

A Máquina Eletrônica a Serviço das Ciências Humanas

DOM JACQUES FROGER

(Tradução de Maria José da Luz V. M. de Oliveira)
Bibliotecária (Universidade de Brasília)

A necessidade de recorrer a máquinas surge com evidência se considerarmos o aumento de documentos (livros ou objetos) dos quais as diferentes disciplinas haurem seus dados.

O domínio das pesquisas, com efeito, não pára de crescer. As ciências antigas alargaram o campo de suas investigações. A lingüística, por exemplo, não se interessa mais sômente por certos idiomas, mas por todos aquêles que foram falados e são falados em todo o mundo. A arqueologia dispõe de meios de prospecção até há pouco insuspeitados e utiliza agora o avião e a fotografia aérea para descobrir locais; vimos nascer ao lado da arqueologia terrestre, uma arqueologia submarina que explora o fundo dos mares. Ciências novas constituíram-se em época bem recente, como a demografia, ou mesmo recentíssima, como a psicologia experimental. As ciências "humanas" intervêm em tôdas as outras, e não sômente nas ciências aplicadas e na indústria, mas até nas ciências puras, tais como a matemática, a metodologia e a heurística, pois, os sábios são homens, integrados em grupos, sua atividade intelectual depende da psicologia.

A proliferação da documentação, não é, por assim dizer, espontânea: resulta de esforços deliberados, quase febris. Tem-se consciência hoje dum fato extremamente inquietante se refletirmos sôbre êle: os dados com os quais as ciências trabalham são não sômente perecíveis, mas perecem de fato à medida que o tempo passa. No domínio da lingüística, por exemplo, os dialetos dos nossos países civilizados, ou as línguas dos povos "selvagens", estão em via de desaparecimento. O mesmo acontece com tudo que se refere ao folclore: costumes, música, artes plásticas, trajes etc. As civilizações "primitivas" se apagam e os etnólogos vêem exaurir-se o objeto de seus estudos. Em sociologia, as observações feitas num dia, no nosso mundo que evolui ràpidamente, perdem sua atualidade em poucos anos e entram no domínio da história. Os arqueólogos lamentam que um grande número de lugares tenham sido pilhados por mãos inexperientes; tomam-se, é verdade, precauções para que as escavações sejam reservadas aos especialistas; mas os próprios especialistas não podem se gabar de extrair dum local tôdas as revelações que êle contém; em arqueologia, descobrir é sempre, de um certo modo, destruir. Muitos manuscritos antigos desapareceram por culpa dos humanistas dos séculos XV e XVI, refugados como inúteis após terem sido publicados em edições impressas, ou dispersos por ocasião da morte do seu proprietário. O cuidado que se dedica hoje para se conservar os velhos pergaminhos e os papiros não os impede de sofrer acidentes que lhes reduzem progressivamente o número. Os monumentos tombam em ruínas, a acidez do ar carcome as estátuas. Co-

nhecemos bastante os desastres que as guerras ocasionam; mas não levando em consideração as catástrofes, basta um acidente banal como um incêndio ou um furacão, para provocar desgastes irremediáveis. Em suma, os objetos de toda espécie sofrem fatalmente a deterioração do tempo. Tem-se portanto suficiente razão para recolher e inventariar os traços do passado antes que eles se dissipem; pode ser mesmo que não nos apressemos suficientemente; é para temer que os séculos vindouros nos acusem de negligência.

Considerado o progresso das ciências antigas e o surto das novas, o número de documentos aumenta com rapidez surpreendente. Todos os anos publicam-se perto de dois milhões de artigos de periódicos e muitas centenas de milhares de livros: é provável que esta produção vá decuplicar-se dentro de alguns anos. A documentação aumenta segundo uma progressão geométrica e sua proliferação apresenta um aspecto exponencial, ou seja, que ela cresce a cada instante de modo proporcional ao estado que alcançou, ou de maneira proporcional à sua função crescente desse estado; ela acompanha uma curva que inicialmente quase horizontal, se eleva cada vez mais até tomar finalmente uma direção quase vertical (1).

Esta curva do crescimento da documentação corresponde muito de perto, como é natural, à das descobertas científicas e à do número de sábios. Como nota Robert Oppenheimer (pelo menos é uma declaração que se lhe atribui), “nove décimos dos gênios que a humanidade engendrou desde seu nascimento estão vivos hoje. Fizemos mais progresso em quarenta anos que em quarenta séculos. 99% dos nossos conhecimentos são devidos a homens vivos atualmente” (2). A população mundial aumenta também de modo acelerado, se bem que a “demografia documentária”, a “demografia das descobertas”, a “demografia dos sábios” e a “demografia”, em si, estejam quase emparelhadas.

Não teríamos razão, na documentação como nos outros domínios, de considerar como desesperador êste crescimento exponencial. Quando se trata de curvas desse gênero é perigoso extrapolar. Acontecimentos imprevisíveis podem sobreviver e modificar a evolução (3). É assim que a ciência grega desde os séculos VII-VI a.C. até os primórdios da era cristã, seguiu uma curva da mesma natureza exponencial que a de nossos dias, aparentemente; em seguida os “bárbaros” romanos vieram abaixar o nível da cultura e moderar o progresso; enfim os outros “bárbaros”, que, invadiram o Império Romano, aproveitando-se de sua debilidade, provocaram uma espécie de derrocada. Reiniciada quase de zero na Europa Ocidental, a curva das principais descobertas científicas não se torna regular senão a partir do século X; ela recupera lá pelos séculos XII-XIII o nível em que os gregos da antiguidade a deixaram; sua ascensão, tornada hoje quase vertical, infletrá talvez para a horizontal, caso se produza uma como que saturação; uma guerra mundial onde se empregasse a bomba atômica fa-la-ia descer de novo como a fizeram durante o primeiro milênio da era cristã, as invasões bárbaras que se seguiram. Mesmo que não sobrevenha catástrofe, o crescimento vertiginoso, a que assistimos, do número de livros, de ciências, de descobertas, de sábios e de vestígios do passado reencontrados, não conservará talvez indefinidamente sua característica exponencial.

Haja o que houver no futuro longínquo, a expansão prodigiosa que verificamos nos nossos dias e que sem dúvida não está próximo de interromper-se faz passar para o primeiro plano o problema da documentação.

O número já grande, constantemente crescente, dos materiais que as diversas ciências utilizam obriga antes de mais nada a colocá-las em ordem, mercê de uma classificação ao menos aproximativa (4).

Os livros são reunidos nas bibliotecas; se seu número é relativamente escasso, podemos classificá-los por matérias nas prateleiras. Todavia, êsse procedimento, na verdade muito cômodo, não satisfaz inteiramente, pois, inúmeros livros interessam a diversas disciplinas ao mesmo tempo; além do mais, é impraticável nas grandes bibliotecas, onde os acervos vêm se ajuntar uns aos outros ao sabor das circunstâncias, de modo que a ordem dos livros é, em grande escala, arbitrária. Em qualquer hipótese, uma biblioteca não pode ser utilizada diretamente: é preciso desdobrá-la numa biblioteca "fantasma" ou catálogo.

O elemento do catálogo não é mais o livro, mas uma nota sinalética que lhe dá as características: autor, título, editor, data etc. Infelizmente um catálogo nunca representa, na prática, o conteúdo integral de uma biblioteca; deixa de lado os artigos de revistas e as monografias públicas em conjunto como partes duma obra mais vasta; êle assinala a existência de publicações com o caráter de coleção, mas não pormenoriza o conteúdo, por que a coleta demandaria um trabalho muito longo.

O catálogo reveste-se da forma duma lista ou dum fichário. A lista, reunida num volume, é cômoda e facilmente publicável, mas tem o inconveniente de ser um catálogo "fechado", ou seja, restrito à data da sua redação; para pô-lo em dia, precisamos juntar-lhe um suplemento, que deverá, êle mesmo, receber mais tarde outro suplemento, e assim, ao infinito. O fichário tem a vantagem de ser um catálogo "aberto"; pode-se mantê-lo constantemente em dia, apenas inserindo nos seus lugares novas fichas.

Seja qual fôr sua apresentação, o catálogo cria outro problema: em que ordem serão dispostos os elementos que o compõem? A mais simples é a ordem alfabética por nomes de autores (ou por títulos para as obras anônimas); mas esta solução cria embaraços ao consulente que deseja compilar uma bibliografia: é necessário saber de antemão quais autores trataram da questão que interessa, o que pressupõe o problema resolvido. Para evitar êsse círculo vicioso, uma classificação metódica, ou por assuntos, impor-se-ia, mas é de tal forma difícil de realizar que se a deixa freqüentemente, mesmo nas grandes bibliotecas, em estado rudimentar.

Os objetos transportáveis são reunidos nos museus ou em coleções. É mais fácil classificá-los "nas estantes" segundo sua natureza, sua proveniência, sua data etc., mas o problema do catálogo aparece tanto para êles como para os livros, ainda que em condições diferentes; a ficha sinalética menciona suas principais características segundo um "código" preestabelecido. Bem entendido, os objetos que se não podem transportar, como os monumentos, permanecem no local e não se pode reunir nos índices outra coisa além dos seus "fantasmas".

Pôsto que os objetos que nos foram legados pelo passado são em número finito, nós podemos levantar uma lista completa e constituir um corpus (*) que

(*) N.T. Corpus (documentário). Todo conjunto de documentos considerados numa experiência de análise e/ou de pesquisa automática (5) de informação. Apud: Cros. R. — C. Gardin, J. — C.; Levy, F. L'Automatisation des Recherches Documentaires; un modèle général "le syntol". Paris, Gauthier — Villars; 1964. p. 237.

reúne num mesmo lugar tudo o que está disperso pelo mundo. Por exemplo os manuscritos antigos e medievais são em número finito, e, se nós não os conhecemos ainda todos, temos pelo menos certeza de que não irão aparecer novos; pode-se portanto fazer-lhes uma enumeração completa. É assim que M. LOWE, nos Estados Unidos, publica nos seus **Codices Latini Antiquiores**, o **corpus** de todos os manuscritos latinos anteriores ao século IX; na França, MM. SAMARAM e MARICHAL, editam os de todos os manuscritos datados, posteriores ao século IX no **Catalogue des manuscrits portant une indication de copiste, de lieu ou de date**; um dia sem dúvida compilar-se-á o repertório de todos os manuscritos latinos, sem exceção. Os manuscritos gregos, muito menos numerosos que os latinos, são mais fáceis de catalogar de modo completo, e é nisto que se aplica o Sr. Abade RICHARD do Instituto de pesquisa e de história dos textos do C.N.R.S. (*) O mesmo acontece com todos os tipos de objetos; o paleógrafo suíço M. BRUCKNER e M. MARICHAL publicam juntos, nas **Chartae latinae antiquiores**, o **corpus** de todos os papiros latinos anteriores ao século IX; Mme. GAUTHIER colige em Limoges o **corpus** dos esmaltes meridionais; a Universidade de Princeton, nos Estados Unidos, reúne no seu **Index of the Christian Art** tudo que se refere à arte e à iconografia cristãs; o **corpus** dos incunábulo foram estabelecidos na Alemanha por HAIN e uma vez que as publicações aparecem desde o século XVI são a cada momento em número finito, pode-se conceber um fichário "universal" que englobaria o conteúdo de tôdas as bibliotecas do mundo e, constantemente atualizada, daria a cada instante o estado da ciência libresca da humanidade. A constituição do **corpus** é uma das características de nossa época; estas compilações, bem entendido, não serão jamais absolutamente exaustivas; tenta-se pelo menos torná-las tão completas quanto possível.

A informação estando assim reunida, resta-nos utilizá-la: a primeira coisa a fazer, assim que se empreende uma pesquisa, é constituir uma bibliografia do assunto, para saber onde se encontra a questão e sobre quais pontos dever-se-ão aplicar os esforços para fazê-la avançar. O problema da bibliografia não se apresenta sob o mesmo aspecto em tôdas as disciplinas. As que têm um caráter "científico" e procuram resolver problemas novos pela formulação de leis não necessitam mais que obras recentes; só lhes importam as últimas descobertas e os livros que as relata; uma publicação perde todo seu interesse assim como a outra renova a questão; a bibliografia se descarta de elementos caducos à medida que a ciência progride; ela se rejuvenesce por um processo de substituição constante, e suas variações dizem respeito mais à sua composição que à quantidade. Acontece diferentemente nas disciplinas "históricas", onde nos esforçamos para reconstituir o passado segundo a evolução e o encadeamento dos fatos: a bibliografia não deve negligenciar coisa alguma, tudo que está escrito tem valor de documento, e não existem livros a serem refugados. Entretanto, se as dificuldades da documentação não são as mesmas nas diversas disciplinas, também não se pode dizer que elas sejam menores em umas que em outras.

Com efeito, os documentos não são classificados a não ser de modo aproximativo nas bibliotecas, nos museus e nos **corpus**. Assim, os **corpus** de manuscritos os alinham pela ordem de suas bibliotecas de depósito; para obter sua classificação por datas, por origens, por tipos de escrita etc., é preciso recortar as páginas para transformar estas publicações em fichários; aliás foi o que previram MM.

(*) N.T. C.N.R.S. É a sigla do Centre National de Recherche Scientifique (Centro Nacional de Pesquisa Científica). É o Centro de Documentação pública da França.
Pronuncia-se: cê-énn-érr-éss.

SAMARAM e MARICHAL, tendo o cuidado de oferecer ao público os “manuscritos datados” sob a forma de folhas soltas impressas de um só lado. Por outro lado, para ajudar os pesquisadores, revistas como *Scriptorium* ou a *Revue d'Histoire ecclésiastique* na Bélgica dão “bibliografias”; mas para utilizá-las comodamente, seria necessário dividi-las também para transformá-las em fichários. Por que essa obstinação de empregar até os dias de hoje o sistema medieval de listas? Não dever-se-ia resolver de uma vez por tôdas a publicar fichas, eventualmente em diversos exemplares, para que cada qual as classifique a seu modo?

Em suma, encontramos-nos presentemente diante de uma documentação muito copiosa, mal classificada e que aumenta prodigiosamente. Estamos esmagados sob a massa dos materiais, e sua abundância é tal que não é mais possível tratá-los apenas “à mão”: é preciso recorrer às máquinas.

* * *

Mas a que máquinas? Elas são de três categorias: o fichário tradicional, a máquina mecanográfica, a máquina eletrônica (5).

O fichário tradicional, com suas fichas alinhadas segundo palavras-assunto ou palavras-chave hierarquizadas que permitem múltiplas classificações é já uma “máquina”: ela é construída e acionada à mão, eis tudo.

A máquina mecanográfica é em suma uma separadora, que manipula fichas. Ela pode ser acionada por não importa qual fonte de energia, e nós podemos fazê-la funcionar, virando uma manivela; de fato, utilizamos um motor elétrico, e é por conseguinte uma “máquina elétrica”, ou seja, onde a eletricidade desempenha apenas o papel de propulsão, enquanto que as operações são efetuadas por “mecânica” (rodas, engrenagens, alavancas etc.). Para que as fichas possam ser agarradas e manipuladas pela máquina, as indicações que nelas se encontram eventualmente escritas, distintamente são representadas de acôrdo com certas convenções, por perfurações no interior, ou entalhes no contôrno etc.

A máquina eletrônica se distingue de uma máquina elétrica pelo fato de a eletricidade não ter mais somente o papel extrínseco de propulsor em relação à uma mecânica que execute o trabalho, mas efetua ela própria as operações: o agente operador é o fluxo de elétrons que constitui a corrente elétrica; o instrumento de trabalho é a impulsão magnética produzida pelos elétrons, e a “mecânica”, em tal máquina não desempenha mais que um papel subalterno e auxiliar. À entrada, as informações (letras ou algarismos, sinais convencionais etc), são incontinenti expressas em perfurações, segundo um código convencional, seja sobre cartões, seja sobre fitas de papel. Estes suportes perfurados deslizam, na máquina, entre uma superfície condutora e as “vassouras” que consistem em um sistema de lâminas de aço; suas partes não perfuradas interceptam a corrente elétrica, enquanto que as perfurações a deixam passar; assim a linguagem da máquina é binária e se comporta por “tudo ou nada”, sinal ou ausência de sinal; é a razão principal porque se utiliza no cálculo, seja um sistema puramente binário (numeração de base dois), seja um sistema decimal binário (em uma numeração de base dez, cada um dos algarismos é expresso binariamente). As informações são registradas em fitas magnéticas introduzidas na “memória” da máquina que se apresenta sob diferentes formas: tambores, discos etc. As operações que a máquina deve efetuar lhe são indicadas sob a mesma forma de perfu-

rações da informação a ser usada, e seu conjunto constitui um "programa". À saída, o resultado das operações é impresso "claramente", ou seja em letras e algarismos comuns, eventualmente sob forma de curvas ou de diagramas. Ao lado das máquinas "digitais", que operam com algarismos (do inglês *digit*) ou grandezas descontínuas, existem máquinas "analógicas" que operam com grandezas contínuas, mais ou menos como as régua de cálculo. Neste domínio da técnica, os progressos são rápidos; um computador em poucos anos está "ultrapassado", e podem-se entrever realizações cada vez mais surpreendentes (6).

Existe hoje em dia uma gama completa de máquinas eletrônicas, desde as pequenas tabuladoras até aos computadores extremamente poderosos. Quando se trata de escolher um tipo de máquina para um trabalho determinado, seria ilusório acreditar que em qualquer hipótese a mais aperfeiçoada seja a mais vantajosa. Dois pontos devem ser levados em consideração: a rapidez e o preço, ou, se se prefere, a despesa de tempo e de dinheiro; adota-se a solução que se mostra mais econômica sob todos os aspectos.

O aparelho mecanográfico manipula as fichas mais rapidamente do que a mão. A máquina eletrônica é ainda muito mais veloz, e é talvez por isso que ela mais confunde a imaginação. Certamente, é exagero dizer que ela trabalha "com a velocidade da luz"; a preparação manual é as vezes bastante longa, eventualmente mais longa que os próprios cálculos; há tempos perdidos, e uma parte de mecânica intervém: desenrolamento de fitas magnéticas, rotação dos tambores ou dos discos de memória, impressão dos resultados etc. Apesar destas limitações, a máquina executa cálculos a uma velocidade que não tem proporção alguma com o trabalho feito à mão. Por exemplo a calculadora IBM 7090 (construída em 1961) pode efetuar, com vírgula fixa, 229.000 adições ou subtrações em um segundo. O cálculo das diferenças entre os valores teóricos e os valores medidos do campo de gravitação terrestre demandam: a mão, com papel e lápis: 1.000 anos. Com uma máquina de calcular de escritório: 5.000 semanas; com uma calculadora eletromagnética (1944): 3.750 dias; com uma calculadora eletrônica IBM (1948): 50 horas; com computador IBM 704 (1957): 75 minutos.

Em contrapartida, o trabalho com a máquina eletrônica custa caro. Os custos não variam da mesma forma nas diferentes fases. As despesas exigidas para a perfuração dos cartões ou das fitas na entrada e a impressão dos resultados na saída são aproximadamente os mesmos para tôdas as máquinas. O estudo e a preparação dos programas são tanto mais custosos quanto mais possante fôr a máquina. Os custos que acarreta o tratamento propriamente dito são inversamente proporcionais à importância da máquina, quer dizer que uma operação elementar custa tanto menos caro quanto mais possante fôr a máquina; o rendimento das máquinas grandes é portanto melhor que o das pequenas. O preço de compra de uma máquina é evidentemente tanto mais elevado quanto maior ela fôr e o mesmo acontece com o preço do aluguel; digamos, para abalzar a escala dos preços, que uma hora-máquina custa: num pequeno tabulador IBM: 20 F; numa máquina mais poderosa: 100 F; numa Gamma 30 Bull: 500 F; nas maiores máquinas que existem hoje: 5.000 F. Mas uma vez que o preço crescente da hora-máquina é compensado por ganho de tempo e melhor rendimento, trata-se somente de encontrar a solução onde o equilíbrio se estabelece. Os cálculos enormes são bem menos caros nas máquinas maiores que nas pequenas; por outro lado, não se tem interesse em fazer trabalhar uma máquina muito

abaixo da sua potência máxima. O problema consiste então, para escolher uma máquina, em encontrar aquela que seja exatamente proporcionada ao trabalho que se pretende fazer executar.

Convém acrescentar que freqüentemente a máquina eletrônica é não apenas mais dispendiosa, mas ainda menos rápida que a máquina mecanográfica, (7) a qual por sua vez pode muito bem não oferecer vantagem alguma sobre o trabalho manual. Experiências realizadas precisamente para esse fim mostraram que, para lidar com número de cartões indo até diversos milhares é vantajoso operar à mão, sobre a mesa, comparando as fichas por transparência ou de uma outra maneira muito simples: faz-se em poucas horas, com despesa mínima, um trabalho que demandará muito mais tempo e acarretaria custos muito mais elevado se o confiássemos à uma máquina.

Em suma, o bom senso manda que nos entreguemos a uma pequena "pesquisa operacional" na hora de executar um trabalho e recorrer ao sistema que convém melhor, dadas as circunstâncias, sem acanhamento de trabalhar à mão se fôr o mais simples.

* * *

Admitamos que a massa das informações justifique o emprêgo de uma máquina eletrônica, ficando bem claro que nossas observações aplicar-se-ão, com as devidas modificações, também às máquinas mecanográficas. O fichário que se organiza então, em "documentação automática" (8) apresenta diferenças mais quantitativas que de natureza, com o fichário tradicional organizado a mão; muito mais rico em dados, oferece, graças às suas entradas múltiplas, possibilidade de combinações muito mais amplas e permite executar classificações que se aplicam a muito maior número de elementos; mas enfim êle também é composto de palavras-chave dispostas hierarquicamente (9).

Os processos da documentação automática não são os mesmos conforme se trate de dados "literais" ou "não literais". Estes últimos são objetos; descrevemo-los segundo um código que assinala suas características, e em seguida o trabalho se assemelha ao cálculo. No momento em que nós encontramos em presença de dados "literaris", ou textos redigidos em uma "língua natural" (língua no sentido dos lingüistas: francês, alemão etc), a pesquisa documentária, ou coleta de documentos, é seguida de uma operação delicada: "a análise documentária" que traduz em "linguagem de informação" o texto escrito em uma língua natural. Procedese a uma "indexação" que conduz a um conjunto formado de dois elementos: duma parte as palavras-chave, cuja lista completa é um "léxico de documentação" compreendendo todos os vocábulos técnicos empregados em dita disciplina de outra parte, os símbolos que representam as relações semânticas e sintáticas das palavras-chave. É claro que um resumo ou uma lista de palavras-chave não pode condensar todo o conteúdo de um livro ou de um artigo de livro e retém deles apenas o principal. Produz-se então, nesta fase, uma perda de informação. É um mal com o qual nos devemos resignar, porque êle é inevitável. Uma vez realizada esta indexação em linguagem documentária, a máquina a registra sob uma forma simbólica que lhe é peculiar e na sua própria linguagem.

A dificuldade consiste em estabelecer a lista das palavras-chave na qual se resume o conteúdo de um livro ou de um artigo, e que permitirá preparar

bibliografias classificando sob uma rubrica tôdas as obras que tratam de um assunto dado. Procurou-se fazer automaticamente esta indexação recorrendo a meios rápidos. Mas o título não basta, pois êle é freqüentemente muito vago e não diz com exatidão aquilo de que se trata. Não podemos nos basear nos vocábulos mais recentes; a experiência mostrou que êles ordinariamente não têm relação alguma com o assunto de uma obra. Tentou-se também um método por citações: admitindo que um autor cite normalmente livros ou artigos que trataram do mesmo assunto antes dêle, pensou-se que agrupando as obras que se citam umas às outras, chegar-se-ia a constituir um conjunto bibliográfico respeitante à mesma matéria; não parece que tal método veja dar bons resultados. De modo geral, todos êsses processos rápidos que tentam "contornar" (*) a indexação, evitam o problema em vez de resolvê-lo. Nenhum dêles traz satisfação plena ou completa. Após diversas tentativas infrutíferas, nesse particular estamos limitados ainda hoje em dia a nos contentarmos com resumos e indexações feitos à mão por algum leitor inteligente. A documentação inteiramente automática será talvez realizada um dia, mas não se pode esperar que isto aconteça em futuro próximo.

* * *

Quando se introduz as palavras-chave na máquina, vê-se desde logo apresentar-se o problema duma tradução automática; não obstante, a documentação fôsse estabelecida ainda que tão perfeitamente quanto possível por uma máquina, não satisfará completamente as necessidades dos pesquisadores a não ser no dia em que a tradução automática a complete. Com efeito, uma vez compilada a bibliografia de um assunto, resta consultar as obras que ela menciona; mas as escritas numa língua que o usuário ignora permanecem-lhe inacessíveis, e é para êle como se não existissem. Êle deve aguardar que sejam traduzidas, mas os tradutores competentes são difíceis de achar, êles trabalham lentamente e custam caro. De fato, as línguas geram uma espécie de barreira intelectual entre as principais partes do mundo.

Apenas uma máquina tradutora poderia suprimir êsses obstáculos muito incômodos e permitir que todos os trabalhos importantes sejam postos sem demora à disposição de todos os pesquisadores, seja qual fôr a nacionalidade dêles. As diferentes disciplinas progredirão mais veloz e desembaraçadamente (10).

O primeiro projeto de tradução automática foi criado em 1933 pelo Russo Smirnov-Trojanskij, mas nem seu país nem os demais acolheram-no favoravelmente. Em 1946, o Inglês A. D. Booth, que provavelmente ignorava os trabalhos de seu predecessor russo, concebeu por sua vez a idéia duma tradução automática; propôs a Warren Weaver, da Fundação Rockefeller, construir uma máquina tradutora, e sua sugestão foi recebida com entusiasmo. Booth imaginou inicialmente a tradução automática como se fôsse a decodificação de uma mensagem cifrada, mas percebeu logo que o problema era bem menos simples e dependia, antes de mais nada da lingüística. Esta disciplina tinha feito então grandes progressos, dos quais se beneficiaram as pesquisas de Booth e Weaver. Eletrotécnicos e lingüistas colaboraram estreitamente e, em 1952, realizaram nos Estados Unidos sua primeira reunião para estudar juntos os problemas da tradução a máquina. Em 1954, teve lugar, na Universidade de Georgetow, a primeira experiência de tradução automática do russo para o inglês; essas duas línguas

(*) N.T. "Court-circuiter", no original.

mantiveram o primeiro lugar nas preocupações dos tradutores, porque elas podem passar tranqüilamente como as mais representativas do "Oriente" e do "Ocidente", as primeiras experiências e os trabalhos mais ativos sendo além disso devidos a pessoas de língua inglesa. Sòmente em 1955, ano em que apareceu o primeiro trabalho sòbre a questão, publicado por William Locke e A. D. Booth, os russos se aplicaram também ao trabalho para não ficar atrás da América. Desde então, as pesquisas prosseguiram não sòmente nos Estados Unidos, na Inglaterra e na Rússia, que são os três países "dirigentes" neste domínio, mas também na França, na Itália, na Escandinávia, no Japão, e, em suma, em todos os países de grande cultura (11).

A despeito dos esforços consideráveis empregados no mundo inteiro, a tradução automática não está ainda realizada até hoje. Conseqüentemente, o público experimenta um sentimento de decepção ao constatar que a máquina tradutora, que êle acreditava quase já realizada ou a ponto de o ser, não funciona ainda e nem mesmo promete esperança de realização a curto prazo. Mas os especialistas, que para dizer a verdade têm passado também por uma fase de otimismo um pouco exagerado, se mostram sempre mais reservados; êles o estão hoje mais do que nunca, porque sua tarefa apresenta-se mais dificultosa à medida que os trabalhos prosseguem. Sua atitude prudente não é entretanto presságio de fracasso. Sem dúvida, se se percorrem as revistas ou as compilações de monografias, as tentativas parecem à primeira vista um tanto desordenadas e dão mesmo a impressão de certa confusão. Mas isto não passa de aparência. Os especialistas divergem em muitos pontos, é claro, mas estão de acòrdo sòbre as linhas mestras do caminho a seguir. Não é pois totalmente exato que a tradução automática esteja hoje em dia "num impasse", como às vèzes se ouve dizer; podemos falar sòmente de uma fase de "estagnação". Não obstante a lentidão do progresso não há nada de surpreendente se considerarmos as dificuldades a serem vencidas.

Para fazer um texto pela máquina, nós lho damos de imediato, como de costume, sob forma de perfurações; ao mesmo tempo abastecemos-la de dicionários e de regras de gramática. A máquina deve reconhecer, na língua de partida, o sentido dos vocábulos e suas relações sintáticas, em seguida encontrar, na língua de chegada, os vocábulos de mesmo sentido e ordená-los segundo a sintaxe que convém. Dessas duas fases, a segunda não é, parece, a mais complicada: assim que a máquina possua o vocabulário de uma língua e as regras de sua sintaxe, não lhe será muito difícil construir frases corretas. Para ela, o mais penoso consiste em "compreender" a língua de partida, e isso porque não conseguimos ainda formular em têrmos discursivos, para uso da máquina, as operações mentais que fazemos em boa parte intuitivamente quando lemos. A função gramatical dos vocábulos numa frase nem sempre nos é indicada por regras precisas, e nós a deduzimos do sentido geral. Nas línguas sem declinação, como o francês, o inglês, o espanhol etc., em princípio é a colocação do sujeito e do complemento em relação ao verbo que os distingue; mas é possível haver inversões aí; elas não nos atrapalham porque o sentido nos guia. Se nós encontramos por exemplo os vocábulos "rói", "osso", "cachorro", sabemos, seja qual fôr a ordem dos mesmos, que é o cachorro que rói o osso, e não o osso que, rói o cachorro; mas como explicar isso para a máquina? As dificuldades semânticas são talvez mais graves ainda. Resultam principalmente da polissemia, ou

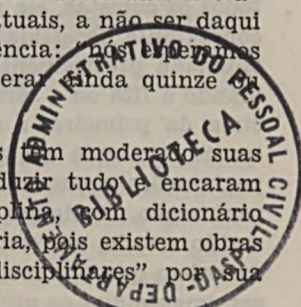
pluralidade de sentidos de que gozam certos vocábulos, os mais usuais sendo por infelicidade aquêles cujas significações são as mais variadas. A presença de homógrafos, ou vocábulos diferentes que se escrevem da mesma maneira, contribui para desnortear a máquina, da mesma forma que os "lexis" ou grupos de palavras que forma bloco (a respeito de) mas que vocábulos intercalares dissociam às vêzes (a respeito precisamente de) ou ainda: (a respeito de... e de...).

Cada língua apresenta problemas específicos; em alemão, por exemplo, existe uma dificuldade quanto às palavras compostas, que é preciso fazer "decompor" pela máquina, o que é particularmente delicado quando um vocábulo apresenta um sentido diferente conforme nós o separemos duma maneira ou de outra atribuindo ao comêço ou ao final do vocábulo composto a letra discriminante; assim **Wachtraum** significa "sonho acordado" se o cortamos **Wach/traum**, ou "sala da guarda" se o cortamos **Wacht/raum**. Em espanhol, é preciso achar o meio de distinguir do verbo o pronome complemento que se lhe junta: por exemplo, **dale**, "dê-lhe", que a máquina deve interpretar **da/le**. Para resolver as ambigüidades semânticas, pede-se à máquina para explorar o contexto; assim **pièce** em francês traduzir-se-á por **gun** em inglês e **cañon** em espanhol se se tratar de artilharia nas proximidades, por **coin** em inglês e **moneda** em espanhol se se trata de finanças, por **room** em inglês e **habitación** em espanhol se o contexto fala de uma casa etc. Apesar de tudo, as dificuldades são tais que a máquina tradutora, ainda hoje, engana-se perto de duas vêzes em três, de modo que as ambigüidades, longe de serem residuais, como pretendem os autores otimistas, são ao contrário o caso geral.

Face a dificuldades assim penosas, e para chegar mais velozmente a resultados, imaginaram-se métodos destinados a poupar para a máquina uma parte de seu trabalho fazendo-o "à mão"; consistem notadamente seja em resolver de antemão as ambigüidades na língua de partida, graças à uma "pré-edição" ou indexação prévia; seja em resolvê-la na língua de chegada, graças à uma "pos-edição", escolhendo entre as diversas soluções propostas pela máquina. Hoje renunciamos a êsses processos. Como diz M. Gardin, a tentativa para servir-se de atalhos e "curto-circuitar" (*) a tradução automática são, aqui como na documentação, uma perda de tempo e desviam a atenção do verdadeiro problema. Mais vale progredir na "estrada real" que consiste em atacar de frente as dificuldades para achar-lhes a solução teóricamente correta. Não devemos ser pessimistas nem nos desencorajarmos: "admitido que o espírito humano faz certas operações, não há razão para que não consigamos fazê-las executar pela máquina". Bastará sòmente analisar com cuidado os passos do espírito humano que lê e compreende um texto e aprofundar então os estudos lingüísticos. Isso será demorado. A tradução automática não desabrochará, segundo as provisões atuais, a não ser daqui a uns quinze anos aproximadamente. Tenhamos pois paciência: "há séculos milênios, diz-me sorridente M. Gardim; bem podemos esperar ainda quinze vinte anos".

Notemos que os construtores de máquinas tradutoras com moderado suas ambições; renunciaram a fazer máquinas capazes de traduzir tudo e encaram sòmente máquinas especializadas em determinada disciplina, com dicionário apropriado. É de se esperar que esta restrição seja provisória, pois existem obras de síntese, e muitas das ciências humanas são "interdisciplinares" por sua

(*) "court-circuiter", no original



própria natureza: a sociologia ou a demografia, por exemplo, recorrem ao mesmo tempo à história, ao direito, à psicologia, à medicina, à matemática (estatísticas) etc. Seja como for, diga-se que a tradução automática não substituirá a tradução "literária", pois não poderá jamais traduzir as sutilezas do estilo: tudo que lhe pedimos é que reproduza exatamente o sentido.

* * *

As vantagens da máquina a que mais freqüentemente damos importância são a rapidez e a segurança. Nós já falamos da primeira; resta apenas fazer algumas reflexões sobre a segunda. A máquina eletrônica é infalível, diz-se, e jamais comete erros: isto é perfeitamente verídico, desde que se tomem certas precauções.

Os erros imputáveis à máquina praticamente não são mais de reear; utilizam-se, para evitá-los, meios tão eficazes que se pode considerá-los como totalmente eliminados. Quando se produzem erros, a culpa é, na maioria das vezes, não da máquina, mas do homem que a utiliza.

Ele pode cometer erros de lógica na maneira de colocar os problemas, na elaboração dos métodos, ou na interpretação dos resultados. A máquina não é evidentemente responsável por isso; ela executa um trabalho, mas não diz se o método que lhe impuseram é bom; responde a perguntas, mas não diz se são pertinentes; apresenta os fatos, mas deles não tira conclusão alguma. Teremos ocasião de voltar sobre este assunto a propósito dos serviços que a máquina pode prestar nestes ou naqueles domínios, por que importa utilizá-la judiciosamente para dela tirar o melhor partido.

Erros materiais podem surgir cada vez que o homem é chamado a intervir. Para remediar isso, automatizamos o mais possível as operações de maneira que uma vez desencadeadas se desenvolvem sozinho. Mas a intervenção humana é indispensável precisamente para iniciar as operações e preparar os trabalhos: é aí que os erros são de temer. A bem dizer, os riscos não dizem respeito ao programa, pois a máquina o verifica ao mesmo tempo que o executa e aponta as incoerências ou os equívocos, se os encontra; eles têm por objetivo principalmente as informações submetidas a execução, e a fase perigosa é a da perfuração.

O aparelho graças ao qual perfuramos os cartões ou as fitas é uma grande máquina de escrever elétrica, que é disposta e funciona como todas as outras, mas cujos toques são ligados a um aparelho anexo que efetua as perfurações segundo o código previsto. Lendo o texto distintamente à medida que perfura, a secretária não trabalha "no escuro"; entretanto, como todas as datilógrafas e em geral todos os copistas, ela pode cometer enganos — e efetivamente os comete. Há por conseguinte necessidade de verificação. Construiu-se para esse fim uma máquina chamada precisamente "verificadora", que funciona da seguinte maneira: retomando a fita ou a série de cartões perfurados, é refeita uma nova perfuração por cima da primeira, e se esta não coincide com aquela, a máquina se recusa a continuar; onde ela parar aí existe um engano, seja numa ou noutra das duas perfurações sucessivas; vê-se em que ele consiste, e se for devido à primeira perfuração, corrigimo-lo. A correção é mais fácil se usamos cartões: refaz-se o cartão errado, anexando se necessário um cartão bis, que traz o mesmo número de ordem que o precedente, com um índice para distingui-lo do outro. Admite-se geralmente que as probabilidades de erro por ocasião da primeira perfuração são

inferiores a 1%; são as mesmas por ocasião da segunda; as duas percentagens se combinam, de modo que depois de ter feito passar um texto na "verificadora", os riscos de êrro seriam cêrca de um sôbre dez mil, o que é desprezível.

Todavia, a experiência mostra que essas estimativas são talvez muito otimistas. Em 1960, tentamos, na Compagnie Bull, realizar na máquina eletrônica as operações de crítica textual que os filólogos fazem à mão. Tratava-se de cotejar imediatamente diversos manuscritos duma mesma obra, para restabelecer por meio de uma comparação "automática" suas diferenças de teor, ou "variantes". (12) Pois a segunda perfuração pela "verificadora" em nada diminui a porcentagem de enganos; ao contrário aumentou-a.

A noção de "margem de êrro tolerável" é talvez admissível em circunstâncias em que se está seguro de que os desvios são mínimos e não podem trazer um prejuízo real; mas, em princípio e de modo geral, os dados fornecidos à máquina devem ser de exatidão absoluta. Entre as vantagens do tratamento automático, a rapidez é apenas secundária em relação à segurança; a velocidade não tem mesmo interêsse algum a não ser que os resultados sejam perfeitamente justos. Para eliminar os erros cometidos pela máquina, precisei recorrer ao único meio verdadeiramente eficaz e que é de regra especialmente na Section d'automatique documentaire de Marseille, dirigida por M. Gardin; cotejar à mão com os originais os textos saídos da máquina e pedir novas provas, como na tipografia, até que tudo seja perfeitamente exato. Com efeito, as correções que eu havia solicitado foram efetuadas de primeira vez de modo impecável; eu me pergunto então se não teria sido mais simples dedicar-me logo em seguida a extração de primeiro jato: teria havido menos enganos para corrigir. A máquina "verificadora" não seria mais prejudicial que útil, dando falsa impressão de segurança? Cabe aos especialistas julgar isso; não obstante êles sabem muitíssimo bem que é necessário proceder ao menos a duas verificações para se estar completamente garantido.

* * *

Entre as vantagens da máquina, existe uma sôbre a qual não se insiste de ordinário bastante, mas que no entanto não é a menor delas: é a sua falta total de inteligência. Incapaz de elaborar um método, faz aquilo que se lhe manda fazer, nada mais, nada menos, de maneira tôda material. Inapta a pensar, obriga aquêle que dela se serve a fazer um esforço bem maior de pensamento para analisar com cuidado as operações que lhes serão confiadas. Mesmo no trabalho a mão, deveríamos em princípio usar sempre métodos rigorosos; mas, enfim, às vêzes evitamos as dificuldades muito aborrecidas e nos contentamos com o mais ou menos. A máquina não permite tergiversar; inexorável, não tolera imprecisão alguma; obriga a definir exatamente os dados do problema e a prever tôdas as eventualidades que se podem produzir, tôdas as dificuldades, por mínimas que sejam, que possam surgir. Eis aí seu valor; ainda que não tivesse nada mais além desta vantagem, a máquina já prestaria serviços imensos.

No domínio da arqueologia, por exemplo, M. Gardin, querendo usar a máquina para classificar a multidão de objetos encontrados nas escavações, precisou incontinentemente levantar o catálogo minucioso de tôdas as particularidades que êles apresentavam. (13) Isto feito, falta classificar efetivamente os objetos, o que é menos simples do que parece à primeira vista, pois trata-se de encontrar, entre tôdas as particularidades que nortearam a lista, as que devem desempenhar papel

preponderante. M. Gardin terá êxito, certamente, e seu trabalho já ultrapassou a fase de estudo para penetrar no da experimentação. Mas suponhamos que as dificuldades de classificação tenham-se revelado insuperáveis e tenha sido preciso permanecer aí: o simples fato de ter precisado organizar um inventário razoável de peças de museu já era algo de muito apreciável, e, não tivesse passado disso, essa tentativa de utilizar a máquina teria produzido um grande proveito.

As mesmas vantagens tornamos a encontrar em muitos outros domínios. Assim o **corpus** da arte cristã empreendida na Universidade de Princeton atinge agora tais dimensões que o tratamento da informação pela máquina tornou-se necessária. Mas então percebeu-se que o catálogo organizado a mão é insuficientemente preciso; vai ser necessário, e aliás já está nos projetos, retomá-lo pela base, para definir com muito maior objetividade em linguagem documentária os elementos que deverão ser classificados pela máquina. Para ter uma idéia nítida das ambigüidades com as quais podemos tropeçar na iconografia, consideremos um exemplo muito simples: uma pintura ou uma escultura representa uma mulher que tem na mão uma cabeça decepada; alguns vêem nisto Judite e Holofernes, outros Erodíades e João Batista: a qual interpretação devemos aderir? Ou mais ainda, como decidir se tal igreja pertence ao estilo gótico ou ao romano? Quem organiza um catálogo à mão recorre, para resolver as ambigüidades dêste gênero, à intuição, ao bom-senso, à experiência; mas tudo isto não basta, e estamos à procura de critérios objetivos, pois a máquina não se contenta com dados aproximativos.

Nós tentamos, na Compagnie Bull de máquinas, em 1960, fazer na máquina eletrônica o trabalho filológico que consiste em classificar os manuscritos segundo suas variantes (descobertos pela comparação automática de que falamos mais acima), para descobrir a natureza de suas relações genealógicas. (14) Esta crítica textual automatizada prestará, ao que parece, grandes serviços, poupando ao filólogo tudo aquilo que depende do cálculo; mas, suposto que ela devesse permanecer para sempre no laboratório sem receber aplicações práticas, eu não teria perdido meu tempo tentando realizá-la, pois foi necessário, para utilizar a máquina eletrônica, tirar a limpo muitas noções que até então estavam imprecisas e formular em termos lógicos questões de método que pareciam depender apenas da intuição.

* * *

Sendo uma calculadora, a máquina eletrônica é usada tôdas as vêzes que devemos manejar algarismos e fazer estatísticas, o que confirma que ela presta serviços em quase todos os domínios, pois quase não há disciplinas que não recorrem hoje em dia à matemática. Ela consegue então as vantagens que nós assinalamos: rapidez, segurança, necessidade de precisar com muito cuidado os dados e os métodos.

Consideremos por exemplo a demografia e a operação aparentemente simples que é o recenseamento. Trata-se de efetuar incontinenti uma simples "contagem" que permita conhecer o número total de habitantes de um país. Mas nós nos aproveitamos disso para registrar dados que serão objeto de estatística: sexo, idade, profissão, situação de família (casado ou celibatário) etc. A operação é extremamente trabalhosa, pois refere-se a dezenas ou centenas de milhões de pessoas (cêrca de 200 nos Estados Unidos, 55 no Reino Unido, 50 na França, 32 na Espanha etc); se usarmos meios mecanográficos, os resultados complexos do

recenseamento não serão conhecidos a não ser muitos anos após. É evidente que em tal caso a máquina eletrônica se impõe, para fornecer aos sociólogos informações atuais, sem aguardar que eles já estejam em parte caducos.

O mesmo acontece em psicologia experimental e em psicossociologia, onde o aproveitamento nos dados fornecidos pelos questionários é muito laborioso. Sem dúvida alguma, é porque tinham numerosas máquinas eletrônicas à sua disposição que os sociólogos americanos realizaram trabalhos tão consideráveis e mesmo tomaram a frente do movimento no que concerne às pesquisas de opinião. (15)

A seu exemplo as disciplinas psicossociológicas dos outros países usam grandemente calculadoras, como o faz (para não citar senão um exemplo) o Laboratório de psicologia social, da Faculdade de letras e ciências humanas de Paris. (16)

A bem dizer, é nestes domínios sobretudo que convém observar as preocupações de que falávamos agora mesmo, no que concerne aos métodos. Manejamos algarismos, e isso é imprescindível, pois não há tratamento científico a não ser que quantifiquemos o qualitativo. Mas até que ponto é lícito representar por algarismos fenômenos complexos onde intervêm fatores afetivos e sentimentais? A abstração à qual somos obrigados a recorrer por acaso não desfigura a realidade concreta que nos propomos a estudar? Eis questões delicadas e passíveis de controvérsias (17). Nós podemos todavia calcular com razão, parece, que os inconvenientes de um método matemático muito brutal serão corrigidos, não pela renúncia à máquina eletrônica, mas, ao contrário, fazendo-a trabalhar ainda mais; desde que ela maneja os algarismos com destreza, podemos fazer intervir um número muito grande de parâmetros para submeter ao cálculo os múltiplos aspectos de um problema e captar assim a realidade de mais perto.

Convém ficar alerta também para o fato de que a máquina eletrônica não diminui as dificuldades de interpretação apresentadas pelas estatísticas, seja qual for seu objeto. Quando fazemos uma estatística à mão, obtemos algarismos e porcentagens que exprimem uma relação entre duas ordens de fatos, dos quais um é levado para a "abscissa" e o outro para a "ordenada", ou entre diversas ordens de fatos cujas curvas são superpostas sobre o mesmo gráfico; mas a resposta dada pelo cálculo não tem sentido a não ser que a questão esteja bem posta e que tenhamos tido razão de relacionar os fatos submetidos à estatística. Os algarismos manifestam eventualmente uma relação, mas não dizem se ela é direta ou indireta e não se aventuram a uma interpretação "causal". Se efetuarmos por exemplo uma estatística da lepra em relação aos rios e os mares, constataremos que esta moléstia é mais freqüente à beira d'água; estaríamos errados se concluíssemos disso que a água é a causa da lepra; se a lepra é mais freqüente nas regiões que se avizinham do mar e dos rios, é porque as aglomerações humanas são mais densas aí; sua "causa" direta é mais a falta de higiene e a promiscuidade, sem que a água tenha participação. Da mesma maneira (18), a estatística revela que o número de estudantes da Faculdade de Direito de Paris e o número dos assinantes de telefone na região parisiense estão em estreita relação (r vizinho de 0,9). Nós não saberíamos concluir disso nem que os assinantes de telefone se inscrevem sistematicamente na Faculdade de Direito, nem que os estudantes de Direito sejam sistematicamente assinantes do telefone. O número de estudantes e o dos assinantes são elevados, um e outro, por um terceiro fator que é o aumento da receita média dos parisienses. É pre-

ciso então interpretar as correlações estatísticas com extrema prudência e não nos apressarmos a concluir uma relação de causalidade. A estatística choca-se contra as mesmas dificuldades de interpretação quando é feita a máquina, e nós nos iludiríamos se acreditássemos que as porcentagens fornecidas pela calculadora fôsem mais significativas que as obtidas a mão.

Se a máquina eletrônica presta serviços preciosos em matéria de estatística, é porque pode tratar velozmente e com muita facilidade dados muito copiosos; permite operar com amostras ampliadas e, por conseguinte, fazer sondagens que dão uma imagem mais fiel da realidade; podemos, graças à facilidade que ela traz, estudar estatisticamente um fenômeno sob múltiplos aspectos e escolher inteligentemente, entre tôdas as correlações que ela propõe, a que tem mais probabilidades de traduzir uma relação direta e casual.

* * *

Demoraria muito passar em revista sistematicamente tôdas as disciplinas nas quais a máquina eletrônica vem em socorro dos pesquisadores. Seja-nos permitido insistir um pouco, a título de ilustração, sôbre a filologia. Neste domínio onde até agora todo o trabalho se fazia a mão e parecia rebelde à automação, os computadores já desempenham papel considerável e assumem uma importância sempre maior; os estudos literários, com efeito, tendem cada vez mais a fundar-se sôbre resumos exaustivos e estatísticas precisas. (19)

A máquina eletrônica pode muito facilmente, quando lhe damos o texto sob forma de perfurações, fazer o levantamento do léxico de um autor, assinalando com suas referências, todos os vocábulos que êle emprega, sob suas diferentes formas. É assim que no Laboratório de análise estatística de línguas antigas, na Universidade de Liège, sob a direção de MM. Delatte e Avard, estabeleceu-se o léxico de diversas obras de Sêneca; na Universidade Tubinga, o Dr. Huebner compilou o dicionário de Goethe; em Gallarate, perto de Milão, o R. P. Busa estuda principalmente o vocabulário de S. Tomás de Aquino etc.

Por mais precioso que seja, o léxico de um autor fornece entretanto apenas esclarecimentos parciais. A "concordância" é um instrumento de trabalho muito mais cômodo, pois que destaca, ao mesmo tempo que o vocábulo (sob tal forma) e sua referência, a frase ou o membro de frase que a acompanha, o que permite perceber o sentido dos têrmos no seu contexto imediato, de ajuntar as passagens paralelas e de estudar o pensamento do autor tanto quanto seu estilo. A concordância "verbal" (a que assenta sôbre os vocábulos) foi imaginada pelos dominicanos franceses no século XIII a propósito da Bíblia Latina; imaginamos o labor imenso que se impõe se quisermos dispor os vocábulos e o membro de frase do qual êles fazem parte, não na ordem do texto, como se fazia inicialmente, mas na ordem alfabética de suas diferentes formas: para a Bíblia, o número de fichas que é necessário fazer e classificar é da ordem de 300.000 pelo menos. Êste trabalho é esmagador se o fazemos a mão, mas não oferece dificuldades muito grandes se o confiamos à máquina. Também vê-se agora a técnica da "concordância verbal" generalizar-se e aplica-se a tôdas as espécies de autores: clássicos latinos como Tibullo (*), teólogos como S. Tomás de Aquino etc.

(*) N.T. TIBULLO (AULUS ALBIUS TIBULLUS). Poeta latino nascido pelo ano de 54 a.c. e morto em 19. Sua obra compõe-se de quatro livros de Elegias. Foi principalmente o poeta do amor, da natureza, das doçuras da vida de família, tudo isso com uma sombra de tristeza.

Além disso existe estreita relação entre uma concordância e êsse léxico desenvolvido que é o dicionário: o *Littre*, como exemplo, não difere de uma concordância a não ser pelo fato de dar somente uma escolha dos empregos característicos de um vocábulo, ao invés da lista exaustiva; é uma espécie de concordância abreviada da língua francesa. Por outro lado, o léxico de um autor particular não apresenta, afinal, nada além de um caráter monográfico. Podemos ter ambições mais vastas e recolher em um *corpus* completo o vocabulário de uma língua inteira. Assim elabora-se em Nancy o “Tesouro da língua francesa”; um grande computador Gama 60 (Bull), trabalha exclusivamente para reunir sob forma de concordância os exemplos extraídos das obras literárias e destinados a fornecer as citações do futuro *Dictionnaire historique national*.

Em matéria de lexicografia, a máquina é capaz de tôdas as acrobacias. Por exemplo, organizou-se, no “Tesouro da Língua Francesa”, um “*Littre* inverso”, onde os vocábulos são dispostos em ordem alfabética não mais a partir da primeira letra, como nos dicionários comuns, mas a partir da última, o que permite estudar as desinências e flexões. Nada impede de arrumar também os vocábulos de acôrdo com o seu tamanho, ou segundo qualquer outro princípio imaginável. Uma vez que a máquina tenha registrado um vocábulo, fará com êle o que quisermos.

Os estudos de lexicografia dos quais acabamos de falar são acompanhados e completados por estudos de estilística. Juntando ao vocabulário de um autor seus processos de escrita, obtemos o quadro de suas características: são seus “hábitos”, na terminologia tradicional, mas poder-se-ia também falar de “espectro” tomando êste vocabulário no sentido usado pelos físicos e químicos; tais “hábitos”, com efeito, revelam a personalidade da mesma forma que certas riscas na luz decomposta pelo prisma indicam a natureza e a composição química de um corpo incandescente. A máquina eletrônica efetuando observações mais completas e estatísticas mais engenhosas, permite determinar com mais precisão os “hábitos” ou “espectro” de um autor; ela faz ressaltar peculiaridades que se escondem por assim dizer na filigrana do texto e que passariam sem dúvida despercebidas num exame “a olho nu”; em todo caso ela dá às constatações caráter mais objetivo.

O emprêgo da máquina eletrônica em matéria de estilística provoca apenas uma objeção, a qual aliás é fácil de responder. Para estudar o estilo e a gramática de um autor, é preciso antes de tudo, indexar o texto, ou seja colocar nêle sinais convencionais que indiquem a natureza dos vocábulos e sua função, bem como tôda particularidade sôbre a qual queremos interrogar a máquina. Eis um trabalho enorme e que, excessivamente difícil, não pode ser feito a não ser por especialista: vale a pena? A esta pergunta convém responder distinguindo duas hipóteses: se se considera na estilística apenas um elemento restrito, como seria por exemplo o lugar do sujeito e do complemento em relação ao verbo para estudar somente inversão, então teríamos mais velocidade executando a mão as anotações de onde tiraríamos as estatísticas, com o único inconveniente de nos limitarmos às pesquisas; em compensação o recurso à máquina tem um interesse incontestável se nos dermos ao trabalho de fazer uma indexação muito completa, que, levantando uma multidão de particularidades, permitirá fazer à máquina uma infinidade de perguntas, as que nós visamos no momento e as que interessarão talvez a outros filólogos.

A análise lexicográfica e estilística com a que podemos fazer à máquina vem em socorro dos filólogos de muitas maneiras. Permite, por exemplo, retrazar a evolução seguida por um escritor no curso de sua atividade literária, e, conseqüentemente, estabelecer, caso o ignoremos, a cronologia de suas obras: determinamos a ordem em que elas se sucederam, de onde resulta uma datação absoluta se dispomos pontos de marca muito aproximados. Examinando por exemplo os **Diálogos** de Platão, pareceu que êste grande escritor evitava cada vez mais os hiatos à medida que ao envelhecer dava mais perfeição ao seu estilo. Estudos feitos a mão por Lutoslawsky, continuados na máquina eletrônica, mostraram que isso era mesmo assim; conseqüentemente, bastou alinhar os **Diálogos** na ordem da porcentagem decrescente dos hiatos para dispô-los pelo mesmo fato na ordem de sua sucessão cronológica, o que permite situar com aquêles cuja data é aproximativamente conhecida, os que não sabíamos onde colocar. Estudos baseados em princípios análogos foram feitos por P. Guiraud sôbre a cronologia de determinados escritos de Rimbaud. (20)

O "espectro" de um autor pode servir também à crítica de autenticidade, quer dizer, permite saber se tal obra é mesmo devida ao autor que lhe imputamos, ou ajudar a descobrir o autor de uma obra anônima. O princípio é o seguinte: escritos cujos "espectros" são o mesmo podem pertencer ao mesmo autor, sem, todavia, têrmos o direito de afirmar algo além de simples possibilidades (visto que pessoas diferentes, mas de mesmo meio cultural, podem ter as mesmas características); aquêles cujo "espectro" fôr distintamente diferente não podem ser atribuídos ao mesmo autor, uma vez admitido que um escritor não muda muito seus hábitos no curso de sua vida. Assim, M. Marichal, que estuda Rabelais no Laboratório de análise lexicográfica de Besançon, propõe-se especialmente a elucidar a questão da autenticidade do **Quarto Livro**; êle julga com razão que um plagiário imitará fãcilmente em Rabelais o que fôr mais vistoso e, sobretudo, seu vocabulário extravagante, mas não poderá prestar atenção a pormenores sutis como a ordem dos vocábulos: a máquina eletrônica tornarã sensível os matizes que atraçoam um imitador ou mostram a mão do autor (*). Podemos, então, aplicando os mesmos princípios, infirmar ou confirmar as atribuições duvidosas, localizar as longas interpolações etc. Utilizaram-se êstes recursos para estudar, por exemplo, as Epístolas de São Paulo, a Imitação de Jesus Cristo, Chaucer; (21) processos análogos levariam talvez luz para o obscuro problema de Shakespeare, que suscitou tantas controvérsias, sem falar de diversos autores medievais como Rogério Bacon, S. Alberto Magno, cuja obra apócrifa ou duvidosa é imensa.

Não obstante o duplo princípio, segundo o qual um mesmo espectro pode corresponder ao mesmo autor, ao passo que espectros diferentes traem uma diversidade de autores, deve ser aplicado com prudência. Até que ponto um escritor permanece fiel a si mesmo? Em que limites seu espectro pode variar conforme os períodos de sua vida e os gênios literários que êle aborda? E, sobretudo, qual é o valor dos métodos empregados para esboçar o esquema que caracteriza ou seu estilo? É isto que precisaríamos saber. O aborrecido, nas disciplinas históricas, é que as cõclusões não são susceptíveis de qualquer fiscalização. As teorias físicas passam pela prova dos fatos; se um engenheiro constrói uma ponte enganando-se nos seus cálculos, a ponte desaba; mas o historiador não se arrisca a ver um autor medieval sair de seu túmulo para desmenti-lo. Não deveríamos, na filologia como em outras disciplinas, submeter a uma espécie

(*) N.T. Acreditamos que isto possa interessar aos "mediuns" autores de obras psicografadas...

de experimentação os métodos que nos propomos a explicar nas circunstâncias em que os resultados são inverificáveis? Para têmos certeza de que um método de cálculo destinado à máquina eletrônica é eficaz, fazemos, antes de mais nada, o cálculo a mão, em seguida confiamos-lo à máquina, conforme um programa que julgamos apropriado; se os resultados não concordam, foi a máquina que errou, ou de preferência o método que lhe impusemos não valia nada. De maneira geral, controlamos um método aplicando-o a um problema cuja solução é conhecida por outros meios. Da mesma forma o filólogo, e sobretudo aquele que se dedica à crítica de autenticidade, não deveria arriscar-se a formular conclusões sobre os autores antigos a não ser depois de terem ensaiado seus métodos com autores modernos para medir sua eficácia. Estudar-se-iam as obras de um autor que as haveria, certamente, escrito êle mesmo, e não por intermédio de um secretário; se a máquina declara que o espectro delas é muito diferente para serem tôdas devidas à mesma pessoa, é que o método é falso e daria resultados ilusórios quando o aplicarmos aos autores antigos. (22)

O estudo exato dos hábitos de um autor pela máquina eletrônica traz também um grande socorro à crítica textual nas operações conjecturais que ela não pode, na maior parte do tempo, dispensar. O melhor método, para restabelecer um texto sob a forma exata que lhe havia dado seu autor, eliminando os erros dos copistas, é garantidamente o método "histórico" ou "genealógico", desde que, entretanto, circunstâncias favoráveis permitam sua aplicação. Com efeito, se tivermos êxito na ordenação dos manuscritos, segundo suas relações de modelo e cópia, ou ancestral e descendente, o texto mais vizinho do original é o do ancestral comum de todos os manuscritos conservados, quer figure entre êles, quer o reconstituamos com a ajuda de seus descendentes. Mas é raro que se possa traçar a árvore genealógica graças, apenas, aos elementos externos; é preciso, sempre, recorrer aos dados internos, quer dizer, ao teor do texto, e às suas variantes. A máquina desempenha, então, o papel preparatório do qual nós já falamos. Comparando os manuscritos, ela descobre suas variantes e delas se serve para estabelecer um esquema que as coloca nas relações puramente "diferenciais", abstração feita do sentido real na qual vão as relações genealógicas. Uma vez realizadas tôdas estas operações materiais e que dependam do cálculo, resta ao filólogo efetuar o trabalho "inteligente" que consiste, para passar do relativo ao absoluto, em considerar não mais simples diferenças ou variantes, mas os enganos; esta apreciação qualitativa é, além do mais, indispensável não somente no método "por grupos" que preconizamos, mas, também, no método dos "enganos comuns", sob sua forma clássica e em todos os métodos críticos, sejam quais forem. O exemplar mais próximo do original, pelo teor de seu texto, é o que contém o mínimo de enganos, e, para distinguir entre diversas correções a que tem mais probabilidade de pertencer ao autor e as que constituem infidelidade de transmissão, é preciso recorrer muito à conjectura. Ora, entre os critérios conjecturais, os "hábitos do autor" ocupam, precisamente, um lugar importante, ao lado dos "hábitos dos copistas", dos quais falaremos dentro em breve. Sendo dadas, com efeito, duas reproduções, ou formas de texto que manuscritos divergentes dão ao mesmo trecho, é natural considerar como autêntico aquele que está conforme com o vocabulário e com o estilo habituais do autor, e como falso aquele que se afasta dêles. Esta apreciação qualitativa é grandemente facilitada pela máquina: fornecendo o "espectro" de um escritor, a estatística permite ao filólogo trazer sobre o valor das reproduções um julgamento ao mesmo tempo objetivo e matizado. A conjectura, então, repousa sobre

bases firmes e não tem mais êsse caráter “divinatório” que faz sua fragilidade e que nós lhe censurávamos, com tãda razão, quando ela se fundava, como antigamente, sôbre a intuição e a perspicácia.

Existe um outro critério conjectural do qual os filólogos fazem grande uso: é o que repousa sôbre os “hábitos dos copistas”, ou sua “psicologia”. O princípio é o seguinte: sendo dadas duas reproduções que rivalizam em certo trecho do texto, temos o direito de considerar como deturpação a que se puder explicar por uma das tendências às quais cedem os escribas quando se enganam? Trata-se, por conseguinte, de descobrir quais “ocasiões” solicitam os copistas a cair em falta e o modo de formação das falhas. Êste procedimento conjectural, codificado (para o latim) pelo Manuel de critique verbale, de Louis Havet, não foi seguido, até aqui, a não ser a mão, e a máquina eletrônica deveria lhe trazer também uma ajuda apreciável. Com efeito, a maioria dos enganos de copistas que examina a “crítica verbal”, são supostos por conjectura, e não constatados pela comparação de uma cópia com seu modelo imediato. Em lugar de se abandonar a especulações aleatórias sôbre a psicologia dos copistas antigos e medievais, não fariamos melhor estudando experimentalmente a dos copistas modernos, que não devem diferir muito? Materiais infinitamente ricos dormem nos arquivos das impressoras, pois os tipógrafos não passam afinal de contas de copistas como os outros; não deveríamos explorar estas minas de informação? Temos lá, lado a lado, o texto do autor e o primeiro jato do tipógrafo antes da correção, quer dizer, um modelo e uma cópia entre as quais não existe intermediário. Os enganos dos copistas diretamente observáveis esperam às centenas de milhares que lhes façamos a estatística. A máquina eletrônica diria de modo completamente objetivo (levando em conta a máquina usada para imprimir e a disposição de seu teclado) quais são as diferentes espécies de enganos, qual sua freqüência relativa, quais ocasiões as provocaram etc.; ela diria mesmo em que medida os enganos dependem da pessoa do copista e de sua psicologia individual, de modo que a lista dos erros cometidos mais habitualmente por certo copista deixaria seu “espectro”, um pouco como os hábitos de vocabulário e de estilo revelam o de seu autor. Baseada sôbre estas bases experimentais, a psicologia dos escribas antigos constituiria o objeto de um estudo mais rigoroso.

A conjectura que repousa sôbre os hábitos do autor e sôbre a dos copistas encontra aplicação, bem entendido, tãda vez que nos achamos em presença de um manuscrito singular, do qual ela é o único meio de descobrir e de corrigir os enganos. Assim acontece quando pudemos identificar ou reconstituir o ancestral comum de muitos manuscritos, caso êsse não seja o original. Ê ainda com mais razão o caso quando uma obra nos chega sob forma de uma cópia única, conservada e descoberta pelo maior dos acasos; o exemplo mais significativo é o dos manuscritos do Mar Morto. Sabe-se que êsses preciosos documentos contêm, além de longos fragmentos da Bíblia hebráica, escritos até agora desconhecidos e que nos informam sôbre as correntes de pensamento na Palestina em uma época vizinha da de Cristo; por infelicidade, êles jazeram durante perto de dois mil anos em jarros no fundo de uma caverna, de modo que as beiradas estão carcomidas e uma parte do texto desapareceu. Para descobrir o que se oculta nas lacunas, o R.P. Busa, graças à máquina eletrônica, estudou os “hábitos do autor” nas partes do texto ainda legíveis. A máquina não restituiu “automaticamente”, como o grande público imagina, os pedaços que fazem falta, mas fornece um elemento precioso de conjectura para restabelecê-los.

As pesquisas de lexicografia e de estilística das quais acabamos de falar não são as únicas aplicações da máquina eletrônica no domínio da filologia: poder-se-iam imaginar outras. Por exemplo, a automatização virá sem dúvida um dia em socorro desta disciplina nova que recebeu o nome de "Codigologia", e à qual é consagrada especialmente a revista *Scriptorium*, publicada na Bélgica sob a direção de M. François Masai. O codicólogo considera os manuscritos antigos como objetos arqueológicos: não poderíamos, por consequência, tratá-los, sob o ângulo codicológico, como M. Gardin trata os objetos encontrados nas escavações. Cada um deles, como uma cerâmica ou um machado de sílex, seria representado por uma ficha ou cartão perfurado que mencionaria suas características: matéria, formato, tessitura, modo de pautar, numeração dos cadernos, disposição do texto etc., bem como o local e a data quando mencionados pelo escriba ou suficientemente seguros. Poder-se-ia, então, por meios mecanográficos ou eletrônicos, estabelecer classificações, reconhecer grupos por zonas culturais e obter sobre a história do livro antes da imprensa tôdas as espécies de informações para datar e localizar os manuscritos, cuja idade ou origem são incertas.

As abreviaturas que se encontram nos manuscritos latinos ou gregos poderiam também ser classificadas pela máquina, o que permitiria, mediante um inventário bastante completo, reconstituir a história delas e de localizar a origem de algumas dentre elas; encontraríamos aí indicações suplementares para datar e localizar os manuscritos. Tais perspectivas são de resto consideradas por MM. Samaran e Marichal, que vão fazer uma relação metódica das abreviações latinas no Instituto de pesquisa e de história dos textos; M. Glénisson, diretor dêste Instituto, propõe-se mesmo a criar aí um centro onde se estudaria o emprêgo dos computadores nos diversos ramos da erudição que dizem respeito aos manuscritos.

A máquina eletrônica presta já tais serviços nos domínios literário e filológico, e vê diante de si um futuro tão rico de promessas que a maior parte das grandes universidades do mundo inteiro recorrem a ela, ou possuem uma para seu uso. Os centros de trabalhos filológicos "automatizados" multiplicaram-se mesmo a tal ponto que disso resulta uma certa desordem: percebeu-se que trabalhos idênticos são, às vêzes, executados simultâneamente por pessoas que se ignoram. Assim acabam de surgir, uma após outra, duas concordâncias do **corpus Tibullianum**, uma publicada na Itália, outra nos Estados Unidos, enquanto que uma terceira foi confeccionada por um membro do Laboratório de análise estatística das línguas antigas de Liège, e vai aparecer próximamente. Para evitar semelhante desperdício de energia e permitir a cada um o aproveitamento da experiência dos outros no que concerne ao método, o Laboratório de Liège acaba de criar (no começo de 1965) uma instituição que leva o nome franco-inglês de **Organisation internationale pour l'étude des langues anciennes par ordinateur** — **International Organization for Ancient Languages Analysis by Computer**, e cuja sede é Liège. Um boletim trimestral fará conhecer os trabalhos filológicos à máquina eletrônica que foram feitos, que estão sendo feitos ou em estado de projeto no mundo inteiro; êle dará também informações precisas sobre os métodos seguidos pelos diversos pesquisadores. Coordenando os esforços, êste Instituto dar-lhe-á um nôvo impulso. Vendo ainda mais afluir adesões dos filólogos de diversas especialidades, notadamente de alguns anglicistas americanos, e considerando que os métodos são os mesmos, quer se trate de línguas antigas ou modernas, o Professor L. Delatte já visa a ampliar os

quadros da organização internacional de Liège e de fazê-la desempenhar o papel de centro mundial de informação para todos os trabalhos filológicos por computador, de maneira geral, e sem distinção de língua.

* * *

Existe finalmente uma técnica que vai pròximamente trazer melhorias consideráveis no emprêgo da máquina eletrônica, e mesmo causar aí uma espécie de revolução: é a leitura automática.

Até o presente, a máquina não pode ler a não ser letras escritas segundo convenções bem definidas: seu traçado constitui um acôrdo entre um código inteiramente convencional e um caráter legível, o que vale dizer que os elementos destinados à máquina são dispostos de modo a imitar suficientemente uma letra comum para serem reconhecidos pelo olho. O fim colimado é menos fazer ler pela máquina letras de tipo corrente que permitir ao público ler um código primeiramente destinado à máquina e impresso além disso muito freqüentemente com uma matéria magnética que faz diretamente impressão sôbre os órgãos receptores da máquina. Faz alguns anos, os grandes bancos utilizam cheques cujos números e referências são escritos desta maneira, o que permite separá-los eletronicamente. Os serviços postais recorrem ao mesmo processo para separar cartas.

Fazemos atualmente pesquisas que se propõem a um fim inverso: fazer ler pela máquina cartas primeiramente destinadas ao olho humano e impressas a tinta de imprensa comum. O professor René de Possel, diretor do Instituto Blaise-Pascal do C.N.R.S., está construindo no momento uma máquina que permite atingir essa finalidade e funciona da seguinte maneira: O espaço destinado a uma linha de texto é dividido idealmente em um quadrilátero do qual cada coluna vertical compreende 72 pequeninos quadrados. Um raio luminoso projetado obliquamente sôbre o papel explora uns após outros, de alto a baixo, todos os pequenos quadrados de uma coluna vertical e tôdas colunas da esquerda à direita. A luz difundida pelo papel é recolhida por um foto-multiplicador, segundo um ângulo muito diferente daquêle que corresponderia à reflexão cartesiana, para evitar os efeitos de reflexos devidos à tinta de imprensa. O papel difunde mais ou menos luz, conforme o ponto luminoso que o explora incide no branco ou no prêto; esta oposição entre as partes do papel que têm ou não escrita traduz-se na linguagem da máquina que não admite senão "tudo ou nada" por sinais sob forma de símbolos zero e um, ou "bits", que são enviados a uma calculadora eletrônica, passando eventualmente pelo intermediário de uma fita magnética. A informação assim recolhida é submetida a um programa de tratamento graças ao qual a máquina "reconhece" as letras por meio de uma equivalência entre os símbolos formados de uma combinação de "bits" e um "alfabeto" que é pôsto à sua disposição. O número de caracteres dêste "alfabeto" reconhecível pela máquina (de 500 a 1.000 por exemplo) compreende as maiúsculas e caixa baixa de caracteres direitos e itálicos dos alfabetos latino, cirílico, grego, gótico, ronde e inglês, bem como numerosas figuras matemáticas e outras. A máquina pode então ler não importa qual página impressa, malgrado a diversidade dos caracteres, os milhares de letras a (por exemplo), as diversas "fontes" utilizadas pelos impressores traduzem-se pelo mesmo resultado; ela poderá, parece, pelo mesmo processo, discernir uma figura entre um número consideravelmente maior, da ordem de bilhões, o que permitiria aplicar a leitura auto-

mática sem dificuldade aos grandes alfabetos extremo-orientais, e também às escritas manuais. (23)

A velocidade da leitura automática é atualmente de mais de 60 letras (ou seja mais de uma linha) por segundo, o que permite já ler por exemplo tôda a Bíblia em menos de dezoito horas. Todavia, a utilização de uma calculadora comercial reduz atualmente de muito a velocidade de coleta de resultados. A construção de uma calculadora especial que permitirá seguir o ritmo da exploração do texto está caminhando, e êstes aperfeiçoamento novos darão o meio de ir bem velozmente ainda.

Graças à leitura automática, a máquina eletrônica será libertada do embaraço que é hoje o mais aborrecido: o gargalo de entrada. Ela funcionará com efeito bem mais velozmente se seu trabalho não fôr freado por dois gargalos, um na entrada, que consiste na longa preparação manual indispensável para introduzir os dados sob forma de perfurações, pelo menos no caso de inúmeros problemas, e outro à saída, onde devem intervir processos mecânicos para a impressão dos resultados. O gargalo de saída é agora alargado de maneira notável, uma vez que as grandes máquinas imprimem ou fotogravam seus resultados a velocidade indo até vinte mil linhas por minuto. Resta ainda o gargalo de entrada; mas no dia em que a máquina de ler intervier, êle desaparecerá de uma só vez: a introdução dos dados na máquina será praticamente instantânea e sem risco de erros. É difícil de prever em que prazo a leitura automática entrará em serviço; a máquina que M. de Possel contrôi neste momento será sem dúvida acabada e posta em forma em dois anos; será necessário em seguida fabricá-la em série para pô-la no comércio, de modo que ela possa ser de uso corrente talvez dentro de cinco ou seis anos. Então ela será o anexo indispensável da máquina eletrônica para todo um conjunto de problemas, a título de entrada automática. Notemos que as duas espécies de máquinas prestam-se mútua ajuda: a leitura automática será cada vez mais rápida à medida que os computadores eletrônicos sejam mais possantes, e êstes por seu turno irão mais velozmente quando forem dotados de uma entrada automática.

É penoso escrever e mesmo imaginar os serviços que prestará a leitura automática, tão vastas são as perspectivas que ela nos abre. É ela, logo de início, que permitirá à documentação e à tradução automáticas de alçar vôo e de sair da espécie de estagnação onde se encontram no momento: tal é a opinião, no que concerne à tradução, expressa primeiramente pelo professor Andréev, de Leningrado, depois por MM. Sestier e Vauquois, do Centro de tradução automática, em tôda sua generalidade.

A máquina de ler prestará um benefício considerável à filologia e à linguística. A bem dizer, ela não será provavelmente de grande ajuda (ao menos num futuro próximo) no que concerne aos manuscritos antigos e medievais. Os mais fáceis de ler automaticamente seriam os que são escritos em uncial latino ou grego e em carolino minúsculo, pois suas letras são bastante regulares e distintas umas das outras, mas são ao mesmo tempo as mais antigas e as mais importantes, de modo que os filólogos não podem dispensar-se de examiná-las "a olho nu" com cuidado minucioso. Por outro lado, os manuscritos mais recentes (século XIII-XV), cujo número é enorme e para os quais haveria conseqüentemente mais interêsse de ler velozmente à máquina, são também os que se recusam à leitura automática, porque suas letras são traçadas de maneira irregular, sobretu-

do no cursivo, unidas por ligaduras e obstruídas por sinais de abreviação de todo tipo. De resto, sejam quais forem sua escritura e data, os manuscritos apresentam correções devidas a mãos diversas: raspaduras, adições marginais ou interlineares, vocábulos ou letras sem pontuação etc., e tôdas essas particularidades, infinitamente preciosas para o filólogo, desnorteariam a máquina.

Haja o que houver com os manuscritos antigos, a leitura automática dos impressos permitirá realizar muito velozmente a edição crítica das obras que nos foram transmitidas apenas pela tipografia e compor os léxicos e as concordâncias das obras antigas, graças a suas edições críticas modernas. O *corpus* das fitas magnéticas e dos léxicos onde figurariam tôdas as obras que algum dia foram escritas desde a Antiguidade até nossos dias perderá, ao menos numa certa medida, seu caráter utópico. Para os estudos estilísticos, a indexação ficará com isso grandemente facilitada: a máquina, por exemplo, lê um texto (impresso ou datilografado) e o imprime à razão de um vocábulo por linha; no espaço livre, o filólogo inscreve (em datilografia, ou com uma escrita caprichada) tôdas as indicações que êle julga úteis; a máquina relê o texto assim indexado e se encontra então capacitada para responder a tôdas as questões, que se lhe quiser fazer. No dia em que a tradução automática fôr também realizada, as coisas serão ainda muito mais simples. Se a máquina traduz, é porque é capaz de descobrir o sentido dos vocábulos e sua função gramatical; podemos então sonhar, se isto não fôr muito quimérico, com uma indexação automática, análoga àquela que não desesperamos de realizar visando à documentação. A partir de um texto impresso, a máquina forneceria ela mesma as indicações necessárias ao estudo estilístico, e um programa estabelecido uma vez por tôdas permitirá tratar não importa qual texto escrito numa língua dada.

Em suma, é na leitura automática que se depositam as maiores esperanças, e é dela que esperamos, em todos os domínios onde se recorre à eletrônica, os progressos mais extraordinários.

NOTAS

- (1) Sobre o desenvolvimento da documentação e das publicações científicas, acharemos indicações bem interessantes em E. CAUDE e A. MOLES (com diversos outros colaboradores), *Méthodologie vers une science de l'action*, Paris, Gauthier-Villars, 1964, especialmente no capítulo *Sociologie de l'action*, por H. MIGEON e A. MOLES, a curva de crescimento das populações de sábios, nas principais partes do mundo (p. 192), aquela que representa o ritmo das invenções importantes desde o século X (p. 195) e no capítulo "Mise en ordre des connaissances", por J. DUBAS, um parágrafo sobre *Développement de la production documentaire* com um gráfico (p. 272). Nos domínios como aqueles dos quais nos ocupamos aqui, não se deve tomar o termo "exponencial" em um sentido rigoroso; explicando (p. 35) o que caracteriza uma progressão geométrica, A. MOLES faz notar: "Quando a curva é bastante imprecisa, freqüentemente é difícil dizer se se tratar de uma parábola $Y = Ax^2$ ou de uma exponencial $Y = Kx$."
- (2) Isso não é verdadeiro a não ser no domínio científico, onde intervêm apenas a observação e o cálculo. Acontece de maneira diversa nos domínios onde reina o qualitativo puro, como nas belas-artes; é evidente que o número das obras primas e dos grandes artistas que nasceram há quarenta anos é ínfimo se o compararmos a tudo que o passado produziu.
- (3) Na obra citada mais acima (nota 1), H. MIGENON e A. MOLES ilustram (p. 193) "os perigos da previsão" apresentando a curva que representa o número dos trabalhadores na indústria da construção elétrica na Inglaterra, com esta reflexão: "Aqui se vê claramente que seria perigoso extrapolar uma curva exponencial exageradamente; ela conduziria ao absurdo resultado de, em 1990 a população trabalhadora inglesa encontrar-se inteiramente ocupada com a construção elétrica, então apareceria desde já evidente que a influência da automação não poderia senão amortecer essa curva motivando uma saturação do emprego". É muito significativo que as curvas dadas pelos mesmos autores para representar o crescimento da população sábia do mundo (p. 192) induzem tôdas (exceto a da China, muito recente) um desvio para a horizontal: "a tendência à saturação já começa a manifestar-se."
- (4) Cf. J. — C. GARDIN: "Problèmes de la Documentation", in *Diogène*, n.º 11, de julho de 1955, pp. 107 a 124.
- (5) As reflexões que se seguem, e grande número das que figuram neste artigo, inspiram-se principalmente numa conferência pronunciada por M. J. — C. GARDIN, a 18 de março de

1965, no salão da Sociedade de encorajamento da indústria nacional, e numa entrevista que ele houve por bem me conceder alguns dias mais tarde. Agradeço-lhe a boa vontade, e espero não deformar seu pensamento.

- (6) A máquina eletrônica nasceu de uma espécie de colaboração entre Inglaterra, Estados Unidos e França. Seu ancestral longínquo, com efeito, não é, propriamente falando, a máquina de calcular de Schickard (1624), de Pascal (1645) ou de Leibniz (cêrca de 1700), que se situam mais no alinhamento de máquinas de escritório, mas o *Analytical Engine* do qual o inglês Charles Babbage desenhou os planos lá por 1830; êste aparelho jamais foi construído, porque, muito avançado sôbre o seu tempo, era irrealizável com os meios técnicos de que dispunha a indústria de então. A idéia de Babbage foi retomada em 1937, nos Estados Unidos, por Howard H. Aiken, que fez construir pela *International Business Machine Corporation* (I.B.M.) uma máquina eletromagnética "Mark I".
- A idéia de uma máquina que utilizaria os recursos da eletrônica foi emitida em 1938, pelo professor francês Louis Couffignal. A segunda máquina de Aiken, chamada "Mark II" era eletromagnética e foi ultrapassada antes mesmo do seu aparecimento pelo "E.N.I.A.C." (*Electronic Numerical Integrator and Computer*), primeira máquina propriamente eletrônica construída por J. P. Eckert e J. W. Mauchly na Universidade de Pensilvânia, lá por 1944. Os Estados Unidos foram seguidos de perto pela Inglaterra, onde uma máquina, cujo estudo foi iniciado em 1946, foi completado em 1951 em Manchester. Na França, a *Compagnie Bull* construiu máquinas aplicando as idéias do professor Couffignal.
- Sôbre a história das máquinas eletrônicas, principalmente nos Estados Unidos, ver: EL. BERKELEY: *Giant Brains, or Machines that Think*; tradução francesa: *Cerveaux géants, machines qui pensent*. Paris, Dunod, 1957. Sôbre o princípio e o funcionamento da máquina eletrônica, ver, por exemplo: N. CHAPIN: *An Introduction to Automatic Computers*, New York, Van Nostrand, 1953, e na coleção "Que sais-je?" (Paris, Presses Univ. de Fr.), J. POYN: *Le langage électronique*, 1960; P. DEMARNE et M. ROUQUEROL: *Les ordinateurs électroniques*, 1959; B. RENARD: *Le calcul électronique*, 1960. A melhor descrição da máquina eletrônica moderna sob suas diferentes formas é a obra de MM. François GAUCHET, Roger LAMBERT et Jacques VIOLET: *Le calcul automatique en psychologie*, Paris, Presses Univ. de Fr., 1965, na coleção "Le Psychologue", n.º 22. O texto é acompanhado de cêrca de 30 figuras, 15 fotografias e 10 tabelas; a última parte dá um exemplo concreto do cálculo automático em psicologia. Os autores tiveram êxito no seu esforço de dar explicações bem pormenorizadas e, portanto, acessíveis aos não-especialistas.
- (7) Muitos trabalhos de amplitude e de precisão notáveis, no domínio da filologia, são realizados com o recurso, ao lado do computador eletrônico, das máquinas mecanográficas; assim acontece especialmente no Laboratório de análise estatística das línguas antigas (*Université de Liège*) dirigido por MM. Delatte e Evrard que descrevem seus trabalhos num artigo da *Antiquité Classiquée*, t. XXX, 1961, fasc. 2, 427-442. Êsses processos, onde a mecanografia desempenha importante papel, foram aplicados recentemente por M. Paul Tombeur, para estudar a língua e o estilo do cronista medieval Raoul de Saint-Trond; êle explica seus métodos num artigo intitulado "Application de méthodes mécanographiques à un auteur médiéval", no *Archivum Latinitatis Medii Aevi*, t. XXXIV, 1964 125-160; seu trabalho será publicado próximamente sob o patrocínio da Comissão real de história da Bélgica.
- (8) Sôbre a documentação automática, citemos sômente J. — C. GARDIN: "États et tendances actuelles de la documentation automatique", na revista *La Traduction automatique*, Paris 5.e année, n.º 1, mars 1964. Ele dá nas suas notas uma abundante bibliografia dos trabalhos realizados nesse terreno nos Estados Unidos e na França.
- (9) A continuidade dos problemas chamados de "classificação", da documentação tradicional à documentação moderna, é ressaltada por B. C. VICKEREY: *Classification and Indexing in Science*, Londres, 1959, e por DE GROLLIER: *Étude sur les catégories générales applicables aux classifications et codifications*, Paris, 1962.
- (10) Eis um exemplo, pouco recente, na verdade, significativo, dos entraves que a falta de tradutores acarreta para o avanço das ciências. Na *Histoire générale des sciences*, publicada sob a direção de R. TATON (Presses Univ. de Fr.), t. II, no capítulo consagrado à *Naissance de la chimie moderne*, M. DAUMAS escrita a propósito dos trabalhos de Lavoisier, cêrca de 1770, a nota seguinte (p. 553): "Muito antes dessa época, o russo M. A. Lomonossov publicou trabalhos contendo notáveis antecipações. Aí se encontra a idéia da constituição atômica dos corpos e a da energia cinética devida à agitação molecular. (...) Infelizmente, seus escritos, publicados em russo, não foram conhecidos dos químicos dos outros países. Disso não se encontra qualquer menção na literatura inglesa, francesa e alemã dêsse tempo. Sem dúvida alguma, a divulgação de suas idéias teria favorecido o advento do sistema químico moderno."
- (11) Nos Estados Unidos, existe pelo menos uma dezena de centros consagrados à tradução automática, notadamente em Harvard, Georgetown, Berkeley, Los Angeles, e uma *Association for Machine Translation and Computational Linguistics*. Na Inglaterra, podemos citar Birkbeck College (*Department of Numerical Automation*). Cambridge (*Cambridge Language Research Unit*). Na Rússia, o organismo mais importante parece ser o Laboratório experimental de tradução automática, na Universidade de Leningrado. Na França, um "Centre d'études de la traduction automatique" funciona em Paris e em Grenoble; a Associação para o estudo e o desenvolvimento da tradução automática (A.T.A.L.A.) em Paris, publica uma revista, *La Traduction automatique*, cuja informação tem caráter internacional. Sôbre a história e a técnica da tradução à máquina, as melhores exposições conjuntas são: Émile DALAVENAY: *La Machine à traduire*, Collection "Que sais-je?" Presses Univ. de Fr., 1959, e A.G. OETTINGER: *Automatic Language Translation: Lexical and technical Aspects*, Harvard Univ. Press, Cambridge, Mass., 1960. Para conhecer em pormenor as pesquisas recentes, nos Estados Unidos sobretudo, mas também na Inglaterra, na França, na Itália e no Japão, consultar as comunicações apresentadas à primeira Conferência Internacional sôbre *La traduction automatique des langues et l'analyse linguistique*, que se realizou no

- National Physical Laboratory, Teddington (Middlesex), de 5 a 8 de setembro de 1961; os relatos e as discussões foram publicadas em inglês por Her Majesty's Stationery Office (Londres, 1962); entre as 36 comunicações apresentadas a esta Conferência internacional, as mais importantes, em número de 14, foram publicadas em tradução francesa sob o título: *Traduction automatique et linguistique appliqué* (Paris, Presses Univ. de Fr., 1964); esse volume é o primeiro de uma coleção que aparecerá sob os auspícios da A.T.A.L.A.
- (12) A descrição pormenorizada desta experiência está publicada no *Bulletin d'information de l'Institut de recherche et d'histoire des textes*, dirigido por M. Glénisson, tome XIII, 1965; Dom Jacques FROGER: "La collation des manuscrits à la machine électronique". O programa de comparação é devido à Mme. Renaud.
 - (13) M. GARDIN expôs seus trabalhos num artigo: "Cartões perfurados e computadores a serviço da arqueologia", *La Nature*, nov. 1962, pp. 449—457.
 - (14) Esta experiência foi apresentada no Colóquio internacional de lexicografia realizada em Bensaçon, em julho de 1961, com uma demonstração sobre máquina Bull na Faculdade de letras de Estrasburgo. Sobre o método seguido "à mão", ver: Dom Jacques FROGER: "A crítica textual e o método dos grupos falíveis" (Comunicação feita em Bensaçon), nos *Cahiers de lexicologie*, n.º 3, 1962, publicados pelo Laboratório de análise lexicológica da Faculdade de letras e ciências humanas de Bensaçon (Dir. M. Quémada). O programa destinado à máquina deve-se a M. Philippe Poré; foi exposto por Mme. Poyen e M. Poré no segundo Congresso da A.F.C.A.L.T.I. em outubro de 1961. Para ter uma idéia do problema no seu conjunto, poder-se-á consultar: Dom Jacques FROGER e Philippe Poré: *La critique des textes et son automatization*. Paris, Dunod, a surgir em 1966, na coleção "Initiation".
 - (15) Sobre recenseamentos e sondagens, métodos e resultados, ver: *Handbook of Population Census Methods* (3 brochuras), publicado pelo Escritório de Estatística da ONU.
 - (16) Este laboratório, cujo diretor é M. Robert Pagès, mantém à disposição dos pesquisadores um repertório cumulativo das publicações de seus membros (última edição: 1965); aí encontramos número apreciável de obras que tratam dos métodos matemáticos e do emprego de computadores.
 - (17) M. Jacques ELLUL, no seu livro intitulado *Propagandes*, Paris, A. Colin, 1962, mostra-se muito céptico sobre a validade dos métodos matemáticos em psicossociologia; ele chega mesmo a recusar-lhes muito categoricamente no parágrafo consagrado à "ineficácia dos métodos para medir a eficácia da propaganda" (Anexo I, pp. 294—295 principalmente. Assinala (p. 286, nota 2) que alguns autores americanos contestam as premissas das sondagens de opinião, por exemplo BLUMER: "Public Opinion and Public Opinion Polling", na *American Sociological Review*, 1948. Sobre o conjunto do problema, ver SOROKIN: *Tendances et déboires de la sociologie américaine*, 1959.
 - (18) Este exemplo é emprestado de Philippe MOUCHEZ: *Démographie*, coleção "Thémis", Paris, Presses Univ. de Fr., 1964, p. 134, nota 1.
 - (19) Sobre os serviços que a máquina eletrônica pode prestar em filologia, podemos encontrar indicações em: Dom Jacques FROGER: "Emploi de la machine électronique dans les études médiévales", *Bulletin de la Société internationale pour l'étude de la philologie médiévale*, Louvain, t. III, 1961, pp. 177—188. Ver também as obras citadas anteriormente, nota 6.
 - (20) Cf. P. GUIRAUD — *Problèmes et méthodes de la statistique linguistique*. Paris, 1960.
 - (21) Cf., G. U. YULE: *The Statistical Study of Literary Vocabulary*, Cambridge, 1944, à respeito do autor da Imitação de Cristo; J. C.; G. HERDAN: "Chaucer's Authorship of the Planetis", em *Language*, 32, 1956, pp. 254—259; Rev. K. GRAYSTINE and G. HERDAN: "The Authorship of the Pastorals in the Light of Statistical Linguistics", no *New Testament Studies*, 6, 1959, pp. 1—15.
 - (22) Lê-se na revista *Science et Vie* n.º 572, t. CVII, maio de 1965, p. 49, sob o título "Briga de máquinas", uma informação que, sob seu aspecto divertido, esconde uma advertência séria: "Em 1963, um teólogo, R.P. Morton, demonstrava graças à uma análise semântica realizada por uma calculadora eletrônica que as quatorze Epístolas de São Paulo não podiam ter sido escritas pela mesma pessoa, mas que seis autores diferentes estavam em causa. Aplicando o mesmo método analítico às obras publicadas do próprio R.P. Morton, um outro eclesiástico, o R.P. Ellison, demonstrou que elas não poderiam ter sido todas compostas pelo autor. De fato, o cérebro eletrônico "provou" que o R.P. Morton logicamente, não existia!
 - (23) M. de Possel estuda também neste momento uma máquina para pronunciar um texto, que poderia prestar diversos serviços, em particular aos cegos. Seria um complemento da leitura automática.