UNIVERSITAS

N.os 8/9
SALVADOR

SEPARATA
JAN./AGÔSTO/1971

Rosa Virgínia Mattos e Silva —
O fichário mecanográfico de um texto antigo

O fichário mecanográfico de um texto antigo

Num momento em que tôda a ciência lingüística se volta sobretudo para as descrições da língua no seu uso atual, dentre o que há de menos renovado nos estudos desenvolvidos sôbre o português, assume posição destacada o conhecimento de sua fase arcaica. Em nosso caso veja-se a preocupação com os vocabulários fundamentais, com o estudo da norma lingüística culta das grandes cidades, dos falares regionais, da língua coloquial nas suas modalidades distensas ou tensas de diferentes comunidades, tudo do maior interêsse e da maior necessidade e urgência; e veja-se ainda tôda a admirável obra que nos legou o Professor J. Mattoso Câmara Jr. O estudo do português antigo, no entanto, continua a ser feito nas mesmas bases de cunho historicista, lançadas por J. Leite de Vasconcelos ou Carolina Michaelis de Vasconcelos no comêço do século. De saída, quem quer que se ponha hoje diante de um texto antigo tem de rever as atitudes tradicionalmente tomadas face ao texto e procurar tirar o maior proveito possível das orientações, posições e instrumental fornecidos pelas tendências lingüísticas e novas téc-

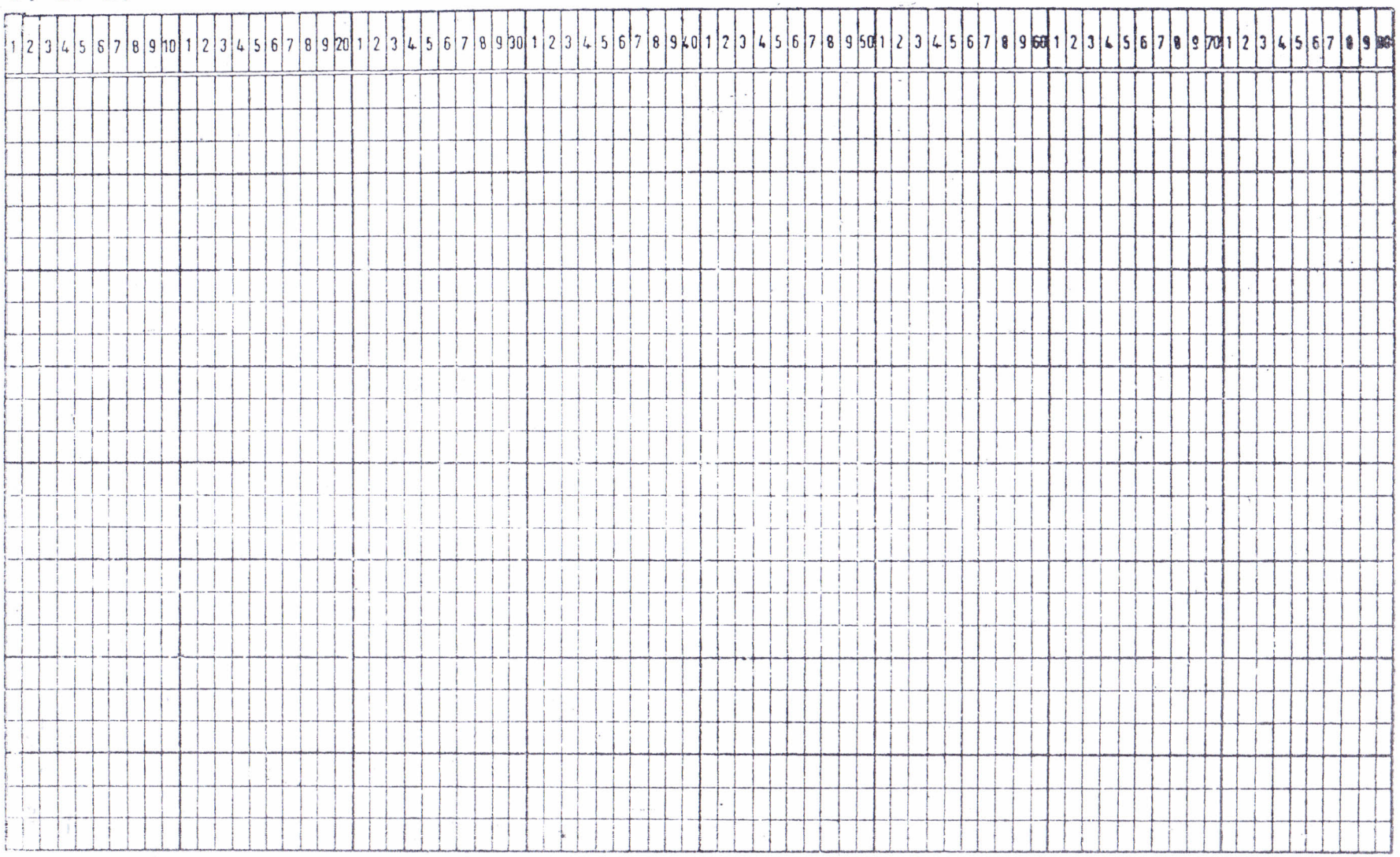
nicas que se vêm aprofundando e difundindo noutros países, desde a década de 20 dêste século, mas nós de língua portuguêsa nunca antes dos anos de 50.

Alguns trabalhos contudo existem que nos animam a pensar que em pouco tempo muito poderá ser feito para o conhecimento de fases passadas do português. Não poderemos ambicionar ser exaustivos na enumeração de trabalhos sôbre a língua arcaica que a procuram encarar, ou algum de seus aspectos, sob um prisma mais rendoso para o entendimento efetivo do funcionamento sincrônico de estados pretéritos: apenas lembramos aqui o que vem desenvolvendo sôbre *A Vida e Feitos de Júlio César* (séc. XV), a professôra Maria Helena Mira Mateus, da Universidade de Lisboa (1); o estudo de vocabulário, segundo a orientação lançada há alguns anos por Wallig-Wartburg (2), da *Perigrinação* de Fernão Mendes Pinto (séc. XVI), feito por Clemente S. Pinho, de S. Paulo (3); e sob orientação semelhante a pesquisa sôbre o vocabulário de D. Duarte (séc. XV) (4), em elaboração pela equipe de Filologia Românica da Universidade da Bahia; temos ainda notícia de um primeiro trabalho de cunho transformativista de Anthony Naro (5). Certamente outros estão sendo feitos sôbre o assunto, fugindo ao método historicista da filologia, mas a falta de informação, lugar comum entre os que fazem pesquisa no Brasil, impede a certeza absoluta sôbre o que se passa atualmente na área em discussão.

Quando ressaltamos os trabalhos acima, como representativos de uma nova orientação que desponta nos estudos do português antigo — deixemos bem claro — não queremos negar mérito ao significante e indispensável labor desenvolvido por filólogos brasileiros, portuguêses e estrangeiros. E aqui nomeamos com o maior respeito a obra desbravadora de Serafim da Silva Neto; o excepcional, por ser único e extraordinàriamente bem feito, estudo de interferência português, ou melhor, galego-português e leonês no séc. XIII, refletida da linguagem dos Foros de Castelo Rodrigo (Beira Alta), do professor L. F. Lindley Cintra; as pesquisas sôbre o Cancioneiro Medieval Galego-Português dos Professôres Celso Ferreira da Cunha e M. Rodrigues Lapa; e ainda a tese que trata da língua de Gil Vicente, de Paul Teyssier (6). Tampouco esquecemos edições críticas, já publicadas, de textos arcaicos ou clássicos. Não seria demais ressalvar que, em têrmos globais, alguns estudos de tendência historicista chegaram a uma precisão e elaboração de dados ainda não alcançada, dentro da sua própria tendência, pelos que seguem uma metodologia voltada para a sincronia e análise descritiva.

Quando iniciamos a edição crítica da versão mais antiga dos *Diálogos de S. Gregório* (séc. XIV), texto em prosa, longo (81 fólios manuscritos, 350 páginas

C. C. C.



C.C.C. – 19-965-5000

datilografadas), com rica documentação para a história da língua, ambicionamos aproveitar ao máximo os dados fornecidos pelo texto para reunir e analisar um primeiro *corpus* representativo da língua portuguêsa no século XIV. Um fichário manual palavra a palavra do texto era o caminho mais natural a seguir, já nosso conhecido (7), e que se vem tradicionalmente tomando desde os primeiros glossários de textos arcaicos, mais ou menos exaustivos. Ao encontro de nossa ambição de explorar ao máximo o texto veio entretanto a oportunidade de usar o conjunto mecanográfico do Centro de Cálculo Científico da Fundação Calouste Gulbenkian em Lisboa (8), apesar de, desde 1961 no congresso internacional sôbre a mecanização das pesquisas lexicológicas, Bernard Quémada do Centre d'Étude du Vocabulaire Français de Bésançon já insistir em que o uso de equipamento emprestado estava ultrapassado, devendo o lingüista dispor de equipamento especializado (9). A contingência histórica a que estamos submetidos, e que lamentamos, não nos tem permitido dispor para o estudo do português de centros com equipamentos eletrônicos, ou adaptados, para a pesquisa lingüística, como não tem permitido tantas outras facilidades de pesquisa, muito mais urgentes sem dúvida, neste e noutros campos. No entanto alguma coisa se pode tentar fazer com os recursos disponíveis, embora quase sempre raros, que surgem.

A ajuda da mecanografia e da eletrônica nos estudos lingüísticos vem sendo utilizada na Europa e na América há mais de duas dezenas de anos aproximadamente, e não cabe aqui destacar tudo que se desenvolve e aperfeiçoa sobretudo na URSS, nos EE. UU., no Canadá, também em certos países da Europa Ocidental, como a França, sôbre a tradução automática, nem centros de documentação lingüística que se multiplicam, com base em equipamentos eletrônicos e eletromecânicos (10).

No Centro de Cálculo Científico de Lisboa já haviam trabalhado, com objetivos diversos dos nossos, os citados Professôres Clemente S. Pinho e Maria Helena Mira Mateus. A boa-vontade dos seus dirigentes e técnicos nos animou a tentar um fichário mecanográfico para o levantamento do material lingüístico oferecido pelo texto em pauta, com vistas à descrição e análise gramatical dos dados.

Não visávamos apenas uma indexação do léxico do texto, mas também, e principalmente, organizar os dados em função de uma descrição gramatical da versão mais antiga dos *Diálogos*, e ainda de seu confronto com as duas versões do século XV.

Sôbre três pontos incidiu o nosso interêsse ao elaborar o material preliminar a ser analisado posteriormente, de forma a ter em mãos:

I. o levantamento exaustivo e total do léxico da versão tre-

centista (que designamos *A*), em ordem alfabética e com a localização das ocorrências no texto.

2. os dados mais completos possíveis para o estudo da linguagem do texto nos níveis que denominamos gráfico-fonético e morfo-sintático.

3. dados relativos às características gráfico-fonéticas e morfo-sintáticas das versões quatrocentistas dos *Diálogos* (que chamamos *B* e *C*), em relação às mesmas características na versão *A*.

Pretendemos, a princípio, utilizar o ordenador eletrônico. À falta de um técnico disponível para programar segundo nossas intenções, nos conformamos com a utilização apenas do conjunto mecanográfico clássico (11), e todo material a ser perfurado e trabalhado por máquinas e técnicos da instituição foi prèviamente por nós codificado.

Nas fichas padrões fornecidas pelo Centro e divididas em colunas (= campos) lançamos todos os vocábulos, que encabeçariam as fichas IBM, como palavra "vedetta", por ordem de ocorrência no texto, seguidos de um contexto mínimo permitido pelo tipo de equipamento com que trabalhamos (dispúnhamos de 44 campos para indicações alfanuméricas e os restantes para indicações numéricas); ao lado direito das fichas lançamos a localização do vocábulo no texto, com indicações numéricas correspondentes ao livro, capítulo e parágrafo da obra, seguida de código numérico estabelecido para os ítens 2 e 3 acima expostos.

Tal código numérico consistia em simples combinação de algarismos de 0 a 9 e sua posição na ficha, isto é, no campo em que os colocamos segundo nossos objetivos. Portanto, cada campo ou grupo de campos correspondia a um aspecto da descrição lingüística (por exemplo: classe de palavra, categoria gramatical, caracterização gráfico-fonética, caracterização morfo-sintática) ou do relacionamento dos dados obtidos em *A* com os de *B* e *C*; dentro dêstes campos a numeração de 0 a 9 correspondia a um traço destacado por nós em nossa classificação (por exemplo: no campo x = classe de palavra, n = substantivo, n^1 = adjetivo, n^2 = numeral etc; no campo x^1 = categoria gramatical, n = gênero, n^2 = tempo, n^3 = modo etc.). Em cada ficha cabiam 24 vocábulos nos seus contextos seguidos dessas indicações numèricamente codificadas, e para cada um dêsses vocábulos seguidos de tais indicações veio a ser perfurado um cartão IBM; portanto a cada ficha elaborada por nós correspondiam 24 cartões IBM.

À primeira vista pode parecer valer mais a pena preparar um fichário manual, palavra a palavra, em que se anotassem as localizações e porventura outras observações, com o cuidado material de não sobrecarregar a ficha, ou também a preocupação de duplicar o fichário, se fôsse o caso, como fazem todos que trabalham com glossário ou índices de vá-

rias naturezas. O tempo gasto na fichagem dos dados a perfurar talvez equivalesse ao gasto em um fichário tradicional; entretanto, o interêsse da codificação que obriga uma reflexão constante sôbre as características a destacar, e sôbre como destacá-las, transforma o trabalho monótono de fichagem em trabalho de elaboração, e as vantagens posteriores de ordem prática e de economia de tempo ultrapassam, sem dúvida, o que possa haver de não-estimulante na etapa inicial de lançamento e codificação dos dados nas fichas.

O caráter cumulativo do fichário mecanográfico permite que, a partir de uma só ficha, graças ao trabalho da máquina seletora, se possam ter listagens diversas, agrupando cada uma determinado aspecto, codificado no processamento inicial do material oferecido pelo texto. Com essas listagens individuais, apresentadas em ordem alfabética e com a localização indicada mecânicamente, impressas pela tabuladora, tem-se um material já ordenado, e fàcilmente manuseável, para a descrição posterior dos ítens lingüísticos que interessam.

Do material fichado segundo os moldes transcritos anteriormente, obtivemos do Centro de Cálculo Científico:

1. o "index verborum" em ordem alfabética e de ocorrência de todos os vocábulos do texto, aproximadamente 54.000 formas: tôdas as ocorrências das palavras lexicais, com suas variantes gráficas e de flexão, e tôdas as ocorrências das palavras gramaticais em um têrço aproximadamente do texto, suas variantes gráficas e flexionais.

2. listagem das formas correspondentes a cada um dos itens selecionados para o estudo lingüístico (ao todo 68 listas), também impressas e apresentadas em ordem alfabética, com sua localização no texto e contagem das ocorrências.

3. listagens referentes aos itens coincidentes e divergentes das versões *B* e *C* em relação ao mesmo item da versão *A* (ao todo 95 listas), e as contagens relativas a cada uma.

O material assim reunido e organizado constitui o *corpus* de que ficamos dispondo para a descrição lingüística da versão trecentista dos *Diálogos* e de suas relações com as duas versões quatrocentistas.

Os recursos mecanográficos, aplicados ao levantamento exaustivo da documentação que uma língua oferece no curso de sua história, podem ser aproveitados com vantagem para a descrição gramatical das sucessivas etapas da língua. Em função do aperfeiçoamento da tradução automática muitos estudos e pesquisas vêm sendo feitos no sentido de aprofundar tal descrição, que fornecerá os elementos básicos para a programação, a ser dada aos computadores que funcionarão como tradutores. Assim, são as línguas de que se fazem traduções automáticas as mais conhecidas e estudadas objetivamente, uma vez que sem tal ponto de

MOSTRA A FIN E O. ENSARRAMENTO 9 .11 5015044

P 44801

COMPANHIA IBM PORTUGUESA

partida se torna impossível a tradução. As dificuldades em adequar aos limites da máquina a complexidade da estrutura da língua têm obrigado os lingüístas a alcançarem maior coerência, economia e simplicidade na análise e interpretação da língua que é seu objeto de estudo. Não só por causa, entretanto, da tradução automática, mas porque as orientações da lingüística moderna se fundam nos princípios aqui de passagem nomeados, qualquer que seja a corrente estruralista a que se filie o pesquisador. A concisão na descrição de um testemunho lingüístico é indispensável para que no pouco espaço de que se dispõe na ficha perfurada se possam lançar tôdas as indicações codificadas que venham a interessar para a caracterização daquele documento, indicações essas que serão trabalhadas com rapidez e perfeição pelo conjunto de aparelhos que compõem o equipamento mecanográfico.

Para a língua portuguêsa nada de semelhante foi ainda tentado, exceto trabalhos esparsos que utilizam ordenadores ou conjuntos eletromecânicos, e que são programados em função de seus objetivos imediatos: nenhum trabalho sistemático de análise da língua em seu conjunto tem sido feito. Estamos certos inclusive de que tal trabalho teria de ser fruto de intensa discussão entre grupos de estudiosos da língua portuguêsa e da constituição de equipes.

Nosso trabalho, que incide apenas sôbre um único testemunho de um estágio da língua, se ressente, sem dúvida, dêsse caráter pioneiro e de nosso autodidatismo na aplicação de tais técnicas. Por exemplo, só tomamos conhecimento do que se vinha discutindo a partir de 1960 no Centre d'Étude du Vocabulaire Français de Besançon (12), sôbre a descrição gramatical da língua francesa utilizando os recursos a que vimos nos referindo, quando já ia nossa codificação bastante adiantada. Tínhamos contudo notícia do trabalho de documentação lexicológica que aquêle Centro, fundado em 1957, realizava, visando um fichário geral do vocabulário francês (13). A exploração do léxico do texto com que trabalhamos foi realizada sem dificuldades e cremos que, apesar do contexto limitado (dispúnhamos de espaço para apenas 44 sinais alfabéticos) que as características do equipamento impunham, tivemos um resultado positivo, que poderá ser aproveitado como documentação relevante para a história do português, em estudos semânticos e lexicológicos. A parte referente à análise gramatical sofreu, sem dúvida, pela inexistência de prévias descrições sincrônicas do português antigo, em que com objetividade, coerência e economia se caracterizasse aquela etapa da língua. Cremos, entretanto, que o material recolhido, com dados exaustivos, pode ser um ponto de partida útil para que se inicie um estudo do funcionamento sin-

crônico do sistema português trecentista (14). Havendo oportunidade de aproveitar equipamentos mecanográficos ou eletrônicos para a descrição gramatical de outros documentos em língua portuguêsa, a experiência que tivemos oportunidade de levar avante, mas cujas imperfeições reconhecemos, valerá sobretudo para que não voltemos a incorrer nas falhas que inadvertidamente cometemos.

Antevemos que uma história da língua portuguêsa feita no século XX não dispensará tais recursos, e muito seria de esperar de um dicionário do português antigo ou do português em geral que contasse com o levantamento exaustivo do vocabulário da língua. Não nos convêm nem convencem as teorias baseadas sôbre dados inexpressivos e insuficientes, pois originados de amostragens muito limitadas e com freqüência subjetivamente constituídas. Precisa-se, ao menos, de uma amostra válida que permita conclusões que condigam com a realidade. Se escolhéssemos textos representativos de cada fase da língua, levantássemos seu vocabulário e ao mesmo tempo os analisássemos gramaticalmente, teríamos, ao menos, com exatidão e em pouco tempo, condições de descrever as diversas sincronias que compõem a história da nossa língua. É que, comparadas, permitiria compreender-lhe melhor a evolução e a dinâmica própria.

ROSA VIRGÍNIA MATTOS E SILVA

1 A Professôra M. H. Mira Mateus tem dado notícia do trabalho que vem realizando sôbre o texto em pauta: V. Perspectivas da Lingüística actual (o método estatístico, a teoria da informação, a utilização de máquinas). Separata da *R. de Portugal*. Série A: Língua Portuguêsa. Lisboa, 31:178-193, 1966 e *Valeurs du plus que parfait de l'indicatif, forme simple dans la langue portugaise du XVe siècle*. Comunicação apresentada ao XII Congresso de Lingüística e Filologia Românicas, Bucarest, 1968, a sair nas *Atas*.

2 WARTBURG, W. von & WALLIG, R. *Système raisonné des concepts pour servir de base à la lexicographie. Essai d'un schème de classement*. Berlin, Akad. Verlag, 1963.

3 PINHO, C. S. *A Linguagem de Fernão Mendes Pinto, segundo um sistema de conceitos. Exame parcial de amostras do vocabulário*. Araraquara, s. ed., 1966. Mimeo. (Tese de Docência à Faculdade de Filosofia Ciências e Letras).

4 Cf. as comunicações apresentadas ao II Congresso da Associação de Lingüística e Filologia da América Latina (ALFAL 3-8 de janeiro de

1969) pela equipe de Filologia Românica da UFBa: *Alguns aspectos do vocabulário da vida anímica (sentimentos e sensações) no "Leal Conselheiro" e Aspectos do vocabulário do "Livro de ensinança de ben cavalgar tôda sela"*.

5 Temos notícia oral e imprecisa sôbre a tese de doutoramento de Anthony Naro apresentada nos últimos anos nos EE. UU., entretanto sabemos que o auor a fêz com metodologia transformacional e analisa o português antigo.

6 Vide principalmente:

SILVA NETO, S. da. *História da língua portuguêsa*. Rio de Janeiro, Livros de Portugal, 1952.

CINTRA, L. F. Lindley. *A Linguagem dos Foros de Castelo Rodrigo. Seu confronto com a dos Foros de Alfaiates, Castelo Bom, Castelo Melhor, Coria, Cáceras e Usagre. Contribuição para o estudo do leonês e do galego-português do sec. XIII*. Lisboa, s. ed., 1959.

CUNHA, C. Ferreira da. *O Cancioneiro de Joan Zorro. Aspectos lingüísticos*. Texto crítico. Glossário. Rio de Janeiro, s. ed., 1949 e *O Cancioneiro de Martin Codax*. Rio de Janeiro, s. ed., 1956.

LAPA, M. Rodrigues. *Cantigas d'escarnho e de mal dizer dos cancioneiros medievais galego-portuguêses*. S. l., Editorial Galaxia, 1965.

TEYSSIER, Paul. *La Langue de Gil Vicente*. Paris, s. ed., 1959.

7 Cf. ROSSI, N., MOTTA, J. A., MATTOS, R. V. e SAMPAIO, V. L. *Livro das Aves*. Rio de Janeiro, I.N.L., 1965, e *Diálogos de S. Gregorio — Livro II (Vida de São Bento), edição crítica segundo os três mss. portugueses conhecidos*. Tese de mestrado por nós apresentada à Universidade de Brasília em 1964.

8 Na altura (1967-1968), éramos bolsista da F. Calouste Gulbenkian em Lisboa.

9 QUÉMADA, B. Introduction. In: Actes du Colloque International sur la Mécanisation des Recherches Lexicologiques. *Cahiers de Lexicol*. Besançon, 3:13-18, 1961.

10 V. o artigo informativo de M. Tereza Camargo Biderman. Lingüística, Matemática e Computacional. Estudos Lingüísticos. *R. Brasil. Ling. Teor. Aplic.* São Paulo, 3 (1-2), jul./dez. 1968. e M. H. MIRA MATEUS. Informática e Lingüística. A Mecanografia nos estudos da linguagem. Separata da *R. de Portugal. Série A: Língua Portuguêsa*. Lisboa, 33, 1968.

11 O Equipamento que utilizamos compunha-se de: perfuradora IBM 026, verificadora IBM 056, separadora IBM 082 e tabuladora IBM 444-A.

12 Cf. os ns. I, II e III (1960) do *Bulletin d'information du Laboratoire d'Analyse Lexicologique*, Besançon, em que vários lingüistas franceses discutem a descrição mecanográfica das formas gramaticais da língua francesa.

13 QUEMADA, V. B. L'Inventaire Lexicographique en vue d'un Thesaurus National. *Cahiers de Lexicol*. Besançon, 3:119-133, 1961 e IDEM La Mécanisation dans les recherches lexicologiques, ibid v. 1, p. 7-46, 1961.

14 Uma primeira notícia de tal estudo que presentemente elaboramos se encontra na comunicação que apresentamos ao II Congresso da ALFAL (3-8 de janeiro de 1969): *O Estudo lingüístico de um texto português do séc. XIV*, a sair nas Atas.