen echter, dat zijn plan tot het samenstellen van een mechno-grafische index onmogelijk was. Het was vreemd, zei men, gezien zijn plannen, dat hij een Italiaan en geen Duitser was⁵⁾. Hij vernam echter ook, dat de IBM te Milaan een belangrijk filiaal had. Na zijn terugkeer uit Amerika, kreeg hij, in het voorjaar van 1951, de gelegenheid zijn in New York begonnen proeven bij I.B.M.-Italia te Milaan voort te zetten en te voltooien⁶⁾.

Hij deed daarbij de ervaring op, dat de problemen waarmee hij bezig was, grote verwantschap vertoonden met die waarvoor zich de techniek der moderne documentatie gesteld zag. Het maken van woordindices en concordanties en het opstellen van woordstatistieken immers is niet alleen belangrijk voor de lexicologie, doch is ook methodologisch van nut voor

1) Zie R. Busa, *Complete Index* 425 [1950]. Een dergelijk woordenboek zou niet alleen aan linguïstische, doch ook aan filosofisch-theologische eisen moeten voldoen. Aan een dergelijk woordenboek zullen, zelfs na het gereedkomen van een index en een concordantie op Thomas, nog vele speciale studies dienen vooraf te gaan.

2) Zie R. Busa, *L'interiorità secondo la scolastica* 674 [1949].

3) R. Busa, *La Terminologia Tomistica dell' Interiorità* 11 [1949].

4) Pater Busa heeft tijdens zijn reis in Amerika ook nog de mogelijkheid overwogen om voor zijn doel gebruik te maken van de „Rapid Selector”, die door Dr. Vannevar Bush van het M.I.T. was ontwikkeld, en die met een snelheid van 10.000 fotogrammen per minuut de gegevens van een microfilm kan overnemen; zie hierover *Varia Specimina* 22 [1951].

5) „When he heard what Busa wanted, IBM Founder Thomas J. Watson threw up his hands. „Even if you had time to waste for the rest of your life, you couldn't do a job like that,” he said. „You seem to be more go-ahead and American than we are!””, *Time, The Weekly Newsmagazine*, Atlantic Edition, December 31, 1956, 40.

6) R. Busa, *Varia Specimina* 22 [1951]; *Mechanisierung* 14 b [1952].

het oplossen der problemen van de „information retrieval” (terugvinden der informatie) in het algemeen, iets waarmee zowel handel, als industrie en defensie te maken hebben ¹⁾.

Busa kwam toen tot de conclusie, dat het maken van een index met ponskaarten slechts een twintigste ²⁾ tot veertigste deel in beslag neemt van de tijd die nodig is bij het traditionele handwerk of wanneer de methode der verveelvoudigingsapparaten wordt aangewend. Dank zij de mechanografie kon de analyse van een tekst voortaan gemakkelijk, snel, accuraat, op grote schaal en vooral economisch verantwoord gebeuren ³⁾.

Bij deze proeven, die op een van Dantes *Canto's* werden genomen ⁴⁾, maakte Busa gebruik van de volgende machines van de IBM:

ponsmachine	type 131
sorteermachine	type 080
collator	type 077
tabelleermachine	type 405 ⁵⁾ .

Toen een bevredigende methode gevonden en voldoende ontwikkeld was, werd deze, bij wijze van inleiding en demonstratie voor de buitenwereld, toegepast op de hymnen *Pange lingua*, *Sacris solemnibus*, *Verbum supernum* en *Lauda Sion* van de hand van Thomas in het werkje *Officium de festo Corporis Christi*.

Intussen had Busa, door bemiddeling van kardinaal Spellman, ook het vertrouwen weten te winnen van Thomas J. Watson Sr., directeur van de IBM te New York ⁶⁾. Tijdens een tweede oponthoud in de Verenigde Staten in 1950, experimenteerde hij verder met allerlei machines, o.m. met de Cardatype ⁷⁾.

In 1951 verscheen van de hand van pater Busa *Sancti Thomae Aquinatis Hymnorum Ritualium Varia Specimina Concordantiarum. Primo saggio di indici di parole automaticamente composti e stampati da macchine IBM a schede perforate*. In de tweetalige, Engels-Italiaanse inleiding, die 49 van de 180 bladzijden beslaat, geeft schrijver een uiteenzetting van de begrippen index, concordantie en lexicon, en een overzicht van wat op dit gebied is gepresteerd (blz. 8-20). In het samenstellen van een concordantie onderscheidt hij vijf etappes:

1. het overschrijven van de tekst waarbij elke „zin” op een aparte kaart wordt gebracht;

¹⁾ R. Busa, *Travaux* 67 [1961].

²⁾ R. Busa, *Mechanisierung* 17 b, noot 2 [1952].

³⁾ R. Busa, *Elektronentechnik* 21 [1957].

⁴⁾ R. Busa, *Varia Specimina* 24 [1951].

⁵⁾ R. Busa, *Mechanisierung* 14 b [1952].

⁶⁾ R. Busa, *Elektronentechnik* 20 a, noot **) [1957]; P. Tasman, *Literature and Document Research Automation* 62 [1958].

⁷⁾ R. Busa, *Mechanisierung* 17 a [1952].

2. het verveelvoudigen van de kaarten, t.w. zoveel kaarten als er woorden zijn;
3. het aanbrengen van de lemmata;
4. het alfabetiseren der lemmata en der woorden;
5. de typografische compositie van het te publiceren werk.

Voor onderdeel twee is reeds eerder, t.w. via bekende systemen van verveelvoudiging, gebruik gemaakt van een zekere mechanisering. Met de ponskaartenapparatuur echter kunnen de verschillende stadia van het werk achtereenvolgens tot stand worden gebracht.

Op blz. 24-34 volgt een beschrijving van de wijze waarop Busa de hymnen van Thomas machinaal heeft behandeld. Aangezien deze mechanische behandeling nu, tien jaar later, voor een groot gedeelte door moderner en praktischer procédés is vervangen, zullen we er niet nader op ingaan. De stadia die thans nog actueel zijn, komen vanzelf aan de orde, wanneer de thans gevolgde werkwijze verderop ter sprake komt. In 1950 moest nog worden afgezien van de interpunctie en het aanbrengen van accenten ¹⁾. Elke kaart bevatte bovendien zowel een woord als een versregel, wat bij de korte strofen der behandelde hymnen toevallig mogelijk was; men werkte dus met een gecombineerde woord-tekstkaart.

In de maanden mei en juni van 1952 experimenteerde Busa in het hoofdkwartier van de IBM World Trade Co verder met de volgende machines: Cardatype, „Tape controlled card punch”, tabelleermachine 407 en „Gang punch” ²⁾.

§ 2. ORGANISATIE

Na een tijd van voorbereiding van tien jaren, werd, op 17 december 1956, in Gallarate, als een onderdeel van de filosofische faculteit van het Aloisianum, het eerste centrum voor gemechaniseerde taalanalyse, „Centro per l'Automazione dell' Analisi Letteraria”, geopend ³⁾. Het Centrum werd ingezegend door aartsbisschop G. B. Montini van Milaan, in tegenwoordigheid van leden van het Italiaanse kabinet, en tal van vooraanstaande personen uit de R.-K. Kerk en de zakenwereld ⁴⁾. Twee jaar later werden op de Expo te Brussel de werkzaamheden van het CAAL gedemonstreerd, terwijl er op 3 september 1958, in België en Duitsland een televisieuitzending aan werd gewijd ⁵⁾.

In mei 1961 waren het personeel, het materiaal en de machines nog verspreid over drie lokalen: 1^o, het Aloisianum, waar het taalkundige

¹⁾ R. Busa, *Varia Specimina* 32 [1951].

²⁾ R. Busa, *Entwicklung* 202-203 [1953]. Het stansen (eng. *gang punching*) gebeurt thans met de „Endprinting Reproducer” 519.

³⁾ R. Busa, *Elektronentechnik* 20 a [1957].

⁴⁾ *Philosophie auf Lochkarten in Italien* 17 [1957].

⁵⁾ R. Busa, *CAAL* 3 [1961].

hand- en hersenwerk (prepareren der teksten, lemmatiseren der ponskaarten) plaatsvindt; 2°. een ruimte in een fabriek waar de machines zijn ondergebracht waarmee de tekstkaarten worden geponst en gecontroleerd; 3°. een kelderruimte in een andere fabriek waar de overige machines zijn en tevens de administratie gevestigd is. Er waren toen plannen om alles onder te brengen in nieuwe en ruime lokalen, waarbij tevens zou worden voorzien in de behoefte aan bergruimte voor de bakken met ponskaarten.

Het secretariaat, waargenomen door pater Busa, die wordt bijgestaan door een secretaris, bevindt zich te Milaan, Via Cerva 35; het zal ook in de toekomst in Milaan, een groot centrum van wetenschap, verkeer en publieke relaties, blijven.

De reorganisatie van het CAAL is echter in volle gang. De bedoeling is, er een zelfstandig lichaam ¹⁾ met rechtspersoonlijkheid (*ente morale*) van te maken, waarbij een commissie van bijstand zich zou belasten met de financiering, zodat Busa zich uitsluitend zou kunnen bezighouden met de wetenschappelijke leiding der organisatie.

Het CAAL wordt gedragen door een nationaal comité van bijstand, met kardinaal Montini als erevoorzitter (de taalstatisticus Professor Marcello Boldrini, directeur van het Nationaal Instituut voor de Statistiek, is een der vice-presidenten), en een uitvoerend comité van vooraanstaande personen uit Lombardije. IBM-Italia verleent geheel gratis technische bijstand, levert kaarten en stelt, volgens een soort jaarlijks puntensysteem, machines ter beschikking tot een huurwaarde van 1.500.000 lire per maand ²⁾. Verder is men afhankelijk van de financiële steun van Italiaanse industriëlen, en van kleine toelagen van het Ministerie van Onderwijs en van het Ministerie van Arbeid ³⁾. Het ontbreken van een vaste financiële grondslag vormt momenteel nog een ernstige handicap voor het werk.

Op het eerste internationale colloquium voor lexicologische automatisering, gehouden te Tübingen in 1960 (24-26 november) en georganiseerd door Busa, werd voorgesteld dat het CAAL zou fungeren als internationaal centrum voor informatie, coördinatie en standaardisering op dit terrein. Zonder een gezonde financiële basis zal het CAAL deze opgave en zijn verdere taken echter niet aankunnen ⁴⁾.

De algemene leiding berust bij pater Busa, die ook thans nog titularis is van de leerstoel voor wijsbegeerte aan het Aloisianum, hoewel hij de

¹⁾ Busa dacht dat er toch een zekere band met het Aloisianum (in de vorm van een seminarium van het Aloisianum) zou blijven.

²⁾ Zie R. Busa, *Centro* 99 [1960]; *CAAL* 3 en 10 [1961]. Andere kosteloze materiaalleveranties aan papier werden door Standard Register Co. (U.S.A.) en Turati-Lombardi in Milaan in het vooruitzicht gesteld.

³⁾ Voor financiële details, zie R. Busa, *CAAL* 8 en 10 [1961].

⁴⁾ R. Busa, *CAAL* 4 en 8 [1961].

colleges niet meer kan geven. Hij wordt bijgestaan door Dr. A. Zampolli ¹⁾, die zich, behalve met het aanwerven, de opleiding en de leiding van het wetenschappelijke personeel, bezighoudt met de „planning”, het coördineren van en het toezicht op de verschillende werken die moeten worden uitgevoerd, alsmede met het scheiden der homografen en het lemmatiseren. Het CAAL beschikt eveneens over een secretaris-boekhouder, belast met de leiding van het bureau te Gallarate en het toezicht op het materiaal.

Het personeel van het CAAL is te onderscheiden in wetenschappelijk en uitvoerend personeel.

Het wetenschappelijke personeel bestaat uit een aantal medewerkers met academische graad (Dr. in de wijsbegeerte of in de letteren). Behalve Dr. Zampolli waren er in mei 1961 drie wetenschappelijke medewerkers; deze sectie was echter nog in volle ontwikkeling. In zijn lezing te Besançon in juni 1961 noemde Busa een aantal van 6 medewerkers voor het filologische gedeelte ²⁾.

Het uitvoerend personeel bestaat thans uit 22 medewerkers, die allen onder de leiding staan van Mej. L. Canestraro, tot op heden de enige vaste uitvoerende kracht in dienst bij het CAAL. Zij houdt permanent toezicht op het ponsen, evenals op het verdere mechanische werk. Bij deze medewerkers zijn ponstypistes en machinebedienden, die allen slechts tijdelijk in dienst zijn van het CAAL. Zij zijn namelijk leerlingen van een onderdeel van deze instelling, t.w. het centrum voor beroepsonderwijs voor machinebedienden van de diensten voor automatisering der comptabiliteit in het bedrijfsleven („Centro di addestramento professionale per operatori di servizi automazione contabilità aziendali”). De leerlingen die tot deze school worden toegelaten ³⁾, moeten minstens lager onderwijs hebben genoten, kunnen typen, en geslaagd zijn voor een psychotechnisch examen bij de IBM in Milaan. Ze krijgen allen een studiebeurs, waarvoor de gelden gedeeltelijk door het Ministerie van Handel, gedeeltelijk door de industrie worden verstrekt. Na een aanvankelijke cursus van drie weken, waarbij ze de machines leren bedienen, wordt een begin gemaakt met het ponsen.

¹⁾ Zampolli promoveerde in 1960 aan de universiteit van Padua op een, met behulp van de machines van het CAAL bewerkte dissertatie *Studi di statistica linguistica eseguiti con impianti IBM*. In dit werk, dat onder invloed staat van de taalstatistische methoden van Marcello Boldrini (Rome) en van P. Guiraud (Groningen), gaf hij, op grond van een Italiaans drama van D. Fabri, *Veglia d'Armi* [1957], behalve een statistiek van de verdeling der fonemen in het Italiaanse fonologische stelsel, een methode om de Italiaanse spelling automatisch fonologisch te transcriberen (zie R. Busa, *Darstellung Exhibit W, X, Y, Z* [1960]) en om de fonetische zin automatisch in lettergrepen te verdelen. In 1960 werd hij assistent van Busa. In verband met de dissertatie van Zampolli, zie C. Tagliavini, *L'automazione nelle ricerche fonetiche* [1961].

²⁾ R. Busa, *Travaux* 66 [1962].

³⁾ „Jamais plus de 20 % des candidats se montrent capables des fonctions d'opérateur mécanographe”, R. Busa, *Travaux* 65 [1962].

De leerlingen blijven twee jaar op de opleidingsschool van het CAAL, gedurende welke tijd ze eveneens praktische cursussen bij de IBM volgen. De ervaring heeft geleerd, dat de leerlingen die de opleidingscursus van het CAAL hebben gevolgd, snel en gemakkelijk hun weg vinden in het geautomatiseerde bedrijfsleven.

In mei 1961 beschikte men over zeventien ponstypistes die, in twee ploegen, ieder vier uur per dag praktisch werk deden: ponsen, controle-ponsen en verifiëren. Daarnaast waren er nog drie machinebedienden werkzaam, die geselecteerd werden uit een aantal kandidaten die zich hadden aangemeld. Ze werkten eveneens vier uur per dag in ploegendienst, zodat ze ook nog gelegenheid hadden de lessen aan een Technische Handelsschool te volgen.

§ 3. TEGENWOORDIGE WERKWIJZE

Automation demands an evolution of the methods of language analysis . . . toward the integral examination of all the material, without allowing the subjectivity of the researcher to bring in his own judgements as to what *is* and *is not* important,

R. Busa, *Darstellung* 4 [1960].

Prepareren van de tekst: Dit geschiedt door het wetenschappelijk personeel onder de algemene leiding van pater Busa, bijgestaan door Dr. Zampolli.

In het algemeen kan men zeggen, dat men bij dit voorbereidend werk dient aan te geven hoe de tekst op de kaarten moet komen te staan. Dit houdt in dat behalve de wijze van citeren (de referenties) ook de begrenzing der paragrafen en eventueel ook der regels moet worden aangegeven. Mogen geen potlood- of inktstrepen in het boek worden aangebracht, dan kan men gebruik maken van bladen doorschijnend papier¹⁾.

Verder stelt het prepareren van de tekst voor elk werk uiteraard zijn eigen problemen. Bij de werken van Thomas b.v. worden bepaalde stukken met kleuren onderstreept; de bedoeling hierbij is, ze als aparte perikopen te onderscheiden van de eigenlijke filosofische leer van de leidsman der scholastiek. Zo worden b.v. onderstreept met blauw: formules die citaten van bepaalde auteurs inleiden (b.v. *Dicit enim Gregorius in homilia quadam* = De Potentia q. 6, a, 4; A. 1); met groen: letterlijke citaten van een bepaalde auteur; met rood: die gedeelten waar deze auteur wordt geparafraseerd of geresumeerd; met zwart: de formules waarmee Thomas naar zijn eigen werk verwijst. Al deze gedeelten, behalve de rood onderstreepte, zullen niet in de concordantie worden opgenomen. Deze naar de kwaliteit door kleur onderscheiden tekstgedeelten worden door speciale ponsingen, b.v. 0, 3, 6 of 9 in kolom 14, van elkander onderscheiden. Dit is nuttig

¹⁾ R. Busa, *Punched Cards* 360 [1958].

niet alleen voor statistische, doch ook voor lexicologische doeleinden. Door deze speciale ponsingen kan men automatisch bepaalde tekstgedeelten b.v. in aparte indices opnemen of ze uit een concordantie weglaten¹⁾. Kastanjebruine onderstreping (= ponsing 5 in kolom 14) is geen aanduiding van de kwaliteit van het op deze wijze onderstreepte tekstgedeelte. Deze kleur kan in combinatie met andere bijzondere tekens (b.v. blauw) voorkomen. Kastanjebruin worden onderstreept: introductoria (schema's of inhoud van hoofdstukken) en zuiver vaktechnisch-scholastische terminologie (syllogismen). De ponsing 9 in kolom 10 is een speciaal teken, waardoor wordt aangegeven dat de volgende woorden, tot aan het volgende speciale teken, geheel van Thomas zijn. Het teken / geeft het wisselen van het speciale teken aan. Vóór een interpunctie vormt het teken # een machinale instructie dat het volgende teken een interpunctie is, en geen grafeem of woord. Hetzelfde teken op het eind van de regel vormt eveneens een machinale instructie²⁾; # # betekent eind van de regel zonder -, # · # echter eind van de regel met interpunctie. Zo kan een formule die een citaat van een andere auteur inleidt (zoals *ut dicit Augustinus*), of een Grieks woord in een tekst van Goethe, worden voorafgegaan door het speciale teken /B en afgesloten door /9³⁾. Deze tekens zijn m.a.w. een aanduiding van de authenticiteit, kwaliteit, grafie enz. van de verschillende woorden en zinnen⁴⁾. Ze verschilden tot nog toe van werk tot werk.

Wat dit punt betreft, zou een zekere standaardisering niet onwelkom zijn. Dat bij de vertolking van de ponsingen die de authenticiteit, kwaliteit, grafie enz. aangeven, ook van interpunctietekens gebruik wordt gemaakt, is weliswaar geen bezwaar bij woordkaarten — immers de vertolking blijft beperkt tot kolom 14, zodat er van verlies van interpunctietekens geen sprake is — doch wel bij tekstkaarten, waar de vertolking nl. niet tot een bepaalde kolom beperkt is en er dus dubbelzinnigheden kunnen optreden. Anderzijds komt het voor dat interpunctietekens (b.v. ; ?) door speciale tekens (t.w. \$ en %) worden vertolkt! Maar dat zijn de nadelen van machines die de lexicoloog wel grote diensten kunnen bewijzen, doch nu eenmaal niet met het oog op zijn speciale behoeften, waaronder in de eerste plaats een groot aantal tekens, werden ontworpen. Daartegenover kan men natuurlijk stellen dat de vertolking voor de machinale verwerking van ondergeschikt belang is.

Ponsen van de tekst: De geprepareerde tekst wordt daarna, regel voor regel, elke regel voorzien van zijn referentie, telkens op een aparte kaart

1) In een concordantie op Thomas van Aquino b.v. heeft het opnemen van citaten uit Augustinus natuurlijk geen zin.

2) R. Busa, *Darstellung* 2 (infra) [1960].

3) Zie R. Busa, *Darstellung*, Exhibit C [1960].

4) R. Busa, *Darstellung* 4 [1960]. Zie in R. Busa, *Erläuterungen* 14 [1960] een uitvoerige codelijst van de speciale tekens die gebruikt zijn bij het ponsen van Goethes *Farbenlehre*; verg. verder ook a.w. blz. 17, 18 en 19.

(de tekstkaart) ¹⁾ geponst. Wanneer de gedrukte regels te lang zijn, worden ze door de typiste verdeeld in twee stukken, b.v. regel 1 a en 1 b; woorden mogen daarbij nooit over twee kaarten worden verdeeld. Ook het nummeren der regels gebeurt door de ponstypistes. Het ponsen geschiedt met de schrijvende ponsmachine IBM 026 — in 1961 waren er drie in gebruik —, die ponst en tegelijk de geponste tekst in de bovenrand van de kaart afdrukt. De ponstypiste corrigeert fouten die ze zelf maakt en onmiddellijk bemerkt; dit betreft 5 à 6 foute kaarten per 100.

De IBM 026 werkt met maximaal drie ponsingen, t.w. een ponsing voor de cijfers, twee voor de letters en drie voor de speciale tekens. De op deze manier geponste kaarten worden daarna geverifieerd met de controleponsmachine IBM 056, waarvan er in 1961 eveneens drie bij het CAAL in gebruik waren. De verificatie bestaat hierin dat men de oorspronkelijke tekst een tweede keer aanslaat op het toetsenbord van de 056 die eveneens de speciale tekens heeft; als de ponsing niet klopt, gaat er een rood lampje branden en stopt de machine. Op deze manier worden de meeste fouten — gemiddeld 6 % van de kaarten bevatten fouten — van de ponstypiste verbeterd, echter niet alle. Immers het kan gebeuren — en het gebeurt ook één op de duizend keer — dat de ponstypiste en de controleponstypiste dezelfde foute toets aanslaan ²⁾. Per uur worden ongeveer 120 kaarten geponst, geverifieerd en verbeterd ³⁾.

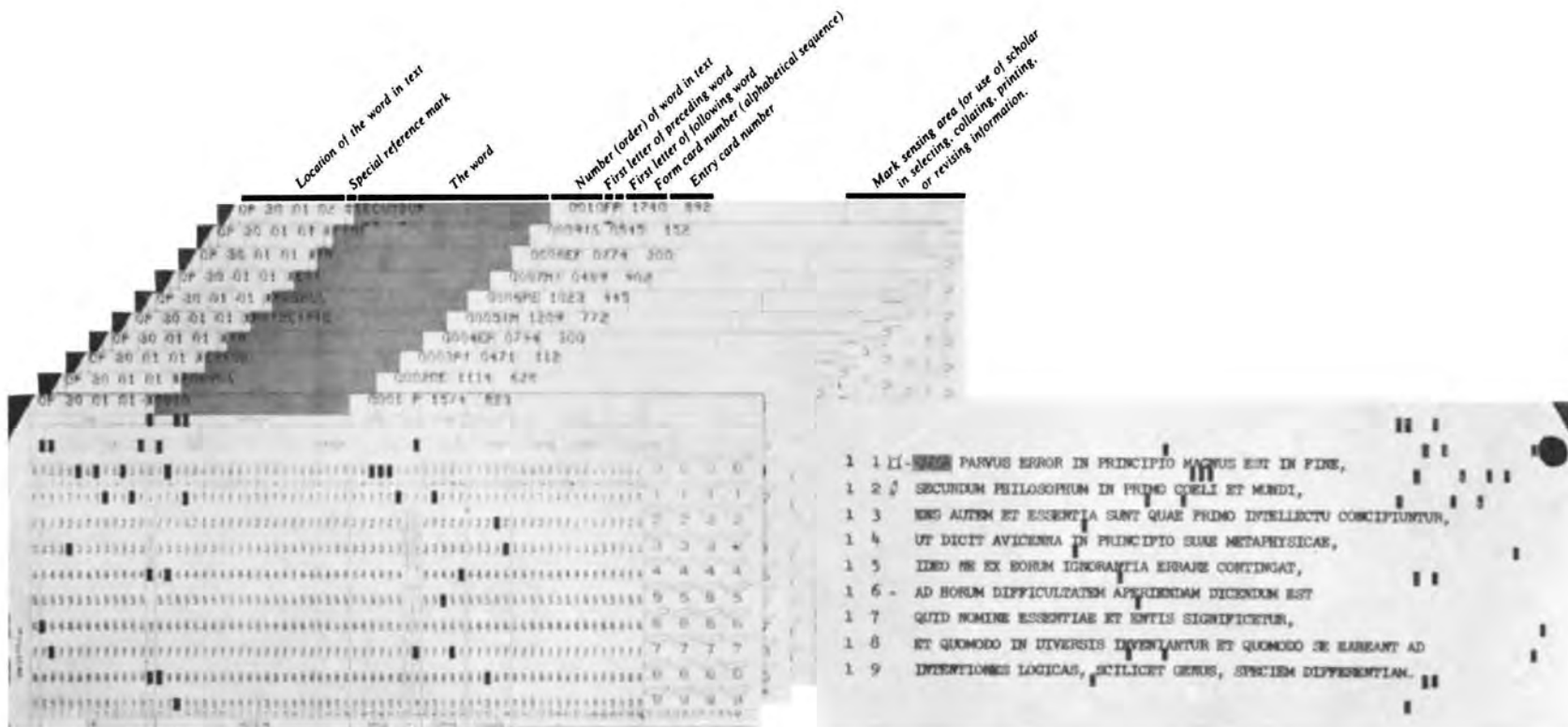
Honderd regels tekst leveren gemiddeld 120 tekstkaarten op, waarvan er 100 correct zijn; van de 20 andere zijn er 5 die fout zijn en bij het ponsen verbeterd worden, 5 die bij de verificatie verbeterd worden, en nog 10 andere waarbij als gevolg van het machinale proces ongewenste afwijkingen (b.v. scheuren in het papier, minder duidelijke afdruk enz.) optreden.

Controle en correctie: Zelfs na het controleponsen kunnen er fouten in de kaarten zitten. Het is absoluut nodig dat ook deze fouten worden gecorrigeerd, omdat ze anders eindeloos herhaald worden. Deze controle bestaat in het vergelijken der, met behulp van de tabelleermachine, van

¹⁾ Men gaat thans voor het werk aan de *Index Thomisticus* steeds uit van de tekstkaarten, op grond waarvan vervolgens woordkaarten worden vervaardigd. Men kan echter ook direct uitgaan van woordkaarten, die elk afzonderlijk van hun eigen referentie worden voorzien. Bij het vervaardigen van een moederblad op grond van de woordkaarten geeft de Cardatype-schrijfmachine 866 dan natuurlijk slechts één referentie per vers. Zie verder blz. 50.

²⁾ Verg. R. Busa, *Varia Specimina* 30-32 [1951]: „This control allows only the following case to pass unobserved, namely two typists make the same error in the same place. This case is very improbable and so much the less probable in as much as the qualities and circumstances of typing and typist are different”. In een ander verband spreekt Busa van „correcting the errors that the tyranny of the laws of chance makes inevitable” (R. Busa, *Darstellung* 4 [1960]).

³⁾ Dit optimale gemiddelde wordt door een goede leerlinge-ponstypiste reeds na drie maanden gehaald.



Afb. 3.

Woordkaarten van het CAAL. Op de achterkant van elk dezer woordkaarten staat de uit 9 regels, met behulp van 9 tekstkaarten (regelkaarten) opgebouwde context *quia parvus error in principio magnus est in fine*, enz. (Illustratie overgenomen uit P. Tasman, *Lit. Data Proc.* 251 [1957], met welwillende toestemming van de IBM).

de tekstkaarten gemaakte lijsten met de tekst van het boek. Het lijsten gebeurt hetzij met de IBM Cardatype 866 Typewriter, die drie ponsingen kan afdrucken, hetzij met de tabelleermachine IBM 441, die slechts op twee ponsingen berekend is, doch veel vlugger werkt dan de Cardatype. Behalve fouten die door de ponstypiste en de controleponstypiste over het hoofd werden gezien, kunnen er uiteraard ook zuiver machinale storingen in de lijst optreden. Deze hebben echter geen gevolgen voor de geponste kaarten zelf.

Verdere mechanografische behandeling

Met behulp van de Cardatype maken de machinebedienden behalve lijsten ook woordkaarten en moederbladen.

De Cardatype bestaat eigenlijk uit drie verschillende, onderling verbonden apparaten, t.w. 1°. Controle-eenheid 858
2°. Schrijfmachine 866
3°. Schrijvende ponsmachine 536.

Deze drie machines verrichten tegelijkertijd drie verschillende werkzaamheden. Het eerste apparaat leest de tekstkaarten, het tweede maakt van deze kaarten lijsten op gewoon, of moederbladen (eng. *master*) op speciaal papier, terwijl het derde, eveneens op grond van de in het eerste ingevoerde tekstkaarten, automatisch woordkaarten ponst en vertolkt ¹⁾.

Deze woordkaarten (één kaart voor elk woord in de tekst) bevatten ook (geponst en vertolkt) de referentie ²⁾, speciale tekens, de voorafgaande en volgende interpunctie, de eerste letter van het voorafgaande en van het volgende woord ³⁾ en een progressief nummer.

¹⁾ Zie R. Busa, *Entwicklung* 203 b [1953]. Oorspronkelijk werden de woordkaarten gemaakt met behulp van een reproducerende ponsmachine; dit was een zeer ingewikkeld procédé en wel op grond van het feit dat de verschillende woorden van de tekstkaart niet alle in dezelfde kolom beginnen. Zie voor een en ander: R. Busa, *Varia Specimina* 24 en 26 [1951].

In *Literature and Document Research Automation* 66 [1958] beschrijft P. Tasman het volgende procédé voor het tot stand komen van woordkaarten uit tekstkaarten: „The phrase cards will be processed through the Cardatype to produce a punched paper tape and a continuous form, Multilith master context plate. The perforated tape . . . will be used . . . to produce the single word cards reproduced from the text”. Van de ponsband werd echter te Gallarate in 1961 geen gebruik gemaakt.

²⁾ Verg. P. Tasman, *Literature and Document Research Automation* 66 volg. [1958]: „In this step, the single word cards will be processed through a reproducing punch which will put into each of the word cards, their respective reference number, special marks, serial number of the position of the word in the text, and the first letter of the preceding word, etc. A second processing of the single word cards in reverse order will provide the first letter of the following word, a reference number of the following sentence, and the reference number of the preceding sentence”.

³⁾ Dit is nuttig om de volgorde der woorden in de tekst te controleren, namelijk als extracontrole naast de progressieve nummering. Bovendien laat dit procédé ook toe, uitdrukkingen bestaande uit twee of drie opeenvolgende woorden (b.v.

Schema woordkaart *Contra Gentes*

ponsing in kolom	1	besturingsponsing van de Cardatype, ponsing 12 : eerste woord van de regel, ponsing 12-9: eerste woordkaart van de context, als er geen commando is voor de Cardatype is de kolom leeg,
" "	2	referentie werk: G,
" "	3	nummer van het boek,
" "	4-5-6	nummer van het hoofdstuk,
" "	7-8-9	progressief nummer van de regel in het hoofdstuk,
" "	10-11	progressief nummer van het woord in de regel,
" "	12	aanduiding dat we te maken hebben met een woord van Thomas zelf,
" "	13	aanduiding dat een fout in de tekst verbeterd is,
" "	14	gereserveerd voor speciale tekens,
" "	15-36	gereserveerd voor het woord dat steeds gevolgd wordt door de besturingsponsing X,
" "	37-42	progressief nummer van het woord in de tekst,
" "	43	eerste letter van het voorafgaande woord,
" "	44	eerste letter van het volgende woord,
" "	45-54	leeg,
" "	55-56	aanduiding van de interpunctie die volgt,
" "	57	leeg,
" "	58-59	aanduiding van de interpunctie die voorafgaat,
" "	60-63	nummer van de context,
" "	64-66	nummer van het woord in de context,
" "	67-80	leeg.

In grote lijnen is de indeling van de woordkaarten voor alle werken dezelfde. Immers men gebruikt altijd een kaart met één bepaalde rubricering. Sommige onderdelen hiervan zijn vrij stabiel, b.v. kolom 15-36 gereserveerd voor het woord, andere kan men uiteraard naar gelang van de behoefte wisselen ¹⁾).

secundum quod) te voorschijn te halen of te reconstrueren, terwijl het ook statistisch bijzonder nuttig is bij tekstlacunes. Het laat immers een waarschijnlijkheidsrekening toe betreffende de vraag welk woord (B) ontbreekt tussen de woorden A en C; in verband hiermee is ook de frequentie der trigrammen, en dat niet alleen binnen het woord, van belang. Voor een statistiek van trigrammen (tevens trifonemen), zie R. Busa, *Darstellung*, Exhibit X [1960]; voor bigrammen en trigrammen in het Nederlands, zie J. L. van Soest, *Informatie-theorie en Communicatie-theorie*, fig. 4 [1955].

¹⁾ Voor ontwerpen van kaarten, zie R. Busa, *Erläuterungen* 15 [1960]. Op de in afb. 4 (blz. 49) afgebeelde woordkaart uit de *Summa Theologiae*, een proefkaart, zijn de kolommen 43-44, bestemd voor de initia, niet ingevuld; op de definitieve kaart is dit echter wel het geval.

Opvallend is dat op deze woordkaart de rubricering van de eerste en die van de tweede regel (die eigenlijk duplieken zijn van elkaar) niet overeenstemmen. Op de bovenste regel worden de woorden vertolkt door de schrijvende ponsmachine IBM 026 (in stippelletters) of door de Cardatype schrijvende ponsmachine 536; op de onderste door de vertolker IBM 552, die een duidelijker, dus groter en ook meer dan één kolom vergende „letter” heeft, zodat hij over de gehele lengte van de kaart niet meer dan 60 tekens kan afdrukken. De woordkaarten van het CAAL zijn m.a.w. berekend op een dubbele vertolking.

Bij sommige werken is men uitgegaan van de woordkaarten, zoals bij *Contra Gentes*, bij andere van de tekstkaarten, waarvan vervolgens woordkaarten werden gemaakt, zoals bij de *Summa Theologiae*. De eerste methode bleek echter, gezien het groter aantal referenties dat daarbij nodig is, tijdrovender dan de tweede, zodat men thans bij voorkeur van de tekstkaarten uitgaat ¹⁾.

Terwijl de woordkaarten op de schrijvende ponsmachine 536 tot stand komen, vervaardigt de schrijfmachine 866 van de Cardatype tevens een moederblad op speciaal papier. Dit blad moet worden gecontroleerd op mechanische storingen (b.v. type-arm die niet of niet goed afdrukt), die 1 à 2 keer per 3.000 à 4.000 woordkaarten voorkomen.

Het met de schrijfmachine 866 vervaardigde moederblad dient om, met behulp van een „Litho Duplicating Machine” ²⁾ Davidson 816, de keerzijde der woordkaarten, t.w. tussen de ponsposities, te voorzien van twaalf of dertien regels context ³⁾. De tekst wordt op de keerzijde afgedrukt, omdat aan de voorkant van de kaart van Gallarate maximaal slechts 9 regels, t.w. tussen de ponsposities 0 tot 9 in, beschikbaar zijn. Deze porties context worden bij het prepareren van de tekst afgebakend door een speciaal teken x . . . x, een teken dat ook in de tekstkaarten wordt geponst, en dat men ook op de moederbladen terugvindt. Een dergelijke portie context is meestal voldoende om de betekenis van een woord te determineren zonder dat men de gedrukte tekst erbij hoeft te halen ⁴⁾.

Daarna worden de woordkaarten machinaal gecontroleerd met de tabelleermachine ⁵⁾; deze telt hoeveel er zijn en controleert daarna of ze alle wel in de vereiste volgorde liggen, na de Davidson te zijn gepasseerd.

¹⁾ Zie blz. 46, noot 1.

²⁾ De Davidson vergt een toegewijd en goed getraind personeel, dat op de hoogte is van het drukprocédé. De door de Cardatype 866 vervaardigde moederbladen zijn duur. Als een master beschadigd is, wat nogal eens voorkomt, moet niet alleen een nieuwe worden vervaardigd, doch moeten ook alle ermee bedrukte ponskaarten worden overgedaan.

³⁾ Een optimale context zou het woord waar het om gaat steeds in het midden van die context moeten hebben, wat nu nog niet het geval is.

⁴⁾ R. Busa, *Punched Cards* 364 [1958].

⁵⁾ Dit kan ook gebeuren met de IBM elektronische statistiekmaschine 101. Het percentage fouten dat op deze manier ontdekt wordt, bedraagt 2.000 op 300.000 kaarten of 2/3 %.

De machine controleert tevens of alle ponsingen wel in de goede kolommen zijn aangebracht en of er geen kolommen bij vergissing ongeponst zijn gebleven.

Is alles in orde, dan gaan de woordkaarten naar de eveneens elektro-mechanische ¹⁾ sorteermachine IBM 082, waarbij twee doorgangen van de kaarten per gesorteerde kolom noodzakelijk zijn ²⁾. De door deze machine tot stand gebrachte alfabetische volgorde wordt vervolgens gecontroleerd door de elektronische statistiekmachine IBM 101, die de kaarten niet in de breedte (verticaal, kolom per kolom dus), doch in de lengte (horizontaal) in één keer leest.

Met behulp van een reproducerende ponsmachine, de IBM 519 „Endprinting reproducer”, worden vervolgens de woordkaarten voorzien van een geponst (en wel in kol. 37-42=ordo) en in een groot cijferbeeld ³⁾ vertolkt (b.v. in kol. 6-8, regel 4-1), totaal progressief nummer van het

woord, links en dwars langs de smalle kant van de kaart, b.v.

21123
21

. In een andere doorgang vertolkt deze machine het nummer van de context (kol. 61-63) en het nummer van het woord in de context (kol. 64-66); dit gebeurt op de achterkant van de woordkaart, eveneens in een groot cijferbeeld links en dwars, langs de smalle kant der kaart, dus samen met de context van 12 à 13 regels ⁴⁾.

Nu worden op grond van de woordkaarten eveneens automatisch vormkaarten vervaardigd; elke vorm, die in een bepaalde tekst hetzij een keer, hetzij zeer vele malen voorkomt, wordt op een aparte kaart vastgelegd. Dit gebeurt op de volgende manier: de woordkaarten worden in alfabetische volgorde in de tabelleermachine IBM 441 gebracht, die „in tab” is ingesteld ⁵⁾. Dit betekent dat de tabelleermachine van deze woordkaarten een lijst van vormen maakt en de ermee verbonden „Endprinting reproducer” IBM 519 tegelijk deze vormen op ponskaarten brengt, met dien verstande dat een vorm die meermalen voorkomt, slechts één enkele

¹⁾ De elektronische sorteermachine IBM 083 gaat vlugger, doch maakt gemakkelijker fouten, waarbij de woordkaarten in verkeerde vakken terecht komen.

²⁾ Deze operatie kan echter bekort worden, b.v. door eerst de kortste woorden van een en twee letters, daarna die van drie letters, enz. eruit te halen; zie R. Busa, *Punched Cards* 361 [1958]. Bij dit geselecteerd sorteren behoeft slechts ongeveer een derde van het totale aantal kaarten tweemaal door de machine te gaan per te sorteren alfabetische kaartkolom; de overige kaarten gaan dan per kaartkolom maar eenmaal door de machine.

³⁾ Dit maakt snel naslaan van de kaart mogelijk.

⁴⁾ Deze gegevens werden overgenomen van een woordkaart van Thomas, *Contra Gentes*; de gegevens op de getekende woordkaart van Goethes *Farbenlehre* in de *Erläuterungen* blz. 15 zijn niet helemaal correct.

In dit stadium zou ook een alfabetische lijst van alle woorden die in een bepaalde wekt voorkomen, voorzien van progressief nummer en referentie, gemaakt kunnen worden.

⁵⁾ Wanneer de tabelleermachine „in tab” is ingesteld, bewerkt ze 9.000 kaarten per uur of 150 per minuut, als ze lijst, slechts 6.000 of 100 per minuut.

keer wordt geregistreerd. Deze tweede serie kaarten (vormkaarten) wordt door de reproducer tevens voorzien van een cijfer dat de frequentie aangeeft ¹⁾.

De frequentie komt ook op de door de tabelleermachine vervaardigde tabel voor; deze heet *Laterculum* ²⁾ *primum verborum* resp. *formarum*.

De vormkaart, die op het eerste gezicht tamelijk overbodig lijkt, is toch bijzonder nuttig. Ze spaart tijd en is een garantie tegen fouten. Als er eenmaal vormkaarten zijn gemaakt en het progressieve nummer van de vormkaart en van de lemmakaart tevens op elke woordkaart ³⁾ is aangebracht, kan het mechanisch sorteren der kaarten veel sneller gebeuren. Immers in plaats van b.v. een tienvoudig herhaalde selectie over tien letters, komt er een veel eenvoudiger sorteren over een getal (waarbij een doorgang per kolom volstaat), dat uiteraard minder dan tien cijfers zal tellen.

Tot zover kan men, alleen beschikkend over kennis van en ervaring met de machine, het werk doen in wat voor taal ook ⁴⁾. Dan begint echter een taak die niet meer geheel automatisch kan worden genoemd. Voortaan is ook taalkundig inzicht vereist in de taal waarin het werk werd geschreven. De machine immers kent de verhouding tussen vorm en „betekenis” der woorden niet. Met behulp van de context die op de achterkant der woordkaarten voorkomt, worden de homografen van elkaar gescheiden ⁵⁾ en op aparte vormkaarten overgebracht (it. *smismamento delle forme omografe*), en worden eventueel vreemde (b.v. Griekse) woorden verwijderd.

1) Deze reproducer kan ook zelfstandig functioneren; dit is het geval, wanneer hij de woord-, vorm- en lemmakaarten nummert.

2) Deze terminologie is overgenomen van Otto Gradenwitz, *Laterculi vocum latinarum* [1904]; de schrijver heeft deze lijst opgezet als hulpmiddel voor de aanvulling van defect overgeleverde teksten. Het woord *laterculum* (sinds Cato) betekent 'lijst, register'. De benaming *Laterculum verborum*, oorspronkelijk door Busa gebruikt in zijn *Varia Specimina* 42 en 65 [1951], werd naderhand terecht vervangen door die van *Laterculum formarum* (zie reeds R. Busa, *Punched Cards* 366 [1958]). In *Darstellung* 1 (infra) definieert Busa deze lijst als volgt: „We call laterculum formarum the listing of the forms in form alphabetical sequence: laterculum primum: alphabetised by the machine (dus zonder daarbij de homografen te onderscheiden); laterculum alterum: after having been lemmatised”.

3) Zie b.v. R. Busa, *Darstellung* 3 (infra) en Exhibit F [1960].

4) Bij het CAAL heeft men de ervaring, dat bij een tekst in een vreemde taal de ponstypistes minder fouten maken, omdat ze elke letter apart moeten bekijken.

5) Zie voor een en ander R. Busa, *Punched Cards* 363 [1958].

Het onderscheiden der homografen is theoretisch en praktisch een moeilijke gelegenheid, omdat hierbij verschillende standpunten (synchronie ↔ diachronie) en verschillende graden van homografie mogelijk zijn. Voor dit probleem, waar we hier niet nader op ingaan, verwijzen we naar R. Busa, *L'Analyse* [1962]. Dit artikel bevat de uiteenzetting die pater Busa hierover heeft gehouden op het lexicologische colloquium te Besançon in 1961; zie in verband hiermee ook F. de Tollenaere, *Het „Colloque international sur la mécanisation des recherches lexicologiques”* [1961]. Zie verder ook blz. 69 en 83.

Vervolgens worden met behulp van de „Endprinting reproducer” 519 (nu niet verbonden met de tabelleermachine) alle vormkaarten genummerd, waarna van deze kaarten, met behulp van de tabelleermachine 441, een alfabetisch geordende lijst wordt gemaakt waarin de homografen gescheiden zijn. Deze lijst noemt men *Laterculum alterum* of *secundum formarum*; de woorden die erin voorkomen, hebben niet alleen een progressief nummer, doch ook een frequentiecijfer ¹⁾.

Nu worden de vormen gelemmatiseerd, d.w.z. de verschillende buigingsvormen der zelfstandige naamwoorden, werkwoorden enz. worden onder een gemeenzaam „kopje” gegroepeerd, b.v. de vormen *vader*, *vaders*, *vaderen* onder het lemma *vader*, of *ben*, *is*, *zijt*, *zijn* onder *zijn*. Dit gebeurt op de volgende manier: in de lijst der vormkaarten (*laterculum alterum* of *secundum formarum*) worden alle vormen aangestreept die als lemma kunnen fungeren (dus b.v. niet *ouder* en *oudst*, doch wel *oud*); bij vormen waarbij het lemma niet voorkomt, wordt de zgn. *titulus* (= lemma) in de lijst bijgeschreven ²⁾. Bij de lemmatisering worden ook geaffigeerde woorden gescheiden (b.v. lat. *vidisti[ne]*, *num[que]*), terwijl anderzijds verschillende woorden tot één lemma kunnen worden samengevat (b.v. *ad instar*, *secundum quod*, *Julius Caesar*) ³⁾. Vervolgens worden van alle lemma's ponskaarten gemaakt die, na verificatie, met de sorteermachine 082 worden gealfabetiseerd en vervolgens gelijst met de tabelleermachine 441, teneinde ze ter controle te kunnen vergelijken met de aangestreepte en met de hand bijgeschreven lemmata in de lijst der vormkaarten. De machinale lijst der lemmata heet *Conspectus lemmatum* ⁴⁾; hierin zijn de lemma's progressief genummerd. Met de „Endprinting reproducer” 519 worden vervolgens de lemmakaarten voorzien van een geponst en vertolkt (it. *intestato*) progressief nummer in groot cijferbeeld, eveneens links en dwars langs de smalle voorkant van de kaart, alsook van hun vorm- en woordfrequentie.

Thans wordt op de vormkaarten, met behulp van de „Endprinting reproducer” 519, fungerend als „multiperforatrice” (Eng. „gang punch”), het nummer van hun lemma gestanst; zo zouden b.v. *geloof*, *gelooft*, *geloofde*

¹⁾ Naderhand wordt deze *Laterculum alterum* ook voorzien van het progressieve nummer van de lemmakaarten; men zou hem dan natuurlijk ook *Laterculum tertium* kunnen noemen.

²⁾ Elke vormkaart die een mogelijke homograaf bevat, wordt bovendien van een speciaal kenteken voorzien, te weten H 70, d.w.z. een H geponst in kolom 70.

³⁾ Bij het lemmatiseren van Thomas wordt het Latijnse woordenboek van Forcellini als leidraad gebruikt (zie R. Busa, *Varia Specimina* 44 [1951]), of, zoals Busa het uitdrukte: „de verantwoordelijkheid wordt op Forcellini overgedragen”, een zuiver praktische beslissing. Met het oog hierop werden alle lemma's van dit woordenboek op kaarten gebracht en de verschillende alfabetten tot één enkel verenigd, waarna alle kaarten links onderaan, met behulp van de „Endprinting reproducer” 519, van een groot progressief nummer werden voorzien.

⁴⁾ Zie R. Busa, *Varia Specimina* 44 en 105 [1951].

het nummer 190 van het lemma *geloven* kunnen krijgen. Hetzelfde gebeurt op dezelfde manier met de woordkaarten.

Vervolgens worden de lemmakaarten met de hand in het kaartenstelsel ingevoegd, namelijk elk lemma vóór zijn eigen vormkaarten. Als de lemmakaarten, de vormkaarten en de woordkaarten op hun definitieve plaats staan, wordt er een lijst van gedrukt, het zgn. *Rationarium verborum*¹⁾, eigenlijk een verbinding van de *Laterculum secundum* en de *Conspectus lemmatum*. In deze lijst worden de progressief genummerde lemmata gevolgd door de eveneens progressief genummerde vormen, welke laatste elk apart en ook gezamenlijk van hun frequentiecijfer zijn voorzien.

Ten slotte wordt met de tabelleermachine 441 een *Index verborum*²⁾ vervaardigd, eigenlijk een combinatie van de *Laterculum secundum*, de *Conspectus lemmatum* en de referenties. Deze index bevat de progressief genummerde lemmata, de progressief genummerde vormen met hun frequentie en de referenties der woorden met hun frequentie³⁾. De op deze manier tot stand gekomen *Index* wordt verder niet meer gecontroleerd.

Men zou dit stroomschema als volgt kunnen weergeven (zie blz. 55, afb. 6).

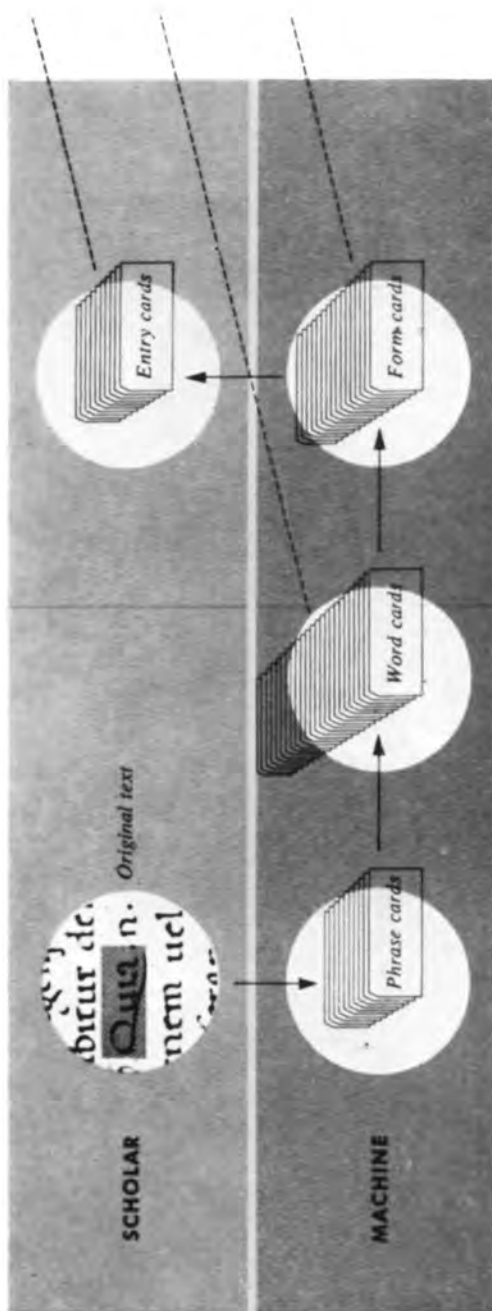
Verdere Indices: De aard en het aantal van de indices die men wenst te maken, kunnen natuurlijk worden gevarieerd. Een en ander hangt af van de taak en van het doel dat men daarbij voor ogen heeft. Het werk van Busa heeft reeds vanaf het begin sterk onder de invloed gestaan van statistische methodes. Deze statistische inslag is nog sterker geworden sinds Dr. A. Zampolli, als assistent van pater Busa, aan het CAAL verbonden is. Sindsdien is er een duidelijke neiging om het aantal indices nog uit te breiden. Zo heeft men bij sommige werken, b.v. bij Goethes

¹⁾ Deze term (zie R. Busa, *Varia Specimina* 44 en 81 [1951]) definieert Busa in *Punched Cards* 366 [1958] als volgt: „The *rationarium verborum* . . . would be the diagram, systematized and with frequencies, of all the same words regrouped according to their meaning, or more exactly, according to the identity of their functional elements. Such a list is the basis of the author's vocabulary” en in *Darstellung* 1 (infra) als „the listing of all the forms grouped under their lemma, with codes, frequencies etc. but not references”. Het woord *rationarium*, dat in het klassiek Latijn reeds gebruikt werd ter aanduiding van een statistisch overzicht, heeft Busa gebezigd, om ermee aan te geven, dat in dit register rekening is gehouden met de gedachteninhoud (*ratio*) der opgenomen woorden.

²⁾ Zie R. Busa, *Varia Specimina* 44 en 117 [1951].

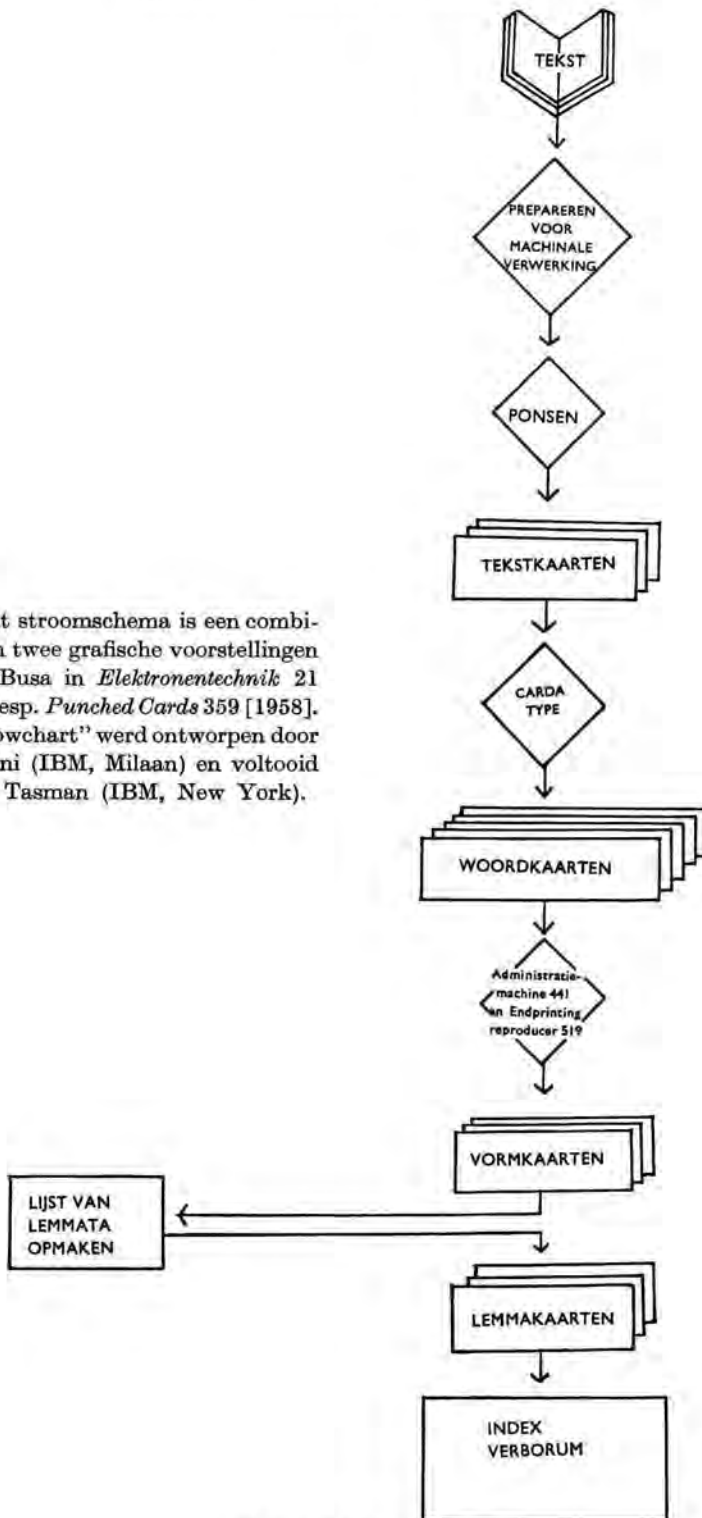
³⁾ In de *Varia Specimina* bestond de *Index verborum* slechts uit de progressief genummerde lemmakaarten, gevolgd door de woordkaarten met hun referentie.

Men zou zich kunnen afvragen, waarom naast de *Index verborum* (of *locorum*), die alle woorden met hun referentie bevat, ook nog een *Rationarium verborum* nodig is, waarin deze woorden tot vormen met hun partiële en totale frequentie worden gereduceerd. Alleen bij een uitvoerig werk is het nut van het *Rationarium* evident; immers het is uiteraard minder uitvoerig dan de *Index verborum*.



Afb. 5.

Stroomschema: De hierbovenstaande uit P. Tasman, *Lit. Data Proc.* 251 [1957], met welwillende toestemming van de IBM overgenomen illustratie geeft een voorstelling van de belangrijkste verwerkingen, waardoor men bij het CAAL komt tot een met behulp van machines vervaardigde *Index Verborum*.



1) Dit stroomschema is een combinatie van twee grafische voorstellingen van R. Busa in *Elektronentechnik* 21 [1957], resp. *Punched Cards* 359 [1958]. Deze „flowchart” werd ontworpen door C. Folpini (IBM, Milaan) en voltooid door P. Tasman (IBM, New York).

Afb. 6. Stroomschema 1)

Farbenlehre, ook lijsten gemaakt van de lemma's naar de toenemende of afnemende frequentie van hun vormen en woorden ¹⁾).

Men stelt zich voor, met het oog op de taalkundige statistiek, de volgende indices te maken:

index a contrario ²⁾)

index naar de woordlengte ³⁾)

index naar de woordfrequentie

index naar de omgeving (interpunctie en andere woorden)

index der suffixen

index der prefixen ⁴⁾).

De *index a contrario* of *contrarius* wordt hier nog apart besproken. Voor het maken van zulk een index — ook wel *Laterculum verborum, contrario litterarum ordine ab ultima ad primam dispositum* ⁵⁾), in het Duits *Konträr-index*, in het Engels *reverse index*, in het Frans *index à rebours* en in onze taal *retrograde index* genoemd —, wordt gebruik gemaakt van een apart stel kaarten ⁶⁾). Daartoe ordent de sorteermachine 082 de vormkaarten volgens de lengte der woorden, waarna de „Endprinting reproducer” 519 het hele stel kaarten reproduceert, en wel zó dat de laatste letter van elk woord in dezelfde kolom komt te staan. Nu worden de op deze wijze geponste kaarten, volgens de normale volgorde van het abc, van achteren naar voren gealfabetiseerd. Hierbij gaat men de kaarten van links naar rechts sorteren, d.w.z. eerst op de beginletter van het langste woord, in tegenstelling met de werkwijze bij gewoon alfabetisch sorteren met de machine, waarbij men van rechts naar links werkt ⁷⁾). Liggen de kaarten

¹⁾ Dr. Zampolli is bij dit initiatief geïnspireerd door het werk van de Franse taalstatisticus P. Guiraud.

²⁾ Zie een bladzijde uit een dergelijke index van Thomas, *De Ente et Essentia* in R. Busa, *Darstellung* 4 (infra) en Exhibit L [1960].

³⁾ Zie voorbeelden van een dergelijke index in *a.w.* 4 (infra) en Exhibit M en N.

⁴⁾ Zie R. Busa, *Travaux* 65 [1962].

⁵⁾ Zie R. Busa, *Varia Specimina* 42 [1951].

⁶⁾ Zie ook J. Štindlová, *Classement inverse* 84 [1961].

⁷⁾ Dat het gewone machinale alfabetiseren verloopt in een richting die precies de omgekeerde is van die welke bij alfabetiseren met de hand wordt gevolgd, laat zich gemakkelijk begrijpen wanneer men in gedachten houdt, dat men bij alfabetisch sorteren dezelfde procedure toepast als bij numeriek sorteren. Bij dit laatste immers wordt allereerst gesorteerd op de minst belangrijke gegevens, d.w.z. op de eenheden, vervolgens op de tientallen, daarna op de honderdtallen enz. Wanneer de sorteermachine numerieke gegevens, b.v. 431, 223, 406, 531, 026 en 606 moet ordenen, dan ontstaan daarbij eerst drie groepen: 406, 026, 606; 223; 431, 531; vervolgens weer drie groepen 431, 531; 026, 223; 406, 606 en ten slotte vijf groepen 606; 531; 431, 406; 223; 026. Zie een voorbeeld van sorteren met ponskaarten in A. D. Booth, L. Brandwood en J. P. Cleave, *Mechanical Resolution of Linguistic Problems* 36 [1958].

eenmaal in omgekeerde alfabetische volgorde — waarbij natuurlijk de normale volgorde der letters binnen de woorden zelf gehandhaafd blijft ¹⁾ —, dan maakt de tabelleermachine 441 er een lijst van. Hoewel men ook met de oorspronkelijk voorhanden vormkaarten een *index contrarius* zou kunnen maken, is het toch economischer hiervoor een apart stel kaarten te gebruiken. Deze laatste immers kunnen, behalve voor de retrograde index, nog voor tal van andere doeleinden worden gebruikt, zoals b.v. voor de studie van en de statistiek over de suffixen (alle of alleen bij bepaalde woordsoorten), een index volgens de lengte der woorden enz.

Concordantie: Al deze diverse indices hebben niet alleen een waarde op zichzelf, doch zijn ook gedacht als voorstadia voor de concordantie. Wanneer de teksten bij het prepareren worden voorzien van speciale kentekens, dan gebeurt dat in hoofdzaak met het oog op de te vervaardigen concordantie. De concordantie verschilt van de *Index verborum* in zoverre als onder de diverse lemmata, i.p.v. de woorden met hun referentie, de woorden in hun context met referentie voorkomen. Deze context kan, al naar gelang de behoefte, bestaan uit een, twee of drie regels; in dit laatste geval zal het woord in kwestie steeds in de middelste regel staan. Met de machinale methode kan een concordantie worden vervaardigd in een dertigste of veertigste deel van de tijd die nodig is om het werk geheel met de hand te verrichten ²⁾.

Kaarten: Om tot machinaal samengestelde lijsten (indices en concordanties) te komen, zijn vier soorten van kaarten nodig ³⁾:

1°. *tekstkaarten*: waarop de tekst per regel in zijn normale volgorde is geponst en vertolkt, samen met zijn referentie en speciale tekens;

2°. *woordkaarten*: waarop de tekst per woord in zijn normale volgorde is geponst en vertolkt, samen met zijn referentie, speciale tekens, progressief nummer en de eerste letter van het voorafgaande en van het volgende woord (de woordkaarten liggen oorspronkelijk in de volgorde van de tekst, naderhand echter alfabetisch, om ten slotte onder hun lemma gegroepeerd te worden);

3°. *vormkaarten*: één kaart voor elk woord dat op enigerlei wijze grafisch verschilt van het voorafgaande, samen met zijn progressief nummer en de frequentie van de vorm, waaraan later het progressieve nummer van het lemma wordt toegevoegd;

4°. *lemmakaarten*: een kaart voor elke grafisch-semanticke woord-

Ten einde het alfabetisch sorteren te versnellen, houdt men voorlopig de kortste woorden (b.v. alle woorden met minder dan zeven letters) buiten het sorteerproces, zodat in eerste instantie slechts de woorden van zeven letters en meer door de machine gaan.

1) Zie J. Štindlová, *Classement inverse* 79 en vooral 84 en 85 [1961].

2) R. Busa, *Punched Cards* 366 [1958].

3) Zie R. Busa, *Darstellung* 1 (infra) [1960].

eenheid, beantwoordend aan de „kopjes” in het woordenboek; deze kaart bevat het progressieve nummer van het lemma, de nummers der vormen die bij het lemma behoren en de frequentie der verschillende vormen ¹⁾.

Terwijl de kaarten 1° en 2° elk voor zich gelijkstaan met de gehele tekst, vormen 3° en 4° een samenvatting van de tekst op het niveau van de vorm en van het lemma.

Van al deze kaarten had in mei 1961 alleen de woordkaart een speciale opmaak. Voor de andere kaarten werden gewone IBM-standaardkaarten gebruikt in verschillende kleuren (b.v. vormkaart: rood; lemmakaart: blauw), welke laatste echter nog niet gestandaardiseerd waren. In een verder stadium zijn waarschijnlijk wel speciale vorm- en lemmakaarten te verwachten ²⁾. In de *Erläuterungen* vindt men althans kaartontwerpen voor vorm- en lemmakaarten ³⁾.

Mark-sensing. Deze methode, in het Nederlands ook wel aanstreepmethode genoemd, bestaat hierin dat men op de ponskaart, in speciaal voor dit doel gereserveerde vakken, met een speciaal potlood streepjes aanbrengt. Op administratief gebied wordt hiervan gebruik gemaakt door personeel in de buitendienst; de streepjes worden daarna op het kantoor door de machine automatisch in ponsgaten omgezet, door middel van elektrische contactmaking. De positie van het streepje of van een ander merk (een kruisje b.v.) bepaalt de vooraf vastgelegde waarde van de later door de machine automatisch aan te brengen ponsing ⁴⁾.

Van deze methode heeft Busa reeds in 1951 gebruik gemaakt voor de aanduiding der grammaticale categorieën bij de lemmata van Bonitz, *Index Aristotelicus* ⁵⁾. Op de IBM-ponskaart ⁶⁾ van Gallarate is rechts een rubriek (vanaf kolom 70) gereserveerd voor dit doel; deze rubriek heeft vier rijen ovale figuurtjes, van negen ponsposities elk. Van deze symbolen kan de onderzoeker gebruik maken om naderhand, t.w. voor reeds op

¹⁾ In R. Busa, *Darstellung 1* (infra) sub 4. "Lemma cards" leest men dat de lemmakaart eveneens bevat "the code representing its grammatical categories". Nu is dit wel eens gebeurd bij een enkel kleiner werk, doch over het algemeen is men nog niet zover, al is er voor dit doel op de lemmakaart een rubriek gereserveerd. Dr. A. Zampolli houdt zich voor het werk aan Thomas bezig met de codificatie der grammaticale categorieën, wat uiteraard verband houdt met de categorieën der homografen.

²⁾ Verg. „The main words must be punched one per card with a special layout designed to accomplish the functions of these cards”, R. Busa, *Punched Cards* 363 [1958].

³⁾ R. Busa, *Erläuterungen* 15 [1960].

⁴⁾ *Vaktermen* 1 [1959].

⁵⁾ R. Busa, *Mechanisierung* 19b [1952]. Blijkbaar is hier alleen maar sprake van een experiment (zie ook R. Busa, *Entwicklung* 203b [1953]); immers in 1960 schrijft Busa dat de lemmata van de *Index Aristotelicus* „at the stage of punching” zijn (*Darstellung 1* [1960]).

⁶⁾ In mei 1961 was alleen de woordkaart — de enige kaart immers met een speciale opmaak — voorzien van een aanstreeprubriek. Ook bij de vormkaart en de lemmakaart zou men t.z.t. een dergelijke rubriek kunnen aanbrengen.

woordkaarten geponste en vertolkte woorden, aan te geven: de woordsoort, het accent, het aantal syllaben enz. ¹⁾).

Met succes is de aanstreepmethode door Dr. Zampolli toegepast bij de verwerking van de woordenschat van een drama van de Italiaanse auteur D. Fabbri. De mark-sensing rubriek was voldoende voor het aangeven van grammaticale gegevens zoals woordsoort, wijze, tijd, persoon, geslacht en getal ²⁾).

Machines: De volgende machines waren in mei 1961 bij het CAAL in gebruik:

IBM Schrijvende ponsmachine 026 (Eng. <i>Printing Card Punch</i> ; It. <i>perforatrice</i>)	3 stuks
IBM Controleponsmachine 056 (Eng. <i>Card Verifier</i> ; It. <i>verificatrice</i>)	3 „
IBM Cardatype Administratiemachine (Eng. <i>Accounting machine</i>)	2 „
bestaande uit: 1 ^o . IBM Controle-eenheid 858 (Eng. <i>Control Unit</i>)	
2 ^o . IBM Schrijfmachine 866 (Eng. <i>Typewriter</i>)	
3 ^o . IBM Schrijvende ponsmachine 536 (Eng. <i>Printing Card Punch</i>)	
IBM Tabeleermachine 441 (Eng. <i>Accounting machine</i> ; It. <i>tabulatrice</i>)	1 „
IBM Endprinting Reproducer 519 (It. <i>riproduttrice, multiperforatrice riepilogativa</i>)	1 „
IBM Elektromechanische sorteermachine 082 (Eng. <i>sorter</i> ; It. <i>selezionatrice</i>)	1 „
IBM Elektronische statistiekmachine 101 (Eng. <i>Electronic Statistical Machine</i> ; It. <i>selezionatrice contabile</i>)	1 „
IBM Alfabetische vertolker 552 (Eng. <i>Card Interpreter</i> ; It. <i>interpretatrice</i>).	1 „

Het volledige ³⁾ machinepark bestaat dus geheel uit IBM-apparaten ⁴⁾, met uitzondering evenwel van de hierboven op blz. 50 genoemde Davidson „Litho Duplicating Machine“; deze laatste gebruikt moederbladen van de Standard Register Co. ⁵⁾.

¹⁾ R. Busa, *Punched Cards* 364 [1958].

²⁾ Zie zijn „Tabella della codificazione delle categorie grammaticali“ in R. Busa, *Darstellung* 9 (infra) [1960].

³⁾ Een IBM collator (Eng. *collator*; It. *inseritrice*) was in mei 1961 nog niet aanwezig, doch werd verwacht.

⁴⁾ Men zie de afbeeldingen achter in het boek.

⁵⁾ R. Busa, *Darstellung* 3 (infra) [1960].

§ 4. WERKEN

Ziet men af van enkele min of meer toevallige, kleinere opdrachten die het CAAL heeft uitgevoerd, en die betrekking hadden op Latijnse, Italiaanse en Duitse teksten¹⁾, en verder van werkobjecten die als noodzakelijke voorstadia, of, als men wil, bijprodukten van de eigenlijke taak van het CAAL kunnen worden beschouwd (zoals b.v. het indexeren van Forcellini's *Lexicon totius Latinitatis*), dan kan men zich, bij de beschouwing van het door het CAAL uitgevoerde werk, beperken tot twee voorname teksten: 1°. de *Opera omnia* van Thomas van Aquino, 2°. de Rollen van de Dode Zee.

Van het werk van Thomas zijn verscheidene boeken in diverse stadia van bewerking²⁾. Het verst is men gevorderd met *De Ente et Essentia* en de *Summa contra Gentiles*, waarvan het ponskaartenmateriaal reeds aan de statistische verwerking toe is. De bedoeling met de bewerking van de *Opera omnia* was oorspronkelijk, zoals reeds gezegd, het aanleggen van een kartotheek van alle woorden in het werk van Thomas.

Thans echter heeft Busa dit oorspronkelijke plan geheel opgegeven. In plaats daarvan wil hij een automatisch gedrukte woordindex, die tevens concordantie is. Hij streeft ernaar om met behulp van de elektronische rekenmachine, een volledige alfabetische index (concordantie) op alle woorden van Thomas te maken, doch slechts in een zeer beperkte „oplage”³⁾ (b.v. tien exemplaren, die, met behulp van carbonpapier, in twee keer door de rekenautomaat te maken zouden zijn). Voor een tweede, minder volledige index (concordantie), zou echter een keus uit de woorden van Thomas worden gemaakt. Bij deze keus zouden al direct machinaal die tekstgedeelten verwijderd kunnen worden die van speciale tekens zijn voorzien (zie hierboven blz. 44 e. v.). Is een stuk tekst kastanjebruin onderstreept, dan is dat voor de computer een bevel, dat dit tekstgedeelte, hoewel het geheel van Thomas is, niet in de concordantie mag worden opgenomen. Worden in het *Rationarium verborum* en in de *Index verborum* alle woorden opgenomen met hun frequenties resp. referenties, in de concordantie zullen slechts de woorden van Thomas zelf worden gehonoreerd, voor zover ze tenminste niet kastanjebruin zijn onderstreept en niet behoren tot een aantal van tevoren vastgestelde, in het geheugen van de computer opgeborgen, als onbelangrijk voor de concordantie beschouwde woorden als *et, non, praeterea* en dergelijke.

Wanneer deze concordanties op Thomas het licht zullen zien, hangt geheel af van de materiële middelen die in de toekomst ter beschikking van het CAAL zullen worden gesteld. Een datum heeft pater Busa in dit verband, bij mijn weten, nog nooit genoemd.

¹⁾ Zie het overzicht bij R. Busa, *Darstellung* 1 en 2 [1960] en G. Hübner, *Goethes Farbenlehre* [1962].

²⁾ Zie R. Busa, *Darstellung* 1 en 2 [1960].

Veel minder omvangrijk, doch niet minder spectaculair, en, wat de elektronische verwerking betreft, baanbrekend, is een tweede werk, waardoor het CAAL zich bekendheid heeft verworven, t.w. het indexeren van de Rollen der Dode Zee, of de teksten van Qumran, die men wel eens de grootste ontdekking van alle tijden op het gebied der handschriften heeft genoemd. Hierbij had pater Busa de medewerking van P. Pietro Boccaccio s.j., van het Pauselijk Bijbels Instituut te Rome, en Professor Millar Burrows van de universiteit van Yale, voor het Hebreeuwse filologische werk; de IBM stelde de nodige materiële middelen ter beschikking en zorgde voor de technische medewerking van de heren Clemente Folpini te Milaan en Paul Tasman te New York¹⁾. Busa publiceerde over dit werk een artikel „The Index of all non biblical Dead Sea Scrolls published up to December 1957”²⁾. Aan dit artikel, evenals aan dat van P. Tasman, getiteld „Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods”, en aan dat van B. W. Secrest, *The IBM Electronic Statistical Machine*, ontleen we een en ander.

Het werk was het eerste project van het CAAL, waarbij zowel de techniek van de traditionele ponskaarten als die der elektronische informatieverwerking gebruikt werd. Men had het indexeerwerk geheel kunnen uitvoeren volgens de methodes die tot dan toe gebruikt waren bij het werk van de *Summa Theologica* van Thomas. Bij de Rollen echter werd sterk de behoefte gevoeld aan een nieuwere methode die minder ruimte zou vergen, een grotere capaciteit zou hebben, gemakkelijker te hanteren zou zijn en ook sneller wat betreft het lezen, verwerken en drukken van de informatie. Het lag voor de hand bij de EDP gebruik te maken van de magnetische band, die een maximale hoeveelheid gegevens kan bevatten, en waaruit deze laatste ook bijzonder snel kunnen worden gelezen, opgenomen en gedrukt³⁾. Pater Busa verwachtte trouwens reeds in 1956 een aanzienlijke tijdsbesparing van de elektronische rekenmachine met magnetische band⁴⁾, zodat het, reeds van zuiver methodologisch standpunt uit, gewenst was de rekenautomaat in te schakelen⁵⁾. De keus viel toen op de elektronische rekenmachine IBM 705, ten einde de verwerking der informatie over te nemen van de conventionele apparatuur.

Eerst vond de filologische voorbereiding plaats⁶⁾. Opzettelijk in de tekst opengelaten ruimtes werden, evenals lacunes, als „woorden” be-

1) R. Busa, *Index* 187 [1958].

2) *Revue de Qumran* 1, 187-198 [1958]. Een korte samenvatting over het indexeren van de Rollen der Dode Zee vindt men ook in R. Busa, *Punched Cards* 373 [1958]. Zie ook: *IBM and the Dead Sea Scrolls* [1958]; *IBM Pioneers Literary Data Processing Program* [1958].

3) P. Tasman, *Lit. Data Proc.* 256a [1957].

4) R. Busa, *I principali problemi dell' automazione dell' linguaggio scritto* 838 [1958].

5) Zie R. Busa, *Index* 187 [1958].

6) Van tevoren werden ook nog de gedrukte teksten met de fotokopieën der handschriften gecollationeerd; zie R. Busa, *Index* 187 [1958].

schouwd; woorden die, geheel of gedeeltelijk, niet konden worden ontcijferd, werden van de andere gescheiden met de bedoeling ze in speciale lijsten onder te brengen. Speciale tekens werden aangebracht in de volgende gevallen: *a.* voor woorden die voorkomen in citaten uit de Bijbel, *b.* voor woorden die geheel of gedeeltelijk geschrapt zijn door de kopiïst, doch leesbaar zijn gebleven, *c.* voor woorden die gedeeltelijk gereconstrueerd zijn, *d.* voor woorden die in het handschrift in een afwijkend alfabet zijn geschreven, *e.* voor woorden met bijzondere paleografische eigenaardigheden, *f.* bij woordvarianten, *g.* ter aanduiding van de taal waartoe het woord behoort.

Daarna werd de tekst geponst met een ponsmachine voorzien van een toetsenbord met de nodige speciale tekens. Hierbij werden echter niet, zoals meestal, tekstkaarten vervaardigd, doch alleen maar woordkaarten ¹⁾ met de nodige referenties, t.w. plaats (deel, bladzijde, kolom, regel, plaats in de regel), eerste letter van het voorafgaande en van het volgende woord, en volgnummer van het woord in de tekst ²⁾. Deze ponskaarten werden vervolgens ook vertolkt.

Het ponsen, controleponsen en vertolken, waarvoor men slechts 27 verschillende Hebreeuwse karakters nodig had (de vocalen ontbreken immers), gebeurde te Gallarate. Hierbij werd een afwijkende, aan de volgorde van de Hebreeuwse teksten (van rechts naar links geschreven) aangepaste ponsmethode gevolgd. De verzameling ponskaarten werd toen per vliegtuig naar Amerika gestuurd ³⁾, waar verdere elektronische verwerking plaatsvond.

Men wilde daarbij zes woordenlijsten vervaardigen:

- 1°. een alfabetische lijst van alle woorden van de tekst samen met de erbij behorende referenties;
- 2°. een alfabetische lijst van alle grafisch verschillende woorden gevolgd door de frequentie en een doorlopend nummer;
- 3°. een lijst waarbij, na ingrijpen van de bewerker, de homografen gescheiden zijn en alle woorden onder hun lemma gegroepeerd;
- 4°. een lijst waarin de gegevens van 2°. en 3°. zijn gecombineerd, en die dus de frequentie van alle vormen en tevens die van het aantal woorden onder elk lemma vertoont;
- 5°. een retrograde index;
- 6°. een concordantie.

Om tot dit doel, de zes indices, te komen, moesten de gegevens van elke

¹⁾ Men had toen namelijk niet de beschikking over de passende apparatuur. Het is praktischer om uit te gaan van tekstkaarten; zie blz. 46 en 50.

²⁾ Later zullen, als gevolg van het proces der elektronische verwerking, bovendien nog de volgende gegevens in de kaarten worden geponst: een doorlopend nummer per vorm, een doorlopend nummer per lemma.

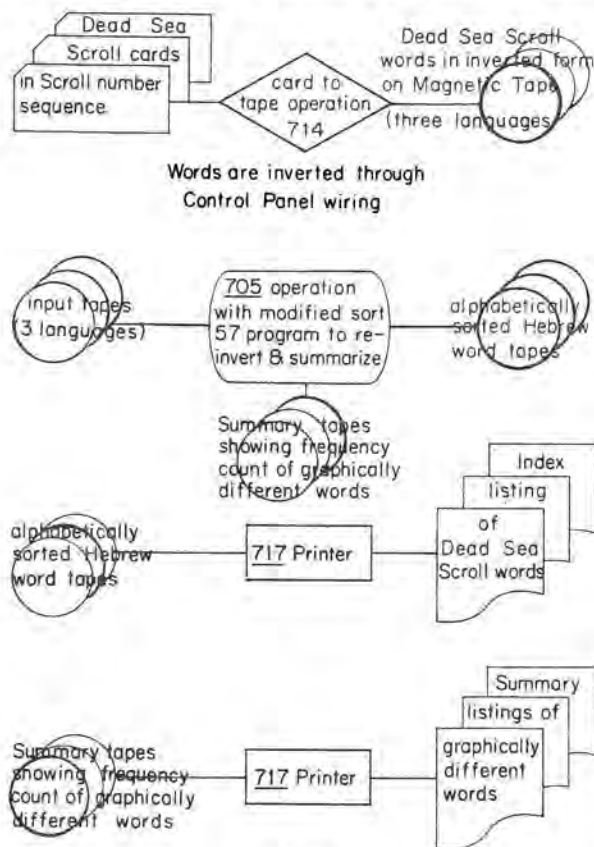
³⁾ *IBM and the Dead Sea Scrolls* 7 [1958].

woordkaart worden overgebracht op de magnetische band. De gegevens die in één kaart zijn opgenomen, beslaan een gedeelte van de magnetische band („unit of tape”) dat 80 tekens omvat. Bij de invoer werden de Hebreeuwse woorden van de rechts-linkse in de links-rechtse volgorde overgebracht.

De verwerking geschiedde door de IBM 705 in vier gangen.

Bij de eerste gang wordt de magnetische band waarop de gegevens der woordkaarten werden overgenomen, in de rekenmachine gebracht, waarbij, met behulp van een sorteerprogramma, alle woorden worden gealfabetiseerd, teruggebracht in hun oorspronkelijke vorm (rechts-links) en overgebracht op een magnetische uitvoerband, die op zijn beurt, met behulp van de drukinrichting, de hierboven onder 1^o. genoemde lijst produceert.

Bij een tweede gang wordt, met behulp van de alfabetische uitvoerband, een tweede band gemaakt, welke de gegevens bevat die zullen eiden tot het verkrijgen van de onder 2^o. genoemde index. Voor het



Afb. 7.

Stroomschema van de elektronische verwerking der Rollen van de Dode Zee. Overgenomen uit R. S. Casey e.a., *Punched Cards* 371 [1958], met welwillende toestemming van Reinhold Publishing Corporation, Technical Book Department, New York.

verkrijgen van deze tweede band dient de machine telkens twee opeenvolgende woorden met elkaar te vergelijken, bij identiteit alleen maar het eerste bij het voorgaande op te tellen, bij verschil echter beide woorden over te nemen.

Bij de derde gang worden twee invoerbanden gebruikt. De eerste band is vervaardigd op grond van de gegevens over homografie en lemmatisering die de onderzoeker in lijst 2^o. heeft aangebracht, waarna de lemmata op kaarten worden geponst en tussen de woordkaarten ingevoegd. De tweede band bevat de oorspronkelijke gegevens van lijst 2^o. De inhoud van beide banden wordt met behulp van een vergelijkingsprogramma gecombineerd en levert lijst 4^o.

Bij de vierde en laatste gang wordt de concordantie tot stand gebracht. Deze wordt verkregen door het collationeren van de band die de gegevens van lijst 4^o. bevat, met een band waarop de doorlopende tekst is opgenomen. Deze collationering bestaat o.m. in het vergelijken van een portie context met een woord; woord en context worden gecombineerd als de referentienummers overeenstemmen. Zodoende verkrijgt men een band waarop elk woord voorkomt voorzien van een of meer regels tekst.

Van de Rollen der Dode Zee werden aanvankelijk alleen 86 niet-bijbelse documenten, t.w. die welke in december 1957 reeds waren gepubliceerd, machinaal bewerkt; alles bij elkaar genomen waren dat ongeveer 30.000 woorden ¹⁾.

Het ponsen en de ponscontrole van deze woorden nam drie à vier weken in beslag, het programmeren echter drie maanden ²⁾. Het overbrengen van de ponskaarten op de band duurde ongeveer drie uur; het vervaardigen van de banden die de gegevens voor de lijsten 1^o. en 2^o. bevatte, ongeveer 90 minuten; het drukken van de twee lijsten ongeveer negen uur.

De indices op de Rollen der Dode Zee hebben het mogelijk gemaakt, woorden die in de sterk beschadigde manuscripten gedeeltelijk, ja zelfs geheel weggevallen waren, te herstellen. De woordindices, de gegevens over frequentie, omgeving der woorden in de tekst en omgeving der letters in de woorden, vormden hierbij een ideaal hulpmiddel. Natuurlijk blijft reconstructie van weggevallen woorden alleen maar reconstructie, doch de machinale methoden vormen hierbij een veel secuurder en bruikbaarder hulpmiddel dan de traditionele. Bij proeven op teksten in de Rollen der Dode Zee heeft men tot vijf in de tekst op elkaar volgende, en opzettelijk weggelaten woorden feilloos kunnen „reconstrueren”, dank

¹⁾ P. Tasman, *Indexing* 11 [1958]. Hetzelfde aantal noemt ook R. Busa, *Punched Cards* 373 [1958]. In 1956 schatte men het totaal nog op maximaal 50.000 woorden; zie R. Busa, *Elektronentechnik* 22b [1957] alsmede P. Tasman, *Lit. Data Proc.* 256b [1957].

²⁾ P. Tasman, *Indexing* 11 [1958]. In 1956 rekende men slechts met drie weken, zie R. Busa, *Elektronentechnik* 23a [1957] en P. Tasman, *Lit. Data Proc.* 256b [1957] en *Literature and Document Research Automation* 71 [1958].

zij de lexicologische gegevens die door de elektronische rekenmachine waren verschaft. Terecht wijst Tasman aan het eind van zijn helder en overzichtelijk artikel erop, dat de moderne machine gecombineerd werd met oude boekrollen ten einde onze theologische erfenis beter te kunnen begrijpen.

De indices op de teksten van Qumran zijn tot op heden nog niet in druk verschenen ¹⁾; ze zullen ongeveer 2000 bladzijden in folio tellen ²⁾. Het werk zal trouwens niet alleen zuivere indices bevatten, doch ook statistieken over de frequentie der letters en lettercombinaties ³⁾.

Demonstraties van de gemechaniseerde bewerking van de woordenschat der Rollen van de Dode Zee vonden in Europa plaats in 1958 en 1959. De eerste werd gegeven in september 1958 op de Wereldtentoonstelling te Brussel; met behulp van de rekenmachine IBM 650 Magnetic Drum Data Processing Machine werden toen lijsten vervaardigd over de distributie der grafemen in de Rollen ⁴⁾. De tweede gebeurde naar aanleiding van de conferentie over gemechaniseerde documentatie: *Automatic Documentation in Action*, die van 9 tot 12 juni 1959 te Frankfurt am Main plaatsvond; met behulp van de rekenautomaat IBM 705 werd toen de indexering der boekrollen gedemonstreerd. Met indices die op twee magneetbanden waren opgenomen, vervaardigde de IBM 705 in enkele minuten alfabetische woordenlijsten van de woordenschat der rollen.

Hoe omvangrijk is nu het tot op heden bij het CAAL verrichte werk, als men dit afmeet aan de hoeveelheid geponste woorden of verwerkte ponskaarten? In 1960 bedroeg het aantal woorden dat op ponskaarten was gebracht meer dan 2.500.000 ⁵⁾, even later werd het cijfer van bijna 3.000.000 woorden genoemd ⁶⁾, in december 1960 was dit aantal gestegen tot 3.500.000 ⁷⁾, terwijl Busa eind 1961 spreekt van 4.000.000 woorden op kaarten ⁸⁾.

¹⁾ In 1958 wachtten de indices nog slechts op de lemmatisering, die, naar Busa hoopte, in 1959 klaar zou zijn (zie R. Busa, *Index* 189 [1958]); in september 1960 echter was het lemmatiseren nog niet voltooid (zie R. Busa, *Darstellung* 1 [1960]).

Professor Busa schreef me (oktober 1961), dat het lemmatiseren nog steeds niet klaar was; intussen werden de teksten geponst die in het laatste jaar verschenen, zodat de publikatie t.z.t. geheel „up to date” zal zijn, en alle niet-bijbelse rollen zal omvatten. De bewerking der bijbelse fragmenten is eveneens in het vooruitzicht gesteld.

²⁾ R. Busa, *CAAL* 5 [1961]: „saranno l'unica pubblicazione scientifica e non puramente divulgativa dell'Italia sui Rotoli del Mar Morto”.

³⁾ Zie hierover B. W. Secrest, *The IBM Electronic Statistical Machine* 3 [1958]; R. Busa, *Darstellung* 5 (infra) [1960]; M. W. Gall, *Die RAMAC und die Philologie* 4c [1961].

⁴⁾ R. Busa, *Darstellung* 5 (infra) [1960].

⁵⁾ R. Busa, *Centro* 98 [1960].

⁶⁾ R. Busa, *Darstellung* 3 [1960].

⁷⁾ R. Busa, *CAAL* 4 [1961]; R. Busa, *Travaux* 64 [1962]. Hiervan waren vermoedelijk slechts 2.000.000 à 2.500.000 ponskaarten verwerkt.

⁸⁾ R. Busa, *L'analisi linguistica* 105b [1961].

§ 5. TOEKOMSTPLANNEN

In een mededeling over de voornaamste problemen der automatisering van de geschreven taal, voorgedragen op het in april 1956 te Milaan gehouden „Convegno Internazionale sui problemi dell' Automatismo”¹⁾, kondigde Busa aan, dat hij voornemens was, zich in de nabije toekomst bezig te houden met de elektronische rekenmachines met magnetische banden. Zijn bedoeling hierbij was, nauwkeurig te onderzoeken in welke mate ze een besparing vertegenwoordigden, gesteld dat ze de mechanische doorvoer van de ponskaarten door de machine overbodig maakten.

Enkele maanden later, op 12 oktober 1956, ter gelegenheid van de 8ste jaarvergadering van de „Deutsche Gesellschaft für Dokumentation”, gehouden te Bad Nauheim, sprak Busa over „Die Elektronentechnik in der Mechanisierung der sprachwissenschaftlichen Analyse”. Hij zette toen de redenen uiteen, die hem ertoe gebracht hadden het gebruik van elektronische rekenmachines voor zijn werk te overwegen. Het verdelen van de tekst in woorden op kaarten en het alfabetiseren van deze woordkaarten was z.i. te omslachtig en te langdurig²⁾. En aangezien taalonderzoek vaak continu, zonder onderbreking door regels of woorden moet kunnen gebeuren, dient men, zo schreef hij, de tekst te kunnen onderzoeken in een richting die horizontaal is ten opzichte van de verticale invoer der kaarten in de machine³⁾. Daarom was het nodig de tekst op een samenhangende band vast te houden.

Busa kondigde toen tevens aan, dat de woordenschat der „Rollen van de Dode Zee”, na het op de traditionele manier ponsen en corrigeren, met behulp van de elektronische rekenmachine IBM 705 gealfabetiseerd zou worden. Volgens zijn berekeningen zou de *Summa Theologiae* op dezelfde manier behandeld kunnen worden, al lag dat niet in de bedoeling. De reeds geponste en geverifieerde ponskaarten (die circa 16.000.000 tekens bevatten) zouden in 13 uur en 20 minuten op een magnetische band kunnen worden overgebracht; de 1.600.000 woorden van de ponsband zouden in 60 uur te alfabetiseren zijn, 1.600.000 woordkaarten zouden in 266 uur en 40 minuten geponst kunnen worden, terwijl het tabelleren van deze woorden 30 uren in beslag zou nemen.

1) R. Busa, *I principali problemi dell' automazione dell' linguaggio scritto* [1958].

2) R. Busa, *Elektronentechnik* 22a [1957]: „So müssen z.B. die 1.600.000 Karten, die die Wörter der *Summa Theologiae* des hl. Thomas von Aquin darstellen . . ., zur alphabetischen Aufzeichnung etwa 20 mal durch die Maschine gehen. Im ganzen gesehen müssen so 32 Millionen Karten durch die Sortiermaschine laufen, die mit einer Geschwindigkeit von bis zu 60.000 Karten je Stunde arbeitet”. Zie ook P. Tasman, *Lit. Data Proc.* 255b [1957].

3) „Large scale data processing systems . . . also scan texts serially — in the swift uni-directional motion characterized by flowing tape — instead of in parallel, as in the card punching method. Serial scanning, which is analogous to the way a person reads, is far superior for literary analysis”, P. Tasman, *Indexing* 3 [1958].

Wat de verschillende types van elektronische rekenmachines betreft: die welke voor handelsdoeleinden waren geconstrueerd (d.w.z. met grote in- en uitvoer en weinig berekeningen) verdienen voor taalonderzoek de voorkeur boven die welke speciaal op wetenschappelijke berekeningen waren ingesteld (d.w.z. met kleine in- en uitvoer en omvangrijke berekeningen).

Een magnetische band (ca. 750 m lang) kan evenveel gegevens bevatten als 24.000 ponskaarten, en drie magneetbanden evenveel als een normale ponskaartenkast; deze banden maken bovendien een snellere in- en uitvoer en een sneller sorteren mogelijk. Bij het alfabetiseren vergelijkt de machine elk woord (dat in de vorm van een getal is voorgesteld) met het voorafgaande woord (getal) en onderzoekt of het hoger of lager is. Hierna schrijft de machine op een tweede band de woorden over die in de goede volgorde liggen en op een derde band die welke dat niet doen. Deze operatie wordt zo dikwijls herhaald tot de juiste volgorde is bereikt¹⁾.

Een bezwaar voor het gebruik van de rekenautomaat voor taalkundige doeleinden waren echter de hoge kosten die eraan verbonden waren.

Busa kwam aan het slot van zijn belangrijke uiteenzetting tot de volgende conclusies:

- 1°. De rekenautomaat dient zijn plaats te krijgen in de verbetering en ontwikkeling der taalwetenschappelijke analyse.
- 2°. de elektronische rekenmachine vervangt de traditionele methode met ponskaarten niet, doch vult ze aan.
- 3°. de traditionele elektromechanische apparatuur blijft voor sommige doeleinden praktischer dan de rekenautomaat.
- 4°. zonder behulp van de elektronische rekenmachine is het gebruik op grote schaal van ponskaarten niet goed mogelijk.
- 5°. een gecentraliseerd gebruik van de rekenautomaat is geboden.
- 6°. de elektronische rekenmachine dient voor het taalonderzoek economisch toegankelijk gemaakt te worden.

Twee jaar later, in 1958, wijst Busa erop, dat in het proces der mechanische verwerking van lexicologische gegevens, twee bewerkingen versneld dienen te worden, t.w. het totstandbrengen van woordkaarten uit tekstkaarten en het alfabetiseren. Door deze twee onderdelen te versnellen zou men veel tijd kunnen sparen en de kosten belangrijk drukken²⁾.

In een kort, samenvattend artikel over het CAAL, verschenen in jaargang 1960 van het tijdschrift *Current research and development in scientific documentation*, blz. 98-100, schrijft Busa: „The practical feasibility of using standard punched-card equipment is being compared with the use of magnetic tape data processing computers”. Hierbij werd niet alleen

¹⁾ Bij dezelfde gelegenheid bracht Busa nog een ander plan ter sprake, namelijk het opstellen van een programma om met behulp van de rekenautomaat de varianten van een tekst te registreren. Zie hierover verder R. Busa, *Punched cards* 369 e.v. [1958].

²⁾ R. Busa, *Punched Cards* 370 [1958].

gelet op de factoren snelheid en kosten, doch o.m. ook op controle en correctie in beide methoden.

In zijn op 1 september 1960 gedateerde samenvatting van de ervaringen van het CAAL, zoals die aan de deelnemers aan het colloquium te Tübingen in november 1960 werd verstrekt, is het duidelijk dat Busa het stadium van het experiment achter zich heeft. Zeer gedecideerd schrijft hij: „experience has proven that the maturity and accessibility of the method require magnetic surface electronic calculators and not punched card machines as the main and fundamental instruments in these applications, even if the punched card machines will still have to condition and integrate the use of the calculators”¹⁾.

Tijdens mijn verblijf te Gallarate kreeg ik de indruk, dat Busa, tenminste voor zijn doel, de verdere verwerking der lexicologische informatie met behulp van de traditionele elektromechanische apparatuur als achterhaald beschouwt. Dit is te begrijpen, gezien het feit dat elk lexicografisch plan hem voorlopig vreemd is²⁾.

Een der redenen die Busa aanvoerde voor het overschakelen in de nabije toekomst van de „U(nit) R(ecord)” op de „E(lectronic) D(ata) P(rocessing)” was, dat het voorbereiden van een index of een concordantie met behulp van de conventionele apparatuur een uiterst gespecialiseerd personeel vereist. Het succes van het werk hangt te veel af van het personeel; dit mag vooral de gemaakte fouten niet verheimelijken. Van dit personeel, dat zich aan de snelheid van de machine moet aanpassen, wordt een voortdurend gespannen aandacht vereist tijdens langdurige operaties. Zo moeten 400.000 te alfabetiseren woordkaarten, gegeven een gemiddelde woordlengte van 10 letters, 15 tot 20 maal (al naar gelang van de machine) in de apparatuur worden ingevoerd en uitgevoerd. De organisatie van het werk is veel moeilijker dan men denkt. Bij de EDP daarentegen kan het organiseren van de werkzaamheden voor een groot deel worden opgenomen in het programma der rekenmachine, wat een eenvoudiger en secuurder verwerking mogelijk maakt. De conventionele apparatuur vergt m.a.w. te veel ingrijpen van de mens, die al te zeer onderworpen is aan de onverbiddelijke wet van het toeval.

Ook de snelheid, die bij de EDP veel hoger ligt dan bij de UR, is een factor die de doorslag kan geven. Men heeft berekend, dat de elektronische verwerking der informatie de tijd benodigd voor de verwerking met de traditionele apparatuur met ongeveer 25 procent bekort. In principe is het dank zij de EDP mogelijk, de tijd die nodig is voor de verwerking der gegevens door handenarbeid, te reduceren tot ongeveer het tweehonderdste deel ervan³⁾.

¹⁾ *Darstellung* 5 [1960].

²⁾ „Nous sommes plus des tisserands que des couturiers”, schrijft Busa (*L'Analyse* 134 [1962]).

³⁾ P. Tasman, *Indexing* 11 [1958] en *Literature and Document Research Automation* 70 [1958].

Een ander voordeel van de EDP is, dat de rekenautomaat waarschijnlijk de oplossing van het probleem van de wisselende context¹⁾ mogelijk maakt. Uitgaande van tekstkaarten als invoer, die in de volgorde van de tekst op een magnetische band worden opgenomen, kan men als uitvoer woordkaarten krijgen. Hierbij kan de computer waarschijnlijk zo worden geprogrammeerd, dat hij tijdens de eerste of eventueel de tweede doorgang van de magnetische band, de woordkaarten voorziet van een veranderlijke context, van zegge 500 tekens voor en na het woord, met dien verstande dat in geen geval halve woorden worden afgedrukt, en dat de machine halt houdt bij een punt of kommapunt na 400 tekens. In deze context wordt het woord in kwestie van een of ander speciaal kenmerk, b.v. een sterretje, voorzien.

Voor zijn eigen specifiek doel, een woordindex en een concordantie op Thomas, hecht Busa trouwens minder belang aan woordkaarten. In de toekomstige EDP wil hij helemaal geen gecontexteerde woordkaarten meer, doch alleen een, via het medium van de magnetische band door de tabelleermachine van de computer vervaardigde woordindex en concordantie op Thomas. De reeds voorhanden woordkaarten zullen zelfs niet meer in Gallarate bewaard blijven, doch opgeborgen worden in een grote bibliotheek te Rome²⁾, waar ze meer diensten kunnen bewijzen dan in de kleine provinciestad van Lombardije. Wie studie maakt van Thomas, zal, volgens Busa, genoeg hebben aan de indices en de concordantie, en verder aan een collectie „schedoni”, d.w.z. losse bladen waarop het hele werk van Thomas met behulp van de tabelleermachine wordt afgedrukt. Mochten later voor een woordenboek op Thomas gecontexteerde woordkaarten nodig blijken, dan kunnen deze, zo zegt Busa, altijd gemakkelijk worden vervaardigd op grond van de magnetische banden.

Busa gelooft zelfs in de mogelijkheid het scheiden der homografen in het EDP-proces te kunnen automatiseren³⁾. Men kan zich een dergelijke automatisering b.v. als volgt voorstellen: op een magnetische band is opgenomen een *rationarium* van een uitvoerig werk, waarbij de lemmati-

¹⁾ Zie hierover ook F. de Tollenaere, *Documentation lexicographique* [1962] en B. Quemada, *L'Inventaire lexicographique* 125 [1962]. Zie in dit werk ook blz. 103.

²⁾ R. Busa, *Centro* 98 [1960].

³⁾ „Such work requires the competent responsibility of the scholar and it cannot be accomplished by machine. However, once such classification has been made, mechanical recognition of different forms of the same word could follow”, R. Busa, *Punched Cards* 363 [1958]. „The grouping of the various inflected forms of the same word under one title-word or lemma and the consequent sorting of the homographs, etc. are operations which can be done by machine, the first time, only to a very limited degree”, R. Busa, *Darstellung* 1 [1960]. Zie ook R. Busa, *L'Analyse* 138 [1962]. Op dezelfde bladzijde wijst Busa ook op de mogelijkheid van automatische scheiding op grond van de omgeving van het woord (b.v. *secundum* is voorz. wanneer het gevolgd wordt door *quod* of *quem*).

Zie thans ook B. Quemada, *L'Inventaire lexicographique* 122 en 127 [1962]. In de praktijk zou dit machinaal-theoretisch vrij eenvoudige probleem wel eens minder eenvoudig kunnen blijken te zijn. Zie ook blz. 52, 83 en 88 in dit werk.

sering op de traditionele manier is gebeurd: elk woord dat homografie kan vertonen, is op de band van een speciaal kenteken voorzien. Met deze eerste magnetische band met aanduiding van de mogelijke gevallen van homografie laat men automatisch een tweede magnetische band vergelijken, waarop in alfabetische volgorde de woorden van een andere tekst voorkomen, die nog niet op homografie is onderzocht. Op deze manier kunnen automatisch uit deze tweede band de meeste gevallen van mogelijke homografie worden gesorteerd. Na een zeker aantal operaties zal men de homografie in een bepaalde taal, althans in teksten met een vast spellingsysteem, geheel hebben geautomatiseerd.

Een programmering van de computer op dit niveau vergt natuurlijk een verregaande systematisering van de homografie in elke taal. Heeft men echter te maken met een niet streng gestandaardiseerde spelling, dan is de mogelijkheid tot automatisering der homografie vrijwel nihil.

Op het tweede internationale colloquium voor mechanisering der lexicologie, gehouden te Besançon in juni 1961, kondigde Busa aan, dat bij het CAAL reeds een begin was gemaakt met de elektronische verwerking der lexicologische gegevens. Programma's voor de rekenautomaten IBM 704 en IBM 7090 waren in voorbereiding, en de eerste proeven werden in augustus 1961 verwacht. Deze programma's waren zo opgezet, dat ze zouden kunnen dienen voor elke tekst in elke taal en voor elk statistisch onderzoek, zelfs voor de automatische fonologische transcriptie! Wat het laatste punt betreft, zal menigeen ongetwijfeld skeptischer gestemd zijn dan pater Busa.

Wenst men echter de EDP daadwerkelijk voor lexicologisch onderzoek te gebruiken, dan zal dat, wat de economische zijde betreft, alleen maar kunnen gebeuren in een gespecialiseerd internationaal centrum. Dit zou instructies geven voor de normalisering bij het ponsen der teksten, en verder ook de kosteloze elektronische verwerking der geponste teksten verzorgen¹⁾. Aangezien het programmeren een langdurige, moeilijke en zeer kostbare aangelegenheid is, zou dit internationale centrum de beschikking moeten hebben over een verzameling van programma's voor de verschillende typen van lexicologisch werk in de verschillende talen. In deze omstandigheden heeft het geen zin om elk lexicologisch centrum volledig te willen uitbouwen. Het ponsen en de verdere werkzaamheden met de conventionele apparatuur (UR) dienen in de nationale centra te gebeuren, de EDP echter dient aan het internationale centrum voorbehouden te blijven.

Voor dit doel rekent Busa op de financiële steun van Euratom. Als internationaal, interacademisch lexicologisch middelpunt ziet hij het nieuwe computercentrum van Euratom te Ispra (Noord-Italië), dat in juli 1961 met een computer IBM 7090 zou worden uitgerust²⁾. Dit centrum zou

¹⁾ R. Busa, *CAAL* 7 [1961].

²⁾ In *CAAL* 4 [1961] spreekt Busa nog van „une consociazione collaborativa, le cui modalità sono in corso di definizione”.

tevens tot taak hebben, niet alleen de codes en methoden, doch ook de terminologie op het gebied van de verwerking der lexicologische informatie te standaardiseren ¹⁾).

Wat de verdere toekomst betreft, houdt Busa zelfs rekening met de mogelijkheid, dat het CAAL zou kunnen uitgroeien tot:

- 1°. een instituut voor de studie van het Latijn als geleerdentaal gedurende de Middeleeuwen en de Renaissance;
- 2°. een hogere school voor de vorming van Italiaanse en buitenlandse onderzoekers op het gebied der lexicologische automatisering ²⁾).

De plannen liggen echter nog te zeer in het speculatieve vlak, dan dat we er hier verder op in zouden behoeven te gaan.

§ 6. SLOTBESCHOUWINGEN

De in de voorafgaande bladzijden gegeven uiteenzetting over het CAAL kan in enkele punten concluderend worden samengevat.

1. Het uitgangspunt is geweest de filosofie en niet de taalkunde. Dit is ook thans nog duidelijk merkbaar: het leeuwedeel der belangstelling gaat naar de lexicologische behoeften der filosofie en der theologie. Zuiver taalkundige taken komen lang niet in die mate aan de orde.

2. Met dit uitgangspunt hangt samen, dat het door het CAAL tot stand gebrachte lexicologische materiaal berekend is op vrij uiteenlopende behoeften, in de eerste plaats echter op die van de filosofie en de taalstatistiek.

3. Het CAAL is niet in de eerste plaats gericht op een lexicologisch-lexicografisch doel. Voorop staat het tot stand brengen van een index en een concordantie op Thomas. Dat voor de toekomst (EDP) wordt afgezien van woordkaarten, bewijst wel hoe het maken van een woordenboek op Thomas slechts als een zeer verfliggende mogelijkheid wordt gezien.

4. Aan taalstatistiek wordt bij het CAAL zeer veel aandacht besteed en de beoefening ervan staat op een zeer behoorlijk peil. In de verschillende indices wordt evenveel aandacht besteed aan de woordvormen, die in de traditionele lexicografie niet voldoende tot hun recht kunnen komen ³⁾, als aan de lemma's waaronder de vormen der woorden gegroepeerd moeten worden.

5. Met het filosofisch-thomistische uitgangspunt hangt ook samen, dat het CAAL, gericht op een onderwerp dat geschreven was in de internationale taal der middeleeuwen, uiteraard meer een internationaal

¹⁾ R. Busa, *Travaux* 67 [1962].

²⁾ R. Busa, *CAAL* 7 [1961].

³⁾ Zie R. Busa, *L'Analyse* 134-135 [1962].

karakter draagt dan een centrum dat gericht is op een moderne nationale taal.

6. De uitrusting met IBM-machines, werkend met 47 geponste en vertolkte alfanumerieke tekens, is wel voldoende voor het bewerken van teksten, geschreven in het geleerdenlatijn der middeleeuwen. Zolang de IBM-machines echter niet over meer tekens beschikken, zijn ze minder geschikt voor de machinale verwerking van moderne Nederlandse of Franse teksten.

7. De economische grondslag van het CAAL voldeed in mei 1961 niet aan redelijke eisen. Dit manco kon echter tot op heden door de energieke, dynamische en invloedrijke persoonlijkheid van de leider der onderneming worden gecompenseerd. Deze compensatie onttrekt echter te veel tijd en werkkraft aan de zuiver wetenschappelijke taak.

8. De personeelsvoorziening was dan ook niet geheel bevredigend te noemen. Wel maakte de in vorming zijnde, kleine wetenschappelijke staf en de leiding der machinale verwerking een zeer goede indruk. Dat de bediening der machines echter wordt toevertrouwd aan leerlingen, die na een tweejarige opleiding het CAAL verlaten voor een werkring in het bedrijfsleven, is geen ideale toestand. Voor een dergelijk centrum is beroepspersoneel, geen gelegenheidspersoneel vereist.

9. Zeer ver is het CAAL gevorderd op de weg der elektronische verwerking (EDP) der langs elektromechanische weg tot stand gebrachte kartotheek. Hier heeft pater Busa, met de verwerking van de schriftrollen der Dode Zee, uiterst belangrijk pionierswerk verricht. Bovendien zijn thans, dank zij het CAAL, mogelijkheden geschapen tot samenwerking in Europees verband (EURATOM te Ispra).

10. Het pionierswerk van Professor Busa op het gebied der automatisering der lexicologische informatie dwingt onze eerbied en bewondering af. Toen hij eenmaal zekerheid had dat de taak, de *Index Thomisticus*, waaraan hij was begonnen, met de traditionele middelen onuitvoerbaar moest worden geacht, kwam hij, geleid door een onmiskenbaar wetenschappelijk instinct, terecht bij de informatieverwerkende machines. Niettegenstaande de aanvankelijk afwijzende houding der technici en fabrikanten, en hoewel hij toen nog generlei kennis van deze machines had, heeft hij deze toch weten in te schakelen. Zodoende heeft hij een weg gebaad voor de automatisering der lexicologische documentatie: zijn enorme verdienste. Zijn gebrek aan materiële middelen bij deze taak werd opgevangen door de manier waarop hij, dank zij zijn psychologisch inzicht, mensen en instellingen onweerstaanbaar weet te winnen en in te nemen voor zich en zijn werk, een werk dat thans gedragen wordt door het CAAL, „il primo nel mondo a introdurre l'automazione nelle ricerche delle scienze del linguaggio" ¹⁾).

¹⁾ CAAL 1 [1961].

HOOFDSTUK III

ELEKTROMECHANISCHE OF ELEKTRONISCHE APPARATUUR?

Les bonnes volontés, les sacrifices même, bref les méthodes traditionnelles ne suffisent plus si nous tenons à l'essentiel qui est d'aboutir, P. J. Wexler, in *Lexicologie et lexicographie* 65 [1960].

§ 1. DE ELEKTRONISCHE REKENMACHINE

De elektronische rekenmachine ¹⁾, dikwijls ook, met een Engelse naam, *computer* genoemd, verwerkt met behulp van elektronische middelen (dioden, transistors enz.) reeksen van gegevens; ze voert rekenkundige of logische bewerkingen uit met de haar toegevoegde informatie. Bij de elektronische rekenmachine onderscheidt men de analogierekenmachine en de digitale rekenmachine. De eerste rekent met gebruikmaking van variabele fysische grootheden als lengte, gewicht, voltage, snelheid enz. ²⁾. Voor toepassingen op het gebied der administratie is vooral de tweede soort van rekenmachine van belang. Dit is ook het geval in de automatisering op taalkundig gebied, aangezien deze zich in oorsprong, en ook nu nog, zeer nauw aansluit bij de automatisering in de sector administratie. Waar hier dus sprake is van de elektronische rekenmachine, is dan ook uitsluitend de digitale of cijferrekenmachine bedoeld. Deze berust op het principe van het al of niet aanwezig zijn van kleine elektrische impulsen. Door een reeks van impulsen of pulsen (een zgn. pulstrein) wordt informatie overgebracht. De logische elementen van de computer kennen slechts twee verschillende stabiele toestanden die door de machine als 1 of 0 kunnen worden geïnterpreteerd. Die logische elementen kunnen door het al of niet aanwezig zijn van een impuls in een van deze twee toestanden gebracht worden. Als gevolg hiervan kan alle informatie gecodeerd worden in een code die voortreffelijk bij het elektronische schakelsysteem past, t.w. de binaire ³⁾. Hierbij wordt zowel de numerieke als de andere

¹⁾ Beschouwt men de elektronische rekenmachine als informatiemachine, dan kan men beter spreken van informatieverwerker (eng. *information processor*, hd. *Informationswandler*).

²⁾ *Vaktermen* 84-85 [1959].

³⁾ In het tweetallig stelsel, dat bij elektronische rekenmachines veelvuldig wordt toegepast, omdat het technisch beter handelbaar is dan het tientallige stelsel, zijn slechts twee alternatieven mogelijk, d.w.z. „ja” of „neen”, d.w.z. in cijferwaarden alleen maar een (1) of nul (0). Het tientallige getal 49 ontbonden in machten van 2 is: $2^5 + 2^4 + 2^0$ ($32 + 16 + 1$) en wordt in het tweetallig stelsel geschreven 110001. De plaatsen van de cijfers in een binair getal vertegenwoordigen dus machten van

informatie door een aantal binaire cijfers of zgn. bits (eng. *binary digits*) voorgesteld. Een letter of een ander niet-numeriek teken, zoals b.v. een interpunctieteken, neemt gewoonlijk 7 bits in beslag. De rekenautomaat of snelteller¹⁾ (eng. *electronic computer*) heeft, via de journalistiek, als zgn. „elektronisch brein” algemene bekendheid verworven.

Een van de belangrijkste organen van de rekenautomaat is het geheugen, d.i. een bewaarplaats voor informatie. Men zou het kunnen vergelijken met vele bladen papier waarop gegevens kunnen worden geschreven, die kunnen worden afgelezen en waar, desgewenst, weer overheen kan worden geschreven.

De invoerorganen coderen de gegevens, b.v. van ponskaarten of ponsband, in een vorm waarmee de rekenautomaat bewerkingen kan uitvoeren. De uitvoerorganen lezen de berekende resultaten af, die in het geheugen zijn opgeborgen, en geven ze door aan een gekoppelde machine. Door het koppelen van verschillende apparaten aan de centrale eenheid wordt het prestatievermogen van de rekenautomaat aanzienlijk opgevoerd. Zo kunnen, schrijft Quemada²⁾, 10 ponsbandleesapparaten 2000 tekens per seconde registreren, terwijl 13 magnetische bandleesapparaten 12.000.000 tekens per minuut lezen, d.w.z. de gehele bijbel in acht minuten.

De toepassing van elektronische rekenmachines in de lexicologie heeft belangrijke voordelen. Hun snelheid is veel groter dan die der conventionele apparatuur; bij deze laatste immers heeft men te maken met mechanische schakelingen. Voor het automatisch alfabetiseren is de traditionele elektromechanische of zelfs elektronische sorteermachine te langzaam. Om 20.000 woordkaarten, met een gemiddelde van 8 tekens per woord, te alfabetiseren, zijn 320.000 (d.i. $20.000 \times 8 \times 2$) doorgangen vereist, wat neerkomt op circa 10 uur³⁾. Zijn de woorden echter opgenomen op een magnetische band, dan gaat het alfabetiseren veel en veel vlugger; het aflezen kan in twee richtingen plaatsvinden, en wel met een snelheid van 20.000 cijfers per seconde. Een ander voordeel is de besparing van ruimte. Een magnetische band kan 7.500.000 cijfers of 5.000.000 letters bevatten. Een archief van 100.000.000 woordkaarten vereist 1540 kasten, terwijl dezelfde hoeveelheid informatie opgenomen op magnetische banden (elke band is 750 m lang)⁴⁾, gewikkeld op spoelen (diameter van de spoel

het getal 2. Het gebruik van slechts twee symbolen (0 en 1) is zeer eenvoudig. Bij het overbrengen van de ponsband op de magnetische band is een zekere conversie nodig.

¹⁾ „De aanduiding „snelteller”, die prof. Bezemer onlangs heeft gebezigd mag . . . als een der meest geslaagde benamingen van de elektronische rekenmachine gelden”, M. Euwe, *Non-numerieke problemen op de computer* 16a [1962].

²⁾ *Mécanisation* 36 [1959]. Tegenwoordig echter is een snelheid van 1000 tekens per seconde voor een ponsbandlezer heel gewoon.

³⁾ B. Quemada, *Mécanisation* 34 [1959].

⁴⁾ Verg. R. Busa, *Elektronentechnik* 24 [1957]: „Eine Magnetbandrolle (ca. 750 m lang, Durchmesser 26,6 cm) entspricht 24.000 in allen Spalten voll ausgelochten Lochkarten”.

26,6 cm), slechts vijf kasten in beslag neemt. Een derde voordeel is de grotere toegankelijkheid van het materiaal; het met het oog op een bepaald gegeven sorteren van 15.000.000 ponskaarten met een uiterst snelle elektronische sorteermachine (60.000 kaarten per uur) zou toch nog 250 uur vergen. Is dezelfde hoeveelheid informatie daarentegen op magnetische banden opgenomen, dan worden de 15.000.000 kaarten of ongeveer 1.200.000.000 tekens, gelezen met de snelheid van 900.000 tekens per minuut, d.w.z. in ongeveer 23 uur.

§ 2. HUIDIGE STAND DER ELEKTRONISCHE LEXICOLOGIE

Voor lexicologische doeleinden is reeds herhaaldelijk gebruik gemaakt van de rekenautomaat. Hier volgt een overzicht van wat er in verschillende landen op dit gebied reeds werd gepresteerd of bezig is verwezenlijkt te worden.

Verenigde Staten

Werd bij het frequentieonderzoek van de Russische schrijftaal (proza-teksten vanaf 1930 tot 1950) dat Professor H. H. Josselson ¹⁾, voorzitter van het „Department of Slavic Languages and Literatures” van de universiteit Wayne, te Detroit in de Verenigde Staten, in 1953 publiceerde ²⁾, nog uitsluitend gebruik gemaakt van de conventionele IBM-ponskaartenapparatuur, zo deed in hetzelfde jaar echter ook de computer zijn intrede in het taalkundig onderzoek.

Een pionier op dit gebied is John W. Ellison, een geestelijke van de Episcopale Kerk, thans rector van de „Parish of the Epiphany” in Winchester (Massachusetts). Reeds in 1945 experimenteerde hij op het gebied der tekstkritiek met behulp van IBM-ponskaartenapparatuur, doch schakelde in 1951 voor hetzelfde doel over op de elektronische rekenmachine ³⁾, en bewees in 1953 de praktische uitvoerbaarheid van elektronische tekstkritiek door, met behulp van de rekenmachine Mark IV ⁴⁾,

¹⁾ *The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian* [1953].

²⁾ Voor de statistische kant van het onderzoek, had Josselson de medewerking van zijn collega B. Epstein, voor de ponskaartentechniek was hij aangewezen op H. Hyatt, van „Modern Business Records” te Detroit. Op blz. 35 van het boek geeft deze laatste een overzicht van de gevolgde werkwijze, met op blz. 37 afbeeldingen van een zgn. „source card” en van een ponskaart. Op deze reproducties ziet men dat het op kaarten gebrachte woordmateriaal ook gegevens bevat over de grammaticale categorieën der verwerkte Russische woorden, gegevens die in de frequentielijsten zelf niet zijn opgenomen.

³⁾ Volgens vriendelijke mededeling van eerw. heer John W. Ellison (17 mei 1962).

⁴⁾ Dit was een opvolger van de eerste moderne computer, de Mark I, die in 1944 in Harvard gebouwd werd. De Mark I was niet elektronisch; hij had elektromechanische relais en mechanische tellers. Pas in 1946 deed de eerste elektronische rekenmachine, de ENIAC (Electronic Numeral Integrator and Computer), haar intrede; zie L. Landon Goodman, *Man and Automation* 61 [1957].

op het „Computation Laboratory” der universiteit van Harvard te Cambridge (Mass.), een overzicht te geven van varianten van vier kapitels in het Evangelie van Lukas in 311 Griekse handschriften ¹⁾.

Nog in 1953, tijdens zijn werk aan het computerprogramma voor zijn studie der handschriften, beseftte hij dat langs dezelfde weg ook op lexicologisch gebied belangrijke resultaten zouden kunnen worden geboekt. Het plan kwam namelijk bij hem op om een elektronische bijbelconcordantie te vervaardigen. Hij ging erover praten met Dr. H. J. Cadbury, die verbonden was aan „Harvard Divinity School”, en die, als lid van het „Standard Bible Committee”, had meegewerkt aan de voorbereiding van de „Revised Standard Version”. De uitgeversmaatschappij Th. Nelson & Sons besloot toen een proef te nemen met twee bijbelboeken (zeven hoofdstukken van Genesis en een twaalftal uit Markus); deze vond plaats in het „Univac Computing Center” te New York en was in 1954 voltooid. Toen de resultaten ervan als bevredigend werden beschouwd, en nadat een contract tussen Remington Rand en de uitgever getekend was, begon Ellison in de eerste maanden van 1955, met behulp van de UNIVAC (Universal Automatic Computer) I computer van Remington Rand ²⁾ te New York, en in samenwerking met de programmeurs J. Graham en A. Bosgang, aan een concordantie op de Engelse bijbel. Reeds in 1957 verscheen *Nelson's Complete Concordance of the Revised Standard Version Bible*, een werk van 2157 bladzijden. Dit was ongeveer vier maanden vóór de auteur met een dissertatie ³⁾ die de resultaten van zijn in 1954 voltooide

¹⁾ *According to Mark IV (Time, The Weekly Magazine, August 9, 1954, 40-41); J. W. Ellison, Can a computer write a book? [1956]; C. M. Cook, Automation comes to the Bible 892 [1957]; R. Buser, Elektronentechnik 22b [1957].*

²⁾ UNIVAC was de eerste rekenautomaat die niet alleen numerieke, doch ook alfabetische tekens kon verwerken; zie *Digitale Informationswandler 663 [1962].*

³⁾ De titel van deze merkwaardig genoeg nog steeds onuitgegeven dissertatie (125 blz. tekst, 600 blz. statistieken) luidt „*The Use of Electronical Computers in the Study of the Greek New Testament. A thesis presented by John W. Ellison to the Harvard Divinity School in partial fulfillment in the requirements for the degree of Doctor of Theology in New Testament.*” Harvard University. Cambridge, Massachusetts. April, 1957.

In augustus 1962 heb ik de auteur in Winchester opgezocht. Hij was toen zo vriendelijk me een exemplaar van de „summary” van zijn dissertatie ter beschikking te stellen; hieruit citeer ik, met toestemming van de auteur, de volgende zinsneden:

“This thesis sets forth a method of using computers in the study of Greek New Testament manuscripts, particularly in the problem of the transmission of the text . . .

Chapter III presents the principal contribution of the thesis: the method of using a numerical code to specify any variant reading in the text and the logical decisions by which a computer can prepare a two-way look-up table of all the differences between all possible combinations of manuscripts . . .

Without recourse to the meaning or significance of the Greek words involved, without drawing up a list of carefully chosen variants known to have great significance in classifying types of text, but only on the basis of the tables prepared

elektronische tekstkritiek bevatte, aan de theologische faculteit van Harvard de doctorstitel behaalde ¹⁾).

Ellisons elektronische concordantie, waarvoor reeds in 1953 met experimenteren en testen werd begonnen, is als volgt tot stand gekomen ²⁾). Vijf typistes kopieerden, met behulp van de Unityper, op vier rollen magnetische band de meer dan 800.000 woorden bevattende tekst van de bijbelvertaling, wat van maart tot augustus duurde. Een tweede groep van tien typistes ponste dezelfde tekst op ponskaarten, wat twee maanden in beslag nam, als men tenminste afziet van de correctie die twee maanden vereiste. Vervolgens werden de gegevens der ponskaarten met behulp van een kaart-naar-band omzetapparaat overgebracht op magnetische band (vier rollen). Deze tweede band werd daarna automatisch, in vijf uur tijds, vergeleken met de eerste, ten einde tikfouten op het spoor te komen. In theorie nu hadden de twee groepen typistes slechts drie of vier keer dezelfde fout mogen tikken op dezelfde plaats ³⁾, zodat na het machinaal vergelijken van beide banden alle tikfouten te voorschijn hadden moeten komen, behalve natuurlijk de drie of vier die, in overeenstemming met de wetten van het toeval, voorzien waren. Niettegenstaande al deze voorzorgen bleken in de uiteindelijke magnetische band toch nog 60 à 80 fouten te zijn ingeslopen. Toen de in te voeren tekst eenmaal correct was, begon het eigenlijke werk van de rekenautomaat, waarbij vijf fasen of gangen (eng. *computer runs*) zijn te onderscheiden.

by the computer (which accounted for *every* variant in each manuscript) the manuscripts listed above have been grouped almost exactly as has been done by students using other methods

If (*this method*) is adopted as a standard procedure in the field, it will release students in the field from countless hours of tedious drudgery. At present, each student compares manuscripts „by hand”, accumulating differences according to his own system of categories and significance of variant readings; no table of the number of differences exists for scholarly use. This is equivalent to each engineer having to work out his own logarithms, square roots, tangents, etc., as he needs them

The possibility of using such methods makes the use of any of the former methods an extravagant waste of scholarly time and talents”.

¹⁾ Vele bijzonderheden over de wetenschappelijke carrière van Dr. J. W. Ellison dank ik aan inlichtingen die hij me mondeling of per brief heeft verstrekt; niet alles wat men over zijn werk in vulgariserende tijdschrift- of kranteaikelen vinden kan, is betrouwbaar.

²⁾ Zie J. W. Ellison, *Can a computer write a book?* [1956]. Zie ook een klein, doch bijzonder overzichtelijk geïllustreerd artikel *Bible labor of years is done in 400 hours* (*Life* 18 februari 1957, p. 92), en verder een artikel *Editing at the Speed of Light — 23 Years’ Work Done in 1000 Hours* (*Publishers’ Weekly. The American Book Trade Journal*, March 3, 1956, p. 1274–1278), William R. McCulley, *Univac compiles a complete Bible Concordance* [1956], en Gardner Soule, *The Machine that indexed the Bible* [1956].

³⁾ „Theoretically the two sets of girls should have made the identical error in the identical spot only three or four times”, J. W. Ellison, *Can a computer write a book?* 3b [1956].

De eerste gang (eng. *keyword run*) had betrekking op de woorden die, op grond van het voorkomen van spaties en interpuncties, alle automatisch werden voorzien van hun identificatie, t.w. de aanduiding van boek, hoofdstuk en vers, die bij het ponsen uiteraard slechts eenmaal in de tekst was aangebracht. Toen deze fase na 50 uur was voltooid, waren er naast de oorspronkelijke 4 magnetische banden 65 bijgekomen, die elk woord van de bijbeltekst vermeldden, samen met het „adres” van dat woord op de stamband (eng. *master tape*). Bij de tweede gang (eng. *elimination run*) kwamen alle voorbeelden van 127 als onbelangrijk beschouwde grammaticaal-technische woorden¹⁾ te vervallen. De tweede fase, die 10 uur tijd vergde, had tot resultaat dat de 65 banden die tijdens de eerste gang tot stand kwamen, tot 26 stuks werden gereduceerd, wat neerkomt op het weglaten van 59,4 % van de tekst. Bij de derde gang, de belangrijkste, werden alle woorden afgestemd op een portie context²⁾. Ten gevolge van tegenslag bij het programmeren nam deze derde gang, waarvoor oorspronkelijk slechts 40 à 50 uur was uitgetrokken, meer dan 250 uur in beslag. Tijdens de vierde gang, de sorteergang, werden alle, nu identificeerbare en contexteerbare woorden van de 26 magnetische banden gesorteerd en op 26 nieuwe banden overgebracht. Hiermee was de computer 50 uur bezig. Bij de vijfde gang werden onder elk kopje de contexten gerangschikt naar hun volgorde in de bijbel. Ten slotte werd dit resultaat door de automatische schrijfmachines uitgevoerd in de vorm van 6282 bladzijden, die niet minder dan 305.242 contexten bevatten, welke naar schatting betrekking hebben op 6000 à 7000 verschillende woorden.

In de bij de uitvoer verkregen vorm was het werk geschikt geweest voor offsetdruk; toch werd het, vermoedelijk uit o.m. esthetische overwegingen, langs de traditionele weg gezet en gedrukt. Van de hierdoor geboden mogelijkheid om voor het drukken nog wijzigingen in de kopij aan te brengen, heeft J. W. Ellison geen gebruik gemaakt, behalve bij

1) B.v. *a, all, are, be, him, his, we, your* enz.; zie de lijst in het voorwerk van *Nelson's Concordance*. Had men deze 127 woorden wel in de concordantie opgenomen, dan was deze twee en een half keer zo groot geworden als ze nu is.

2) Hoewel het trefwoord zowel aan het begin, in het midden als aan het eind van een zin kan voorkomen, en hoewel verder de zinnen zelden of nooit even lang zijn, moest toch gewerkt worden volgens een methode waarbij de portie context steeds volgens een stel strakke regels werd geselecteerd, en toch een min of meer afgerond, betekenisdragend geheel diende te blijven. Wie *Nelson's Concordance* opslaat, zal geredelijk erkennen dat de auteur, met bijzonder korte contexten van zelden meer dan acht woorden, een zeer bevredigend resultaat heeft weten te bereiken.

De regels die ten grondslag liggen aan de begrenzing van de porties context waren de volgende: 1. geen context mag verder gaan dan een punt in elk van beide richtingen; 2. geen context mag het begin of het eind van een vers overschrijden; 3. elke context moet het aantal van 35 tekens zo dicht mogelijk benaderen, doch dit bij voorkeur niet overschrijden; 4. geen context mag midden in een woord beginnen of eindigen.

kleine contexten van minder dan 18 tekens, als b.v. het woord *amen* aan het eind van een psalm of bij tussenwerpsels als *behold* aan het begin van een zin ¹⁾.

In totaal heeft de computer 400 uur aan het maken van deze concordantie besteed ²⁾.

Insgelijks in 1957 verscheen bij de „University of California Press” (Berkeley en Los Angeles) de *Concordance to the Poetical Works of John Dryden*. De titel van deze publikatie is eigenlijk misleidend; immers wat uitgegeven werd, is geen echte concordantie, doch slechts een index, waaruit bovendien nog allerlei woorden zijn weggelaten die men als onbelangrijk beschouwde, zoals b.v. *a, after, all, am, and* enz. De index berust echter op een echte concordantie, namelijk op die welke door Professor Guy Montgomery bij zijn dood in 1951 is nagelaten. De omvang van het werk, de drukkosten en de omslachtigheid der controle vormden een ernstige belemmering voor de publikatie. Deze werd echter mogelijk langs machinale weg, en wel met behulp van IBM-machines. Voor elk woord werd een kaart gepost met bijvoeging der referentie, zoals die op de kaarten van G. Montgomery voorkwam, en van de ponskaarten werd vervolgens een machinale tabel vervaardigd die gecorrigeerd werd. Daarna werden de kaarten automatisch gesorteerd naar de volgorde van de gedichten en de versregels, zodat de tweede proef met de oorspronkelijke tekst kon worden vergeleken. Na correctie en opnieuw alfabetisch sorteren kon de definitieve tabel worden vervaardigd, die ten slotte verknipt en in drie kolommen op speciaal gelinieerde bladen werd geplakt, waarna deze klaar waren voor fotolithografische druk. De citaten uit de oorspronkelijke concordantie van G. Montgomery werden weggelaten ³⁾, wat de omvang van het werk natuurlijk aanzienlijk verminderde; toch bevat deze index in drie kolommen nog altijd 722 bladzijden. Ook dit werk is een vorm- en geen woordindex; homografen zijn niet onderscheiden,

¹⁾ Volgens mededeling mij per 'brief (6 april 1962) verstrekt door de eerw. heer J. W. Ellison.

²⁾ R. Busa, *Elektronentechnik* 22b [1957] schrijft dat Ellison, samen met Prof. L. B. Kiddle en J. W. Carr van de universiteit in Michigan, eveneens heeft gewerkt aan een concordantie op Cervantes. J. W. Ellison deelde me echter op 6 april 1962 mee, dat dit project, bij gebrek aan de nodige geldmiddelen, in het vergeetboek was geraakt.

Volgens Quemada, *Mécanisation* 44 [1959] (overgenomen door J. B. Bessinger, *Computer Techniques for an Old English Concordance* 229a [1961] en E. G. Fogel, *Electronic Computers and Elizabethan Texts* 21, noot 12 [1962]) zou Ellison zich hebben voorgenomen de 800 verschillende versies van de Griekse bijbel te vergelijken met behulp van de rekenautomaat. Dit berust echter, naar J. W. Ellison me op 17 mei 1962 schreef, op een misverstand. In het artikel van C. M. Cook, *Automation comes to the Bible* 892 [1957], is trouwens sprake van 4.600 Griekse handschriften van het Nieuwe Testament.

³⁾ In de inleiding van deze uitgave schrijft Josephine Miles: „Quotations could not be handled by machines”; in principe is dit echter zeer goed mogelijk.

terwijl anderzijds een enkelvoud *field* gescheiden is van het meervoud *fields*. Apostrofs ter aanduiding van een genitief en accenten werden met de hand aangebracht.

Het is niet duidelijk of voor deze woordindex op Dryden — want meer dan een woordindex, en dan nog een onvolledige, is deze zgn. concordantie niet — gebruik werd gemaakt van een elektronische rekenmachine. Men krijgt echter de indruk dat zulks niet het geval is geweest, wat klaarblijkelijk te wijten is aan het feit, dat men in dit geval niet voor de taak stond langs elektronische weg een index of een concordantie te vervaardigen, doch dat men alleen maar met behulp van de conventionele ponskaartenapparatuur een woordindex had te exciperen uit een bestaande, met de hand vervaardigde concordantie. Wilde het werk verbonden blijven met de naam van Guy Montgomery, dan lag de gevolgde procedure min of meer voor de hand ¹⁾.

In het voorjaar van 1957 begon de anglist Stephen Maxfield Parrish met proefnemingen die moesten leiden tot het vervaardigen van een concordantie met behulp van een computer ²⁾. Reeds twee jaar later verscheen, bij „Cornell University Press”, zijn elektronische concordantie op de gedichten van de Engelse schrijver Matthew Arnold (1822–1888). Deze concordantie werd vervaardigd met de IBM 704 Processing Machine, een computer die 42.000 logische beslissingen per seconde neemt en van de magnetische band 15.000 tekens per seconde leest.

De versregels van Arnold werden, zonder interpunctie, geponst op ponskaarten en wel één regel per kaart, wat 69 uur duurde. De nummering der regels gebeurde automatisch, en wel zo dat het stel ponskaarten van elk gedicht samen met een (van 1 tot 999) genummerd stel kaarten door een IBM Reproducer werd gevoerd; op dezelfde manier werden de pagina's bij elke versregel gestanst. Verder werd een titelkaart geponst voor elk gedicht ³⁾. Deze bewerkingen, samen met het ponsen der versregels met

¹⁾ In de voorrede spreekt Josephine Miles haar dank uit „to many under the guidance of Mr. Gordon Morrison and Mr. Boyd Judd at the Computer Laboratory' van de universiteit van Berkeley (Californië). Uit haar beschrijving in de voorrede van de manier waarop de woordindex tot stand kwam, kan men echter alleen maar opmaken dat daarbij gebruik is gemaakt van de conventionele, elektromechanische machines. J. Miles heeft in haar dankwoord blijkbaar slechts de perifere, bij de computer behorende conventionele apparatuur op het oog gehad en niet de computer zelf; de door haar gebruikte aanduiding „IBM machines” is een beetje vaag. S. M. Parrish, *Computer Concordances* 3–4 [1962] schrijft: „the word-index to Dryden . . . was . . . checked and printed by means of IBM accounting machines”. Zie ook de kritiek van E. G. Fogel, *Electronic Computers and Elizabethan Texts* 17 [1962].

²⁾ S. M. Parrish, *Computer Concordances* 3 [1962].

³⁾ Dit was nodig omdat een regel tekst samen met zijn referenties te lang bleek voor een enkele kaart. Met behulp van de titelkaart (bevattend een identificatie als titelkaart en de naam van het gedicht) was het echter mogelijk de tekstkaarten die erop volgden te kenmerken als behorend tot het op de voorafgaande titelkaart vermelde gedicht. Zie J. A. Painter, *Computer Preparation* 92a [1960].

varianten, kostte 80 uren. Nadat het gehele stel ponskaarten was gelijst door de IBM-printer en een proef was gelezen, werden kaarten die fouten bevatten door correcte vervangen. Het totale stel van 17.000 ponskaarten werd daarop een IBM Card Reader toegevoerd, die de gegevens der ponskaarten overbracht op een magnetische band (line-dictionary tape), wat slechts 1 uur duurde. Met behulp van een inmiddels klaargekomen programma ¹⁾ werd aan de computer opgedragen:

1°. Alle woorden van de magnetische band te onderzoeken, ten einde 151 in het geheugen van de rekenautomaat opgeslagen zgn. „non-significant words” (b.v. *also, and, are, be, but* enz.) eruit te sorteren en de overblijvende woorden, samen met het serienummer van hun regel, naar een andere band over te brengen met het oog op het alfabetiseren;

2°. de circa 64.000 overblijvende „significant words”, elk voorzien van het serienummer van de regel waartoe ze behoren, te alfabetiseren ²⁾;

3°. met behulp van de op de versregels afgestemde serienummers der gealfabetiseerde woorden, voor elk woord van de tot stand gekomen woord-index de desbetreffende versregels te halen uit de „line-dictionary tape” en ze vervolgens klaar te maken voor het lijsten.

Deze versregels met hun referenties werden hierna op een magnetische band overgebracht, in aan pagina's beantwoordende porties verdeeld en ingevoegd onder die woorden van de alfabetische index waarbij ze behoorden, en wel in de volgorde van de „line-dictionary tape”, d.i. in die der gebruikte tekstuitgave. Deze paginaband werd gelijst op de IBM-printer, wat 10 uur in beslag nam. De paginagewijs ingedeelde concordantie werd gefotografeerd ³⁾, en was daarmee klaar om door offsetdruk te worden vermenigvuldigd.

Alles bij elkaar had het maken van deze concordantie van 936 pagina's slechts 198 uren gekost, waarbij echter de tijd besteed aan het nemen van proeven, het testen der werkwijze en het opstellen van het programma buiten beschouwing is gelaten ⁴⁾.

¹⁾ Dit werd opgesteld door J. A. Painter van het „IBM Scientific Computation Laboratory” te Endicott (New York).

²⁾ Dit proces komt neer op het vergelijken van een enorm aantal binaire getallen. De computer, die ongeveer 2.500 vergelijkingen per seconde maakt, had hiervoor 25 uur nodig, waarvan echter een gedeelte in beslag werd genomen door een ingewikkeld controleprogramma. Dank zij een verfijnde techniek gebeurt het alfabetiseren thans in minder dan de helft van de tijd die hiervoor oorspronkelijk nodig was.

³⁾ J. A. Painter, *Computer Preparation 936* [1960].

⁴⁾ Men dient hierbij te bedenken, dat één programma dienst kan doen voor een willekeurig aantal werkobjecten. Aan zijn verslag in *Computer Concordances* [1962], dat veel uitvoeriger en duidelijker is dan dat in de voorrede van de concordantie op M. Arnold, voegt S. M. Parrish de volgende woorden toe: „Both Mr. Painter's account of the program and mine, however, now fall in the realm of history, for the techniques have been wholly revised for forthcoming concordances” (blz. 6, noot 4).

Voor het, uiteraard meer technische verslag van de programmeur, zie J. A. Painter, *Computer Preparation* [1960].

De computer had de mogelijkheid tot automatische controle van elke versregel die hij uitzocht, waardoor bij het optreden van fouten bij het afdrukken (wanneer b.v. het afdrukwiél niet in de goede positie zou zijn gekomen) de machine stopte, en op een volgende regel het vers correct afdruckte. Een nog belangrijker automatische controle bestond hierin, dat vóór het afdrukken elke versregel werd afgetast, ten einde te controleren of hij inderdaad het trefwoord bevatte. Bij een totaal van circa 17.000 versregels, die meer dan 65.000 trefwoorden opleverden, bleken negen trefwoorden voorzien te zijn van een context die niet het goede trefwoord bevatte¹⁾. Verder werden alle trefwoorden met een frequentie van 1 (er waren er niet minder dan 4.358)²⁾ geverifieerd aan de hand van de gedrukte tekstuitgave³⁾, om er zeker van te zijn dat ze wel authentieke woorden voorstelden en geen spookwoorden die het gevolg waren van fout ponsen.

Deze eerste door Parrish bewerkte elektronische concordantie vertoont echter ook schaduwzijden, die men eerlijkheidshalve niet mag verzwijgen. In de eerste plaats is er geen verschil tussen hoofdletter en kleine letter, kursieve en niet-kursieve letter, maar wat dat betreft is waarschijnlijk binnen afzienbare tijd geen betere apparatuur te verwachten.

Een tweede nadeel is het ontbreken van interpunctietekens, behalve dan het koppelteken; hier ligt de schuld aan het gebruiken van een voor dit werk minder geschikte apparatuur, d.w.z. een die over te weinig tekens beschikte. Dit heeft tot komische consequenties geleid; zo had het ontbreken van de apostrof tot gevolg dat *she'll* en *shell*, *I'll* en *ill* onder één hoedje gevangen zijn. „It could have been worse”, schrijft Parrish, „Arnold never used „he'll” ”⁴⁾.

Een derde en vierde nadeel hadden eveneens verholpen kunnen worden, namelijk het ontbreken van elke lemmatisering en het ongescheiden zijn der homografen. Bij Parrish zijn b.v. de verleden deelwoorden *fixt*, *fixd* en *fixed* apart opgenomen, en wordt er geen verschil gemaakt tussen *rose* als zelfstandig naamwoord en *rose* als werkwoord. Het scheiden der homografen vóór het ponsen is in de meeste gevallen onpraktisch en teleurstellend en de lemmatisering eenvoudig onuitvoerbaar. Men kan in nagenoeg alle gevallen beter wachten tot er een eerste, zuiver alfabetische

1) J. A. Painter, *Computer Preparation* 95b [1960].

2) Als appendix heeft de concordantie op Matthew Arnold van S. M. Parrish ook nog een frequentie-index (blz. 937-965); hierin heeft *I* de hoogste frequentie (1045).

3) Zie J. A. Painter, *Computer Preparation* 95b [1960].

4) *Concordance M. Arnold* x [1959]. Men vraagt zich af of het niet beter was geweest *I'll* in twee woorden te scheiden: *I* en *ll*. Door een zonering der drukwielen (120 wielen: 80 zonder numeriek doch met interpunctietekens en 40 standaard) wil men in de toekomst dit euvel vermijden; zie S. M. Parrish, *Computer Concordances* 7 [1962].

index klaar is gekomen. Men kan dan aan het volgende procédé ¹⁾ denken.

Men vervaardigt eerst een genummerde index ²⁾. Met behulp hiervan gaat men na, welke woorden ook als lemma bruikbaar zijn en welke nieuwe lemmata ingevoegd moeten worden, waarna men deze laatste op hun alfabetische plaats in de lijst schrijft. Nadat men de nieuwe lemmata in de nummering heeft ingepast, zet men in de gelemmatiseerde index bij elke referentie die hiervoor in aanmerking komt, het nummer van het lemma waaronder het desbetreffende woord valt ³⁾.

Het scheiden van de homografen gebeurt op een soortgelijke manier. Met behulp van de referenties in de woordindex ⁴⁾ gaat men in de tekst na welke woorden nu nog als homografen gescheiden moeten worden ⁵⁾, waarna men de daarmee corresponderende referenties van de nodige diakritische cijfers voorziet. Nu corrigeert men de oorspronkelijk uitgevoerde indexband, waarna de computer de bewerkingen uitvoert die moeten leiden tot een nieuwe band met lemmatisering en gescheiden homografen ⁶⁾.

¹⁾ Dit werd, in overleg met mij, opgesteld door de heer J. A. Th. M. van Berckel van het Mathematisch Centrum te Amsterdam.

Een algehele automatisering bij het ordenen der woordvormen onder hun lemma's, t.w. met behulp van een in het geheugen van de computer opgeslagen woordenboek der stammen (A. D. Booth, *Use of a computer* 19a [1957]), resp. der buigingsvormen (B. Quemada, *L'Inventaire lexicographique* 122, 125 en 128 [1962]) is theoretisch natuurlijk de oplossing. In de praktijk is dit echter vooralsnog niet zo gemakkelijk te verwezenlijken, om de eenvoudige reden dat een dergelijk woordenboek niet a priori, doch slechts a posteriori is op te stellen, iets wat Booth trouwens uitdrukkelijk erkent.

²⁾ In dit stadium is een index waarschijnlijk al voldoende.

³⁾ Het is niet nodig een herkenningsteken bij het lemma te plaatsen daar in de machine een lemma gedefinieerd is als:

of een woord waar geen referentie bij staat (dit zijn de nieuwe lemmata);

of een woord waar een of meerdere referenties bij staan en dat bovendien een verwijzing naar zichzelf heeft (d.w.z. een lemma met meer woordvormen);

of een woord met referentie maar zonder verwijzing (d.w.z. een lemma met slechts één, al of niet toevallige, woordvorm).

⁴⁾ Eventueel met behulp van een gedeeltelijke concordantie op alle woorden met een frequentie groter dan 1. Zie over het scheiden der homografen ook blz. 52, 69 en 94.

⁵⁾ Eng. R O S E (znw.) en (ww.) b.v. zijn immers al door de lemmatisering gescheiden.

⁶⁾ In *Computer Concordances* 9 [1962] beschrijft S. M. Parrish een verwante techniek als volgt: „Discriminations (of homographs) can be made after indexing has been completed and variant meanings have been drawn together under a single index word. This technique involves listing all lines as they are to occur in the finished print but with no page spacing; each line is given an arbitrary serial number. Working with this unpagged list the editor can punch instructions into cards and feed them into a separate tape, which is played against the main tape to produce the final paged print. The instructions may be of two sorts: (1) delete line X (2) between lines X and Y insert the following line. By combining deletions and insertions the editor can thus either drop all non-significant meanings from an index entry or separate the lines representing one entry into two entries, each with its

De hierboven beschreven procedure is gebaseerd op het gebruik van een computer die geen in- of uitvoer met ponskaarten kent, zoals de X 1 van het Mathematisch Centrum in Amsterdam. Kan en wenst men voor een bepaald doel de door de computer gealfabetiseerde woorden ook op woordkaarten uit te voeren, dan ligt het gebruik van tussengevoegde lemmakaarten bij het lemmatiseren en het scheiden der homografen meer voor de hand.

Hoe dan ook, het is zeer wel mogelijk, de rekenautomaat een concordantielijst te laten maken, waarin onder elk lemma de diverse vormen, gevolgd door een portie context van een of meer regels tekst, worden opgenomen. Natuurlijk zal een dergelijke werkwijze iets meer tijd kosten, maar de elektronische concordantie zal er stellig ook beter door worden. Die op Matthew Arnold is bedoeld als de eerste van een reeks; in de volgende delen, die gewijd zullen zijn aan William Butler Yeats, Ben Jonson, Emily Dickinson enz., zal met het probleem der lemmatisering en der homografie wel degelijk rekening worden gehouden.

In de „Cornell Concordance Series” zal eveneens een concordantie op *The Anglo-Saxon Poetic Records*¹⁾ verschijnen. De voorbereidende werkzaamheden hiervoor bevinden zich in een vergevorderd stadium, maar aan het eigenlijke mechanische werk was men in het najaar van 1962 nog niet toe. Naar de auteur van deze concordantie, J. B. Bessinger, hoogleraar aan de universiteit van Toronto, me schreef, hoopt men daarbij gebruik te kunnen maken van de nieuwe IBM 7090 computer. Over zijn project heeft Bessinger voor de groep Oudengels van de „Modern Language Association of America”, in december 1960 te Philadelphia, een lezing gehouden, die in juli 1961 werd gepubliceerd²⁾.

Bij de voorbereiding van de tekst moet de oorspronkelijke interpunctie genormaliseerd worden, omdat de interpunctie in de concordantie tot een minimum zal worden beperkt. De delen der samenstellingen zullen bij het ponsen worden verbonden door een koppelteken, zodat ook het tweede lid in de concordantie kan worden opgenomen, en wel met een verwijzing naar het eerste lid. De beschrijving die Bessinger in zijn artikel van het mechanische proces geeft, is uiteraard vereenvoudigd en zeer summier. Interessant is de passage waarin de groei van de concordantie, op basis van de door de computer geproduceerde, niet genormaliseerde alfabetische lijst van woordvormen wordt beschreven. Deze lijst wordt namelijk door de bewerker voorzien van de nodige aanvullingen, t.w. voor het verwijzen naar de spellingvarianten en voor het scheiden der homografen. Ze wordt dan in haar geheel weer overgeponst, waarna ze overgebracht wordt op een magnetische band. Deze laatste, die de als

own index word. If desired, the index words can be followed with grammatical identifications: ROSE (noun), and ROSE (verb)”.

1) Uitgegeven door G. P. Krapp en Elliott van Kirk Dobbie [1931→].

2) In *American Documentation* 12, 227-229 [1961].

definitief te beschouwen alfabetische woordenlijst bevat, wordt dan, samen met de band van de tekst der gedichten, in de computer ingevoerd, die op grond hiervan de tekst der echte concordantie uitvoert. Hierin kunnen nog, door middel van een zgn. mutatieband, correcties en aanvullingen worden aangebracht.

Bessinger hoopt een volledige concordantie te maken, dus een waarin ook alle zgn. grammaticale woorden zijn opgenomen. Met behulp hiervan zou b.v. een studie der syntactische eenheden in de traditioneel opgebouwde Oudengelse poëzie mogelijk worden. De verschillende banden die voor het vervaardigen van deze concordantie gebruikt worden, zullen worden bewaard en ter beschikking staan voor verdere mechanische verwerking in verband met allerlei taalkundige doeleinden. Dit is trouwens eveneens het geval met alle andere banden van de „Cornell Concordance Series”.

Aan de „Cornell Concordance Series” werkt ook mee E. G. Fogel, die, samen met Parrish, bezig is aan een concordantie op Ben Jonson. In een in 1962 verschenen artikel, *Electronic Computers and Elizabethan Texts*, heeft hij een plan ontworpen om alle Engelse teksten uit de zgn. Elizabethaanse periode, d.i. de tijd van Wyatt tot Milton, op magnetische banden vast te leggen die, naar gelang de behoefte zich voordoet, voor allerlei onderzoek te gebruiken zouden zijn. Van deze banden kunnen gemakkelijk duplicaten worden vervaardigd. Met een dergelijke collectie „master-tapes” of bestandsbanden zou men b.v. alfabetische woordenlijsten, frequentielijsten en concordanties kunnen maken. Ook een volledig woordenboek over het Elizabethaanse Engels zou erdoor mogelijk worden. Verder zouden deze banden dienstbaar kunnen worden gemaakt aan de tijdrovende collatie van teksten met vele varianten en aan de attributiekritiek ¹⁾.

Th. A. Sebeok maakt bij zijn studie van linguïstische en andere problemen die verband houden met de volkspoëzie der Tsjeremissen, gebruik van de mogelijkheden die het „Research Computing Center” der universiteit van Indiana hem biedt. Dit instituut heeft de beschikking over een IBM 650, uitgerust met vier magnetische-bandeenheden en de daarbij

¹⁾ In noot 25 (blz. 30) schrijft Fogel: „Since I wrote these remarks, I have learned that Professors Frederick Mosteller and David L. Wallace, using computers, have applied statistical methods to the determination of the authorship of the disputed *Federalist* papers. A report on their work will appear shortly in a book on the Harvard Computer Symposium”.

In hetzelfde artikel pleit Fogel voor het samenstellen van een mechanische bibliografie op het gebied van de studie der Engelse literatuur, dit naar het voorbeeld van de door H. P. Luhn mogelijk gemaakte *Chemical Titles*. Een dergelijke mechanische bibliografie zou moeten bestaan uit twee gedeelten: 1. een lijst van publikaties alfabetisch gerangschikt op de auteursnaam, 2. een concordantie van de sleutelwoorden in de titels (dus sleutelwoord met context) met een verwijzingscode naar lijst 1.

behorende perifere apparatuur. In 1961 publiceerde hij, in samenwerking met Valdis J. Zeps, een *Concordance and thesaurus of Cheremis poetic language*. Dit werk heeft betrekking op de woordenschat van 1.200 dichterlijke teksten in het Tsjeremissisch, ook wel Mari genoemd. Het bevat zeven woordenlijsten alle over een en dezelfde woordenschat, doch geordend volgens verschillende criteria. Iets meer dan de helft van het boek wordt ingenomen door wat de auteurs een concordantie noemen, een term die evenwel misleidend is. Wat ze bieden, is immers niets anders dan een index van woorden, elk gevolgd door het symbool van het dialect waartoe het behoort, de aard der woordsoort, de Engelse vertaling en de vermelding der vindplaatsen ¹⁾. Na deze zgn. „concordantie” volgt een lijst van woorden geordend naar de woordsoorten, die op hun beurt weer ingedeeld zijn naar de verschillende dialecten van het Mari. Dan komt het „tergum”, d.w.z. een retrograde index, waarbij het woordmateriaal uiteraard beperkt werd tot één woord per grafische eenheid. Een vierde index bevat de woorden geordend naar de eerste vocaal, zulks met het oog op de vergelijking der structuur van het Fins-Oegriscche vocalisme. Als vijfde krijgt men een index der beginwoorden met de erbij behorende referenties. Het boek wordt besloten met twee Engelse woordenlijsten: een eerste, door de schrijvers „thesaurus” genoemd, waarin de Engelse „glossen” der Mari-woorden geordend zijn naar het model van de classificatie die C. D. Buck heeft gevolgd in zijn *Dictionary of selected synonyms in the principal Indo-European languages* [1949], en een tweede die de Engelse „glossen” bevat met een verwijzing naar de zgn. concordantie en thesaurus. Niet alle bewerkingen kwamen voor rekening van de computer; het maken van de index volgens de eerste vocaal gebeurde uitsluitend met behulp van de perifere apparatuur; noodzakelijke vereenvoudigingen in de genoemde index alsmede in de lijst der woordsoorten werden overigens

¹⁾ In A. D. Booth, L. Brandwood en J. P. Cleave, *Mechanical Resolution of Linguistic Problems* 34 [1958] vindt men echter de volgende definitie van een concordantie: „in its simpler form it is merely a list of the words contained in the text in alphabetical order, each word being accompanied by the page and line numbers for its occurrence. In the more complex form the words are reduced to basic types, so that, for example, all of the forms of the verb ‚to have’ would be entered under the same heading”. Deze definitie, waarvan men de kern reeds aantreft in A. D. Booth, *Use of a computer* 18 en 19 [1957], past op een niet-gelemmatiseerde, resp. een gelemmatiseerde index. A. J. T. Colin, *The Automatic Construction of a Glossary* [1960] definieert *glossary* als „a list, in alphabetical order, of all the different words used in any text, together with a statement of how many times each word occurs in the text” (211), dus wat wij een frequentielijst zouden noemen; *concordance* omschrijft hij als „a list, in alphabetical order, of all the different words in a text; each word being followed by a number of references, consisting of page and line numbers” (224), wat gelijkstaat met onze woordindex. In een later artikel spreken A. D. Booth en A. J. T. Colin zelfs over „dictionaries of texts” (*Dictionary Construction* 327 [1960]), waar kennelijk niets anders dan woordindices zijn bedoeld.

Dit alles is in strijd met de traditionele, linguïstische en meer algemene opvatting, die een duidelijk verschil maakt tussen woordindex, concordantie en woordenboek.

met de hand aangebracht. Het gehele werk werd ook op de conventionele wijze gedrukt.

Het Amerikaanse „Summer Institute of Linguistics”, dat zich specialiseert in het vastleggen van talen die nooit eerder werden opgetekend, ontwikkelt onder leiding van J. E. Grimes een methode om te komen tot een automatische concordantie van tekstmateriaal van die talen. De bedoeling hierbij is, degenen die studie maken van een of meer der over de tweehonderd door het S.I.L. bestudeerde talen, de nodige gegevens te verschaffen voor elk gewenst onderzoek. J. E. Grimes gebruikt voor dit doel een elektronische rekenmachine IBM 650, die hiervoor echter niet bijzonder geschikt is. Programma's voor de IBM 1410 en 1620 worden thans echter reeds beproefd ¹⁾).

In augustus 1962 stelde de uitgeverij Mouton en Co. te Den Haag de inschrijving open op een groots opgezette serie publikaties over de Romaanse talen en hun structuren, onder leiding van Professor A. Juilland. Deze serie is voortgekomen uit een project over het gebruik van computers bij het bestuderen van linguïstische structuren. Juilland begon dit project aan het „Department of Romance languages” van de universiteit van Pennsylvania tussen 1956 en 1961; hij zette het naderhand voort aan de universiteit van Stanford. Zijn bedoeling hierbij was, de elektronische rekenmachine in te schakelen bij de beschrijvende, vergelijkende en historische studie van de verschillende niveaus der linguïstische structuren, d.w.z. zowel op fonologisch, grammaticaal als lexicaal gebied. In het bijzonder wil hij hierbij de vijf voornaamste Romaanse talen betrekken, t.w. Frans, Italiaans, Portugees, Roemeens en Spaans, alsmede het Latijn, waarvan ze afstammen. Hij wenst namelijk binnen korte tijd een grote hoeveelheid exacte en gedetailleerde informatie te verkrijgen over de fonologische, grammaticale en lexicale structuur van de genoemde talen, zoals die vertegenwoordigd zijn in een groot aantal als representatief beschouwde teksten. Voor elke taal gaat hij uit van ongeveer 20.000 tot 25.000 zinnen met in totaal ongeveer 500.000 woorden. De gekozen teksten zijn homogeen in tijd en ruimte; ze liggen alle tussen de jaren 1920 en 1940, en dialect wordt uitgesloten. Wel zijn deze teksten gekozen uit en gelijkelijk verdeeld over een vijftal genres: toneel, romans, didactisch proza en essays, journalistiek en techniek. De gekozen teksten worden eerst in ponskaarten geponst, daarna overgebracht op elektromagnetische band, en door de computer volgens bepaalde programma's verwerkt. De resultaten hiervan zullen in de loop van de eerstvolgende tien jaren worden gepubliceerd in niet minder dan 50 boekdelen. Frequentiewoordenboeken van woorden, morfemen, en lettergrepen krijgen echter de voorrang.

¹⁾ De heer Grimes is zo vriendelijk geweest me per brief (15-XI-1962) deze inlichtingen te bezorgen.

Engeland

In het derde bulletin van het „Institute of Classical Studies” van de universiteit van Londen, publiceerde L. Brandwood, in 1956, een artikel getiteld „Analysing Plato’s style with an electronic computer”. Dit werd in een enigszins omgewerkte vorm, als hoofdstuk 4, „Stylistic Analysis”, opgenomen in A. D. Booth, L. Brandwood en J. P. Cleave, *Mechanical Resolution of Linguistic Problems* [1958]¹⁾, een boek dat een overzicht geeft van de resultaten die, aan het „Birkbeck College Computational Laboratory” der Londense universiteit, werden verkregen op het gebied van de toepassing der digitale elektronische rekenmachine op linguïstische problemen. „Birkbeck College” is het centrum waar in 1947 de gedachte van de mechanische vertaling is ontstaan; het eigenlijke machinale werk op dit gebied kon echter pas in 1955 op gang komen, toen de „Nuffield Foundation” de nodige financiële middelen voor dat doel ter beschikking had gesteld²⁾.

In zijn bijdrage wijst Brandwood op de wenselijkheid, de computer in te schakelen in het onderzoek van Plato’s stijl, ten einde te komen tot een preciezer, sneller en accurater chronologische ordening³⁾ der verschillende werken van de Griekse filosoof. De stijl, zo schrijft hij, is even individueel als de vingerafdrukken, indien hij tenminste niet opzettelijk gewijzigd wordt. Veranderingen in de woordenschat van een schrijver treden daarbij nog gemakkelijker op dan meer subtiele stijlaspecten: de structuur der zinnen en de woordorde in de zin. De statistiek van deze veranderingen nu kan alleen door een computer binnen een redelijke tijd en met de nodige exactheid worden behandeld; hierna kan de analyse van een overzichtelijk geordende materie door de onderzoeker beginnen en kunnen daaruit de nodige conclusies worden getrokken.

Het onderzoek van Plato’s stijl, schrijft Brandwood, zou kunnen gebeuren volgens een drievoudig schema, dat betrekking heeft op de woordenschat, het zinsritme en de syntaxis. Wat de woordenschat betreft, kan de computer, op grond van een ingevoerde band die de te bestuderen tekst bevat, gemakkelijk woordindices en woordstatistieken vervaardigen. Brandwood wijst erop, dat de computer zonder meer niet in staat is om verschillende buigingsvormen tot hetzelfde grondwoord terug te brengen, noch om gelijke woorden te scheiden naar hun verschillende betekenis. Dit zijn dus de problemen van het lemmatiseren en van het scheiden der

¹⁾ Ten onrechte laat E. Delavenay (*La machine à traduire* 122 [1959]; *Machine Translation* 113 [1960] in dit geval Booth als auctor intellectualis optreden.

²⁾ Zie A. D. Booth, L. Brandwood en J. P. Cleave, *Mechanical Resolution of Linguistic Problems* v en 4 [1958]. Men houdt zich aan het „Birkbeck College” thans bezig met de mechanische vertaling Duits→Engels en Frans→Engels; zie b.v. E. Agricola, *Die automatische Sprachübersetzung* 164 [1962].

³⁾ Verg. R. Busa, *L’analisi linguistica* 108a: „La cronologia delle opere platoniche è stata a suo tempo ricostruita, e oggi resta fuori di discussione, appunto con la statistica degli stilemi, pur condotta senza sussidio di macchine automatiche” [1961].

homografen, een terminologie die Brandwood niet gebruikt. Voor het lemmatiseren meent hij te zijn aangewezen op het traditionele handwerk, als de machine eenmaal de lijst van woordvormen heeft geproduceerd, tenzij men op de band, samen met de buigingsvorm, ook de vorm van het lemma vermeldt, waar de eerste vorm hiervan afwijkt¹⁾. Voor het scheiden der homografen²⁾ wil hij een polyvalent woord op de band van een diakritisch teken of van zijn vertaling voorzien.

Voor de mechanische analyse van het prozaritme stelt Brandwood een schema van verschillende ritmische patronen voor, die met behulp van een dertigtal symbolen op de in te voeren band dienen te worden aangebracht. Voor het onderzoek der syntactische structuur ontwerpt Brandwood eveneens een schema, dat echter in dit verband onbesproken kan blijven.

Over de uitvoering van het project blijft Brandwood echter vaag. Wel laat hij doorschemeren dat het mechanisch onderzoek van Plato's stijl een te grote taak zou zijn voor één enkel persoon, zodat men ten slotte toch nog op „team-work” zou zijn aangewezen.

Op het „Department of Numerical Computation” van Birkbeck College is men thans bezig met het maken van woordindices³⁾ op Koning Alfreds Oudengelse vertaling van Orosius, die ongeveer 45.000 woorden bevat. A. J. T. Colin schreef voor dit werk een programma, dat hij in 1960 in het tijdschrift *Information and Control* publiceerde. Uit de programma-beschrijving blijkt, dat de gebruikte computer, een Ferranti Mercury⁴⁾, niet is berekend op woorden die langer zijn dan 16 symbolen. Bij het ponsen van de tekst heeft men sommige woorden gekenmerkt door een letter die in de behandelde tekst nooit aan het begin van een woord voorkomt. Zo worden de voorzetsels door een J, de bijwoorden door een K, de eigennamen door een Q, de voegwoorden door een X, de aanwijzende voornaamwoorden door een V en de persoonlijke voornaamwoorden door een Z voorafgegaan. Het is namelijk niet zeker, of men in de definitieve frequentie-index („glossary”) al dexe woorden wil opnemen; wèl wil men de mogelijkheid openhouden om ze bij de linguïstische analyse van

1) De oplossing die Brandwood hier aan de hand doet, lijkt me in vele gevallen niet aan te bevelen.

2) Eigenlijk gaat Brandwood veel verder, aangezien hij spreekt van het onderscheiden „between the different meanings which a word may have in different contexts” (blz. 55). Dit is veel ingrijpender dan het eenvoudiger probleem der homografie, dat overigens al gecompliceerd genoeg is. Voor het onderscheiden der betekenissen, een overtalrijke verzameling van moeilijk af te scheiden grootheden, lijkt me de door Brandwood aangeprezen methode a priori zeer bezwaarlijk, ja zelfs geheel verwerpelijk.

3) Colin gebruikt in dit verband, zoals ik hierboven op blz. 86, noot 1 al vermeldde, de misleidende termen „glossary” en „concordance”. R. Wisbey, *Concordance Making* 167 [1962] schrijft zelfs dat Birkbeck College „a dictionary and concordance to Alfred's translation of Orosius” maakt.

4) In- en uitvoer gebeurt met 5-kanaalsband; zie M. Levison, *The Application of the Ferranti Mercury Computer to Linguistic Problems* 233 [1960].

de tekst te achterhalen. Afgezien van de tijd die de computer nodig had voor het verwerken der gegevens, kostte het maken van de frequentie-index („glossary”) op Orosius 3 weken aan ponsen van de tekst en ongeveer 20 uur aan schrijven en testen van het programma. Men schat dat het 1 jaar zou hebben geduurd om hetzelfde werk met de hand te verrichten. In het artikel van Colin is ook de beschrijving van het programma voor de alfabetische woordindex met referenties („concordance”) opgenomen; daaruit blijkt echter niet hoever het staat met dit tweede stadium van het werk.

FAMULIS is de afkorting van „Faculty of Arts Manchester University Lexicological and Indexing Seminar”. Dit „seminar”, dat zich ten doel stelt langs elektronische weg allerlei indices te vervaardigen, koos voor zijn publikaties een veelzeggend motto, ontleend aan de Latijnse dichter Marcus Annaeus Lucanus (39-65 na Chr.): *vix famulis audenda parat* „hij speelt dingen klaar die nauwelijks door slaven zouden worden aangedurfd”. FAMULIS werd in 1959 opgericht na het slagen van een experiment, t.w. het met een elektronische rekenmachine vervaardigen van een woordindex op Racines *Mithridate*. Nadat P. J. Wexler, van het „Department of French”, de tekst had geprepareerd, kwam de woordindex, met behulp van Miss Cicely M. Popplewell van het „Computing Machine Laboratory”, tot stand op de Ferranti Mercury computer, werkend met 5-kanaalsband, van de universiteit van Manchester. In deze eerste index zijn de homografen niet onderscheiden, terwijl tevens elke lemmatisering ontbreekt. Men zou in dit geval dus eigenlijk beter niet kunnen spreken van een woordindex, doch van een vormindex¹⁾. Voor het omwerken van de tekst van Racines *Mithridate* tot een index had de rekenautomaat 3 uur nodig, voor het tabelleren van de index, d.w.z. 100 bladzijden van een enkele kolom, eveneens 3 uur. Daar moet nog aan toegevoegd worden de tijd benodigd voor het ponsen van de tekst op de ponsband.

Het is intussen niet bij een index op *Mithridate* gebleven, doch er zijn ook andere indices op Racine, alsmede op Mallarmé en Plato tot stand gekomen, en verder ook een inventaris van de eigennamen bij Victor Hugo en een vergelijking van de woordenschat in de *Sophonisbe's* van Mairet, Corneille en Voltaire. Op Frans gebied werkt FAMULIS nauw samen met het CEVF te Besançon; het staat echter niet uitsluitend ten dienste van de Franse taal- en letterkunde. Ook onderzoeken op het gebied van het Engels (o.m. een studie over de verbale prefixen in het Oudengels) en het Minoïsch A zijn in voorbereiding²⁾. Sinds 1962 heeft men de beschikking over een Ferranti Atlas computer met 7-kanaals-invoerband. Voor het ponsen is een toetsenbord ontworpen dat het mogelijk maakt zonder speciale code alle Europese talen te schrijven, in hun oude of moderne vorm, met uitzondering evenwel van de Skandinavische talen.

1) Ook Ch. Muller, *Les index de vocabulaire* [1961] wijst hierop.

2) Zie P. J. Wexler, *Le programme de recherches de Famulis* [1962].

Manchester is niet de enige universiteitsstad in Engeland waar men experimenteert met de elektronische verwerking van lexicologisch materiaal. Op het mathematisch laboratorium van de universiteit van Cambridge werkt R. Wisbey, lector in de Duitse taal, in samenwerking met P. Swinnerton Dyer, lector in de wiskunde, aan het tot stand komen van een permanente verzameling gecodeerde teksten voor linguïstisch onderzoek door de rekenautomaat.

De teksten worden, evenals te Manchester, in ponsband geponst, aangezien de rekenmachine der universiteit, de EDSAC (Electronic Delay Storage Automatic Calculator) 2, niet met ponskaarten als invoermateriaal werkt. Oorspronkelijk werd gebruik gemaakt van ponsband met 5 kanalen waarbij slechts 32 combinaties, of met de omschakeltoets, 64 combinaties mogelijk zijn. Dit had o.m. tot gevolg dat de interpunctie in de eerste tekst die men bewerkte, in cijfers gecodeerd moest worden (zo vertegenwoordigde 1 b.v. een komma, 2 een punt enz.). Thans is men echter overgestapt op band met 7 kanalen en het toetsenbord van de Flexowriter, een automatisch ponsende schrijfmachine.

De controle van de band gebeurt voorlopig op de traditionele wijze, d.w.z. door tweemaal proeflezen van de uit de band getypte tekst. De band wordt gecorrigeerd door hem voor de correcte stukken automatisch te dupliceren en voor de foutieve tekstgedeelten opnieuw met de hand te ponsen. In de toekomst wil men echter de correctie van de oorspronkelijke band laten gebeuren, door hem in de machine te lezen en dan een band in te voeren die de gecorrigeerde regels tekst bevat. Is men eenmaal in het bezit van een correcte ponsband, dan worden de gegevens die hij bevat ter verdere verwerking op magnetische band overgebracht.

Zoals uit de titel van het in 1962 in *The Modern Language Review* verschenen verslag van Wisbey blijkt, ligt het in zijn bedoeling concordanties te vervaardigen. In een memorandum dat in 1960 aan de deelnemers van het internationale colloquium in Tübingen werd uitgedeeld, werd gezegd dat voorlopig alleen maar woordindices gemaakt zouden worden. Begin 1962 was het stadium van het eigenlijke concordantiewerk intussen nog niet bereikt, aangezien slechts de volgende fasen van het eerste project voltooid waren: het alfabetisch sorteren der woordvormen en hun identificatie met behulp van de vindplaatsen. De derde fase, die van de zgn. „parsing”, d.w.z. het scheiden der homografen en het lemmatiseren (alles met de hand), was nog niet afgesloten.

De teksten waarmee Wisbey zich bezighoudt, zijn Hoogduitse teksten uit de middeleeuwen, inzonderheid uit de elfde en twaalfde eeuw. Als eerste taak koos hij de zgn. „Wiener Genesis”, een Vroegmiddeleeuws, zowel taalkundig als literair-historisch belangrijke tekst van meer dan 6.000 verzen. Terwijl in 1960 het verschijnen van een index op de „Wiener Genesis” voor het begin van 1961 in het vooruitzicht werd gesteld, schijnt het echter thans eerder in de bedoeling te liggen een con-

cordantie op deze tekst te publiceren¹⁾. Reeds is een tweede Middelhoogduitse tekst, het „*Speculum Ecclesiae*”, geponst. Wisbey streeft ernaar te komen tot de oprichting van een centrum voor gemechaniseerde lexicologie aan de universiteit van Manchester²⁾.

Italië

Na sinds november 1949 met mechanografische ponskaarten te hebben geëxperimenteerd, is pater Busa zich reeds in 1957 gaan oriënteren op het gebruiken van elektronische middelen. Thans is, bij de verdere verwerking van de geponste tekstkaarten, de omschakeling van de traditionele mechanografische methode naar een nieuwere, waarbij van de magnetische band en de rekenautomaat gebruik wordt gemaakt, reeds begonnen. We kunnen hiervoor met een verwijzing naar hoofdstuk II, § 5 volstaan.

De werkgroep Olivetti voor de toepassing der elektronische rekenmachine op de verwerking van niet-numerieke informatie, heeft voor het „*Instituto di Filologia Romanza*” der universiteit te Rome een filologisch werkproject verwezenlijkt³⁾. Op voorstel van Professor A. Roncaglia werd besloten de rijmschema's der oude Provençaalse dichtkunst, zoals die te vinden zijn in het werk van I. Frank, *Répertoire métrique de la poésie des troubadours* [1953–1957], en de rijmschema's der middeleeuwse Franse lyriek, opgenomen doch niet structureel gegroepeerd in de alfabetisch (op het rijm van het eerste vers) geordende *Bibliographie des altfranzösischen Liedes* van G. Raynaud en H. Spanke [1955 →], samen te smelten in een alfabetisch, naar de structuur der metrische schema's geordende eenheidsindex. Hiertoe werden deze schema's, alsmede een overzicht van het aantal syllaben van alle versregels der in de werken van Frank en Raynaud-Spanke vermelde gedichten, op 80-koloms formulieren genoteerd. De gegevens werden daarop in ponskaarten geponst, vertolkt, en gecollationeerd met de teksten. Vervolgens werden de aldus geponste, vertolkte en geverifieerde gegevens op magnetische band overgebracht, daarna, met behulp van een computer ELEA 9003, met een snelheid van 45.000 tekens per minuut geordend, en ten slotte uitgevoerd op overzichtslijsten, waarop voor elk gedicht het rijmschema en het aantal syllaben per regel voorkomt. Het geheel vormt een index van ongeveer 14.000 regels in meer dan 500 bladzijden. De verwerking van de gegevens door de computer nam slechts enkele tientallen minuten in beslag. Dit eenheidsrepertorium van het Instituut voor Romaanse Filologie der universiteit van Rome ligt ter beschikking van alle belangstellenden die het voor studiedoeleinden willen raadplegen. Het maakt de vergelijkende studie der Provençaalse en Oudfranse poëzie, en van hun verwantschap en wederzijdse beïnvloeding mogelijk.

¹⁾ Zie R. Wisbey, *Concordance Making* 171 [1962].

²⁾ *a.w.* 172 [1962].

³⁾ Zie M. Pacifico, *Filologia elettronica* [1961].

Sovjetunie

In januari 1961 werd het hiërogliefenschrift der Maja's door drie jonge mathematici, E. V. Evreinov, J. G. Kosarev en V. A. Ustinov, in het Siberische rekencentrum (Novosibirsk) van de Sovjetakademie van Wetenschappen ontcijferd. Dit werd te Leningrad, ter gelegenheid van het Mathematisch Congres der Sovjetunie, wereldkundig gemaakt door S. L. Sobolev, directeur van genoemd instituut. Het materiaal (hiërogliefen en Latijnse letters van Majateksten uit de Spaanse tijd) werd in ponskaarten en op magnetische trommels vastgelegd. Toen daarna met behulp van de computer de frequenties der hiërogliefen en die der letters berekend en met elkaar vergeleken werden, bleek er geen overeenstemming tussen beide te bestaan. Deze was tenslotte wel te vinden tussen de frequentie van de letterparen en die der hiërogliefen; hiermee was de oplossing van de puzzel gevonden. De computer had 48 werkuren over de berekeningen gedaan¹⁾. De resultaten van het onderzoek werden reeds in 1961 gepubliceerd onder de titel *Primenenie elektronnych vyčislitel'nych mašin v issledovanii pis'mennosti drevnich majja*. Het eerste deel van dit werk behandelt het handschrift van Madrid, het tweede dat van Dresden, terwijl het derde deel bestaat uit een systematische catalogus der Majahiërogliefen.

Tsjecho-Slovakije

Het Oosters Instituut van de Akademie is bezig met de voorbereiding van een vergelijkend woordenboek der Semitische talen. Als eerste etappe heeft men de noordwestelijke talen (Ugarit, Hebreeuws, Fenicisch, Aramees) op het oog. Een voorlopig rapport over dit project werd uitgebracht op het vijftiengste internationale Oriëntalistencongres te Moskou in augustus 1960; het is verschenen in *Archiv Orientální* 28, 470-487 [1960]. Men stelt zich voor het materiaal te gaan ordenen met behulp van een elektronische rekenmachine EPOS²⁾ van het Instituut voor Rekenmachineonderzoek te Praag; de invoer gebeurt met 8-kanaalsponsband. Elk kopje van het vergelijkend woordenboek dient te bevatten: de nodige gegevens over het woord, alsmede zijn wortel en zijn verondersteld protosemitisch schema. Men wil het woordenboek met behulp van de rekenautomaat drukken op losse bladen, die met het oog op verbeteringen en aanvullingen gemakkelijk door nieuwe bladen vervangen kunnen worden. Het project wordt voorbereid door Dr. Stanislav Segert, verbonden aan het Oosters Instituut te Praag; hij heeft hierbij de medewerking van ingenieur Miroslav Valach, van het Instituut voor Rekenmachineonderzoek³⁾.

¹⁾ Mijn kennis van dit werk berustte aanvankelijk uitsluitend op een artikel van F. Shirokov in *The Unesco Courier* van maart 1962, waarvan een vertaling is verschenen in *De Groene Amsterdammer* van 7 april 1962, alsmede op een bericht van E. Mater in *Einfluss der Technik* 27 [1962].

²⁾ In *Digitale Informationswandler* 676 [1962] is sprake van „der in Entwicklung befindliche elektronische Rechenautomat EPOS”.

³⁾ Voor deze inlichtingen ben ik niet alleen dank verschuldigd aan Dr. S. Segert, doch ook aan Dr. E. Mater te Berlijn voor zijn vriendelijke bemiddeling in deze.

België

Aan de universiteit van Luik bestaat sinds oktober 1961 een „Laboratoire d'Analyse statistique des Langues anciennes”¹⁾, afgekort: L.A.S.L.A. De leiding berust bij Professor L. Delatte, die wordt bijgestaan door zijn werkleider É. Évrard, twee doctorandi in de klassieke letteren en een ponsstypiste. Dit laboratorium maakt gebruik zowel van de conventionele mechanografische apparatuur als van de elektronische rekenmachine. Het laboratorium zelf bezit alleen een ponsmachine, waarmee gegevens in ponskaarten worden geponst, en een vertolker die deze gegevens vervolgens op de kaarten vertolkt. Het sorteren der ponskaarten met de sorteermachine gebeurt op het „Centre Interdisciplinaire de Calcul” der universiteit. Dit rekencentrum zorgt ook voor de verdere verwerking der in de ponskaarten opgenomen gegevens, en wel door middel van een elektronische rekenmachine met een magnetische trommel als geheugen met zeer grote capaciteit.

De volgorde der werkzaamheden omvat vier stadia. Allereerst wordt de te bestuderen tekst in kaarten geponst, d.w.z. één vormwoord per kaart, waarbij dit woord steeds in kolom 33 begint. Hierbij worden door aparte codeponzingen (in kolom 53) woorden aan het eind van een zin, van een hoofdstuk of van een vers speciaal onderscheiden. Als het stel ponskaarten in de juiste volgorde van de tekst ligt, kan de computer daarna, met behulp van deze drie codes, automatisch de referentie van elke woordvorm in de kaarten ponsen, alsmede zijn volgnummer in de zin, in het vers of in het hoofdstuk. Als dit gebeurd is, noteren de medewerkers van het L.A.S.L.A. met de hand op elke kaart het lemma waartoe een woordvorm behoort²⁾, en verder, volgens een speciale alfanumerieke code, allerlei morfologische en syntactische kenmerken³⁾, lengte van het vormwoord (in letters en lettergrepen), en eigenaardigheden in de handschriftelijke traditie (variant, crux enz.). Als vierde en laatste stadium worden vervolgens op elke ponskaart, in speciaal voor dit doel gereserveerde posities (voor het ponsen van het lemma is b.v. een ponsveld tussen de kolommen 54 en 70 gereserveerd), deze met de hand aangebrachte gegevens geponst. Hiermee zijn de kaarten klaar.

Met deze kaarten maakt men indices die, door het aanbrengen van lemmatisering en het scheiden der homografen, meer zijn dan alleen maar indices van woordvormen. De ponskaarten, voorzien van de hierboven genoemde gegevens, stellen bovendien de computer in staat om recapitula-

1) Zie L. Delatte en É. Évrard, *Laboratoire d'Analyse statistique* [1962].

2) Het lemmatiseren gebeurt, evenals het scheiden der homografen, naar het model van Forcellini's *Lexicon totius Latinitatis*. Over de moeilijkheden die zich bij het lemmatiseren kunnen voordoen, zie L. Delatte en É. Évrard, *Sénèque, Consolation à Polybe* VI volg. [1962].

3) Hoe dit gebeurt, ziet men op een modelkaart in *a.w.* III; voor het overzicht van de kenmerken die gesorteerd worden, zie eveneens *a.w.* 165 [1962].

tieve lemmakaarten te maken, die voorzien zijn van de aanduiding van het totaal der woordvormkaarten waarop elk der verschillende lemmata voorkomt. Dit totale getal nu is niets anders dan de woordfrequentie. Met behulp van het stel woord- of lemmakaarten kan nu een lijst van woorden in af- of toenemende frequentie worden afgedrukt. Wanneer de computer verder alle woord- of lemmakaarten gaat groeperen volgens hun frequentiecijfer, verkrijgt men een lijst van de frequenties der woorden gevolgd door het aantal woorden dat aan de verschillende frequenties beantwoordt. Behalve een woordindex komt zodoende ook een frequentielijst en een overzicht der distributie tot stand. Hiermee is men echter nog niet tevreden. Men wil ook beschikken over grafieken en overzichten met allerlei grammaticale gegevens die betrekking hebben op de behandelde woordvormen. Verder moeten nog gegevens volgen over de lengte der zinnen, en wel: een lijst volgens hun volgorde in de tekst, een overzicht van hun distributie — b.v. zinnen van 2 woorden: 7; zinnen van 50 woorden: 1 —, en een grafiek over de lengte der zinnen zoals die in de tekst op elkaar volgen. Dit laatste is mogelijk doordat elk woord aan het eind van de zin een speciaal codeteken heeft. Sorteert men alle woorden aan het eind van een zin uit het kaartbestand, dan beantwoordt het rangnummer van deze woorden in de zin aan de zinslengte. Door deze woordvormkaarten te sorteren naar de toenemende waarde van hun rangnummer, maakt de computer een overzicht van de distributie der zinnen volgens hun lengte¹).

Als eerste werk van het L.A.S.L.A. verscheen in 1962 een index op L. A. Seneca's *Ad Polybium, De Consolatione*, uitgegeven door de Faculteit van Wijsbegeerte en Letteren der universiteit. Hierbij werden de door de printer uitgevoerde bladen via het offsetprocédé gereproduceerd. Het boek bestaat uit drie gedeelten. Het eerste (blz. 1-163) heeft betrekking op het vocabularium van de *Consolatio ad Polybium*, het tweede (blz. 165-193) bevat grammaticale gegevens, terwijl het derde (blz. 195-203) gewijd is aan de lengte der zinnen. Verreweg het grootste stuk van het eerste gedeelte wordt ingenomen door de Index Verborum (blz. 1-143). Hierin komt voor: het lemma van elk woord gevolgd door de referenties met hun vormen, en ten slotte nog een aanduiding of we bij de woordvormen te maken hebben met een variant, en of de woordvorm voorkomt aan het eind van een zin.

Nederland

Dank zij de belangstelling die Professor Dr. Ir. A. van Wijngaarden, directeur van het Mathematisch Centrum te Amsterdam, bleek te koesteren voor taalkundige problemen, kon ik reeds aanvang 1961 een begin maken

¹) Elke statistiek over de lengte der zinnen in oude teksten, waarvan de interpunctie immers allesbehalve vaststaat, heeft iets hachelijks. Professor Delatte wijst daar trouwens zelf ook op; zie *Laboratoire d'Analyse statistique* 439 [1962].

met de toepassing der elektronische rekenmachine op het gebied der Nederlandse lexicologie.

Als oefenstof diende P. Apherdianus, *Tyrocinium linguæ Latinæ* [Antwerpen, 1552], een tweetalig Latijns-Nederlands schoolboekje, een der bronnen waaruit C. Kiliaan zijn beroemd *Etymologicum Teutonicæ Linguæ* [1599] heeft samengesteld. In januari en februari 1961 heb ik de tekst filologisch geprepareerd, die daarna in maart, met behulp van de Flexowriter, in een papieren ponsband geponst werd. Op grond van de ervaringen opgedaan bij het prepareren en ponsen, konden in de loop van de volgende maanden de definitieve ponsconventies worden vastgesteld. Na herhaald proeflezen werden in de geponste tekst de nodige veranderingen aangebracht, waarbij de Flexowriter het mogelijk maakte die tekstgedeelten waarin geen wijzigingen noodzakelijk waren, automatisch te dupliceren. In het najaar van 1961 begon J. A. Th. M. van Berckel, medewerker van het Mathematisch Centrum, met de opbouw van een sorteerprogramma, waardoor, met behulp van de X 1, een elektronische rekenmachine¹⁾ van het Centrum, een viervoudige woordindex zou worden vervaardigd. Deze moest bevatten:

1. een alfabetische lijst van alle Nederlandse woorden die in Apherdianus voorkomen met verwijzing naar hun plaats in de tekst;
2. een frequentielijst van deze woorden gegroepeerd naar afnemende frequentie;
3. een lengtelijst van deze woorden gegroepeerd naar afnemende woordlengte;
4. een retrograde index: de woorden werden hierbij van achteren naar voren gealfabetiseerd.

Midden 1962 was dit programma klaar; in 3 uur tijd werd hiermee een voorlopige alfabetische index gemaakt. Dit duurde echter te lang en een verdubbeling van het aangeboden materiaal zou ongetwijfeld veel meer dan 6 uur computertijd gaan vergen. Daarom werd besloten een geheel nieuwe versie van het programma te schrijven²⁾. Nadat eerst nog de proef van de voorlopige index was gelezen en de daardoor in de invoerbanden ontdekte fouten waren verbeterd, kon eindelijk, met behulp van de inmiddels klaargekomen tweede versie van het programma, met de definitieve uitvoer der verschillende lijsten worden begonnen; dit laatste was begin 1963 voltooid³⁾.

Bij het sorteerproces leest de machine, met behulp van een snelle foto-elektrische ponsbandlezer (1.000 symbolen per seconde), de aangeboden

¹⁾ Deze beschikt over een magnetisch kernengeheugen met 12.288 adressen van elk 27 bits.

²⁾ Zie J. A. Th. M. van Berckel, *Sorting by Chain Formation* [1962].

³⁾ De resultaten zullen, samen met een fotografische reproductie van de tekst van het oude boek, door Mouton & Co. te 's-Gravenhage worden gepubliceerd.

7-kanaalsband, heptade voor heptade, en vertaalt deze heptaden in een zodanige code, dat woorden, die in de machine als getallen gerepresenteerd worden, door eenvoudige aritmetische bewerkingen gealfabetiseerd kunnen worden. Na deze bewerkingen worden de diverse lijsten door middel van een ponsband uitgevoerd; de daarin geponste gegevens kunnen ook weer door een Flexowriter uitgetikt worden.

De homografen in de Nederlandse tekst van Apherdianus werden van tevoren, bij het prepareren van de tekst gescheiden, door ze van een diakritisch teken te voorzien. De ervaring heeft echter geleerd dat dit procédé te omslachtig en te tijdrovend is; daarom zal in het vervolg het scheiden der homografen niet meer vooraf plaatsvinden, doch achteraf, met behulp van een voorlopige alfabetische referentielijst. In de toekomst zullen de indices trouwens ook gelemmatiseerd worden. Dit laatste kan bovendien, evenals het scheiden der homografen, tot op zekere hoogte geautomatiseerd worden¹⁾. In het programma werd eveneens rekening gehouden met de door de uitgever gewenste „opmaak” van de door de computer uitgevoerde „pagina”, zodat deze zonder meer fotografisch te reproduceren is.

Het maken van de alfabetische woordindex met referenties uit een tekst van 11.150 woorden kostte de computer 1 uur en 12 minuten; terwijl de frequentielijst, de lengtelijst en de retrogradelijst respectievelijk 50 minuten, 36 minuten en 1 uur vergden. Het prepareren en ponsen van de tekst, alsmede het opbouwen van het programma namen natuurlijk veel meer tijd in beslag.

In de loop van 1961-'62 werd op het Mathematisch Centrum eveneens een begin gemaakt met een statistisch woordonderzoek over een miljoen woorden der geschreven taal²⁾. Als proef werd het woordmateriaal van een aantal krantekunsten in tien Nederlandse dagbladen gekozen. Bij het voorcoderen van de teksten werden, door middel van een gekleurde onderstreping, slechts vier „woordsoorten” onderscheiden, t.w. werkwoorden: groen; bijvoeglijke naamwoorden en bijwoorden: blauw; zelfstandige naamwoorden: rood, en een restgroep: geen kleur. Bij het ponsen werden de kleurverschillen omgezet in codecijfers³⁾, t.w. resp. 3, 2, 1 en 0. Na het tellen en ordenen volgens frequentie door de X 1, bleken er in totaal 50.661 woorden te zijn verwerkt, waarvan 11.063 verschillende woorden en 6.949 met frequentie 1. Dit woordmateriaal is gepubliceerd in rapport R 642/2 van de rekenafdeling van het Mathematisch Centrum; hierbij staan de woorden naar afdalende frequentie, terwijl deze met gelijke frequentie alfabetisch geordend zijn. De woorden met frequentie 1 werden

¹⁾ Zie blz. 82 volg.

²⁾ Het plan van dit statistisch woordonderzoek dateert reeds van begin 1955; zie voor een en ander: F. de Tollenaere, *Automatisering in dienst van de taalkunde* 90 volg. [1962].

³⁾ Zie J. A. Th. M. van Berckel, *Onderzoek Woordfrequentie* [1961].

niet opgenomen; ze zijn echter wel in een archiefexemplaar beschikbaar. Vaste verbindingen als *van heinde en verre* werden als één woord beschouwd, terwijl anderzijds buigingsvormen als even zoveel aparte woorden geteld en gesorteerd werden. De verwerkingsduur der ca. 50.000 woorden bedroeg 12 uur¹⁾.

In de toekomst zal de voorcodering van het woordmateriaal volgens minder simplistische taalkundige maatstaven gebeuren, zodat aan het onderzoek der woordfrequentie tevens een volwaardig statistisch onderzoek der Nederlandse woordsoorten zal worden verbonden²⁾.

Euratom

Ook bij het centrum van Euratom te Ispra schijnt men zich bezig te houden met elektronisch lexicologisch onderzoek. Op het internationale colloquium te Besançon in 1961 deelde Ihm mede, dat bij Euratom 350 versies van het Onze Vader in verschillende talen waren gecollationeerd, dat de persoonlijke voornaamwoorden van 300 verschillende talen in groepen waren geordend, en dat men, met behulp van de omgekeerd alfabetische classificatie de suffixen in het Etruskisch had bestudeerd³⁾.

§ 3. DE ELEKTRONISCHE WOORDINDEX

We keren nu terug tot het maken van een woordindex. Uit wat voorafgaat, is het reeds min of meer duidelijk geworden hoe dat met behulp van een elektronische rekenmachine kan gebeuren. Allereerst moet de te indexeren tekst taalkundig geprepareerd worden. Hierbij moet b.v. worden uitgemaakt wat in de tekst als één woord (b.v. een oude spelling *appel boom*) of als twee of meer woorden zal worden beschouwd. Het maken van een elektronische index is uiteraard niet uitsluitend het werk van de taalkundige.

1) Zie J. A. Th. M. van Berckel, *Onderzoek Woordfrequentie. Resultaten Kranten* [1962]. Bij het nemen van een proef bleek, dat het oorspronkelijke programma, beschreven in het rapport van 1961, veel te tijdrovend was, zodat zeer ingrijpende wijzigingen moesten worden aangebracht. Ondanks deze wijzigingen was de tijdsduur toch nog de genoemde 12 uur. In de toekomst zal ook voor dit project de werkwijze gevolgd worden die besproken wordt in J. A. Th. M. van Berckel, *Sorting by Chain Formation* [1962]. Een publikatie over de statistische verwerking van het krantemateriaal is in het vooruitzicht gesteld.

2) Zie F. de Tollenaere, *Automatisering in dienst van de taalkunde* 91 [1962].

3) Zie *Cah. de Lex.* 3, 72 [1962].

In aansluiting hierbij kan nog een ander project worden vermeld, dat in 1961-'62 door J. C. Gardin, in samenwerking met P. Braffort van Euratom, op de IBM computer 650 werd geprobeerd. Dit project, dat echter niet van lexicologische doch van historische aard is, betreft Akkadische spijkerschrifttabletten van Kültepe (eertijds Kanish) in Anatolië. De bedoeling hierbij is een overzicht te verkrijgen over de handelsbetrekkingen die in Kanish gevestigde Assyriërs in de 19de eeuw v. Chr. onderhielden met de hoofdstad Assur en met de steden van Kappadocië. Thans is men overgestapt op een nieuw programma, t.w. voor de IBM 7090. Zie J. C. Gardin, *Mécanisation recherches historiques* [1962].

Deze is aangewezen op de hulp van de programmeur, die de betrokken rekenmachine grondig dient te kennen, en die de reeksen van opdrachten voor deze machine moet opstellen. Hiertoe maakt de programmeur eerst een stroomschema (eng. *flow-chart*), d.i. een stroomvormige grafische voorstelling van de belangrijkste processen bij de verwerking van de informatie. Na aldus de grote lijn voor de werkwijze vastgelegd te hebben, worden de diverse onderdelen nog in detail uitgewerkt tot het niveau van de opdrachten. Daarna moeten de opdrachten worden omgezet in de taal der machine, d.w.z. in de voor de betrokken rekenmachine geldende codecijfers of letters, en wel door middel van een programma, d.w.z. een reeks van gecodeerde opdrachten voor het oplossen van een probleem. Daarna ziet het schema van de werkzaamheden er in principe uit als volgt ¹⁾:

- 1°. Ponsen van de tekst in ponskaarten of in ponsband.
- 2°. Toevoer van de gegevens naar ²⁾ en verwerking ervan door de elektronische rekenmachine: de woorden worden alfabetisch gesorteerd³⁾, waarbij elk woord wordt voorzien van zijn referenties; bij deze referenties kan desgewenst door een speciaal teken, b.v. een sterretje, ook worden vermeld waar een woord in rijmpositie (voor de machine: aan het eind van elke dichtregel) staat.
- 3°. Uitvoer van de elementen.
 - a) Ponsen van de woordkaarten (deze behoeven niet meer alfabetisch gesorteerd te worden, aangezien ze in alfabetische volgorde uit de rekenmachine komen), en tegelijkertijd drukken van een alfabetische woordenlijst met referenties op de met de rekenmachine verbonden tabelleermachine, of
 - b) Afdrukken van de alfabetische woordindex met referenties op het drukwerk van de computer.

Wanneer men voor lexicologische doeleinden de beschikking wil hebben over woordkaarten, dan kan men zich uiteraard niet tevredenstellen met de onder 3°, b) vermelde procedure. Gaat het echter alleen maar om een vormindex, waarbij dus geen lemmatiseren en geen scheiden der homografen plaatsvindt, dan is de uitvoer in de vorm van ponskaarten overbodig en kan men volstaan met het drukken van de resultaten op het drukwerk. Een andere mogelijkheid bestaat in uitvoer op ponsband, die gelezen wordt door een teletypemachine (b.v. Flexowriter), die tevens automatisch de tekst tikt op een schrijfmachine met een snelheid van 550 tekens per minuut. Ook hierbij wordt een niet te verwaarlozen besparing aan kaarten,

¹⁾ Verg. B. Quemada, *Mécanisation* 38 [1959].

²⁾ Hierbij wordt de ponskaart elektromechanisch, de ponsband foto-elektrisch, d.i. met behulp van fotocellen, gelezen.

³⁾ Heeft men bij het sorteren de beschikking over magnetische band, dan dienen de in ponskaarten of in ponsband geponste gegevens eerst met behulp van een elektromechanisch, resp. foto-elektrisch leesapparaat en een transcodeuse, of met behulp van de computer, op magnetische band te worden overgebracht.

kaartenbakken, bergruimte en tijd verkregen. Wenst men bij het maken van een concordantie of een index een aantal woorden als onbelangrijk uit te schakelen (b.v. de zgn. „mots-outils” of grammaticaal-technische woorden), dan kan dit gebeuren tussen de operaties 2 en 3 in; de tekst passeert hierbij door een vergelijkingsinrichting waarin van tevoren een lijst is opgenomen van de woorden die verwaarloosd kunnen worden¹).

Deze werkwijze heeft grote voordelen voor werk waarbij het resultaat direct verkregen wordt bij de uitvoer der machine, en waarbij geen opslag van materiaal, geen archief nodig is, zoals b.v. bij de statistische inventaris. Nemen we b.v. een enquête betreffende een natuurwetenschappelijke tekst van 1.000.000 woorden, die 500.000 onbelangrijke woorden, en in totaal 40.000 verschillende woorden bevat, dan krijgen we successievelijk de volgende werkzaamheden:

1. Ponsen van de band (doorlopende tekst, zonder referenties).
2. Overbrengen op magnetische banden.
3. Uitschakelen van de onbelangrijke woorden (50 %).
4. Alfabetisch sorteren.
5. Ponsen van een kaart per afzonderlijk woord, met aanduiding van het totale aantal keren dat dit gebruikt wordt en de berekening der frequentie.

Deze langs elektronische weg tot stand gekomen statistische inventaris bevat slechts 40.000 kaarten, terwijl er met de traditionele mechnografische techniek misschien wel 1.000.000 of 25 maal zoveel nodig waren geweest. Men kan trouwens, desgewenst, de uitvoer van 40.000 kaarten uitsparen, door deze te vervangen door een ponsband die daarna machinaal wordt gelezen en automatisch getikt op een schrijfmachine²). Ook vormstatistieken, aparte lijsten van rijmwoorden enz. kunnen gemakkelijk aan de elektronische index worden toegevoegd.

Van al deze operaties met de elektronische rekenmachine is het stadium vóór de invoer het enige dat werkelijk tijd kost. Het vervaardigen van een ponsband kan beschouwd worden als de basisinvestering. Uit een ponsband-archief waarin een hoeveelheid taalmateriaal is opgeslagen, kunnen, naar gelang van de behoefte, telkens de gewenste gegevens uitgesorteerd worden, hetzij direct, hetzij na overbrenging op een magnetische band.

Voor een gedetailleerde beschrijving van het maken van een gemechaniseerde woordindex, moge schr. verder verwijzen naar zijn behandeling van het onderwerp „Indices” in zijn artikel *Automatisering in dienst van de taalkunde* 94-99 [1962].

§ 4. DE TOEKOMST

Het dienstbaar maken van elektronische middelen aan het lexicologisch onderzoek zal in de toekomst ongetwijfeld nog aanzienlijk vereenvoudigd

¹) Zie B. Quemada, *Mécanisation* 38 [1959].

²) t.a.p. [1959].

en versneld kunnen worden. Wanneer men inderdaad binnen zeer korte tijd – er is zelfs sprake van een maximum van vijf jaar¹⁾ – in staat zal zijn zeer snel een gedrukte tekst foto-elektrisch te lezen zoals hij is, dan betekent dit dat het handwerk van het ponsen komt te vervallen, zodat men óf de gedrukte tekst in codevorm op de magnetische band kan overbrengen, óf van de gedrukte tekst direct leesbare (geponste) kaarten kan vervaardigen. Dit zou betekenen dat de ponstypiste overbodig wordt, wat het hele proces zou vereenvoudigen en versnellen; in dit verband spreekt men wel van de ontbrekende schakel in de automatische verwerking van gegevens. Verder zou, en dat is van lexicologisch standpunt het allerbelangrijkste, de collatie van de (pons)kaarten met de oorspronkelijke tekst en de tijdrovende en vervelende correctie (een eerste maal door de ponstypiste, een tweede maal b.v. door de wetenschappelijke „bewerker” of liever preparator van de tekst) geheel komen te vervallen. De ponskaarten, voor zover die voor bepaalde lexicologische doeleinden nog gewenst zouden zijn, kunnen dan veel vlugger en accurater tot stand komen en dus ook belangrijk goedkoper worden. Is echter het machinaal lezen mogelijk, dan is ook het machinaal sorteren van niet-geponste kaarten denkbaar²⁾, zodat de ponskaart op den duur misschien geheel door de steviger niet-geperforeerde machinale kaart kan worden vervangen.

De ontwikkeling der door diverse grote wereldfirma's geconstrueerde elektronische machines voltrekt zich in een snel tempo; de hoeveelheid rekenwerk die nodig is voor allerlei wetenschappelijk onderzoek, maakt steeds machtiger elektronische rekenmachines noodzakelijk. Daarnaast is reeds een ontwikkeling begonnen in de richting van machines die speciaal voor bepaalde doeleinden worden geconstrueerd, zoals speciale „information retrieval” machines of vertaalmachines³⁾. Van deze ontwikkeling kan ook de lexicologie, die tot voor kort nog geheel als handwerk moest worden bedreven, profijt trekken. De snelle ontwikkeling der mechanisatie behoeft ons niet af te schrikken, en wel in deze zin dat we niets zouden doen, omdat alles toch zo vlug veroudert. Immers de vertolking in mechnografische taal, zoals die bij de ponskaart of de ponsband aanwezig is, kan steeds weer automatisch in een andere, een nieuwere code worden omgezet.

De grote mogelijkheden die de rekenautomaat ons op lexicologisch gebied aan de hand doet, mogen ons echter niet blind maken voor enkele bezwaren. Bij een betrekkelijk geringe hoeveelheid informatie is het gebruik van zulk een enorm rekenapparaat niet praktisch; immers het programmeren zou wel eens meer tijd kunnen vergen dan de behandeling

1) Zie *Inf.* 12, 14a [1961]. Dit is echter vermoedelijk rijkelijk optimistisch.

2) „Eind november 1960 annonceerde IBM de 1418, een sorteermachine (400 documenten per minuut) met optische leesmogelijkheid die in staat is 480 cijfers per seconde te lezen”, *Inf.* 12, 18a [1961].

3) Zie S. Ceccato, *La storia* 124a [1961]; F. de Tollenaere, *Automatisering in dienst van de taalkunde* 86 [1962].

met de klassieke elektromagnetische middelen. Bij de lexicologische analyse zijn soms slechts enkele delen van de grote elektronische rekenmachine nodig; men moet nu eenmaal geen kanon gebruiken om een vlieg te doden¹⁾! Het gebruik van de rekenautomaat is erg duur, namelijk ongeveer f 500. — tot f 1000. — per uur. Het gebruik van elektronische rekenmiddelen is m.a.w. slechts te rechtvaardigen bij werkzaamheden waar een groot aantal woorden mee is gemoeid. Een ander probleem is dat de lexicograaf slechts af en toe de beschikking krijgt over de rekenmachine, b.v. een uur per week 's avonds, terwijl men vaak een gesloten werkprogramma heeft, waarbij men ongehinderd over de machine moet kunnen beschikken. Voor onze elementaire lexicologische gegevens is de computer dikwijls niet voldoende productief te maken. Voor lexicologische doeleinden zou daarom een kleine, speciale, elektronische rekenmachine, waarover men permanent zou kunnen beschikken, de voorkeur verdienen boven een machine die slechts onregelmatig, en dan vaak op ongelegen komende werkuren verhuurd wordt. Met een dergelijke kleine rekenmachine zou b.v. snel gefabetiseerd of gesorteerd kunnen worden²⁾.

Het is in elk geval thans reeds voldoende duidelijk, dat voor het maken van een enigszins uitvoerige concordantie of woordindex, de elektronische rekenmachine de voorkeur verdient boven de conventionele apparatuur³⁾. De capaciteit en de snelheid der elektromechanische machines zijn voor deze taak toch eigenlijk te gering⁴⁾. De context echter zoals die voorkomt in de elektronische concordantie van John W. Ellison op de Engelse bijbel, t.w. per woord slechts een korte regel, voldoet wel aan theologische behoeften, doch is van lexicologisch-lexicografisch standpunt uit geheel onvoldoende. En de woordindex die, voor zover hij althans volledig is, ideaal materiaal bevat voor de taalstatistiek, is voor de lexicografie weinig anders dan ruwe grondstof, die niet zonder meer te gebruiken is. Het valt echter niet te betwisten, dat een concordantie of een woordindex ook voor lexicografische doeleinden indirect bijzonder nuttig kan zijn. Daarom mag en kan de lexicograaf die studie maakt van de mechanisering in de lexicologie, niet zonder meer voorbijgaan aan de prestaties op het gebied van de elektronische concordantie⁵⁾ en de elektronische index.

Kan de elektronische rekenmachine nu ons het ideale lexicografische

1) R. Busa, *Elektronentechnik* 25b [1957].

2) De Gamma 60-BULL geeft bij de sorteermachine een besparing van 90 %, bij de tabelleermachine een besparing van 37 %.

3) „Experience has proven that the maturity and accessibility of the method require magnetic surface electronic calculators and not punched card machines as the main and fundamental instruments in these applications, even if the punched card machines will still have to condition and integrate the use of the calculators”, R. Busa, *Darstellung* 5 [1960].

4) Zie E. G. Fogel, *Electronic Computers and Elizabethan Texts* 16 [1962].

5) In 1960 heeft P. Meile voorgesteld de *Vedic Concordance* van M. Bloomfield [1906] met behulp van de rekenautomaat te ordenen in retrograde volgorde.

materiaal leveren, d.w.z. de „lexie”¹⁾, zoals die als zgn. „bewijsplaats” in de grote woordenboeken voorkomt²⁾? In principe is de elektronische rekenmachine hiervoor even geschikt of ongeschikt als de conventionele ponskaartenapparatuur, alleen is ze waarschijnlijk sneller en vaak ook duurder.

Het lijkt geen twijfel, dat de gemechaniseerde lexicologie in de toekomst in toenemende mate van de rekenautomaat gebruik moet gaan maken. Hoe groter het met de klassieke mechanografische middelen verzamelde materiaal wordt, des te onontbeerlijker wordt de elektronische rekenmachine. Alleen deze immers kan een enorm materiaal, een kartotheek van miljoenen fiches snel en economisch toegankelijk maken³⁾. Beide stadia der mechanisatie, de traditionele mechanografische en de elektronische, vormen geen tegenstelling, doch vullen elkaar aan⁴⁾. De taalkundige analyse, het fijne wetenschappelijk-lexicologische werk kan alleen maar gebeuren op kaarten; dit blijft de grondslag⁵⁾. Daarnaast zal in toenemende mate een beroep moeten worden gedaan op geheugens in de vorm van magnetische banden, schijven of trommels, waarop een zeer grote hoeveelheid informatie kan worden vastgelegd.

Het beschikbaar stellen van een enorme hoeveelheid taalkundige informatie laat in de lexicologie toe profijt te trekken van kwantitatieve methodes, die een exacter en vollediger resultaat garanderen dan totnogtoe voor mogelijk werd gehouden. Het zijn de machines, mechanografische en elektronische, die de taalkundige informatie, het lexicologische materiaal ter beschikking stellen. Het is echter de onderzoeker, de taalkundige, die met dit enorme materiaal zal moeten werken en het, zij het met behulp van elektromechanische of elektronische middelen, de baas zal moeten blijven. Aan het slot van zijn fundamenteel artikel „La mécanisation dans les recherches lexicologiques”, schrijft Professor Quemada de volgende woorden: „Cet aspect de la question touche au cœur même du problème, qui demeure essentiellement humain. . . . Ce sont des hommes qu’il faut, pour assurer la préparation et l’exploitation des matériaux”.

1) Over dit begrip, dat afkomstig is uit de Kopenhaagse School, zie A. J. Greimas, *Problèmes* 53 [1959].

2) Zie F. de Tollenaere, *Documentation lexicographique* [1962], waarin het probleem behandeld wordt, hoe men met behulp van de conventionele apparatuur, en in de toekomst ook met de computer, een context verkrijgt die zoveel mogelijk de traditionele „lexie” benadert.

Ik vraag me echter af of, uit praktische overwegingen, een enigszins uitvoerige, vaste context toch niet de voorkeur verdient.

3) P. Imbs, *Au seuil de la Lexicographie* 3 [1961]: „des centaines de millions de fiches sont à rassembler, à normaliser, à vérifier, à trier”.

4) Men kan hier nauwelijks van een tegenstelling: elektromechanisch ↔ elektronisch woordenboek spreken; verg. L. Kukenheim, in *Lev. Talen* 1960, 33. Een zuiver elektronisch woordenboek is m.i. alleen denkbaar bij de automatische vertaling; als zodanig is dit elektronisch vertaalwoordenboek dan, evenals elk ander tweetalig woordenboek, eigenlijk weinig anders dan een glossarium.

5) Zie ook B. Quemada, in *Cah. de Lex.* 3, 14 [1962].

HOOFDSTUK IV

EEN CENTRUM VOOR LEXICOLOGIE IN NEDERLAND?

Nous ne pouvons pas attendre vingt ans, sous prétexte qu'il y aura de nouveaux perfectionnements; il y a un moment où il faut savoir commencer, G. Matoré, in *Lexicologie et Lexicographie* 64 [1960].

§ 1. WOORDENBOEK EN THESAURUS

Wörterbücher einer lebenden Sprache werden nicht für die Ewigkeit geschrieben. Sie sollen von einer späteren Generation überholt werden, G. Roethe, in *Das Institut für Deutsche Sprache und Literatur* 132 [1954].

Het bewerken, het maken van een woordenboek mag niet te lang duren; toelaatbaar is een periode van 25 tot 40 jaar, waarbij de noodzakelijke eenheid van wetenschappelijk-lexicografische opvatting, een sluitende methode en een zekere stijl, d.w.z. een uniform uiterlijk der diverse artikelen, nog mogelijk zijn. Een woordenboek mag vooral niet permanent zijn, het moet evenals een revolutie niet alleen begonnen, doch ook beëindigd kunnen worden. Het begin van een woordenboek is meestal een moeilijke zaak, een kwestie van veel overleg en veel zorg. Als het kind eenmaal met veel moeite ter wereld is gekomen, wil het vaak niet zo vlug en voorspoedig groeien als men wel zou wensen. Het heeft in de regel last van kinderziekten en groeistuipen, en de oorspronkelijke verwachtingen die men ten opzichte van de nieuwe telg koesterde, worden niet altijd vervuld. Een woordenboek heeft ook veel moeite om het eindpunt te bereiken; er zijn tal van voorbeelden te noemen van woordenboeken, die voltooid werden meer dan een eeuw nadat het eerste deel verscheen. Dit komt in feite hierop neer, dat ze min of meer, zoals levende fossielen, zichzelf wetenschappelijk en methodologisch overleven.

Wanneer de meeste grote woordenboeken veel te lang duren, is dat in de regel het gevolg van verschillende factoren. Een voorname factor is b.v., dat men bij de meeste grote historische woordenboeken, zoals die in de 19de eeuw, naar het voorbeeld van dat van Grimm, tot stand kwamen, heeft vergeten dat naast een „terminus a quo” ook een „terminus ad quem” of „post quem non” een absolute noodzaak is. De voornaamste oorzaak is waarschijnlijk echter toch wel deze, dat het verzamelen van het apparaat, d.w.z. het lexicologische materiaal waarop het werk berust, een veel te zware belasting vormt. Alleen al het voorbereidende werk, t.w. het bij elkaar brengen van dit, in den beginne lang niet volledige, materiaal pleegt

verschillende decennia in beslag te nemen. Het tot stand brengen van dit apparaat, van de miljoenen kaarten met citaten, geschiedt verder, als vanouds, geheel met de hand. Hierbij komen de fiches, die volledig met de hand geschreven worden, zeer langzaam tot stand ¹⁾ en ze zijn bijzonder duur. Dit oude lexicologische handwerk, dat de basis vormt van elke lexicografische onderneming, heeft zich tot op onze dagen in principe vrijwel onveranderd weten te handhaven ²⁾. Nu zijn de eisen die men aan de uitvoerigheid en volledigheid der grote woordenboeken is gaan stellen, steeds meer toegenomen. Het gevolg is een discrepantie tussen het groeiende, evoluerende doel, en de beschikbare materiële middelen. Men mag dus gerust zeggen, dat wanneer de grote woordenboeken zo lang duren, dit voor een groot deel — ik zeg niet uitsluitend — te wijten is aan het feit dat men het enorme benodigde materiaal met de traditionele methode van het handwerk niet meer aankan.

De conclusie die zich opdringt, is deze: de lexicologische taak moet van de zuiver lexicografische gescheiden worden. De zorg voor het materiaal behoort niet langer de zorg van de lexicograaf te wezen en nog minder zijn doel. Woordenboek en woordarchief dienen gescheiden te worden. Naast tijdelijke, liefst niet al te uitvoerige, hanteerbare woordenboeken van allerlei aard, geconcipeerd en uitgevoerd volgens de wetenschappelijke eisen en methoden van een bepaalde tijd, behoort een permanent woordarchief te staan waarin de woordenschat van een taal is en wordt ondergebracht ³⁾. Als het waar is dat men, zij het met enige overdrijving, mag zeggen dat het enig blijvende van een historisch woordenboek ten slotte de citaten zijn, dan kan men omgekeerd ook zeggen dat de citaten blijvend moeten zijn. Belangrijker dan elk woordenboek is dan het taal museum ⁴⁾,

¹⁾ „Het WNT... werkt (niet) met een kant en klaar apparaat, maar met materiaal dat pas beschikbaar komt naarmate het werk vordert”, C. Kruyskamp, in *Wet. Tijd.* 20, 148-149 [1960].

²⁾ „Die Materialsammlung für ein Wörterbuch ist zeitraubend und kostspielig. Die Exzerption von fremdsprachlichen und sondersprachlichen... Texten kann im allgemeinen nur von wissenschaftlich ausgebildeten Kräften vorgenommen werden... Mit den handschriftlich hergestellten Wörterbuchbelegen steht unsere Lexikographie zum Teil noch auf dem Stande des mittelalterlichen Handschriftenkopierens”, G. Ising, *Zur Erprobung eines fotomechanischen Exzerptionsgerätes* 90a [1962].

³⁾ In Duitsland heeft B. Beckmann in 1954, in zijn rede *Das Deutsche Wörterbuch*, gehouden bij de opening van *Das Institut für Deutsche Sprache und Literatur*, gewezen op de noodzaak van een permanente instelling voor lexicologie, en op de taak om, onafhankelijk van het gebruik dat ervan kan worden gemaakt, een blijvend archief van de Duitse woordenschat op te bouwen: zie vooral blz. 127 en 135. Aan de mogelijkheid der automatisering op lexicologisch gebied heeft de schr. echter niet gedacht; wel aan het bekende procédé waarbij elke halve bladzijde van een tekst zo dikwijls vermenigvuldigd wordt als er woorden op staan (en dat in duplo), gevolgd door het aanbrengen der trefwoorden en hun alfabetisering (zie blz. 129-130).

⁴⁾ M. de Vries, *WNT* I, Inleiding LXXIV [1882]. Onlangs heeft B. Quemada gesproken van „Archives Générales de la Langue Nationale” (*L'Inventaire lexicographique* 121 [1962]).

waarin zijn ondergebracht de rijkdommen der moedertaal. Deze thesaurus is principieel niet voor publikatie bestemd. Hij moet gebruikt kunnen worden, niet alleen als materiaalverzameling waaruit in de toekomst elk woordenboek van de taal, groot of klein, moet kunnen putten, doch bovendien als grondslag voor lexicologische studies in de ruimste zin van het woord. Hij moet, om enkele concrete voorbeelden te noemen, te gebruiken zijn, niet alleen voor het tot stand brengen van een woordenboek over de taal van de zestiende eeuw en een woordenboek der middeleeuwse jachttermen, doch evenzeer voor een studie over het voegwoord in de eerste helft der zeventiende eeuw. Zijn materiaal, dat gemakkelijk volgens allerlei criteria te rangschikken is, moet niet alleen gebruikt kunnen worden voor zuiver lexicografische doeleinden, doch evenzeer voor de studie van de semantiek en de vormleer. Als b.v. de prozateksten van de 14de eeuw eenmaal in hun geheel geficheerd zijn, is de proportionele verhouding van de imperatief t.o.v. de conjunctief b.v. in een minimum van tijd te constateren. Bij een volledige fichering kan het materiaal ook belangrijke diensten bewijzen voor de onomastiek. Men kan niet alleen uit een materiaal van de veertiende eeuw alle voornamen sorteren, doch ook door middel van het rekenwerk van de tabelleermachine het totale aantal keren dat ze gebruikt worden; deze gegevens worden door de totaalkaartenponsmachine in totaalkaarten vastgelegd, zodat men een overzicht krijgt van de onderlinge proporties. Ook zuiver onomastisch werk kan mechanografisch gebeuren. Van een door een toponymist geprepareerde tekst b.v. kan snel een index, resp. een concordantie van toponiemen worden aangelegd.

Dat het woordenboek dan secundair wordt t.o.v. de mechanografische kaartenverzameling, dat de thesaurus de voorrang krijgt op het woordenboek of op de woordenboeken, is geen nadeel; integendeel. Het woordenboek als thesaurus wordt geheel overbodig; het behoeft niet langer tevens als materiaalverzameling te fungeren, en kan dus beknopter worden. Immers wenst de gebruiker meer voorbeelden van een bepaald woord in een bepaalde betekenis, dan kan hij bij de thesaurus terecht. In plaats van een onoverzichtelijk, soms afschrikwekkend taalmagazijn krijgt de gebruiker een soberder, duidelijker overzicht, waarmee hij meestal meer gebaat is.

Dat een al te eenzijdige binding tussen woordenboek en materiaalverzameling, of, om het op een geleerde manier te zeggen, tussen lexicologie en lexicografie moet worden vermeden, heeft het comité voor het Noorse woordenboek, *Norsk Riksmålsordbok*, reeds dadelijk bij het begin van de werkzaamheden duidelijk ingezien. De miljoenen citaten die men bij elkaar bracht, moesten niet alleen dienst doen voor het woordenboek, maar het begin betekenen van een zo volledig mogelijke woordverzameling met zelfstandig bestaansrecht¹⁾.

De waarde van de woordverzameling, van het lexicologisch apparaat

¹⁾ T. Knudsen, *Det norske litterære ordboksverk i 25 år, 1921-1946 (1922-1947)* 6 [1948].

op zichzelf, los dus van het woordenboek dat eventueel op grond daarvan tot stand komt, is men in de kringen van het *WNT* pas zeer laat gaan inzien. Hoewel reeds in 1939 C. H. A. Kruyskamp, in stelling XI van zijn dissertatie, gepleit had voor het bewaren van het materiaal in zijn geheel, is men toch nog geruime tijd doorgegaan met het vernietigen van het kostbare en dure apparaat, nl. zowel van de citaten die in de kolommen van het *WNT* werden opgenomen als van de andere, naarmate de klaar-gekomen artikelen werden afgedrukt. Pas op 5 mei 1953 werd besloten tot het bewaren van het materiaal, zowel van de in het woordenboek verwerkte fiches als van de niet-gehonoreerde briefjes. Tegelijk viel ook het besluit om het formaat van de toenmalige grote fiches te halveren.

Wanneer de lezing, *Het Nederlandsch Woordenboek in de twintigste Eeuw*, die Blancquaert in 1938 voor de Leidse Maatschappij ¹⁾ en voor de Koninklijke Vlaamse Academie ²⁾ heeft gehouden, tot op heden, bij mijn weten, geen enkel concreet resultaat heeft opgeleverd, ligt dat stellig niet aan een gebrek aan inzicht of visie. Het tegendeel is eerder waar; de visie was zo groots, dat vrijwel iedereen moest twijfelen aan de mogelijkheid tot praktische verwezenlijking ervan. Als men Blancquaert een verwijt kan maken, is het misschien dit, dat hij niet voldoende heeft beseft dat een dergelijke onderneming, wilde ze slagen, niet alleen maar kon berusten op „het nuchtere wikken en wegen van ieder onderdeel”, alsmede „op den vasten wil, op de hardnekkigheid, op een soort van bezetenheid om steeds met het einddoel voor oogen, over alle voorziene en onvoorziene hinderpalen heen, recht op het doel af te gaan” ³⁾. In werkelijkheid was dit plan op de grondslag van het traditionele handwerk onmogelijk te verwezenlijken. Dat Blancquaert verder het zuiver lexicologische en het lexicografische gedeelte van de taak niet van elkaar gescheiden houdt, en het geweldige te verzamelen materiaal slechts ziet in functie van „het Nederlandsch woordenboek der twintigste eeuw”, is een erfenis, of, als men wil, erfelijke belasting van het lexicografische verleden. Daartegenover staat dat hij t.o.v. het materiaal de eis stelt van volledigheid ⁴⁾ (het moet o.m. ook de onomastische gegevens bevatten) ⁵⁾, van een maximum van produktiviteit ⁶⁾ en van alzijdigheid ⁷⁾. Al deze progressieve eisen zijn echter niet te verenigen met de methodes waarover men tot voor kort de beschikking had.

¹⁾ Op 9 december 1938; zie *Jaarboek van de Maatschappij der Nederlandsche Letterkunde 1938-1939*, blz. 141.

²⁾ Op 21 december 1938; zie *VMKVA 1938*, 777-778. De titel van de gedrukte lezing luidt *Het Nederlandsch Woordenboek van de twintigste eeuw*; in 1935 (*VMKVA 1935*, 10 (1052)) sprak Blancquaert reeds van het „Nederlandsch taalkundig woordenboek van de twintigste eeuw”.

³⁾ *VMKVA 1938*, blz. 23 (799).

⁴⁾ *a.w.* 7 (783).

⁵⁾ *a.w.* 8 (784).

⁶⁾ *a.w.* 17 (793).

⁷⁾ *a.w.* 13 (789).

§ 2. MODERNE LEXICOGRAFIE

Er bestaat geen enkele machine die woordenboeksartikelen kan maken; het fijne speurwerk van de wetenschappelijke onderzoeker kan nu eenmaal niet door de machine worden overgenomen. Wat dat betreft, behoeft men dus niet direct een revolutionaire ontwikkeling te verwachten. Toch zal men er goed aan doen, zich opnieuw op de theoretische grondslagen van het historische woordenboek te bezinnen.

Van het nieuwe, grote, Spaanse historische woordenboek, *Diccionario Histórico de la Lengua Española*, waarmee in 1948 onder de leiding van Julio Casares een begin werd gemaakt, is in 1960 de eerste aflevering verschenen. Het is opgezet naar het model van het grote Engelse woordenboek van Oxford. Dit laatste, op zijn beurt weer geïnspireerd door het werk der gebroeders Grimm, is echter een typisch produkt van de eenzijdig-historisch gerichte lexicografie der negentiende eeuw, waarbij taalperioden die meer dan duizend jaar uit elkaar liggen, toch als één geheel worden beschouwd. Dit laatste moge dan juist zijn, het is toch alleen maar juist in zuiver historische zin. Synchronisch is het Oudengels een taal die van het moderne Engels minstens evenveel verschilt als het Oudengels van het Oudnederlands. Het is, m.i., niet meer in overeenstemming met de tegenwoordige stand en ontwikkeling der taalwetenschap, historische woordenboeken op touw te zetten waarin de woordenschat van een taal als het ware van de wieg tot het „graf” wordt uitgespreid.

Men zal er goed aan doen, het ene grote negentiende-eeuwse historische woordenboek te vervangen door een reeks van woordenboeken waarin de woordenschat der verschillende taalperioden wordt behandeld¹). Welke deze perioden moeten zijn, is echter niet alleen een kwestie van theorie, doch tevens een aangelegenheid waarin de eisen der praktijk een woordje zullen meespreken. Zo spreekt het vanzelf dat een apart woordenboek voor b.v. elke periode van vijftig jaar, iets dat misschien theoretisch gewenst zou kunnen zijn, in de praktijk onuitvoerbaar is. Men zal dus voor een taal als het Nederlands een stel woordenboeken moeten maken: een of misschien zelfs liever twee voor de tijd vóór 1500 (I. van het oudste Middelnederlands tot 1410 en II. van 1390 tot 1510); een voor de 16de eeuw, een voor de 17de eeuw, een voor de 18de eeuw en een gedeelte der 19de, met een duidelijke grens „post quem non”. Hierdoor wordt de „slechte oneindigheid” der historische woordenboeksondernemingen uit de 19de eeuw, het hopeloze en onbevredigende gevecht met de stortvloed der elke dag verschijnende publikaties tevens vermeden. Deze verschillende historische woordenboeken mogen niet tegen elkaar aan liggen, doch ze moeten elkaar overlappen; zo moet een woordenboek van het Nederlands

¹) Zie ook B. Beckmann, *Das Deutsche Wörterbuch* 133-134 [1954], die verwijst naar uiteenzettingen hierover van W. Craigie [1919] en W. von Wartburg [1939], en verder P. Imbs, *Au seuil de la Lexicographie* 4 [1961].

der zestiende eeuw tevens een of meer decennia van de voorafgaande en van de volgende eeuw omvatten¹⁾.

De woordenschat van de tegenwoordige taal moet uitsluitend synchronisch behandeld worden en gescheiden blijven van de historische woordenboeken. De tegenwoordige taal met haar rijke en schijnbaar onbeperkte, hoewel natuurlijk niet eindeloze woordenschat²⁾, met haar overvloed aan geschreven bronnen, vereist nu eenmaal een andere opzet en een andere behandeling dan een historisch woordenboek³⁾.

De lexicografie der negentiende eeuw, zoals die in Duitsland door de gebroeders Grimm was begonnen, en die in Engeland, Nederland en Scandinavië navolging vond, berustte op de historische taalwetenschap en de triomfen van het taalhistorisch comparatisme. De vernieuwing der lexicografie die zich thans baan breekt, berust op twee factoren: in de eerste plaats op de herwaardering der linguïstiek zoals die van De Saussure is uitgegaan en waardoor men synchronie en diachronie in een duidelijk perspectief is gaan zien; in de tweede plaats op de mechanisering van het lexicologisch apparaat, dat niet langer aan één dimensie, de alfabetische, is gebonden. Is de vernieuwing der lexicografie in de 19de eeuw uitgegaan van Duitsland, in de twintigste daarentegen gaat ze uit van Frankrijk en Frans-Zwitserland, bepaaldelijk van het land aan beide zijden van het Juragebergte; van Genève, waar F. de Saussure doceerde, en van Besançon, waar B. Quemada zijn mechanografisch laboratorium oprichtte, daarmee bewijzend dat een mechanografische thesaurus in de praktijk mogelijk, economisch realiseerbaar en rendabel is. Het is op de door hem gelegde grondslag en langs de door hem gebaande weg dat een groot project der Franse taalwetenschap, de door het „Centre National de la Recherche Scientifique” te Parijs gepatroneerde „Fondation Nationale du Trésor Français”, concrete vormen aan gaat nemen. In Besançon, met de door Quemada geïnaugureerde en aldaar aan de praktijk getoetste methoden, wordt de grondslag gelegd voor de enorme materiaalverzameling, een reusachtige kartotheek, die de Franse thesaurus mogelijk moet maken. T.z.t. zal daarvoor het bescheiden machinepark van het lexicologische „proefstation”, nl. het „laboratoire” van het „Centre pour l’Étude du Vocabulaire Français”, natuurlijk niet kunnen volstaan; een veel en veel groter centrum, ten dele uitgerust met kleine, uiterst snelle elektronische apparatuur, zal daarvoor nodig zijn.

1) „Les „raccords” entre différents fichiers devraient être sérieusement étudiés; dans le premier cas (*la description du seul français moderne*), une „marge de sécurité” comprenant une assez longue période et empiétant sur le moyen français, devrait être laissée, permettant ultérieurement une délimitation, fondée sur les faits, de deux états de langue”, A. J. Greimas, *Problèmes* 71 [1959].

2) L. Országh, *Problems and Principles* 221 [1960].

3) Zie hierover het in noot 2) genoemde werk. Verg. „L’usage contemporain est le premier et principal objet d’un dictionnaire. C’est en effet pour apprendre comment aujourd’hui l’on parle et l’on écrit, qu’un dictionnaire est consulté par chacun”, E. Littré, *Dictionnaire de la langue française, Préface au premier tome* III [1867].

§ 3. GEMECHANISEERDE LEXICOLOGIE

Het CAAL, van pater R. Busa te Gallarate, en het CEVF, onder leiding van Professor B. Quemada te Besançon, zijn tot op heden, bij mijn weten, de enige centra in Europa waar de gemechaniseerde lexicologie het stadium van het experiment achter zich heeft. Hoewel in beide centra teksten uit verschillende talen behandeld werden, houdt men zich te Gallarate toch wel in hoofdzaak bezig met de Latijnse teksten van Thomas van Aquino, terwijl de taak van Besançon geconcentreerd wordt op het tot stand brengen van een documentatiemateriaal van honderden miljoenen ponskaarten, bestemd voor de „Trésor de la langue française”¹⁾.

Jongere centra waar aan gemechaniseerde lexicologie wordt gedaan, zijn nog bescheiden van omvang, en verkeren nog in het experimentele stadium²⁾.

De Akademie der Hebreeuwse Taal te Jeruzalem heeft het plan opgevat een historisch woordenboek samen te stellen van het Hebreeuws. Op het internationale colloquium voor gemechaniseerde lexicologie, gehouden te Tübingen in november 1961³⁾, deelde Professor Z. Ben-Ḥayyim mede, dat hij zich beraadde over de mogelijkheid, over te gaan tot het aanleggen van een mechanografische kartotheek⁴⁾. In de eerste helft van 1961 heeft men te Jeruzalem reeds een proef genomen met de mechanische vervaardiging van indices en concordanties op een kleine Hebreeuwse tekst⁵⁾, geheel volgens de methode die te Gallarate gebruikelijk is⁶⁾.

Bij de „Deutsche Akademie der Wissenschaften” in Berlijn bestaat een „Komitee zur Einführung technischer Hilfsmittel und maschineller Informationsverarbeitung”⁷⁾. Een eigen volledig machinepark had men in september 1961 nog niet; wel had men bij de experimenten de beschikking over een eigen sorteermachine, werkend met kaarten van 90 kolommen (Aritma). Voor grotere taken diende men een beroep te doen op een reken-centrum, dat echter met ponskaarten van 80 kolommen werkte. Een en ander had tot gevolg, dat de projecten op het gebied der automatisering in hoofdzaak alleen nog maar in concept aanwezig waren. Afgezien van de natuurwetenschappen, hadden deze projecten betrekking op documentatie,

1) Zie P. Imbs, *Au seuil de la Lexicographie* 3 [1961].

2) Dit geldt ook voor de elektronische lexicologische centra, die in Hs. III ter sprake zijn gekomen.

3) Zie Z. Ben-Ḥayyim, *The International Meeting on Mechanical Methods of Literary Analysis and Lexicography* [1961].

4) De mogelijkheid daarbij is het stenografische karakter van het oude Hebreeuwse schrift. Zie over een en ander Z. Ben-Ḥayyim, *The historical Dictionary of the Hebrew Language* [1960].

5) Hiervoor werd gekozen *Barayta di-Melekhet Ha-Mishkan*.

6) Zie G. Sarfatti, *The Mechanization of Linguistic Research* [1961].

7) Thans „Komitee für technische Hilfsmittel und maschinelle Informationsverarbeitung”.

bibliografieën, registers, indices, grammatica en woordenboeken¹⁾.

Het *Wörterbuch der deutschen Gegenwartssprache*, dat onder leiding van Ruth Klappenbach en Wolfgang Steinitz bij de Akademie-Verlag in Oost-Berlijn verschijnt, werkt niet met een apparaat van citaten op ponskaarten. Naarmate de artikelen echter klaarkomen, worden er woordkaarten van gemaakt, waarop de bewerker van het artikel de nodige numerieke gegevens aanbrengt, die, volgens een bepaalde code, verschillende taalkundige en zakelijke sferen aanduiden. Hierna worden deze woordkaarten geponst en gesorteerd²⁾.

In het kader van het Tsjechisch Instituut der Akademie te Praag is mevrouw Jitka Štindlová actief in verband met het inventariseren der Tsjechische technische en speciale termen. Voor deze projecten wil zij gebruik maken van de traditionele mechanografische apparatuur.

In Nederland dient ten slotte nog te worden vermeld dat aan het Instituut voor Italiaanse Taal- en Letterkunde der Utrechtse universiteit, op initiatief van M. L. Alinei, in 1960 een begin werd gemaakt met het vervaardigen van mechanische indices op Italiaanse auteurs (o.m. Verga en Dante). Hierbij maakte men oorspronkelijk uitsluitend, en maakt men thans nog in hoofdzaak – sinds 1962 beschikt men namelijk over een eigen ponsmachine – gebruik van de conventionele IBM-ponskaarten-apparatuur van de administratie der universiteit. Het verschijnen van een *Indice inverso della lingua italiana, con indice di frequenza delle terminazioni* werd in 1962 in het vooruitzicht gesteld.

Internationaal gezien is het thans duidelijk, dat in de nabije toekomst in elk land een of meer mechanografische centra zullen ontstaan, met het oog op de lexicologische taken die men zich heeft gesteld. Het is ondenkbaar dat de lexicologie in de verschillende landen, bepaaldelijk de grote lexicologische of lexicografische projecten, in de toekomst aan de mechanisering zouden kunnen voorbijgaan. In 1958³⁾ schreef E. Delavenay in zijn in mei 1959 gepubliceerd boek *La machine à traduire*, dat de mechanische analyse der linguïstische gegevens nog in de kinderschoenen stond⁴⁾, een opvatting die hij een jaar later in de Engelse vertaling van zijn boek herhaalde⁵⁾. Het is echter duidelijk dat zijn uitspraak thans reeds ruimschoots door de feiten is achterhaald.

1) Ik dank deze inlichtingen aan dr. Erich Mater, van het „Institut für deutsche Sprache und Literatur“ (5-IX-1961). Zie ook in *Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin* 6 [1960], Heft 6, de artikelen van E. Mater, H. Geiss en M. Bierwisch.

2) Zie voor een ander R. Klappenbach, *L'emploi des cartes perforées dans le dictionnaire de langue allemande contemporaine* [1960-61], en M. Blumrich en E. Tellenbach, *Die Anwendung von Lochkarten im Wörterbuch der deutschen Gegenwartssprache* [1962].

3) *Machine Translation* 119 [1960].

4) *La machine à traduire* 122 [1959].

5) *Machine Translation* 113 [1960].

In ruil voor een relatief bescheiden nieuwe investering biedt de mechanisering de lexicologische onderzoeker een enorm rendement. De snelle ontwikkeling van de machines en van de elektronische techniek maakt de tegenwoordige realisaties niet waardeloos, aangezien de mechanografische resultaten van vandaag steeds automatisch kunnen worden vertolkt in elke toekomstige codetaal.

§ 4. EEN CENTRUM VOOR LEXICOLOGIE TE LEIDEN?

Nothing could be more erroneous than to assume that the first step in applying automation is selecting the machines. . . . The whole concept of systems analysis and design, basic to automation, requires a careful and detailed plan for the entire organization if the benefits realized are to be more than marginal, *John Diebold Answers Twenty Questions* (10).

In zijn hierboven genoemde lezing heeft Blancquaert de progressieve eis geformuleerd, dat er een „permanente centrale”¹⁾ als verzamelplaats voor de woordenschat van de Nederlandse taal tot stand zou komen.

Nu leeft sinds jaren bij de redactie van het WNT de wens, dat het Leidse scriptorium zou worden omgevormd in een centrum, een centraal instituut voor Nederlandse lexicografie²⁾. Dit verlangen mag zich verheugen in de belangstelling van de Commissie van Bijstand. Na wat voorafgaat, zal het duidelijk zijn, dat het principieel en tactisch de voorkeur verdient te spreken van een „Centrum voor Nederlandse Lexicologie”. Principieel, omdat de bruikbaarheid der mechanografische kaartverzameling niet beperkt is tot de lexicografie, tactisch omdat een dergelijk centrum niet alleen de kleine schare der lexicografen moet interesseren, doch de belangstelling van alle taalkundigen zal wekken.

Wenst men werkelijk binnen afzienbare tijd te komen tot een centraal instituut, een centrum voor de studie van de Nederlandse woordenschat, dan ligt het voor de hand, dat men aan de revolutie die zich in het laatste decennium in de lexicologische techniek heeft voltrokken, niet voorbij zal kunnen gaan. Ja, er is meer: met het al dan niet overnemen van de moderne mechanografische methode staat of valt m.i. het hele instituut; het is eenvoudig een kwestie van zijn of niet-zijn. Zonder de hulp van machines is het gestelde doel een utopie, niet alleen omdat het omvangrijke materiaal van een Nederlandse thesaurus onmogelijk binnen een redelijke tijd en

1) VMKVA 1938, blz. 12 (788) en 17 (793). Vergelijk wat de hoofdredacteur van *Die Afrikaanse Woordeboek*, P. C. Schoonees, schrijft in *Rondom die Woordeboek* 72 [1958]: „Van die alleruiterste belang is . . . die feit dat 'n Woordeboekkantoor 'n permanente instelling behoort te wees waar jonger kragte opgelei kan word om die werk volgens vaste plan voort te sit”. Evenmin als Blancquaert echter scheidt Schoonees het lexicologische materiaal van de lexicografische taak.

2) Zie *WNT* XVIII, vii [1958].

tegen een economisch verantwoord bedrag kan worden tot stand gebracht, doch ook omdat aan de eis der volledigheid niet kan worden voldaan langs de traditionele weg, doch alleen door het toepassen van kwantitatieve methodes.

Waarom zal men de mens trouwens „mechanisch” werk laten verrichten, dat beter, accurater, sneller en goedkoper door machines kan worden gedaan? Voor zover de taak van de assistenten van „zuiver mechanische” aard is, kan ze door de machine worden overgenomen; de lexicograaf zelf wordt bevrijd van lexicologische zorgen, zodat zijn mogelijkheden worden vergroot. De machine, die de onderzoeker onnoemelijk veel zuiver materieel, tijdrovend en weinig interessant werk bespaart, herleidt de lexicografie tot menselijker, intellectueler verhoudingen. Tevens behoeft het buiten alle proporties aanzwellen van de bouwstoffen tijdens de bewerking niet meer plaats te vinden. Het eerste deel en het twintigste deel van een woordenboek kunnen dan berusten op een materiaal dat grosso modo gelijkwaardig is.

Bovendien is er ook nog de factor tijd. Wanneer men met machines in één jaar kan bereiken wat met het oude handwerk twintig of veertig jaar duurt¹⁾, is alleen de mechanische methode in elk opzicht verantwoord. De mechanisering in de lexicologie der twintigste eeuw afwijzen, staat gelijk met het afwijzen van de boekdrukkunst op het eind van de 15de, of, wenst men een meer moderne vergelijking, met het niet-gebruiken van de naaimachine op een modeatelier van deze tijd, of van de elektrische wasmachine in een modern gezin.

Vraagt men zich af, waar het Nederlands Centrum voor Lexicologie eventueel gevestigd zou kunnen of moeten worden, dan ligt het voor de hand in de eerste plaats aan Leiden te denken. Geen andere stad in Nederland immers is in dergelijke mate verbonden met de Nederlandse lexicologie en haar grote monumenten: het *WNT* van De Vries en Te Winkel en het *MNW* – oorspronkelijk eveneens een plan van M. de Vries – van Verwijs en Verdam, waarnaast trouwens, zij het op een ander plan, met ere ook het *Rhetoricaal Glossarium* van Mak mag worden genoemd. De lexicologische traditie zoals die te Leiden is gegroeid, de langdurige lexicografische ervaring waarover men beschikt, het apparaat en het scriptorium van het *WNT* – dat binnenkort ruimer en moderner gehuisvest zal worden –, en ten slotte het feit dat het plan tot oprichting, aanvankelijk van een Centrum voor Nederlandse Lexicografie, naderhand van een Nederlands Centrum voor Lexicologie, in de schoot van de redactie van het *WNT* is gegroeid, dat alles maakt dat Leiden haast vanzelfsprekend in aanmerking komt als de enige plaats waar het Centrum zou kunnen worden gevestigd.

¹⁾ „Je nachdem wie genau ein Ergebnis gewünscht wird, verringert die Methode des Lochens und der Aufstellung von Verzeichnissen die Zeit für die bisher manuell oder mit Vervielfältigungsapparaten durchgeführte Arbeit auf den 20. bis 40. Teil der früheren Bemühungen”, R. Busa, *Elektronentechnik* 21 [1957].

Besluit men tot de oprichting van het Centrum voor Lexicologie te Leiden, dan zal de allereerste opgave bestaan in de verdere studie van de problemen der mechanisering, inzonderheid met het oog op de speciale behoeften van de Nederlandse lexicologie en de mogelijkheden van een brede, internationaal gerichte opzet. Hoe belangrijk het voorbeeld van het CEVF ook moge wezen, in de praktijk zal het Leidse Centrum toch meer moeten zijn dan alleen maar een kopie van het CEVF te Besançon of van het CAAL te Gallarate. De lengte van de Nederlandse woorden alleen al maakt een andere rubricering van de ponskaart noodzakelijk. Het Nederlandse toetsenbord vereist gedeeltelijk andere tekens dan die welke te Besançon worden gebruikt. Men zal zich dienen af te vragen, of men een apparatuur wil die slechts op de behoeften van de Nederlandse lexicologie is afgestemd, dan wel of een machinepark gewenst is dat ook voor Franse, Engelse en Duitse teksten gebruikt kan worden. Wanneer de constructeurs zullen kunnen voldoen aan een op het Tübinger colloquium (1960) naar voren gekomen, en op de „Colloque International” te Besançon (1961) opnieuw besproken wens, namelijk die van een apparatuur beschikkend over een groot aantal tekens en tevens ook met hoofd- en kleine letter¹⁾, dan zou dit probleem vanzelf zijn opgelost. De mechanisering ten behoeve van het Centrum voor Lexicologie vergt in elk geval veel voorbereiding en studie. Herhaald contact met de constructeurs der machines, overleg met en eventueel advies van de reeds over de nodige ervaring beschikkende, gemechaniseerde lexicologische centra is ten zeerste gewenst. Verder is de mogelijkheid van combinatie der elektromechanische met de elektronische machine of computer een studie op zichzelf. Het zou in elk geval hoogst onverantwoord zijn, zonder degelijk en grondig voorbereidend werk tot het aanschaffen van machines te besluiten.

De lexicoloog behoeft niet alles van de machines af te weten. Ten slotte is de constructie, het onderhoud en de eventuele reparatie van de mechanografische apparatuur een zaak van en een taak voor de ingenieurs en de machinebedienden. Wel is het van belang te weten waartoe de machine kan dienen en wat voor gebruik men van de mogelijkheden der machine wenst te maken. Wanneer men precies weet wat men wil, legt men zijn wensen voor aan de ingenieurs, die dan uitmaken of deze in principe te verwezenlijken zijn met de bestaande mechanografische of elektronische apparatuur. Als dat inderdaad het geval blijkt te zijn, kunnen aan bestaande machines wijzigingen worden aangebracht met het oog op de behoeften

1) Verg.: „Still lacking to a practical perfection of the method are: the insufficient number of characters and marks (e.g. the 36 or 47 characters of the normal machines do not permit distinction of upper and lower case letters); the technical difficulties involved in changing type on the same machine (e.g., different alphabets)”, R. Busa, *Darstellung* 5 [1960].

De modernste machines (BULL) hebben 59 tekens; om hoofdletters en kleine letters te hebben zou men er minstens 26 tekens aan toe moeten voegen, wat een ernstig technisch probleem betekent.

der gemechaniseerde lexicologie. Een van de belangrijke resultaten van de internationale colloquia (Tübingen (1960), Besançon (1961)) bestaat hierin, dat de grote fabrieken van mechanografische en elektronische machines thans op de hoogte zijn van de belangstelling en de desiderata van de zijde der lexicologen.

§ 5. *Het LLC*

a) *Taak en doel*

Het doel van de automatische of, zo men wil, mechanografische linguïstische analyse bestaat in het tot stand brengen van een verzameling van mechanografische kaarten, die kan dienen als uitgangspunt van, als grondslag voor lexicologisch onderzoek in de ruimste zin. Toegepast op onze taal betekent dit, dat een centrum in de eerste plaats een lexicologische taak heeft, namelijk het tot stand brengen van een thesaurus van de Nederlandse taal op ponskaarten¹), het verzamelen en op peil houden van een unieke en volledige verzameling van de Nederlandse woordenschat. Deze verzameling, waarlijk „het monument van den Nederlandschen woordenschat in al zijn rijkdom en wetenschappelijke volledigheid”²), moet voor iedereen toegankelijk zijn, duplicaten van onderdelen ervan moeten tegen een geringe prijs geleverd kunnen worden.

Het tot stand brengen van deze permanente verzameling kan in etappes gebeuren. Bepaalde onderdelen, bepaalde kleinere taken kunnen succesievelijk verwezenlijkt worden of, wanneer het machinepark voldoende uitgebreid is, zelfs simultaan. Als vanzelf zullen bepaalde taken, de meest dringende, voorrang krijgen. Zo zou een Nederlands mechanografisch centrum zich bij wijze van aanloop b.v. kunnen bezighouden met het ficheren van de complete woordenschat van de dichterlijke werken van Hooft en Vondel. Ik beperk me met opzet tot de dichterlijke werken, omdat deze praktisch zonder enige voorbereiding, zonder enige verdeling in perikopen, door de typiste op de band kunnen worden geponst. Wanneer men eenmaal alle woorden van onze beide klassieken op kaarten heeft gebracht, en deze alle heeft voorzien van de passende portie context, is men in het bezit van een complete concordantie.

Men zou op grond van dit apparaat een concordantie op Hooft en Vondel kunnen uitgeven, maar dat is minder interessant. Interessanter zou een woordenboek zijn op beide dichters³). Welnu het concordantie-apparaat kan de grondslag voor lexica zijn. Hierbij worden de reeks

1) Verg.: „Het groot modern Nederlandsch woordenboek zal dus geen boek meer zijn, zal geen reeks boeken meer zijn, maar een verzameling van gedrukte steekkaarten”, Blancquaert, *Het Nederlandsch Woordenboek van de twintigste Eeuw* 18 (794) [1938].

2) Blancquaert, *Vlaamsch Neerlandicus op den uitkijk* 11 (1053) [1935].

3) Zie over de „lexicologie littéraire”: H. Mitterand en J. Petit, *Index et concordances dans l'étude des textes littéraires* 167 [1962].

contexten van een bepaald woord, b.v. *verbergen*, die immers bij een concordantie in de volgorde van de tekst liggen, door de lexicograaf gewijzigd in een classificatie naar semantische criteria, daarna worden de definities zelf op kopjeskaarten („cartes-vedettes”) geponst en vertolkt, waarna de tabelleermachine een proef van het woordenboek drukt. Na correctie, die uiteraard alleen maar de kopjeskaarten of de volgorde der citaten betreft — zetfouten zijn immers uitgesloten —, kan een definitief exemplaar van het woordenboek worden vervaardigd, dat dan, na fotografische reductie tot het gewenste formaat, als offsetdruk kan worden uitgegeven. Als men weet dat het totale aantal woorden van Vondel in de editie van de W.B. ongeveer 1.600.000 bedraagt, terwijl Hooft in zijn *Ned. Hist.* ongeveer 650.000 en in zijn *Baeto* ongeveer 13.000 woorden gebruikt¹⁾, dan heeft men een idee van de omvang van het concordantie-apparaat dat aan een woordenboek op Vondel of Hooft ten grondslag zou moeten liggen.

Een andere dringende taak voor de Nederlandse lexicologie is het tot stand brengen van een woordinventaris van de zestiende eeuw. Dat is immers de eeuw die tussen twee lexicografische stoelen, het *MNW* en het *WNT*, in de as zit. Na het gereedkomen van het *WNT* zal er allereerst behoefte bestaan aan een woordenboek van het Nederlands der zestiende eeuw, om hiermee de gaping tussen beide grote werken op te vullen.

De noodzaak van een supplement op het *MNW* doet zich hoe langer hoe meer gevoelen. Verdam zelf had een supplement willen maken, en bij de voltooiing van het werk heeft J. W. Muller nog eens op de wenselijkheid ervan gewezen²⁾. Even dringend bijna als een supplement is een nieuw Middelnederlands woordenboek, berustend op betere tekstuitgaven, minder eenzijdig gericht op literaire bronnen en zonder het onoverzichtelijke en thesaurusachtige karakter dat de artikelen van het *MNW* kenmerkt. Maar aangezien één lexicografische vogel in de hand per slot van rekening beter is dan tien in de lucht, zou, op Middelnederlands gebied, de taak van het centrum allereerst kunnen bestaan in het mechanografisch verwerken, ficheren of exciperen van Middelnederlandse publikaties die niet in het *MNW* tot hun recht komen, en van die welke verschenen zijn na het voltooiën van het grote werk. Hierbij zou ook te overwegen zijn, in welke vorm het in de zgn. „loketkast” van Verdam aanwezige supplement op het *MNW* in de nieuwe verzameling zou zijn te integreren. Aangezien deze kast weinig anders bevat dan vindplaatsen zonder context, zou de meest praktische werkwijze waarschijnlijk hierin bestaan, dat men deze vindplaatsen achter elkaar op de band ponst, waarna er woordkaarten met referenties van gemaakt kunnen worden. Men zou dit materiaal natuurlijk ook in de vorm van filmstroken in ponskaarten kunnen opnemen, doch

¹⁾ Deze berekening ben ik verschuldigd aan de heer A. H. Goedhart, assistent bij het *WNT*.

²⁾ *MNW* IX, xxxii [1929].

daarvoor is de waarde ervan waarschijnlijk niet groot genoeg. Zijn eenmaal de hierboven genoemde Middelnederlandse publikaties geficheerd of geëxcerpeerd, dan kan worden overwogen of de bewerking en de uitgave van een supplement op het *MNW* nog wel gewenst is, en of men niet liever aan een nieuw Middelnederlands woordenboek moet beginnen. Dit laatste, dat het thesauruskarakter van Verdam kan missen, doch aan de andere kant ook een grotere woordenschat zou bestrijken, behoeft m.i. niet uitvoeriger te worden dan het bestaande *MNW*¹⁾.

Een andere taak van het centrum zou bestaan in het excerpereen voor het *WNT*. Nieuw materiaal voor het *WNT* zou voortaan uitsluitend langs mechanografische weg kunnen worden aangebracht.

Het centrum zou verder ook bijzondere diensten kunnen bewijzen aan de Nederlandse filologie in de engere zin. Wie een oude Nederlandse tekst wil uitgeven, b.v. uit de 16de of 17de eeuw, zou bij het centrum terecht kunnen voor het mechanografisch bewerken van een woordindex. Wanneer men weet dat een typiste 10.000 woorden per dag in de band ponst, dan begrijpt men dat het maken van een volledige of gedeeltelijke (b.v. zonder de zgn. grammaticaal-technische woorden) woordindex binnen het economische bereik van elke tekstuitgever ligt. Het centrum zou dan meteen voor eigen gebruik een duplicaat van de kaartverzameling kunnen laten vervaardigen.

Ook bij de verzorging van tekstuitgaven op het gebied van andere talen zou het centrum belangrijke diensten kunnen bewijzen. Zo zou het Leidse instituut tegen passende vergoeding, b.v. ten behoeve van een Nederlandse romanist, een index kunnen maken op een Franse tekstuitgave. De band waarmee de index werd vervaardigd, zou bovendien ter beschikking van romanistische centra in het buitenland gesteld kunnen worden, zodat ook deze er hun profijt mee kunnen doen. Bij een lexicologisch centrum dat zo ruim is opgezet, dat het b.v. ook de belangstelling van theologen en kunsthistorici kan wekken, is het aantal geïnteresseerden natuurlijk groter dan bij een centrum dat zuiver en alleen op de behoeften der Nederlandse taalkunde is berekend; de kansen dat het centrum er inderdaad komt, worden er uiteraard ook groter door. Tegenover één investering staat dus een veelheid van mogelijkheden.

b) *Excerperen of ficheren?*

De vraag kan worden gesteld, in hoeverre een volledige lexicografische kaartverzameling noodzakelijk is. Is het nodig, bij het aanleggen van een mechanografische kartotheek, geen enkel woord over te slaan, doch integendeel van elk apart woord in de tekst ook een aparte woordkaart te maken? Moet men m.a.w. de tekst in zijn geheel, integraal ficheren, of kan men volstaan met excerpereen, dus met het maken van een zekere keus?

¹⁾ Principieel kan men trouwens de vraag stellen of men wel tot één Middelnederlands Woordenboek moet komen, danwel of verdeling over twee aparte werken wenselijk is. Zie hierover blz. 108.

Voor oudere perioden met een afgesloten en niet te omvangrijke produktie, laten we zeggen voor Middelnederlandse teksten, voor belangrijke werken of voor klassieke auteurs, b.v. Vondel en Hooft, is in de regel een integrale fichering, waarbij dus geen enkel woord wordt overgeslagen, inderdaad gewenst. Immers het excerpereen is een zeer subjectief en onvolledig procédé, het toeval speelt een veel te grote rol en het resultaat is niet veel meer dan een pluk. Zo kon het gebeuren dat het oudste citaat van *Venter*, dat in deel XVIII, kol. 1827 van het *WNT* wordt vermeld, niet ouder is dan 1841, terwijl bij Six van Chandelier reeds een plaats van 1657 te vinden is¹). Voor minder belangrijke werken, vooral uit latere tijd, voor latere perioden waarin de produktie zeer groot is, en vooral voor perioden die nog niet zijn afgesloten, zal men zich noodgedwongen hebben te beperken tot excerpereen, omdat de teksten dan zo talrijk en overvloedig zijn, dat een ficheren „ad absurdum” zou leiden. Immers het kaartmateriaal zou dan zo ontstellend groot worden, dat het niet meer toegankelijk, dus onbruikbaar en steriel zou worden en bovendien economisch niet verantwoord zou zijn. Men stelle zich even de omvang voor van een mechanografische kartotheek waarin alleen maar de jaarlijkse produktie van de Nederlandse journalistiek zou zijn opgenomen! Naast de fichering is dus eveneens de excerpereing niet alleen gewenst, doch eenvoudig geboden. Bij dit excerpereen, dat een keuze veronderstelt, zijn tal van variaties mogelijk. Zo kan men zich b.v. beperken tot het opnemen van de autosemantische woorden (substantiva, adjectiva, verba en de meeste adverbia) en de synsemantische woorden²) (voorzetsels, voornaamwoorden, lidwoorden, voegwoorden, hulpwerkwoorden en telwoorden) verwaarlozen. Dit procédé garandeert een besparing van minstens 50 %. Hierbij dient men bij het prepareren van de tekst die woorden te onderstrepen, die men in de vorm van woordkaarten in de kartotheek wenst opgenomen te zien. In beide gevallen zal de typiste toch een bandje moeten ponsen waarin de tekst integraal is opgenomen, doch dat tevens is voorzien van codetekens die aangeven welke gedeelten door de machine bij het maken van de woordkaarten kunnen worden overgeslagen³). Bij dit procédé krijgt men toch

¹) Zie *WNT* XVIII, blz. vi [1958].

²) Zie voor deze terminologie, die aan een reeds op Aristoteles teruggaande dichotomie beantwoordt: S. Ullmann, *The Principles of Semantics* 58 [1951] en *Semantics* 44 [1962]. De terminologie *full words* tgv. *form-words* is afkomstig van H. Sweet. L. Országh, *Problems and Principles* 246 [1960] stelt *content words* tgv. *form-words* (logischer zou hier zijn: *empty words*). In het Frans spreekt men van: *mots pleins*: m. *vides*; *mots notionnels*: m. *outils*; *mots lexicographiques*: m. *grammaticaux*; *mots de signification*: m. *de structure*; *mots autonomes*: m. *accessoires*. Zie S. Ullmann, *Précis de sémantique française* 84 volg. [1952]. Kukenheim noemt de synsemantische woorden „grammaticaal- of taaltechnische woorden” (*Van Glosarium tot Thesaurus* 22 [1960]).

³) Wanneer men gebruik maakt van de elektronische rekenmachine, kan het uitschieten der synsemantische woorden automatisch gebeuren, namelijk door alle woorden van de tekst te vergelijken met een in het geheugen der rekenmachine op-

tekstkaarten van de volledige tekst, zodat het aanbrengen van de context op de woordkaarten mogelijk wordt. Het exciperen kan natuurlijk nog veel beperkter zijn, als de ponstypiste slechts een zeer beperkt aantal woorden uit een bepaalde tekst behoeft op te nemen; in dit geval wordt natuurlijk niet de gehele tekst op tekstkaarten gebracht, doch alleen de zinnen waarin de bedoelde woorden voorkomen.

Een andere begrenzing, van zuiver praktische aard, zal m.i. hierin bestaan dat men in principe alleen maar gedrukte, dus toegankelijke en door anderen controleerbare teksten verwerkt¹⁾.

c) *Enkele of dubbele kaartenverzameling?*

De vraag is gesteld, of men bij een lexicologisch centrum niet een dubbele kaartenverzameling moet bezitten: een waarbij alle kaarten, zonder onderscheid van schrijver en tijd, in alfabetische volgorde gerangschikt zijn, een tweede waarbij alle kaarten per werk alfabetisch worden opgeborgen, met dien verstande dat de verschillende werken van een en dezelfde auteur bij elkaar staan²⁾. Voor de gemechaniseerde lexicologie is het probleem minder belangrijk, omdat men de woordenschat van elk werk en van elke auteur gemakkelijk machinaal uit de kartotheek kan lichten. Hoe groter de kartotheek echter wordt, hoe langer het mechanisch sorteren zal duren. Daarom zou het, ook bij een gemechaniseerde ponskaartenkartotheek van grote omvang, een niet te verwaarlozen gemak zijn als men over een dubbel exemplaar kon beschikken, opgesteld op de hierboven aangegeven manier. Wat de kosten en het werk betreft, is het dupliceren der kaarten geen probleem; het enige werkelijke probleem is dat van de ruimte, het opbergen.

Wanneer men een bepaald gedeelte der kaartenverzameling voor het maken van een woordenboek gaat gebruiken, verdient het misschien aanbeveling ook van dat gedeelte een duplicaat te vervaardigen. Immers als ponskaarten niet met zorg behandeld worden zodat er beschadigingen optreden, gaan ze in de kaartenbaan vastlopen en scheuren. Men kan er gerust van uitgaan, dat ponskaarten die door de handen van de redacteur, de typiste en een of meer correctoren moeten gaan, minder geschikt zullen zijn voor mechanografische behandeling.

d) *Woordstatistiek*

Dat statistisch onderzoek van de taal, ook wel lexicometrie genoemd,

geslagen lijst van synsemantische woorden; zie B. Quemada, *L'Inventaire lexicographique* 125 en 128 [1962]. Aangezien deze woordenlijst van tevoren moet worden opgesteld, zal dit automatisch uitschieten m.i. slechts mogelijk zijn bij moderne, streng genormaliseerde teksten.

¹⁾ Verg. „Grundsätzlich wurden nur gedruckte Texte in das Programm einbezogen, da die Exzerpierung von Handschriften zu zeitraubend gewesen wäre“, O. Prinz, „*Mittelateinisches Wörterbuch*“ *Entstehung und Anlage* 297 [1960]. Zie echter B. Quemada, *L'Inventaire lexicographique* 120 [1962].

²⁾ Z. Ben-Hayyim, *The historical Dictionary of the Hebrew Language* 16 [1960].

kan bijdragen tot de kennis en het begrip der taalverschijnselen, is wel nauwelijks aan twijfel onderhevig. De Franse linguïst P. Guiraud verwacht van de taalstatistiek, dat ze de vaagheid van veel taalkundige conclusies zal kunnen vervangen door exacte metingen en gegevens¹⁾. Men vraagt zich trouwens af, waarom het meten in de taalkunde niet evenzeer op zijn plaats zou zijn als in de economische, juridische, sociologische en psychologische wetenschappen.

Taalstatistiek die deze naam verdient, wordt echter nog maar sinds kort bedreven. Wel werden in de negentiende eeuw reeds woorden opgeteld, doch statistisch interpreteren omvat veel meer dan alleen maar optellen²⁾. Wat op het gebied van het woordfrequentieonderzoek van onze taal werd gepresteerd, beperkt zich tot een publikatie over het onmisbare basisvocabulary van het Nederlands³⁾.

Dat de taalstatistiek geheel als traditioneel handwerk en dus bij voorkeur aan taalmateriaal van geringe omvang werd beoefend, verklaart voldoende dat ze niet de gewenste vlucht kon nemen. Dit is thans anders geworden. De traditionele mechanografische apparatuur en vooral de elektronische rekenmachine openen op dit gebied grote perspectieven. Men kan dan ook gerust zeggen, dat de ontwikkeling der taalstatistiek geheel afhankelijk is van de mate waarin van mechanische hulpmiddelen gebruik zal worden gemaakt⁴⁾.

Wanneer men in een instituut voor mechanisch-lexicologische analyse teksten in hun geheel, woord voor woord, op kaarten gaat brengen, dan lijkt het gewenst daarbij ook aandacht te besteden aan de statistische verwerking van het verkregen materiaal. De bedoeling hierbij is de redacteur van het woordenboek niet alleen gecontexteerde woordkaarten te verschaffen die beantwoorden aan de traditionele normen, doch hem bovendien de nodige statistische hulpmiddelen aan de hand te doen die de klassieke lexicologie en lexicografie nooit hebben gekend⁵⁾. Dank zij de machine nu kan het ficheren van een bepaalde tekst voortaan gepaard

1) P. Guiraud, *Problèmes et méthodes de la statistique linguistique* 22 volg. [1959].

2) Zie P. Guiraud, *Bibliographie critique de la statistique linguistique* 1 [1954].

3) *De meest voorkomende Woorden en Woordcombinaties in het Nederlandsch* [1937]. Men zie ook een schema van woordfrequentie in J. L. van Soest, *Informatie-theorie en Communicatie-theorie*, fig. 5 [1955].

4) „Mechanization is required because the collecting, grouping, comparing, correlating, and counting the elements of the language is a time consuming task not suitable to manual methods and certainly not on a very large scale. Mechanization is required for accuracy and flexibility. Any manual collection of statistical analyses of linguistic analysis may not easily be checked and verified by others. This is not so in mechanization. Through mechanization, the text material becomes explorable by machine”, P. Tasman, *Literature and Document Research Automation* 64-65 [1958]. Verg. ook G. Gougenheim, *Statistique linguistique* 32 [1961]: „Les dépouillements lexicographiques effectués à l'aide de machines vont modifier la situation du tout au tout. Dans peu d'années des index complets de textes étendus vont être publiés”.

5) Zie B. Quemada, *L'Inventaire lexicographique* 120 [1962].

gaan met een frequentieonderzoek der woordvormen; het vervelende telwerk dat hieraan vastzit, wordt geheel door de machine verricht. Dit betekent dat men gegevens over de frequentie der vormen met weinig moeite in de schoot geworpen kan krijgen. In deze omstandigheden zou het niet verantwoord zijn, de taalkundige statistiek, in ons geval dan de statistiek der woorden, althans der woordvormen, te verwaarlozen. In dit verband dient erop te worden gewezen, dat men, met P. Guiraud, de frequentie mag beschouwen als een onderdeel van de woorddefinitie¹⁾. Terecht schrijft P. Tasman: „The objectives in literary data processing require that due consideration be given to the statistical treatment of the language data. In fact, linguistics always require some degree of statistical classification”²⁾.

e) *Organisatie*

Installations should be planned as systems, realistically evaluated, and put „on stream” with a minimum of productive shock, J. Harrington, *Summation of the Conference on Automation Systems* 173 [1958].

Wat de organisatorische vorm van het Leids Lexicologisch Centrum betreft, lijkt een geheel zelfstandig instituut, zo mogelijk in nauw contact met de Leidse universiteitsbibliotheek, de meest aangewezen vorm van organisatie.

Het Leids Lexicologisch Centrum zou moeten omvatten:

- I. Een afdeling voor lexicologische analyse, die zorgt voor het mechanisch tot stand komen van lexicologische bouwstoffen (zgn. Thesaurus).
- II. Een afdeling voor lexicografie, waaronder ressorteren:
 1. Voorlopig alleen het *WNT* dat, als zelfstandig onderdeel van het Centrum, in hoofdzaak op de traditionele manier zou worden voltooid. Alle verzamelingen van het Centrum zouden uiteraard ook ten dienste staan van het *WNT* en omgekeerd.
 2. In de toekomst, b.v. na de voltooiing van het *WNT*:
 - a) Het nieuwe woordenboek van het Middelnederlands. Met het oog hierop zou de afdeling voor lexicologische analyse op korte termijn een aanvullende kartotheek op het *MNW* tot stand kunnen brengen.
 - b) Het zestiende-eeuwse woordenboek.
 - c) Het zeventiende-eeuwse woordenboek.

De personeelsbezetting van de „Thesaurus” zou in de eerste tijd vrij bescheiden kunnen zijn. Men zou b.v. kunnen denken aan twee typistes voor het ponsen en controleponsen, en aan een of meer machinebedien-

¹⁾ *Problèmes et méthodes de la statistique linguistique* 29 [1959].

²⁾ *Literature and Document Research Automation* 62 [1958]; *Index and Concordance* 142 [1961].

den. Voor dit personeel is een vorming aan te bevelen bij de fabriek die de machines heeft geleverd.

Verder zou men binnen niet al te lange tijd de beschikking dienen te hebben over lexicologisch geschoolde assistenten die de teksten voor de onstypiste prepareren, b.v. de woorden aanstrepen die men opnemen wil, sommige teksten in perikopen verdelen met het oog op de tekstkaart enz.; zij zouden zich ook met de correctie der getabelleerde woordenlijsten hebben bezig te houden. Ook de medewerking van een taalkundige met de nodige mathematisch-statistische vorming is dringend gewenst ¹⁾.

¹⁾ „Dans une équipe de lexicologie, il faut avoir un mathématicien”, schrijft J. Dubois in *Cah. de Lex.* 3, 158 [1962].

BESLUIT

Tegenover de automatisering in de lexicologie zijn drie houdingen denkbaar¹⁾. In de eerste plaats een afwerende houding, berustend op het postulaat der onverenigbaarheid van geesteswetenschappen en automatisering. Hierbij gaat men er a priori van uit, dat automatisering zou neerkomen op deshumanisering van de humanistische vakken²⁾. De techniek gevoelt men als een bedreiging, en de vernuftige uitvindingen der ingenieurs ziet men als een leger van gevaarlijke robots. Als twee Britse theologen, Professor G. H. C. Macgregor van de universiteit van Glasgow en zijn leerling Andrew Morton, met behulp van de rekenautomaat het evangelie van Johannes formeel ontleden, en in het aldus gerangschikte materiaal het statistisch bewijs menen te ontdekken, dat dit evangelie uit twee stilistisch verschillende gedeelten bestaat en dus ook door twee auteurs geschreven is, dan geeft dat aanleiding tot zuiver emotionele reacties als de volgende: „Eine Theologie der Maschinen ist ein Widerspruch in sich selbst. Tasten weg von den Dokumenten des Christentums!“³⁾. In zijn opwinding vergeet men echter dat de rekenmachine, als ze tenminste deugdelijke gegevens krijgt ingevoerd en deze volgens een logisch sluitend programma heeft verwerkt, alleen maar objectief materiaal uitvoert, dat weliswaar anders geordend, doch in wezen gelijk is aan wat werd ingevoerd. Met de conclusies die theologen uit deze objectief statistische gegevens willen trekken en met de manier waarop ze deze conclusies aan de man willen brengen, heeft de rekenmachine als zodanig natuurlijk niets te maken; de interpretatie immers blijft volkomen een zaak van de denkende mens.

Een tweede geesteshouding wortelt in een soort van blind, mysticistisch geloof in de almacht der machine. Men gaat zich verbeelden dat een druk op de knop ons de schitterendste resultaten in de schoot werpt. Nu neemt de machine ons weliswaar een groot gedeelte van het vervelende en weinig interessante werk uit handen, maar daarmee zijn we er nog lang niet. Voor het zover is, dient men zich meestal heel wat en dan nog zelfs niet uitsluitend geestelijke inspanning te getroosten. Als men leest dat een bepaalde rekenautomaat de 15.000 woorden van een tragedie van Racine in 4 minuten kan lezen, in 3 minuten alfabetisch sorteren, en er in 5 minuten een alfabetische lijst van kan afdrukken, d.w.z. alles bij elkaar: een index

1) Zie ook F. de Tollenaere, *Automatisering in dienst van de taalkunde* 106 [1962].

2) Zie K. Baldinger, *Automation und Lexikologie* 546 [1959].

3) *Die Christengemeinschaft* 33, 319a [1961]. Nog fraaier is de reactie waarmee John W. Ellison, toen hij een elektronische concordantie op de bijbel bewerkte, werd bedacht in de *Boston Post* van 3 juni 1956: „The Bible as a best seller has been subjected to every kind of indignity. Now it is being consumed by a machine to save the human mind the necessity of pondering its sacred mysteries”.

in minder dan een kwartier¹⁾, dan mag men vooral niet vergeten dat de daaraan voorafgaande filologische en technische voorbereiding van het project heel wat tijd vereist. Een weloverwogen, logisch doordachte analyse van het gestelde probleem komt steeds op de allereerste plaats; de machine immers gedooft noch slordigheden noch vergissingen²⁾; ze is bovendien slechts in zoverre een kostbaar hulpmiddel als er een denkend mens achter staat³⁾. Verder vergt een woordindex die deze naam waardig is — en dat is meer dan een dom mechanische vormindex —, nog heel wat bijkomstig hand- en hersenwerk tussen de bedrijven van de machine door.

Tussen beide uitersten is nog een derde houding mogelijk. Het is die waarbij men zich nuchter en zakelijk de vraag stelt: wat heeft de automatisering aan de lexicologie te bieden? Wie deze laatste geesteshouding tot de zijne maakt, is niet bang voor de machine, doch evenmin blind voor haar beperkingen. Nuchter wil ook zeggen: zich terdege bewust dat de resultaten ongetwijfeld kwalitatief zullen toenemen met de stijging van de kwantiteit van het feitenmateriaal, doch niet proportioneel⁴⁾.

De feiten zijn intussen duidelijk: in toenemende mate wordt de automatisering te hulp geroepen op verschillende taalkundige gebieden, niet het minst op het gebied der lexicologie in de ruimere⁵⁾ of engere zin. Wanneer de Franse linguïst G. Matoré schrijft: „C'est seulement quand la lexicologie aura réalisé de grands progrès que des ponts pourront être jetés utilement entre elle et les sciences mathématiques et physiques”⁶⁾, dan vraag ik me af, of we de volgorde in deze uitspraak niet precies moeten omkeren. De lexicologie zal slechts dan grote vooruitgang kunnen boeken, als ze, de brug slaande naar de mathematische logica⁷⁾ en de toegepaste natuurkunde, gebruik wil en weet te maken van de vernuftige apparatuur die door de ingenieurs werd uitgedacht en vervaardigd, een apparatuur die weliswaar niet voor de taalkunde en de taalkundigen werd geconstrueerd, maar die, met enige moeite en goede wil, toch door de taalkundigen met succes kan worden gebruikt. Een grote vooruitgang, ja een echte revolutionaire doorbraak in de lexicologische methode is slechts te bereiken door wat S. M. Parrish, in zijn concordantie op de gedichten van Matthew

1) F. de Tollenaere, Het „*Colloque international sur la mécanisation des recherches lexicologiques*” 12 [1961].

2) E. Garfield, *Printed Indexes* 76b [1955].

3) „Alle Methoden und Maschinen (bleiben) nach wie vor stets nur Werkzeug und Hilfsmittel . . . hinter denen der denkende und fühlende Mensch stehen muss, wenn sie ihren Sinn behalten und ihren Zweck erfüllen sollen”, M. Scheele, *Lochkartenverfahren* v [1954].

4) K. Baldinger, *Automation und Lexikologie* 541 volg. [1959]. Vergelijk de uitspraak van E. R. Vincent, *Mechanical Aids* 59 [1955]: „although quantity cannot explain quality, it can at least never be harmful in questions of style, as in other matters, to know as much as possible about what we are talking”.

5) Zie voor deze definitie van de lexicologie, hierboven blz. 13.

6) *La méthode en Lexicologie* 48 [1953].

7) Verg. P. Tasman, *Index and Concordance* 144 [1961]: „In fact, we are transposing prose into a series of mathematical relationships”.

Arnold, heeft genoemd „The wedding of literature and technology”¹⁾).

Bij dit „huwelijk” zal men voorlopig, misschien zelfs nog wel de eerste tien jaar, zo niet langer, zijn aangewezen op „huisraad” dat werd ontworpen met het oog op de behoeften van de technische partner alleen. Voor beide deelgenoten is het gebrek aan standaardisering op het gebied der mechanische hulpmiddelen een handicap, voor de lexicologische partner echter is vooral het nijpende tekort aan voldoende alfanumerieke tekens een grote hinderpaal. Het wachten op „de volmaakte machine” zou echter even irreëel als onvruchtbaar zijn.

De grote belangstelling waarin de mechanische vertaling zich mag verheugen, de energieke voortvarendheid waarmee de problemen die ermee samenhangen worden onderzocht, en de grote bedragen die ervoor worden uitgetrokken, wekken de hoop, dat de stem der lexicologen meer en meer tot de ontwerpers en constructeurs zal doordringen, zodat ze meer rekening zullen gaan houden met de wensen en verlangens die van taalkundige zijde naar voren komen. Wie de overtuiging is toegedaan, dat de geautomatiseerde lexicologie in de toekomst meer en meer op de elektronische rekenmachine zal zijn aangewezen, ziet de uiterst snelle ontwikkeling en voortdurende vervolmaking van de computer als bijzonder hoopvol.

Reeds in 1952, in de inaugurele rede gehouden ter gelegenheid van zijn benoeming tot bijzonder hoogleraar aan de universiteit van Amsterdam, schreef Dr. Ir. A. van Wijngaarden: „Het wonderinstrument in de handen van de rekenaar heeft hem . . . het centrum gemaakt van een gezelschap van allerhande personen, vanaf de boekhouder tot de taalkundige, die zich verdringen om zijn gunst”²⁾. Nu ziet het er bepaald niet naar uit, dat dit zich verdringen in de toekomst minder zal worden, noch dat de lexicoloog hierbij verstek zou laten gaan, integendeel. Als gevolg van de automatisering zal het lexicologisch instituut van de toekomst, althans in figuurlijke zin, zijn natuurlijke plaats krijgen in de buurt van het rekeninstituut, en dat misschien niet alleen in figuurlijke, doch zelfs in letterlijke zin. Bij dit laatste denkt men aan een voor de hand liggend gemeenschappelijk gebruik van de apparatuur.

Ook dit verschijnsel is typisch voor de wetenschappelijke ontwikkeling in deze tijd. Lijnrecht tegenover het ononderbroken voortschrijdende proces der specialisering staat thans het merkwaardige fenomeen, dat vertegenwoordigers van zulke uiteenlopende richtingen als b.v. taalkunde, techniek en wiskunde gemeenschappelijk bepaalde problemen bespreken en tot een oplossing brengen. De automatisering gaat in deze tijd tot op zekere hoogte een verbindende rol spelen, die men zou kunnen vergelijken met die welke door de middeleeuwse filosofie in het verleden werd vervuld³⁾.

¹⁾ *Concordance M. Arnold* xx [1959].

²⁾ *Rekenen en Vertalen* 11 [1952].

³⁾ Zie F. de Tollenaere, in *Inf.* nr. 16, 13b [1961]. Ook M. Pacifico, *I nuovi Gutenberg* 103a [1961] en E. Mater, *Einfluss der Technik* 26 [1962] wijzen op dit verschijnsel.

„The object of automation”, schrijft L. Landon Goodman¹⁾, „is to make the best possible use of available resources: man, material, money and machine”. Als we deze uitspraak betrekken op de Nederlandse lexicologie, of ruimer nog, op de lexicologie in Nederland of in de Nederlanden, dan kan men poneren dat het wellicht niet uitgesloten moet worden geacht dat de bevoegde instanties de nodige geldmiddelen ter beschikking zullen willen stellen van een groots project: een instituut voor gemechaniseerde lexicologie. Maar zelfs dan blijft de aanschaf der machines meer dan alleen maar een kwestie van geld, aangezien er in ieder geval een zekere aanpassing van de apparatuur aan de gestelde doeleinden is vereist, iets waarvoor geen enkele fabrikant bijzonder enthousiast is.

Het materiaal, d.w.z. de teksten die door de machine moeten worden omgezet in een taal die de machine kan lezen²⁾, en de informatiedragers waarmee dit moet gebeuren (ponskaarten, ponsbanden, magnetische banden en derg.), vormen in den beginne nauwelijks een probleem. Op den duur echter kunnen het opbergen en het eventuele transport van dit materiaal — men denke b.v. aan een kartotheek van tientallen miljoenen ponskaarten — een ernstig probleem worden. Als het geld er eenmaal is, en de machines een reusachtig materiaal moeten gaan verwerken, dan blijft nog de belangrijkste kwestie. Ook, of liever zelfs, ja vooral bij de automatisering kan men zeggen dat geen enkele machine knapper is dan de mens die erbij staat. De uitspraak: „no computer is smarter than the man running it”³⁾ geldt mutatis mutandis ook voor de conventionele apparatuur. De beste apparatuur kan geen waardevolle resultaten bereiken, als de mensen die haar moeten bedienen niet berekend zijn op de taak die wacht. De geautomatiseerde lexicologie stelt aan haar beoefenaars strengere eisen dan de traditionele lexicologie. Van hen die de machine laten functioneren wordt verwacht: een volkomen openheid ten opzichte van gemaakte fouten en het maximum aan acribie dat, menselijk gesproken, bereikbaar is. Immers voor de automatisering in het algemeen, en stellig niet minder voor de automatisering in de lexicologie, staat onomstotelijk vast de fundamentele waarheid van de in Gallarate in tekst- en woordkaart vastgelegde uitspraak van Thomas van Aquino: *quia parvus error in principio magnus est in fine*.

Aangezien het er voorlopig niet naar uitziet dat het lexicologisch werk in zijn geheel, van A tot Z, volkomen geautomatiseerd zou kunnen worden, en aangezien geen enkele machine tot vrij, creatief denken in staat is — de machine kan en moet geprogrammeerd worden, de mens niet⁴⁾ —, behoeft men volstrekt niet bang te zijn dat de lexicologie haar menselijk aspect zal verliezen. Wel is het waarschijnlijk dat in de toekomst haar

1) *Man and Automation* 26 [1957].

2) P. Tasman, *Index and Concordance* 143 [1961].

3) Zie J. W. Ellison, *Can a computer write a book?* 4 [1956].

4) S. M. Parrish, *Computer Concordances* 11 [1962] schrijft: „machines are easier to program than men”.

traditioneel, uitsluitend filologisch karakter minder op de voorgrond zal staan, en dat ze zich in toenemende mate zal ontwikkelen tot een linguïstische discipline met een sterk technische en mathematische inslag. Sommigen zullen dit betreuren, anderen toejuichen, allen echter zullen rekening hebben te houden met een evolutie die niet is tegen te houden en die perspectieven opent met vele voorheen ongekende mogelijkheden.

SUMMARY

The introduction deals with the definitions of the terms *lexicography* and *lexicology*. As a rule the first one is used to denote the making of dictionaries, although the latter also has been used in this acceptation. A good and generally accepted definition of the word lexicology however is still lacking. Sometimes it was used to denote a separate part of grammar, viz. semantics and etymology, or semantics and morphology. The French scholar Matoré even considers lexicology as a branch of sociology, not of linguistics.

It would be desirable to attach to lexicology the general sense of theory of the vocabulary, including of course both semantics and morphology, word-formation, etymology, synonymics etc., and to reserve the word lexicography for the making of dictionaries. To denote the theoretical foundations of dictionary-making it is preferable to speak of theoretical lexicography.

Lexicography is never possible unless the necessary preliminary work, the collecting of the material to be made use of by the lexicographer, is done by lexicology. Until recently the processing of linguistic data for the making of dictionaries was done entirely in the traditional, manual way. Pioneers in this field, however, became convinced that the amount of linguistic material had grown too fast to be mastered any longer by the traditional devices. They looked forward to data processing machines, both conventional and electronic, as allies. Automation had come to lexicology, as was already the case in many fields of science and scholarship. In this field this amounts to a revolution, as great as the one which occurred towards the end of the fifteenth century: the invention of printing. The apprehension that automation would mean the end of the human side of the humanities is without foundation, as the intellectual aspect of the work, thinking, planning and programming rests entirely on man, whereas the machine gets only the tiresome, irksome and boring work that requires neither fantasy nor intellect. The necessary precision, exactitude and completeness in future lexicology are only possible through the application of quantitative methods. That is why lexicological automation will continue to advance at a steadily faster pace.

Chapter I is devoted to the „Centre d'Étude du Vocabulaire Français" at Besançon (France).

The first paragraph deals with the origin, organization and objectives of the Center. At the colloquium on lexicology, held at Strasbourg in 1957, professor B. Quemada delivered a lecture on the possibility of mechanizing lexicological inventories, and announced that a center for treating French literary information was being established at Besançon University with

the aim of bringing about a general and complete mechanized inventory of the French vocabulary.

The second paragraph gives a description of the machinery and the work done by it. This equipment, made by the French BULL Company, deals normally with 36 alphanumeric characters; it had however to be adapted to 59, or 23 more. Therefore a new codification of the signs, comprising up to three punches per column, had to be provided for, and adaptations had to be made to the collator and the sorter, while the capacity of the printing wheels of both the reporteuse-printer and the tabulator were increased to 45 characters. Next attention is paid to the punch cards used by the CEVF. Of the two chief card types used, the text card (sentence card) and the word card, the latter, which is the most important with regard to the building up of a lexicological inventory, is described. The first step in literary data processing is pre-editing the text. The next step is the punching of the text on an 8-channel paper tape by means of a typewriter connected to a tape punch. Up to 59 characters, together with the necessary control punches, can be punched. Punched tape is used as a temporary information carrier. The paper tape reader makes it possible for the connected card punch automatically to punch the same text twice, once on word-cards, i.e. word by word, and a second time on text cards, i.e. line by line. The punched cards are interpreted by the reporteuse, which permits interpreting more characters than a standard interpreter. The next step is that of verifying both word and text cards; this is done first by comparing the pre-edited text visually with the cards and replacing faulty cards by correct ones, and secondly by printing the text from the cards by means of the tabulator, and comparing this printing with the pre-edited text as well. Now the words on the word cards are provided with their running number in the line or verse; this is done by a reproducer on the basis of a special line-end code, which has previously been punched automatically in column 71, line 9. After that the word cards are fed into the sorter, which arranges them into alphabetic sequence. Next the tabulating machine prints an alphabetic word list, the latter permitting making corrections to a strict automatic alphabetic sequence. Once a conclusive order of the word cards has been arrived at, the definitive word index can be printed, in which every one of the 59 characters previously punched in the tape is present. This is made possible by a special device, in which a card verifier is linked to a receiving typewriter. From this typewritten text an offset index can be printed and issued. The total machine time involved in the making of a mechanical index is about 36 hours; pre-editing the text and verifying the cards and the preliminary word lists, however, takes much more time. If word cards are used to make a word index, the combination of word cards and text cards makes it possible to elaborate concordances, every word card being followed by one or more text cards. The text cards also permit providing the word cards with a certain number of lines (context).

This is done by the reporteuse in the following way: the text cards are laid on the reading path, the word cards on the printing path, and the lines from several text cards are automatically printed on the word card. Word cards which in this way have been provided with a portion of context can now be used as the basic material in dictionary making. The price of a word card provided with a context is about 0,12 N.Fr. In 1960 the staff consisted of two keypunch operators, two half time verifiers and two machine operators, who at the end of the same year had produced a total card file of about two million cards. At present however the personnel has been more than doubled.

In the third paragraph an enumeration is given of the work already done by the CEVF. Indexes and concordances, rime word indexes, frequency indexes and reverse indexes have been developed, none of which however has yet been issued in print. A thesaurus of the words contained in older French dictionaries, as well as a historic repertory of the French vocabulary, are also in preparation. The first part of a series *Datations et Documents Lexicographiques*, comprising words beginning with *A*, was published in 1959, the second one (*B*) in 1960. The center has also done and is still doing some indexing and processing work for linguistic centers and institutions in France and abroad, for example a statistical and indexing project of „Grunddeutsch”, under the direction of Professor J. Alan Pfeffer (University of Pittsburgh) with an eye to a forthcoming dictionary of basic German and the revision of the *English-German, German-English Dictionary of Everyday Usage*. Quemada, director of the Besançon center, was the first to draw attention to the fact that the mechanization of concordances and indexes with the help of punched cards had opened the way towards a complete renovation of traditional lexicography, which has now been changed from traditional handwork in an „industrial” enterprise. The CEVF issues two reviews: a cyclostyled *Bulletin*, of which by now six issues have been published, and the printed *Cahiers de Lexicologie*, the first part of which was issued in 1959, the second in 1960 and the third in 1962.

Chapter II, which is devoted to the „Centro per l'Automazione dell' Analisi Letteraria” at Gallarate (Italy), deals first with the history of the CAAL. While a professor at the philosophical faculty of the Aloisianum at Gallarate near Milan, R. Busa s.j. became conscious of the need of creating a broader philological basis for the research in scholastic thinking. A *thesaurus verborum* of Thomas Aquinas however would necessitate a collection of 13 million words on cards, not to speak of an index, a concordance or even a future dictionary of the language of the „doctor angelicus”. On a tour in the USA in 1949, Professor Busa got in touch with IBM, where at that time his plans for making a mechanographical index of Thomas were met with skepticism. Back in Italy, however, he came in touch with IBM-Italy, and got the opportunity of continuing his research.

He came to the conclusion that making an index on punched cards would cut the time required to a twentieth or fortieth of that of the traditional manual work. After some experiments on cantos of Dante, he started working on some hymns of Thomas. In 1951 his first publication in the field appeared, *Sancti Thomae Aquinatis Hymnorum Ritualium Varia Specimina Concordantiarum*, with a brief description, both in English and Italian, of the stages of the mechanical process, five indices and a concordance.

The second paragraph gives a description of the organization of the CAAL, which was opened on December 17th, 1956. In May 1961, at which time the present author had the opportunity of visiting the Center, the personnel, the material and the machines were scattered over three localities, while the secretariat was in Milan. A reorganization was planned in order to provide the CAAL with corporate capacity, while a special board should be set up to provide for the financial aspects of the work, in order that Busa might devote his attention exclusively to the scholarly side of the work. Since its creation the CAAL has been supported by a national committee with cardinal Montini as honorary president, the members being chiefly prominent persons from Lombardy. IBM-Italy provides for technical assistance, card material and machines, all of them free of charge. For financial assistance the Center was dependent upon Italian industrialists, and rather small allowances from the Ministry of Education and the Ministry of Labour. The general direction of the CAAL rests with professor Busa. The staff of the Center has two sections, a scholarly and a technical one. The former consisted of Busa, his assistant Zampolli and three collaborators, some of them part time. The latter section numbered twenty-two persons, of whom however only one was in permanent service. All of them were typists (seventeen) and machine operators (three), apprentices of a training school for administrative automation, attached to the Center.

In the third paragraph an exposition is given of the working method followed at Gallarate. It deals with the way in which the texts are prepared for punching. Not only the manner of quoting, but also the delimitation of the paragraphs and in some cases also the lines has to be provided for. In preparing the works of Thomas Aquinas, the „quality” of the text has to be indicated by underlining with different colours the parts of different origin such as introductory formulas (blue), literal quotations from authors (green), paraphrased or summarizing parts from authors (red), references of Thomas to his own work (black). All these parts, which, with the exception of the red ones, are not to be included in the concordance, are to be distinguished by special punching (e.g. 0, 3, 6 or 9 in column 14). This is useful not only for statistical, but also for lexicological purposes; it allows automatic excluding or including in the indices of the words contained in these special parts of the text. A punching 9 in column 10 is a special mark indicating that the words that follow are entirely Thomas. The text,

having been carefully prepared, is punched and interpreted by the IBM 26 Printing Card Punch on sentence cards, each of them comprising a single line together with its reference and line-number. Now the punched cards are verified by the IBM 56 Card Verifier. Errors still remaining are detected by listing the sentence cards either on the IBM 866 Cardatype Typewriter, or on the IBM 441 Accounting Machine. Now on feeding the sentence cards into the IBM 858 Cardatype Control Unit, the IBM 866 Cardatype Typewriter makes masters of the different contexts, while at the same time, the IBM 536 Cardatype Printing Card Punch is punching and interpreting each different word of the sentence cards on separate word cards, together with the reference, special marks, the preceding and following punctuation marks, the first letter of the preceding and of the following word and a progressive number. With help of the master, each word card is provided with twelve to thirteen lines of context printed on the reverse side of the card by a Davidson 816 Litho Duplicating Machine. The portions of context are automatically delimited, just as they were staked out by special marks $\times \dots \times$ while pre-editing the text: these marks are also punched in the sentence cards. When this has been done, the word cards pass through the IBM 441 Accounting Machine, which counts them and checks them for the right sequence and for their having been punched in the proper columns. If everything is right, the word cards pass through the IBM 082 Sorter, which arranges them into alphabetical order, which after that is checked by the IBM 101 Electronic Statistical Machine. Next the IBM 519 Endprinting Reproducer provides the word cards with a progressive number, punched in column 37-42 and end-printed on the left, narrow front side of the cards. In a second run the Reproducer end-prints the number of the context and the number of the word in the context. Now the machine produces form cards, i.e. different word cards, from the word cards. For this purpose the alphabetically arranged word cards are fed into the 441 Accounting Machine working in tab, in which process only graphically different words, together with their frequency, are listed; at the same time the connected 519 Endprinting Reproducer punches another set of cards, one being punched for every different form, which is automatically accompanied by its frequency in the entire work. There follows the separating of the homographs by the scholar, after which the 519 Endprinting Reproducer numbers all form cards and the 441 Accounting Machine prints an alphabetical list of the different forms with separated homographs. Next the scholar examines this list and groups different forms under the words that will serve as entries (lemmatizing). When separate cards have been punched for each of these, the entry cards are arranged in their alphabetical order, provided with a punched and endprinted progressive number on the left front side of the card, together with their form frequency and word frequency. By gang punching, the form cards and the word cards are now provided with their lemma number. When both word cards and form cards have been grouped in the right

order under their respective lemma cards, a *Rationarium Verborum* is printed, i.e. a list of the entries followed by their forms with both their individual frequency and the total frequency of all forms of a single entry. Now a second list, *Index Verborum*, is printed, containing the sequentially numbered entries, the forms and the words with the references and their respective frequency. On the basis of the same material several indexes can be made, e.g. a reverse index, an index according to the length of the words, a frequency index, an index of prefixes, of suffixes, etc.

After reviewing the problem of the reverse index and of the concordance, together with the use which is being made at Gallarate of the mark-sensing method, the paragraph is concluded with a list of the IBM machines of the Center.

The fourth paragraph is devoted to the tasks being undertaken by the CAAL. They concern mainly the *Opera Omnia* of Thomas Aquinas and the Dead Sea Scrolls.

Originally Busa had in view a complete punched card file of all words of Thomas. At present he is tending towards an automatic complete word index and concordance, which however would not be printed, but of which only a very limited number of copies (e.g. ten) would be listed by the computer. In a second, less complete, but printed index and concordance, some common words such as articles, conjunctions, prepositions, would be omitted. In this regard however the plans of Busa have not yet reached their ultimate form.

A second, less extensive, but not less epoch-making task of the Gallarate Center is the indexing of the famous Dead Sea Scrolls. This project was the first one of the Center in which the use of traditional punched card equipment was combined with that of electronic data processing machines. The major objective in this application was a bigger capacity, the saving of space, an easier access and a greater speed in reading and printing. After its philological pre-editing the text was punched, interpreted and control punched on word cards; a modified card punch with Hebrew characters was produced to permit the cards to be punched in the unusual right to left sequence corresponding to the reading of the text. The whole file of cards was prepared at Gallarate and then flown to New York for electronic processing. The data were converted to magnetic tape and to the left to right pattern. Data processing was made by the IBM 705 EDP Machine in four runs. The first run produces an each-word tape and list; the second run creates a different-word tape and list; in the third run the machine, comparing and merging into a single tape a lemma-tape and a different-word tape, produces what is called a lexicon tape; finally in the fourth run a concordance is produced by collating the lexicon tape to a tape containing the contexts. The punching and verifying of the cards took only four weeks, but computer programming not less than three months; creating the different tapes was a question of about ninety minutes, while printing took about nine hours. The concordance and the other preliminary

indexes however are not the only word lists which are produced; also a reverse index is made. The different indices of the Dead Sea Scrolls have proved a most valuable tool to textual criticism, making it possible to restore words that had partially or even completely disappeared. The indices of the Qumran texts have up until now not been published; besides the indexes, statistics about frequency of letters and letter combinations are also to be edited.

The total number of punched cards processed up to the present in the center of Gallarate is estimated at about two and a half million.

The fifth paragraph deals with future prospects of the CAAL. As early as 1956 Busa announced that he was studying the use of electronic data processing machines. Reduction of the texts to single words on separate word cards and alphabetizing them had proved too cumbersome and too protracted a task. Moreover serial scanning on swift flowing magnetic tape, analogous to the ordinary way of reading a text, was to be considered superior to the parallel scanning, as is the case in the traditional punch card method. Notwithstanding its expensiveness the computer was to take its place in the mechanization of linguistic analysis, not replacing but supplying the standard punched card equipment. In 1960, however, at the Tübingen colloquium on mechanization of literary analysis, it was sufficiently apparent from Busa's account of the experience of the CAAL that at that time he was considering magnetic surface, digital electronic calculators and not punched card machines as the fundamental instruments in literary data processing methods. During my stay at Gallarate in May 1961, my impression was strengthened that punched card machinery was considered only as an aid to the main computer equipment, but in itself out of date. Another reason why Busa wished to replace unit record by electronic data processing was that the latter enables us to organize much of the work by means of programming, whereas the conventional equipment requires too much human interference, subject to the inexorable laws of chance. Moreover using large-scale data-processing methods will reduce to about 25 % the time required when using conventional punched-card methods. Furthermore EDP will probably make it possible to solve the crucial problem of the variable context by programming the computer in such a way that it provides every word card with a context of about 500 signs before and after the word in question, with the understanding that a word will always be printed in full and that the machine will stop when coming across a full stop or semicolon after 400 signs.

The exclusive predilection of Busa for the electronic data processing machine is easy to understand, if one considers that lexicographical work is alien to him. As he only intends to make a word index and a concordance to Thomas, in his future EDP he will abandon making word cards provided with a context. Those already present at Gallarate are even to be moved to a large library in Rome, where they are likely to be more useful than in a little town of Lombardy. According to Busa, for studying Thomas Aquinas

a word index and a concordance will do, together with a collection of loose pages which will contain the text of the works printed by the printing mechanism of the computer. If at a later stage a dictionary of Thomas were to be made, the necessary word cards could always be punched from the magnetic reels. Busa even believes in the mechanization of lemmatizing in electronic data processing. In my opinion such lemmatizing will only be feasible to a very limited degree, viz. only in connection with standardized spelling.

At the second international colloquium on lexicological mechanization at Besançon in 1961, Busa proclaimed the need of a specialized international center for large scale linguistic and literary data processing, in order to standardize the punching of the texts and allow the following stages of the processing to be done, free of charge. This international center would also have at its disposal a collection of programs for different types of lexicological work in different languages. This would be preferable to having complete EDP centers in every country. National centers would only have to deal with conventional unit record. For this international center Busa is looking forward to the financial assistance of Euratom; in the European computer center of Ispra he sees the possibility of developing an international, interacademical lexicological institute, which would standardize both methods and terminology.

In the sixth paragraph a presentation is given of the conclusions arrived at. The main interest of the literary data processing center of Gallarate is focused on philosophy; linguistic statistics come in the second place, while any lexicographical aim is remote. The indices produced are of high quality. Being concentrated on the works of Thomas Aquinas, which are written in Latin, the international language of the Middle-Ages, the center has much more an international than a national character. The IBM machinery, with 48 punched and interpreted characters, can be regarded as satisfactory for the object in view; more characters however would be necessary for treating modern languages such as Dutch and French. The economic basis of the CAAL was not quite satisfactory, nor was the non-professional part of the personnel. Notwithstanding a certain lack of material expedients and finances, Busa, through his insight in the problems and their possible solutions, has done pioneer work in the field of lexicological information retrieval, and is still ahead in the special sector of EDP of lexicological information.

Chapter III deals with the question whether electromechanical or electronic equipment should be used in future lexicology.

The first paragraph contains some very general, preliminary remarks about the digital computer.

The second paragraph gives a general survey about electronical lexicology to-day in different countries.

United States. Whereas in his comprehensive Russian word count, published in 1935, H. H. Josselson only used punched card equipment, John W. Ellison, the pioneer in the field, back in 1935 proved the feasibility of electronic text-criticism by comparing 311 Greek manuscripts of Luke by means of a computer. In the following years he succeeded in producing a concordance to the Revised Standard Version of the Bible by electronic computing equipment; the concordance was published in 1957 (Nelson, New York). In 1959 there appeared an electronic concordance by S. M. Parrish of the poems of Matthew Arnold; the work was done on the IBM 704 computer. This concordance will be followed by several others (William Butler Yeats, Ben Jonson, Emily Dickinson etc.) in the Cornell series. For this series J. B. Bessinger (Toronto) is preparing a concordance to *The Anglo-Saxon Poetic Records* and E. G. Fogel one to Ben Jonson. Fogel has also set forward a plan for recording on magnetic tape all Elizabethan texts; all sort of lexicological information could be gathered from such tape files. Using the facilities of the Indiana University Research Computing Center, T. A. Sebeok and Valdis J. Zeps compiled a *Concordance and Thesaurus of Chereemis Poetic Language* [1961]. Actually the book is not a real concordance, as quotations are lacking, but it contains several indices verborum. At the Summer Institute of Linguistics J. E. Grimes is trying to develop the automatic concordance of textual material recorded in the languages the Institute specializes in. This automatic concordance is intended as a general purpose research tool which will facilitate answering questions that are asked in a number of areas. A. Juilland (Stanford University) is preparing computer-made frequency dictionaries of words, morphemes and syllables in the five major Romance languages.

England. In 1956 L. Brandwood (University of London) put forth the idea of analyzing Plato's style with increased speed and accuracy with the aid of an electronic digital computer. Dealing with three main aspects, namely lexicology, rhythm and syntax, this would enable us to know in what order Plato's dialogues were written. At Birkbeck College (London University) word-indexes to Alfred's translation of Orosius are being compiled. The „Faculty of Arts Manchester University Lexicological and Indexing Seminar“ has begun making electronic indexes; the seminar, which is under the direction of P. J. Wexler, is working in close connection with the Besançon center. A first index of Racine's *Mithridate* has been compiled with the help of the Ferranti Mercury computer of Manchester University. In this index no attempt at separating the homographs or at lemmatizing has been made, nor was this done in the concordance of Parrish. This is to be viewed as a shortcoming. At Cambridge University R. Wisbey, lecturer in German, is working together with the university's institute of mathematics to process, with the help of the Edsac 2 computer, German texts from the Middle Ages, especially from the eleventh and twelfth centuries.

Italy. Since 1957 Busa has directed his attention to electronic data processing, and at the time is changing over from traditional punched card to computer processing. In Rome the Institute of Romance Philology has merged the metrical schemes of both Provençal and Old French poetry with help of a computer.

Soviet Union. At the Soviet Institute of Mathematics, Siberian Division, at Novosibirsk, three mathematicians, E. V. Evreinov, J. G. Kosarev and V. A. Ustinov succeeded in deciphering the hieroglyphic script of the ancient Mayas with the help of a digital computer.

Czechoslovakia. In Prague S. Segert is using a computer for his work on a comparative dictionary of Semitic languages.

Belgium. At the university of Liège L. Delatte organized in 1961 a „Laboratoire d'Analyse statistique des Langues anciennes”, using both punch card equipment and a computer. An index to Seneca, *Ad Polybium, De Consolatione* was published in 1962.

The Netherlands. At the beginning of 1961 the present author made preparations for computer indexing P. Apherdianus, *Tyrocinium lingue Latinæ* [1552], a bilingual Latin-Dutch schoolbook. The text was punched at the Mathematical Center in Amsterdam, where a program for the X 1 computer was developed by J. A. Th. M. van Berckel. The results, a word-index, frequency list, length list and reverse index, were ready by the beginning of 1962. They will be published by Mouton & Co., together with a new edition of the old manual. In future indexes an attempt will be made to automatize the separating of the homographs and the re-grouping of inflected forms under their lemmas. The Mathematical Center is also busy with a frequency count of one million Dutch words, taken from texts in every field of the written language. The first product, dealing with 50.000 words of newspaper texts, is already available.

The third paragraph is devoted to the technique of making an electronic word-index. Making a punched tape is considered the basic operation, permitting any possible retrieval afterwards.

In the fourth paragraph the view is expressed that in the future it will be possible to simplify and speed up the use of electronic equipment in lexicology. The moment when photoelectric machine reading is perfect enough to read very quickly a printed text just as it is will be of the utmost importance. From that time on any text can be converted to tape or card without human intervention, while at the same time the rather tedious and protracted verifying processes will become superfluous. In a more remote future, optical machine reading could be followed by optical sorting of non-punched cards. The rapid evolution of electronic computing equipment justifies the expectation that some day even special machines

for linguistic analysis, permitting access to a vast quantity of lexicological data, could be used. For the time being, the use of the computer in lexicology with an ultimate lexicographical aim is only practical and only pays once the stage has been reached where a vast amount of lexicological data has to be processed. For making concordances or indexes the use of the computer is already to be preferred to conventional electro-mechanical equipment. The amazing possibilities of electronic equipment in the field of lexicological information retrieval, however, should not make us blind to the fact that linguistically significant exploitation of that information must still be carried out by man.

Chapter IV deals with the possibility of mechanized lexicological center in the Netherlands.

In the first paragraph a distinction is made between dictionary and thesaurus. One of the reasons why the compilation of most of the large national dictionaries lasted too long is that collecting the material, which is done by hand in the traditional way, is a task in itself which encroaches on the lexicographical part of the work. In the future national dictionaries will have to be based on permanent word archives on punched cards. On the basis of these files, general or special dictionaries planned will always have the material ready at hand; moreover this material will also be available for lexicological studies in the widest sense of the word.

The second paragraph deals with some principles of modern lexicography. The great historical dictionaries of the 19th century, covering the national language from its earliest beginnings up to the present time, are the results of a one-sided historical view in linguistics. In the future there will be more need for a series of dictionaries, covering the different periods of the language. Present-day vocabulary is to remain separate from the historical dictionaries, since it will have to be treated synchronically and not historically. The rejuvenation of lexicography was made possible, in the first place by a theoretical reevaluation in linguistics, due mainly to the theories of F. de Saussure, and in the second place by the development of mechanical data processing. Whereas lexicography in the 19th century bore the stamp of linguistics in Germany, that of the future will be greatly indebted to French-speaking countries: French-Switzerland with F. de Saussure and France with B. Quemada.

In the third paragraph a bird's-eye view is given of mechanized lexicological centers. The CAAL at Gallarate and the CEVF at Besançon are the only centers in Europe where mechanized lexicology has left the experimental stage. The other centers are more recent; most of them have only started or are still in the developmental stage. The Academy of the Hebrew Language (Jerusalem), which is planning a historical dictionary of the national language, has made elaborate plans for the establishing of a mechanographic center. In the first half of 1961 there were made indexes and concordances from a short Talmudic text, all of them according to

the Gallarate method. The German Academy of Sciences in Berlin has a committee for planning and experimenting with lexicographical indexing and grammatical projects by means of punch card machinery. The Czech Institute in Prague has plans for making a mechanical inventory of Czech technical and special terms. In the Netherlands, at the Italian Institute of Utrecht university, word-indexes of Italian literature are made by conventional punch card equipment. It is not going too far to predict the establishing in every country of one or more centers dedicated to mechanized lexicology.

The fourth paragraph deals with a possible lexicological center in the Netherlands. The wish to establish it originated before automation in this field was known. The goal however can only be achieved if automation is made use of, because a lexicological thesaurus which would answer to modern qualitative and quantitative standards cannot be achieved within a reasonable time and at a reasonable price if it is to be compiled in the conventional way. Moreover there is a human factor which pleads for mechanical lexicology: man should not be forced to perform „mechanical” and uninteresting tasks which a machine can do in a much more accurate, speedy and inexpensive way. Starting from automatically processed and readily available lexicological material, the various volumes of a dictionary could be much more uniform than has been the case until now. Just as Renaissance lexicography accepted the devices of the newly discovered art of printing, today lexicography will have to avail itself of the possibilities offered by modern information retrieval. A Dutch lexicological center however cannot be a mere copy of that of Gallarate or Besançon, since the circumstances and the needs are not wholly identical. No mechanized lexicological center should or even can be established without careful preparatory work resulting in a detailed plan.

In the fifth and last paragraph the potential problems and objectives of a Dutch lexicological center are treated. One question is whether in the texts to be treated every word should be taken separately, or whether a choice has to be made; for older periods, with only a limited number of documents, the first way is to be preferred. Another question now debated is whether it is advisable to have a double card file: one in which the words are arranged according to the general alphabetic order, another in which they are arranged according to their alphabetic order within the scope of each individual work. Another aspect is that of the possibilities in the field of word statistics, automation bringing the statistical treatment of language data within close reach. The paragraph ends with a rough plan for organizing a Dutch center for lexicological automation.

In the conclusion three attitudes towards automation in lexicology are distinguished. The first one is a hostile one based on the so-called incompatibility of humanities and automation, which considers the technique a threat. A second one, the belief in the omnipotence of the machine, is to be

considered even more dangerous as it leads to oversimplification of the problems involved, forgetting that a machine permits neither sloppy thinking nor mistakes. The only positive attitude towards the question is a critical one that neither fears the machine nor is blind to its limitations. Future progress in lexicology necessitates what might be called the wedding of technology and lexicology. The lexicological institute of the future will have close relations to the mathematical institute, which makes it possible to draw upon common machinery. Man, however, will always remain more important than the machine.

REGISTER VAN PERSOONSNAMEN

- Agricola, E. 88
 Alinei, M. L. 111
- Baldinger, K. 19, 34, 123, 124
 Barry, M. I. 38
 Beaujean, A. 12
 Beckmann, B. 105, 108
 Ben-Hayyim, Z. 110, 119
 Berckel, J. A. Th. M. van 83, 96, 97, 98, 137
 Bessinger, J. B. 79, 84, 85, 136
 Bierwisch, M. 111
 Blancquaert, E. 107, 112, 115
 Blumrich, M. 111
 Boccaccio, P. P. 61
 Boldrini, M. 42, 43
 Bonnard, H. 36
 Booth, A. D. 9, 56, 83, 86, 88
 Bosgang, A. 76
 Braffort, P. 98
 Brandwood, L. 9, 56, 86, 88, 89, 136
 Buck, C. D. 86
 Burrows, M. 61
 Busa, R. 19, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 64, 65, 66, 67, 68, 69, 70, 71, 72, 74, 76, 79, 88, 92, 102, 110, 113, 114, 130, 131, 133, 134, 135, 137
 Bush, V. 39
- Cadbury, H. J. 76
 Canestraro, L. 43
 Carr, J. W. 79
 Carroll, J. B. 9, 11
 Casares, J. 9, 11, 13, 108
 Casey, S. 63
 Ceccato, S. 101
 Cleave, J. P. 9, 56, 86, 88
 Colin, A. J. T. 86, 89, 90
 Cook, C. M. 76, 79
 Craigie, W. 108
- Deferrari, R. J. 38
 Delatte, L. 94, 95, 137
 Delavenay, E. 37, 88, 111
 Delavenay, K. 37
 Diebold, J. 17, 112
- Dubois, J. 122
 Dyer, P. S. 91
- Ellison, J. W. 75, 76, 77, 78, 79, 102, 123, 126, 136
 Epstein, B. 75
 Euwe, M. 74
 Évrard, É. 94
 Evreinov, E. V. 93, 137
- Fogel, E. G. 79, 80, 85, 102, 136
 Folpini, C. 55, 61
- Gall, M. W. 65
 Gardin, J. C. 98
 Garfield, E. 124
 Geiss, H. 111
 Goodman, L. L. 75, 126
 Gougenheim, G. 34, 36, 120
 Gradenwitz, O. 52
 Graham, J. 76
 Greimas, A.-J. 17, 36, 103, 109
 Grimes, J. E. 87, 136
 Grimm, J. en W. 108, 109
 Guiraud, P. 43, 56, 120, 121
- Haeringen, C. B. van 12, 13
 Halliday, M. A. K. 10
 Harrington, J. 121
 Heger, K. 36
 Hübner, G. 60
 Hyatt, H. 75
- Ihm, P. 98
 Imbs, P. 19, 103, 108, 110
 Ising, G. 105
- Jespersen, O. 10
 Josselson, H. H. 75, 136
 Judd, B. 80
 Juilland, A. 87, 136
- Kiddle, L. B. 79
 Klappenbach, R. 34, 111
 Klerk, D. de 23
 Knudsen, T. 106
 Kosarev, J. G. 93, 137
 Kruyskamp, C. 105, 107
 Kukenheim, L. 12, 13, 16, 103, 118

- Levison, M. 89
 Lewicka, H. 34
 Littré, E. 10, 11, 12, 13, 32, 33, 109
 Loey, A. van 9
 Luhn, H. P. 85
- Mc. Culley, W. R. 77
 Macgregor, G. H. C. 123
 Mak, J. J. 9, 113
 Marouzeau, J. 9, 10, 11
 Mater, E. 32, 36, 93, 111, 125
 Matoré, G. 12, 104, 124, 128
 Meile, P. 102
 Meillet, A. 16
 Migliorini, B. 9, 10, 11
 Miles, J. 79, 80
 Mitterand, H. 36, 115
 Montgomery, G. 79, 80
 Montini, G. B. 41, 42, 131
 Moreau, R. 34
 Morrison, G. 80
 Morton, A. 123
 Mosteller, F. 85
 Muller, Ch. 90
 Muller, J. W. 116
 Möller, K. 11
- Naarding, J. 12
 Natalis, G. 36
 Noreen, A. 13
- Odendal, F. F. 10
 O'Reilly 38
 Országh, L. 109, 118
- Paardekooper, P. 9
 Pacifico, M. 92, 125
 Painter, J. A. 80, 81, 82
 Parrish, S. M. 15, 80, 81, 82, 83, 85, 124, 126, 136
 Petit, J. 115
 Pfeffer, J. A. 34, 130
 Popplewell, C. M. 90
 Prinz, O. 119
- Quemada, B. 17, 19, 20, 25, 34, 35, 69, 74, 79, 83, 99, 100, 103, 105, 109, 110, 119, 120, 128, 130, 138
- Rivenc, P. 34, 82
 Roethe, G. 104
 Romein, J. 13
 Roncaglia, A. 92
- Sarfatti, G. 110
 Saussure, F. de 11, 109, 138
 Scheel, H. L. 12, 14
 Scheele, M. 124
 Schoonees, P. C. 112
 Sebeok, T. A. 85, 136
 Secrest, B. W. 61, 65
 Segert, S. 93, 137
 Shirokov, F. 18, 93
 Sledd, J. 13
 Smrčková, J. 34
 Sobolev, S. L. 93
 Soest, J. L. van 48, 120
 Sohoni 34
 Soule, G. 77
 Spellman, F. J. 40
 Steinitz, W. 111
 Štindlová, J. 32, 34, 36, 56, 57, 111
 Sweet, H. 10, 118
- Tagliavini, C. 43
 Tappolet, E. 11
 Tasman, P. 37, 40, 46, 47, 54, 55, 61, 64, 65, 66, 68, 120, 121, 124, 126
 Tellenbach, E. 111
 Tilander, G. 17
 Tollenaere, F. de 30, 36, 52, 69, 96, 97, 98, 101, 103, 123, 124, 125, 137
- Ullmann, S. 9, 11, 12, 118
 Ustinov, V. A. 93, 137
- Valach, M. 93
 Venzac, G. 12
 Verdam, J. 113, 116, 117
 Verwijs, E. 113
 Vincent, E. R. 17, 124
 Vries, M. de 105, 113
- Wallace, D. L. 85
 Wartburg, W. von 9, 108
 Watson, T. J. 39, 40
 Weinreich, U. 9, 10
 Wexler, P. J. 10, 73, 90, 136
 Wijngaarden, A. van 95, 125
 Winkel, L. A. te 113
 Wisbey, R. 89, 91, 92, 136
 Worth, D. S. 10
- Zampolli, A. 43, 44, 56, 58, 59, 131
 Zemanek, H. 15
 Zeps, V. J. 86, 136

LITERATUUR

PUBLIKATIES ¹⁾

- According to Mark IV (*Time, The Weekly Magazine*, August 9, 1954, 40-41).
- AGRICOLA, E., Die automatische Sprachübersetzung (in *Forschung. Lehre. Praxis. Schriftenreihe der Gewerkschaft Wissenschaft. Heft 4*, 150-171) Berlin [1962].
- Automatisering en Taalkunde (Stichting Studiecentrum voor Administratieve Automatisering) Amsterdam [1961].
- BALDINGER, K., Automation und Lexikologie (*ZRPh.* 75, 540-548) [1959].
- BECKMANN, B., *Das Deutsche Wörterbuch in Gegenwart und Zukunft* (in *Das Institut für Deutsche Sprache und Literatur. Vorträge gehalten auf der Eröffnungstagung* 125-136. — Deutsche Akademie der Wissenschaften zu Berlin. Veröffentlichungen des Instituts für Deutsche Sprache und Literatur ²⁾) [1954].
- BEN-HAYYIM, Z., *The historical Dictionary of the Hebrew Language*, a plan (The Academy of the Hebrew Language and The Bialik Institute) [1960].
The International Meeting on Mechanical Methods of Literary Analysis and Lexicography (*Lěšonenu. A Journal for the Study of the Hebrew Language and Cognate Subjects. Vol. XXV, No. 1-2* (Oct. 1960-Jan. 1961), 81-89).
- BERCKEL, J. A. TH. M. VAN, Onderzoek Woordfrequentie (Rapport R 642, Rekenafdeling, Stichting Mathematisch Centrum, Amsterdam) [1961].
Onderzoek Woordfrequentie. Resultaten Kranten (Rapport R 642/2, Rekenafdeling, Stichting Mathematisch Centrum, Amsterdam) [1962].
Sorting by Chain Formation. A Program for a Computer Handling Punched Tape (Report R 628, Rekenafdeling, Stichting Mathematisch Centrum, Amsterdam) [1962].
- BESSINGER, J. B., Computer Techniques for an Old English Concordance (*American Documentation* 12, 227-229) [1961].
- Bible labor of years is done in 400 hours (*Life* February 18, 1957, p. 92).
- BIERWISCH, M., Die Verwendung von Lochkarten bei der grammatischen Analyse (*Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin* 6, 255-258) [1960].
- BLANCQUAERT, E., Vlaamsch Neerlandicus op den uitkijk (*VMKVA*, Nov.-Dec. 1935, 1045-1054).
Het Nederlandsch Woordenboek van de twintigste Eeuw (*VMKVA*, Dec. 1938, 781-799).
- BLUMRICH, M. und E. TELLENBACH, Die Anwendung von Lochkarten im Wörterbuch der deutschen Gegenwartssprache (*Forschung. Lehre. Praxis. Schriftenreihe der Gewerkschaft Wissenschaft. Heft 4*, 88-93). Berlin [1962].
- BOOTH, A. D., Use of a computer for certain operations of classification (*Computers and Automation* 6, no. 4, April) [1957].

¹⁾ Van de in dit werk aangehaalde publikaties werden in deze lijst uitsluitend die vermeld welke betrekking hebben op de theoretische lexicografie en de automatisering der lexicologische analyse. Artikelen die in kranten en dergelijke zijn verschenen, werden slechts bij uitzondering opgenomen.

²⁾ In de titels zijn afgekorte aanduidingen van publikaties (b.v. BECKMANN, B., *Das Deutsche Wörterbuch* i.p.v. *Das Deutsche Wörterbuch in Gegenwart und Zukunft*) door kursivering als zodanig signaleerd.

- BOOTH, A. D., L. BRANDWOOD and J. P. CLEAVE, *Mechanical Resolution of Linguistic Problems*. London [1958].
- BOOTH, A. D. and A. J. T. COLIN, On the Efficiency of *Dictionary Construction* (*Inf. & Contr.* 3, 327-334) [1960].
- BRANDWOOD, L., Analyzing Plato's style with an electronic computer (University of London, Institute of Classical Studies, Bulletin Number 3, 45-54) [1956].
- BUSA, R., s.j., A proposito di immanenza e trascendenza. *L'interiorità secondo la scolastica*. — L'interiorità della scolastica alla filosofia moderna (Comunicación. *Actas Congreso internacional de Filosofía*, Barcelona, 4-10 Octubre 1948, II, 673-689). Madrid [1949].
- La Terminologia Tomistica dell' Interiorità*. Saggi di metodo per un' interpretazione della metafisica della presenza (Archivum philosophicum Aloisianum a cura della Facoltà di Filosofia dell' Istituto Aloisianum S. J. Serie II, 4). Milano [1949].
- Complete Index Verborum of Works of St. Thomas (Speculum XXV, 424-425)* [1950].
- Sancti Thomae Aquinatis Hymnorum Ritualium *Varia Specimina* Concordantiarum. Primo saggio de indici di parole automaticamente composti e stampati da macchine IBM a schede perforate. Milano [1951].
- Rapida e meccanica composizione e pubblicazione di indici e concordanze di parole mediante macchine elettrocontabili (*Aevum XXV, 479-493*) [1951] (= de tekst der inleiding van *Varia Specimina*).
- Mechanisierung der philologischen Analyse (Nachr. Dok. 3, 14-19)* [1952].
- Entwicklung der Mechanisierung der sprachlichen Analyse (Nachr. Dok. 4, 202-204)* [1953].
- La Cibernetica (*Ingegneria Meccanica*, Fasc. 2 — Febbraio 1954, 3-8).
- I principali problemi dell' automazione dell' linguaggio scritto (Convegno Internazionale sui problemi dell' Automatismo*. Milano 8-13 aprile 1956. Vol. 1, 831-841 = Supplemento a *La Ricerca scientifica* anno 28) [1958].
- Die *Elektronentechnik* in der Mechanisierung der sprachwissenschaftlichen Analyse (*Nachr. Dok. 8, 20-26*) [1957].
- The *Index* of all non biblical Dead Sea Scrolls published up to December 1957 (*Revue de Qumran*, I, 187-198) [1958].
- The use of *punched cards* in linguistic analysis (in Robert S. Casey e.a., *Punched Cards* 357-373) [1958].
- Centro per l'automazione dell' analisi letteraria (*Current research and development in scientific documentation* 6, 98-100) [1960].
- Zusammengefasste *Darstellung* der Erfahrungen des Centro per l'Automazione dell' Analisi Letteraria des Aloisianum, Gallarate (Varese) Italien (Kolloquium Maschinelle Methoden der literarischen Analyse und der Lexikographie. Tübingen, 24.-26. Nov. 1960) [1960].
- Erläuterungen* zu den lexikographischen Arbeiten zu Goethe, Farbenlehre, Bd. 3 (Kolloquium Maschinelle Methoden der literarischen Analyse und der Lexikographie. Tübingen, 24.-26. Nov. 1960) [1960].
- C(entro) A(utomazione) A(nalisi) L(etteraria) 20 Gennaio 1961.
- L'analisi linguistica* nell' evoluzione mondiale dei mezzi d'informazione (in *Almanacco Letterario Bompiani* 1962, 103-117) [1961].
- Les *Travaux* du „Centro per l'Automazione dell' Analisi Letteraria“ de Gallarate (*Cah. de Lex.* 3, 64-68) [1962].
- L'Analyse* des homographes et des formes flexionnelles (*Cah. de Lex.* 3, 134-139) [1962].
- CASARES, J., Introducción a la Lexicografía moderna (*Revista de Filología Española*. — Anejo LII) Madrid [1950].

- CASEY, R. S., J. W. PERRY, A. KENT and M. BERRY, *Punched Cards*. Their Application to Science and Industry, Second edition. New York [1958].
- CECCATO, S., *La storia di un modello meccanico dell' uomo che traduce* (in *Almanacco Letterario Bompiani* 1962, 122b-134b) [1961].
- COLIN, A. J. T., The Automatic Construction of a Glossary (*Inf. & Contr.* 3, 211-230) [1960].
- COOK, C. M., Automation comes to the Bible (*The Christian Century* July 24, 1957, 892-894).
- DEFERRARI, R. J. and M. I. BARRY, *A complete Index of the Summa Theologica of St. Thomas Aquinas* (Catholic University of America Press) [1956].
- DELATTE, L. et É. ÉVRARD, Un Laboratoire d'Analyse statistique des Langues anciennes à l'Université de Liège (extrait de *L'Antiquité Classique* XXX, 427-442) [1962].
Sénèque, Consolation à Polybe. Index Verborum. Relevés statistiques. Liège [1962].
- DELAVENAY, E., La machine à traduire. Paris [1959].
 An Introduction to *Machine Translation*. London [1960].
- DELAVENAY, E. et K., Bibliographie de la Traduction Automatique. Bibliography of Mechanical Translation. 's-Gravenhage [1960].
- Digitale Informationswandler*, herausgegeben von W. HOFFMANN. Braunschweig [1962].
- ELLISON, J. W., Can a computer write a book? (*Univac in the News*, Remington Rand Univac; overdruk uit *The Living Church* September 23, 1956, p. 7).
 The Use of Electronical Computers in the Study of the Greek New Testament. A thesis presented by J. W. E. to the Harvard Divinity School in partial fulfillment in the requirements for the degree of Doctor of Theology in New Testament. Harvard University. Cambridge, Massachusetts. April, 1957 (onuitgegeven).
 Nelson's Complete Concordance of the Revised Standard Version Bible. New York [1957].
- EUWE, M., Non-numerieke problemen op de computer (*Inf.* nr. 20, 16-17) [1962].
- EVREINOV, E. V., J. G. KOSAREV, V. A. USTINOV, Primenenie elektronnykh vyčislitel'nykh mašin v issledovanii pis'mennosti drevnich majja (Akademija Nauk SSSR, Sibirskoe otdelenie). Novosibirsk [1961].
- FOGEL, E. G., *Electronic Computers and Elizabethan Texts* (*Studies in Bibliography*. Papers of the Bibliographical Society of the University of Virginia, 15, 15-31) [1962].
- FRIELINK, A. B., Mogelijkheden van automatisering in de taalkunde (*Automatisering en Taalkunde* 5-15) [1961].
- GALL, M. W., Die RAMAC und die Philologie (*IBM Deutschland*, Januar 1961, 4-5).
- GARDIN, J. C., La mécanisation de certaines recherches historiques, à partir de textes préalablement analysés (*Cah. de Lex.* 3, 177-184) [1962].
- GARFIELD, E., The preparation of printed indexes by automatic punched-card techniques (*American Documentation*. A quarterly review of ideas, techniques, problems and achievements in documentation VI, 68-76) [1955].
- GEISS, H., Aufgaben und Perspektiven der Arbeitsgruppe „Minoische Texte" des Instituts für griechisch-römische Altertumskunde (*Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin* 6, 250-254) [1960].
- GOODMAN, L. L., Man and Automation. London [1957].
- GOUGENHEIM, G., La Statistique linguistique et l'histoire du vocabulaire (*Cah. de Lex.* 2 (1960) 31-40) [1961].
- GOUGENHEIM, G. et P. RIVENC, La préparation du vocabulaire scientifique général (*Cah. de Lex.* 3, 94-98) [1962].

- GRADENWITZ, O., *Laterculi vocum latinarum* Voces latinas et a fronte et a tergo ordinandas curavit. Leipzig [1904].
- GREIMAS, A.-J., Les problèmes de la description mécanographique (*Cah. de Lex.* 1, 47-75) [1959].
- Grosse Werke der Weltliteratur werden maschinell bearbeitet (*IBM Deutschland, Beilage zur IBM World Trade News*, November 1956, 12 en 15).
- GUIRAUD, P., *Les caractères, statistiques du vocabulaire*. Essai de méthodologie. Paris [1954].
Bibliographie critique de la statistique linguistique (Comité International Permanent de Linguistes. Publications du comité de la statistique linguistique II). Utrecht-Anvers [1954].
Problèmes et méthodes de la statistique linguistique. Dordrecht [1959].
- HARRINGTON, J., Summation of the Conference on Automation Systems (*Proceedings of the E(lectronic) I(ndustries) A(ssociation) Conference on Automation Systems for Business and Industry* 171-173). New York [1958].
- HEGER, K., Zu den Möglichkeiten einer *Automation in der Sprachwissenschaft* (*ZRPh.* 76, 537-546) [1960].
- HÜBNER, G., Die Exzerption des dritten Bandes von *Goethes Farbenlehre* mit Hilfe von Lochkarten und Maschinen (*Cah. de Lex.* 3, 99-100) [1962].
- IBM and the Dead Sea Scrolls (*IBM World Trade News*, May 1958, 7).
- IBM Pioneers Literary Data Processing Program (*IBM World Trade Corporation Data Processing Review* December 15, 1958 Volume 1, Number 12, blz. 1-2).
- IMBS, P., Au seuil de la Lexicographie (*Cah. de Lex.* 2 (1960), 3-17) [1961].
- ISING, G., Zur Erprobung eines fotomechanischen Exzerptionsgerätes (*Spektrum* 8, 90-92) [1962].
- John Diebold Answers Twenty Questions (reprinted from *Automatic Data Processing*, March 1959).
- JOSSelson, H. H., The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian. Detroit [1953].
- KLAPPENBACH, R., L'emploi des cartes perforées dans le dictionnaire de langue allemande contemporaine (*Cah. de Lex.* 2 (1960), 87-97) [1961].
- KNUDSEN, T., Det norske litterære ordboksverk i 25 år, 1921-1946 (1922-1947) (*MoM* 1948, 1-11).
- KUKENHEIM, L., *Van Glossarium tot Thesaurus*. Lexicologische Theorien en lexicographische Realiteiten (*Lev. Talen* 1960, 15-34).
 Codering van grammaticale Gegevens (*Automatisering en Taalkunde* 35-45) [1961].
 Esquisse historique de la linguistique française et de ses rapports avec la linguistique générale. Leiden [1962].
- LEVISON, M., The Application of the Ferranti Mercury Computer to Linguistic Problems (*Inf. & Contr.* 3, 231-247) [1960].
- LEWICKA, H., Metody mechaniczne w leksykologii (*Kwartalnik Neofilologiczny* (Polska Akademia Nauk) 6, 343-348) [1959].
- Lexicologie et Lexicographie* Françaises et Romanes. Orientations et exigences actuelles. Strasbourg 12-16 Novembre 1957 (Colloques internationaux du Centre National de la Recherche Scientifique). Paris [1960].
- MCCULLEY, W. R., Univac compiles a complete Bible Concordance (*Univac and the Bible Concordance*, Remington Rand Univac = overgedrukt uit *Systems* XX, 2 (March-April), p. 22-23) [1956].
- MATER, E., Über die Verwendung von Maschinenlochkarten für wissenschaftliche Aufgaben (*Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin* 6, 248-250) [1960].
Internationales Kolloquium über „Maschinelle Methoden der literarischen Analyse und der Lexikographie“ in Tübingen (*Spektrum* 7, 81-83) [1961].

- Der *Einfluss der Technik* auf die wissenschaftliche Arbeitsmethodik (*Forschung. Lehre. Praxis. Schriftenreihe der Gewerkschaft Wissenschaft*. Heft 4, 26-29). Berlin [1962].
- Herstellung eines rückläufigen Wörterbuches (*a.w.* 122-125) [1962].
- MATORÉ, G., *La méthode en Lexicologie*. Domaine français. Paris [1953].
- MEILE, P., A propos des concordances (*Trad. Aut.* 1, 6-9) [1960].
- MIGLIORINI, B., *Che cos' è un vocabolario?*. Firenze [1951].
- MITTERAND, H., *Lexicologie et machine (Vie et Langage* 1961, 659-664).
- MITTERAND, H. et J. PETIT, Index et concordances dans l'étude des textes littéraires (*Cah. de Lex.* 3, 160-172) [1962].
- MONTGOMERY, G., *Concordance to the Poetical Works of John Dryden*. Berkeley and Los Angeles [1957].
- MØLLER, K., *Leksikologi og Leksikografi* (Udvalg for Folkemaals Publikationer. Serie A Nr. 19). København [1959].
- MULLER, CH., Les index de vocabulaire (Extrait du *Bulletin des jeunes romanistes* IV, 9-14) [1961].
- NATALIS, G., *Internationale Tagung über Maschinelle Methoden der literarischen Analyse und der Lexikographie in Tübingen*, 24. bis 26. November 1960 (*Nachr. Dok.* 11, 219) [1960].
- ODENDAL, F. F., *Leksikografiese Probleme I* (*Standpunte* XIV, Nr. 6, Augustus 1961, 53-61), II (XV, Nr. 1, Oktober 1961, 49-55), III (XV, Nr. 2 en 3, Desember 1961-Februarie 1962, 48-55).
- ORSZÁGH, L., *Problems and Principles of the new Dictionary of the Hungarian Language* (*Acta Linguistica Academiae Scientiarum Hungaricae* X, 211-273) [1960].
- PACIFICO, M., *I nuovi Gutenberg*. *Linguistica ed elettronica nel mondo, oggi* (*Almanacco Letterario Bompiani* 1962, 100-103) [1961].
- Un esperimento di *filologia elettronica*: l'omogeneizzazione degli indici del Frank e dello Spanke (*Almanacco Letterario Bompiani* 1962, 135-142) [1961].
- PAINTER, J. A., *Computer Preparation of a Poetry Concordance* (*Comm. ACM* 3, 91-95) [1960].
- PARRISH, S. M., *A concordance of the poems of Matthew Arnold* (Cornell University Press) [1959].
- Problems in the Making of *Computer Concordances* (*Studies in Bibliography*. Papers of the Bibliographical Society of the University of Virginia, 15, 1-14) [1962].
- PFEFFER, J. A., *Grunddeutsch: Werden und Wesen* (*The German Quarterly*, Vol. XXXV (No. 2-March 1962), 179-186) [1962].
- La mécanisation dans la préparation statistique de l'allemand fondamental (*Cah. de Lex.* 3, 90-93) [1962].
- Philosophie auf Lochkarten in Italien (*IBM Deutschland, Beilage zur IBM World Trade News* 9 (1957), Af. 1, blz. 17).
- PRINZ, O., „Mittelateinisches Wörterbuch“ Entstehung und Anlage (Estr. dagli *Studi medievali*, 3a Serie, I, 1, 1960, 296-303).
- Problems in Lexicography. Edited by F. W. Householder and S. Saporta (Part IV *International Journal of American Linguistics*, Volume 28, Number 2, April) [1962].
- QUEMADA, B., *La technique des inventaires mécanographiques* (*Lexicologie et Lexicographie* 53-63) [1960].
- La Mécanisation dans les recherches lexicologiques (*Cah. de Lex.* 1, 7-46) [1959].
- Datations et Documents lexicographiques, Ire Série. Publiés par B. Quemada avec la collaboration de P. J. Wexler (*Annales Littéraires de l'Université*

- de Besançon, Publications du Centre d'Étude du Vocabulaire Français). A [1959], B [1960].
- L'Inventaire des dictionnaires bilingues (Cah. de Lex. 2 (1960), 67-78) [1961].*
- L'Exploration mécanique des dictionnaires bilingues: index français-flamand du Vocabulaire de N. de Berlaimont (Bull. d'Inform. 4, 9-49) [1961].*
- L'Inventaire lexicographique en vue d'un Thesaurus National (Cah. de Lex. 3, 119-131) [1962].*
- SARFATTI, G., The Mechanization of Linguistic Research (*Lěšoněnu. A Journal for the Study of the Hebrew Language and Cognate Subjects*. Vol. XXV, No. 1-2 (Oct. 1960-Jan. 1961), 90-95).
- SHEEL, H. L., Neuere Arbeiten zur Lexikologie (*GRM 36, 253-263*) [1955].
- SCHHEEL, M., Die Lochkartenverfahren in Forschung und Dokumentation mit besonderer Berücksichtigung der Biologie. Stuttgart [1954].
- SCHOONEES, P. C., Rondom die Woordeboek. Kaapstad [1958].
- SEBEOK, TH. A. and V. J. ZEPS, Concordance and thesaurus of Cheremis poetic language (*Janua linguarum, Studia memoriae Nicolai van Wijk dedicata, Series maior VIII*). 's-Gravenhage [1961].
- SECRET, B. W., *The IBM Electronic Statistical Machine Applied to Word Analysis of the Dead Sea Scrolls* (IBM World Trade Corporation, November 17, 1958). New York.
- SHIROKOV, F., Computer deciphers Maya hieroglyphs (*The Unesco Courier*, March 1962, 27-32).
- SOEST, J. L. VAN, Informatie-theorie en Communicatie-theorie (Handleidingen bij het Onderwijs aan de Technische Hogeschool te Delft — Onder redactie van de Centrale Commissie voor Studiebelenen) [1955].
- SOULE, G., The Machine that indexed the Bible (*Popular Science 169, 173-175 en 242 en 246*) [1956].
- SMRČKOVÁ, J., Mechanografické metody ve francouzské jazykovědě (*Časopis pro moderní filologii 43, 121-123*) [1961].
- ŠTINDLOVÁ, J., Mechanizace a automatizace v jazykovědě (*Věstník ČSAV 69, 453-458*) [1960].
- Retrográdní slovníky (*Slovo a slovesnost 21, 270-272*) [1960].
- Sur le classement inverse des mots et sur ce qu'on appelle „Dictionnaire inverse” (*Cah. de Lex. 2 (1960), 79-86*) [1961].
- Využití technických prostředků mechanizace a automatizace při organizace archívních sbírek odborného názvosloví (*Naše řeč 45, 23-32*) [1961].
- Stroje na zpracování informací a jejich význam pro jazykovědu (*Slovo a slovesnost 22, 208-215*) [1961].
- Uplatňování metod mechanizace a automatizace v lexikologické práci v zahraničí (*Slovo a slovesnost 22, 67-71*) [1962].
- Projet de mécanisation d'un inventaire des termes techniques et spéciaux (*Cah. de Lex. 3, 73-83*) [1962].
- TAGLIAVINI, C., L'automazione nelle ricerche fonetiche (*Almanacco Letterario Bompiani 1962, 120-122*) [1961].
- TASMAN, P., *Literary Data Processing (IBM Journal of research and development, Vol. 1, No. 3, July 1957, 249-256).*
- Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods (International Business Machines Corp.) New York [1958].*
- Literature and Document Research Automation (*Proceedings of the Electronic Industries Association Conference on Automation Systems for Business and Industry 61-72*). New York [1958].
- Index and Concordance development for literary documentation and information retrieval (ADIA-Proceedings (Beiheft zu den Nachrichten für Dokumentation Nr. 8). Proceedings of the ADIA-Conference (Automatic*

- Documentation in Action), Frankfurt/Main vom 9.-12. Juni 1959. Im Auftrage der Veranstalter herausgegeben von E. H. Erich Pietsch. Deutsche Gesellschaft für Dokumentation. Frankfurt/Main) [1961].
- TOLLENAERE, F. DE, Verslag betreffende het „Kolloquium über maschinelle Methoden der Literarischen Analyse und der Lexikographie“ in Tübingen 24-26 november 1960 (*Inf.*, Nr. 12, 13-15) [1961].
- Automatisering in de Lexicologie (*Automatisering en Taalkunde* 24-34) [1961].
- Het „Colloque international sur la mécanisation des recherches lexicologiques“ (*Inf.*, Nr. 16, 10-13) [1961].
- Automatisering in dienst van de taalkunde (*Taalonderzoek in onze Tijd* 83-112) [1962].
- La *documentation lexicographique* et ses propres besoins (*Cah. de Lex.* 3, 101-115) [1962].
- Vaktermen met betrekking tot de administratieve automatisering. Verklarende Woordenlijst (Uitgave van de Stichting Studiecentrum voor administratieve Automatisering). Amsterdam [1959].
- VINCENT, E. R., *Mechanical Aids for the Study of Language and Literary Style* (in *Literature and Science*. International Federation for Modern Languages and Literatures. Proceedings of the sixth triennial congress. Oxford. 1954, 56-60) [1955].
- WARTBURG, W. VON, *Betrachtungen über das Verhältnis von historischer und deskriptiver Sprachwissenschaft* (in *Mélanges de Linguistique offerts à Charles Bally* 3-18). Genève [1939].
- WEXLER, P. J., Le programme de recherches de Famulis (*Cah. de Lex.* 3, 71) [1962].
- WIJNGAARDEN, A. VAN, Rekenen en Vertalen. Delft [1952].
- WISBEY, R., *Concordance Making by Electronic Computer: some experiences with the 'Wiener Genesis'* (*The Modern Language Review* LVII, 161-172) [1962].
- De meest voorkomende Woorden en Wordcombinaties in het Nederlandsch*. Verslag van een onderzoek in opdracht van het Departement van Onderwijs en Eeredienst. Batavia [1937].
- Wörterbuch der deutschen Gegenwartssprache*. Herausgegeben von RUTH KLAPPENBACH und WOLFGANG STEINITZ (1. und 2. Lieferung. *A-annahmen*). Berlin [1961].
- ZAMPOLLI, A., Studi di statistica linguistica eseguiti con impianti IBM (Dissert. Padua) [1960].
- ZEMANEK, H., Automaten und Denkprozesse (in *Digitale Informationswandler* 1-66) [1962].

TIJDSCHRIFTEN ¹⁾

- Aevum*: Rassegna de scienze storiche linguistiche e filologiche. Pubblicata per cura della Facoltà di Lettere dell' Università Cattolica del Sacro Cuore. Milano 1927→
- Bull. d'Inform.*: Bulletin d'Information du Laboratoire d'Analyse Lexicologique. (Publications du Centre d'Étude du Vocabulaire Français. Faculté des Lettres et Sciences Humaines). Besançon 1960→
- Cah. de Lex.*: Cahiers de Lexicologie. Publiés par B. Quemada (Publications du Centre d'Étude du Vocabulaire Français. Faculté des Lettres de Besançon). Besançon 1959→
- Comm. ACM*: Communications of the ACM (Association for Computing Machinery). Baltimore 1958→

¹⁾ Alleen tijdschriften die verkort zijn aangehaald werden hier opgenomen.

- GRM*: Germanisch-Romanische Monatschrift. Heidelberg 1909→
- Inf.*: Informatie van de Stichting Studiecentrum voor administratieve Automatisering. Amsterdam 1959→
- Inf. & Contr.*: Information and Control Published quarterly at Mt. Royal and Guilford Aves. Baltimore 2, Md. for Academic Press Inc. New York 1958→
- Lev. Talen*: Levende Talen. Groningen 1915→
- MoM*: Maal og Minne. Norske Studier. Utgit av Bymaals-laget ved M. Olsen. Kristiania 1909→
- Nachr. Dok.*: Nachrichten für Dokumentation. Frankfurt am Main 1950→
- NTg.*: De Nieuwe Taalgids. Tweemaandeliks Tijdschrift. Groningen 1907→
- Speculum*: A Journal of Mediaeval Studies. Published quarterly by The Mediaeval Academy of America. Cambridge, Massachusetts 1926→
- Spektrum*: Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin (oorspronkelijk alleen *Mitteilungsblatt*; jaargangen 1, 2 en 3 gehektografeerd) 1955→
- Trad. Aut.*: La Traduction Automatique. Bulletin bimestriel de l'Association pour l'Étude et le Développement de la Traduction automatique et de la Linguistique appliquée (ATALA). La Haye-Paris 1960→
- VMKVA*: Verslagen en Mededeelingen der Koninklijke Vlaamsche Academie voor Taal- en Letterkunde. Gent 1887→
- Wet. Tijd.*: Wetenschappelijke Tijdingen (sinds 1951 nieuwe titel van *Wetenschap in Vlaanderen*). Gent 1935→
- ZRPh.*: Zeitschrift für romanische Philologie. Halle 1877→

AFBEELDINGEN BEHOREND BIJ
HOOFDSTUK I



Handponsmachine P. De kaart wordt met de hand in de machine gelegd. De vingerdruk op een der numerieke toetsen drukt mechanisch de ponsstift door de ponskaart op een der numerieke (0 t/m 9) of zoneregels (11 en 12) van een bepaalde, door de kolomaanwijzer op een modelkaart aangegeven kolom. Dit toestel wordt alleen voor incidenteel ponswerk, zoals b.v. het aanbrengen van correcties in de kaarten, gebruikt.



Ponsbandleesapparaat. Leest de ponsband met behulp van een leeskop die is aangepast aan het aantal in de band geponste kanalen. De decodering van de band gebeurt door elektrische circuits, met behulp van een verwisselbaar schakelbord. Het leesapparaat wordt gekoppeld aan een ponsmachine.



Ponsmachine Pelérod MC (Mémoire Contrôlée). Heeft behalve een numeriek, ook een alfabetisch toetsenbord, zodat per letter slechts één toets behoeft te worden aangeslagen. De kaarten worden automatisch in- en uitgevoerd. Dit apparaat, waarop de kaart onder een rij van 12 ponsstiften doorgaat die, kolom voor kolom, de door het leesapparaat uit de ponsband gelezen gegevens in de kaart ponsen, is uitgerust met een aantal (20/30) geheugenposities voor semi-constante alfabetische en numerieke gegevens.



SCR (Sélectionneuse, Contrôleuse, Retourneuse). Rangschikt de kaarten, die bij het ponsen met de bedrukte kant naar boven doorgevoerd, en dus in omgekeerde volgorde afgelegd worden, in de volgorde waarin ze geponsd werden, zodat ze goed gerangschikt door de tabelleermachine kunnen worden afgedrukt. Kan ook selecteren en de nummervolgorde controleren.



Photo-lecteur. Leest automatisch tekens die met de hand in een bepaalde voorbedrukte indeling op een ponskaart zijn geschreven. De aldus gelezen gegevens worden geponst door de reproducerende ponsmachine PRD.



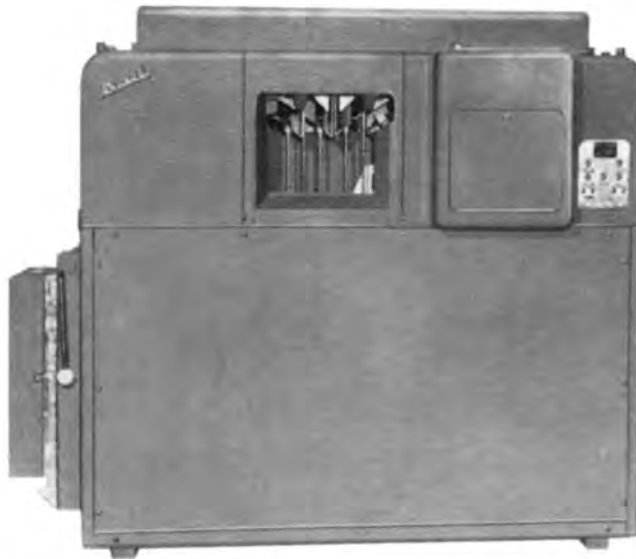
Vertolker. Drukt in gewone letters op de bovenste rand van de ponskaart de gegevens (b. v. woorden) af die in de kaart in de vorm van ponsingen door de ponsmachine zijn aangebracht.



Reproducerende ponsmachine PRD (Poinçonneuse, Reproductrice, Duplicatrice). Voor het reproduceren van gegevens van een kaart in een andere kaart, het dupliceren van bestaande kaarten of het in serie ponsen van gegevens die uit een moederkaart worden gelezen. Het programma der werkzaamheden wordt vastgelegd in verwisselbare schakelborden.



Tabelleermachine BS. Hiermee wordt een staat (b.v. woordenlijst) afgedrukt van de in de ponskaarten geponste gegevens (b.v. woorden), in de volgorde (b.v. alfabetische) waarin de kaarten de leesborstels passeren. Deze gegevens kunnen ook automatisch worden opgeteld (b.v. woordstatistiek). De tabelleermachine werkt met in verwisselbare schakelborden vastgelegde programma's voor het automatisch besturen van de functies. De resultaten van de uitgevoerde optellingen kunnen gelijktijdig door de aan de tabelleermachine gekoppelde totaalkaartenponsmachine PC (Poinçonneuse Connectée) in totaalkaarten worden geponst.



Reporteuse. Ingericht voor het lezen en vertolken van in de ponskaarten geponste gegevens en het afdrukken (maximaal 19 regels per kaartzijde) van de gegevens van verschillende ponskaarten op één enkele kaart (zgn. contexteren).



Sortermachine D 3. Elektronische machine voor numeriek en alfabetisch sorteren. Bij alfabetisch sorteren wordt meer dan de helft der kaarten gesorteerd in de eerste doorgang, de overige gaan naar het restvak en worden gesorteerd in de tweede doorgang. Door een schakelbord kan willekeurig het verband worden gelegd tussen de uit de kaarten af te tasten ponsingen en een bepaald sorteervak.



Tussen- en uitsorteermachine (Interclasseuse). Ingericht voor het in één doorgang samenvoegen van twee stellen kaarten (b.v. de gealfabetiseerde woordvoorraad van een boek met de gealfabetiseerde woordvoorraad van een ander boek) of het uitsorteren van kaarten met bepaalde gegevens. De bewerkingen worden vastgelegd in verwisselbare schakelborden.



Controleponsmachine VIN (Vérificatrice, Intégrale, Numérique). Controleponsmachine met automatische in- en uitvoer. Hiermee worden de door de ponsmachine in de kaarten geponste gegevens automatisch vergeleken, doordat een tweede ponsstypiste de oorspronkelijke tekst een tweede keer op de toetsen aanslaat. Bij ongelijkheid wordt de machine geblokkeerd. Door het indrukken van een speciale toets wordt de blokkering opgeheven, doch tegelijkertijd wordt een foutteken onder de kolom afgedrukt; de foutieve kaarten worden uitgeworpen.

AFBEELDINGEN BEHOREND BIJ
HOOFDSTUK II



Schrijvende ponsmachine (Printing card punch) 026. Combineert het ponsen van de gegevens in de ponskaart met het gelijktijdig vertolken ervan aan de bovenrand van de kaart, en wel direct boven het geponste gaatje. De 026 werkt met 1 ponsing per cijfer (10), 2 ponsingen per letter (26) en 3 ponsingen voor speciale tekens (11) (totaal 47 tekens).



Controleponsmachine (Card verifier) 056. Door middel van toetsen die worden ingedrukt bij het een tweede keer aanslaan van de oorspronkelijke tekst, vergelijkt de controleponsmachine de oorspronkelijke gegevens met de reeds door de ponsmachine geponste gaatjes in de kaarten. Indien de ingedrukte toets geen geponst gaatje vindt, wordt de machine geblokkeerd.



Cardatype Controle-eenheid (Control Unit) 858 (de machine rechts). De 858 leest de kaarten. Aanvullende gegevens worden op het toetsenbord aangeslagen en via de verbinding op de 536 in een nieuwe kaart geponst.



Cardatype Schrijfmachine (Typewriter) 866. Deze schrijfmachine, verbonden met de controle-eenheid 858, maakt, met behulp van de in de 858 ingevoerde tekstkaarten, moederbladen (op speciaal papier) van de verschillende porties context.



Cardatype Schrijvende Ponsmachine (Printing card punch) 536. Deze schrijvende ponsmachine, verbonden met de controle-eenheid 858, pons en vertolkt automatisch woordkaarten, op grond van de in de 858 ingevoerde tekstkaarten.



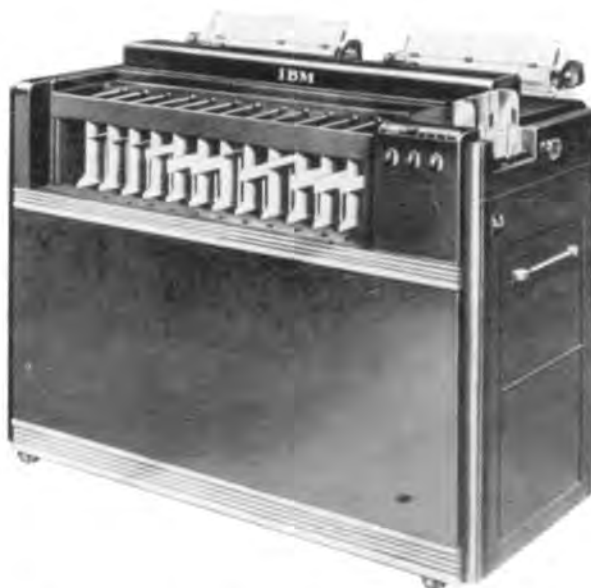
Tabelleermachine (Accounting Machine) 441. Werkt met maximaal 2 ponsingen per kolom. Telt de kaarten en controleert of ze in de goede volgorde liggen. Drukt de gegevens der ponskaarten zonder meer op een lijst af (zgn. lijsten), of voorziet deze gegevens van bepaalde indicaties en totalen (zgn. tabelleren).



Endprinting Reproducer 519. Reproduceert bestaande kaartstellen en voorziet allerlei kaarten van een in grote cijfers vertolkte nummering op de linkerhelft van de ponskaart.



Sorteermachine (Sorter) 082. Rangschikt automatisch de ponskaarten in numerieke alfabetische of alfanumerieke volgorde. Bij het alfabetisch sorteren zijn twee werkingsgangen per kolom noodzakelijk, aangezien een letter door twee ponsingen in één kolom wordt voorgesteld.



Elektronische Statistiekmaschine (Electronic statistical machine) 101. Rangschikt de kaarten in elke gewenste volgorde. De kaart wordt niet kolom voor kolom (in de breedte) maar over alle 80 kolommen tegelijk (in de lengte dus) gelezen; hierdoor kan de machine b.v. tevens de kaarten op woordlengte splitsen. Telt zowel de kaarten als de getallen geponst in deze kaarten. Bijzonder geschikt voor taalstatistisch werk.



Alfabetische Vertolker (Card interpreter) 552. Drukt de in de kaarten geponste gegevens in leesbaar schrift op de ponskaarten af, echter met een lettertype dat groter is dan dat van de schrijvende ponsmachine 026, zodat, in plaats van 80, slechts 60 kolommen op één regel kunnen worden vertolkt.



Card Collator 077. Vergelijkt twee stellen kaarten, voegt ze in een bepaalde volgorde samen, en controleert tevens de volgorde der kaarten. Selecteert kaarten die bepaalde gegevens bevatten.