

**A ARTE DE FAZER O MELHOR USO DE MAUS DADOS
ESTUDOS DIACRÓNICOS NO PORTUGUÊS ANTIGO**

Rosa Virgínia Mattos e Silva

1. APRESENTAÇÃO DO PROBLEMA

Diacronia é a sucessão de sincronias que constitui o processo histórico de mudança em qualquer língua: a uma definição desse tipo, que poderemos encontrar em dicionários de terminologia da Lingüística, preferi a de Labov:

**"A arte de fazer o melhor uso de maus dados"
(1982:20)**

não só por ser mais expressiva - de muito bom gosto até! -, mas

porque já aí transparece o tipo de trabalho que enfrentará todo aquele que se propuser abordar o estudo de uma língua no seu constante fazer-se. Lembro que estou me reportando à diacronia ao longo dos séculos ou no tempo real, já que os modernos estudos diacrônicos no tempo aparente podem compor os instrumentos adequados para a recolha dos seus dados, a partir dos princípios metodológicos que a Sociolinguística contemporânea vem afinando da década de sessenta para cá.

Nesta exposição me dedicarei a reportar alguns aspectos dessa "arte" com base em alguma experiência que venho adquirindo nos estudos diacrônicos da fase arcaica da língua portuguesa, que, convencionalmente, se estabelece entre os inícios do século XIII, com o primeiro documento (entre os até agora conhecidos e estudados) escrito em português - o Testamento de Afonso II - de 1214 e os fins do século XV, terminus ad quem determinado por fatos externos à história da língua portuguesa, mas significativos para a história cultural e política de quem a utilizava.

Todos sabemos que a Linguística chamada moderna deu prioridade absoluta quase, desde 1916, aos estudos sincrônicos, o que se pode avaliar como um corte teórico e metodológico que só foi benéfico para os avanços na compreensão do fenómeno linguagem humana. Se essa hegemonia dos estudos sincrônicos, ou mesmo acrônicos, perdura até hoje, não quer isso significar que a face sócio-histórica desse fenómeno não deva ser englobada na ciência da linguagem. No meu modo de ver, os estudos

Lingüísticos só atingirão a sua maturidade - se se pode falar em maturidade em ciência -, quando equilibrarem nos seus objetivos a busca da explicitação da linguagem humana tanto na sua face bio-psíquica, campo prioritário da Lingüística Teórica de hoje, tanto na sua face sócio-histórica, quer sincrônica quer diacrônica, que considero o campo da Lingüística Histórica no seu sentido amplo. Vale dizer que situo os estudos diacrônicos, aqueles que se dedicam à constituição das línguas através do tempo, como o campo de Lingüística Histórica no seu sentido estrito. Desenvolvi em outro trabalho esse ponto de vista (MATTOS E SILVA, 1988).

Se a Lingüística Histórica no seu sentido amplo já ocupa hoje grande espaço na pesquisa lingüística, a Lingüística Histórica no seu sentido estrito, hegemônica no século XIX, começa a renascer com vigor novo, já que alimentada pelos refinamentos teóricos e metodológicos da Lingüística do século XX. Reforçada tanto pelo desenvolvimento dos estudos de diacronia sincrônica da Sociolingüística que tem mostrado que a análise das mudanças em curso podem indicar caminhos para explicitar as mudanças através do tempo (LABOV, 1982), como pela recente "theory of grammar" que já admite que dos dados fornecidos pela mudança lingüística se podem obter "informações sobre o limite de gramáticas, sobre o momento em que o ambiente lingüístico muda de tal maneira que desencadeia um diferente tipo de gramática" (LIGHTFOOT, 1984: 149).

2. OS "MAUS DADOS" NOS ESTUDOS DIACRÔNICOS DA FASE ARCAICA DO PORTUGUÊS

Quando Labov define o estudo da mudança através do tempo como "a arte de fazer o melhor uso de maus dados" refere-se ao fato de que "os fragmentos da documentação escrita que permanecem são os resultados de acidentes históricos para além do controle do investigador". (1982:20).

No caso do português antigo, de que documentação naturalmente fragmentada dispomos?

Começarei pelo que não dispomos: qualquer trabalho de mudança lingüística em curso, de diacronia sincrônica, portanto, que se faça hoje deverá levar em conta a avaliação dos falantes da língua como uma das variáveis que fornecerá indicações sobre o peso que determinado elemento lingüístico em variação, e que esteja sendo observado, tem entre os usuários que constituem o grupo social em que se observa o fenômeno em causa. Na diacronia real não se contará, portanto, nem com possibilidade de avaliar socialmente a mudança que se observa, nem se contará, também, com as intuições do analista que, na melhor hipótese, será usuário da mesma língua, mas em outro estágio de sua história e deverá estar treinado no conhecimento lingüístico daquele estágio passado sobre que esteja pesquisando.

Para minorar tal situação, o analista poderá contar, no estudo de estados passados da língua, com testemunhos indiretos e rarefeitos de gramáticos ou de outras fontes

assemelhadas que podem fornecer indícios sobre as mudanças que se tem sob a mira.

Se a primeira situação nunca concorrerá nos estudos de mudanças em curso em estados lingüísticos pretéritos, a segunda poderá ocorrer. No caso da fase arcaica da língua portuguesa ela não concorrerá também, já que não dispomos de informação de tal natureza. Delas só se poderá dispor a partir do século XVI, quando surgem os primeiros trabalhos sobre a língua.

Diante disso, só contará o investigador com os fragmentos da documentação escrita resultado de acidentes históricos e com os princípios teóricos e metodológicos que lhe fornecerá a Lingüística em geral e, especialmente, a Lingüística Histórica.

O conjunto da documentação fragmentária que constitui o corpus total do português arcaico - do século XIII ao XV - pode ser classificado, e o farei aqui de maneira sumária, da seguinte forma:

a) documentação não-literária, oficial ou não: documentos jurídicos, notariais, administrativos que começam a aparecer nos inícios do século XIII e se vão geometricamente multiplicando a partir dos meados daquele século;

b) documentação literária, que se pode sub-dividir em dois grandes sub-grupos: documentação poética, corporificada no que se denomina de modo geral de O cancionero medieval português conjunto de poemas que não chegarão a dois mil, mas ultrapassarão mil e quinhentos, que se situa entre os inícios do

século XIII e os meados do XIV: documentação em prosa, que, para uma análise lingüística, deve ser também considerada em dois sub-grupos: os textos em prosa traduzida, em geral do latim, mas também de outras línguas, castelhano, francês, incipientes nos fins do século XII e crescentes nos subseqüentes; e a prosa originalmente escrita em português, que começa a surgir com muito vigor só nos começos do século XV.

Esse conjunto de textos é o universo total de que dispõe o investigador: os dados brutos, sobre os quais trabalhará, em geral, com a intermediação das edições paleográficas e/ou críticas, que devem ser previamente avaliadas como trabalhos filológicos fidedignos.

A par desse conjunto de textos - os dados brutos, há o que chamaria de dados secundários, constituídos pela bibliografia disponível que, sob várias formas, explorou essa documentação medieval, mas que, como o CORPUS natural remanescente, é também fragmentária, embora fragmentação de outra natureza, como veremos, e de qualidade extremamente variável.

O português antigo foi trabalhado, não há como negar, embora não tanto quanto as outras línguas românicas, na sua fase arcaica. No entanto, a natureza desses trabalhos, com raras exceções, é tão atomizada que, só com muita arte, se poderá reconstruir partes do quebra-cabeças, que é uma sincronia pretérita de uma língua.

Sobre tal bibliografia também se pode fazer uma classificação:

a) gramáticas históricas que pinçam exemplos para ilustrar suas regras, que seguem a tradição neo-gramática de análise histórica, sem preocupação de identificar as fontes e, sobretudo, sem preocupação de datá-los. Faça-se exceção entre essas gramáticas à obra de Joseph Huber (1986), que apresenta em apêndice o corpus sobre o qual trabalhou e é a única que aborda problemas sintáticos do português arcaico:

b) Introduções a edições críticas de textos medievais portugueses que informam, assistematicamente, em geral, sobre fatos isolados - fônicos, mórficos, sintáticos, lexicais - que, por razões diversas, interessaram ao filólogo editor crítico:

c) manuais de filologia que também, assistematicamente, fornecem informações muitas vezes preciosas e imprescindíveis:

d) glossários apensos a edições críticas, ou independentes delas, de natureza também diversa, que podem ser exaustivos ou seletivos, predominando os do último tipo, enquanto os do primeiro são de grande utilidade para estudos diacrônicos, como o de Ramón Lorenzo da versão galega da Crônica Geral e da Crônica de Castela (1975):

e) monografias raras e de qualidade variável sobre problemas específicos dessa fase da língua. Entre essas vale destacar a Linguagem dos Foros de Castelo Rodrigo (1959) de Lindley Cintra que, sistematicamente, explora um conjunto de documentos em galego-português e leonês e galego-português leonês

do século XIII, nos níveis fônicos e mórficos.

f) dicionários etimológicos, que, na maioria dos casos, não fornecem dados cronológicos sobre os vocábulos que historicam.

O investigador do português arcaico não disporá de gramáticas dessa fase da língua, não disporá também de um dicionário cronológico do português medieval. Encontrar-se-á numa imensa floresta, com sinalizações dispersas, caso deseje organizar seus dados para rastrear mudanças em curso que se refletem na documentação remanescente.

Para reforçar a importância de dados organizados, sistematicamente, a partir da documentação fragmentária original, ressalto, como exemplo, o destaque de David Lightfoot nos seus Principles of diachronic syntax (1979) para a obra de F. Th. Visser, An historical syntax of the English Language (1963-1973), alentado corpus sobre a sintaxe inglesa organizado em várias sincronias, que Lightfoot considera fundamental para os estudos diacrônicos que têm sido feitos sobre a sintaxe diacrônica do inglês.

Para o português antigo destaco a recente obra de M. Clarinda Maia, História do galego-português (1986) em que, pela primeira vez, no caso do português, se tem uma análise sistemática, ao nível fônico e mórfico, de uma documentação seriada, datada e localizada, do século XIII ao XIV, constituída de mais de 200 documentos provenientes da área galega e da área portuguesa até os limites do rio Douro. Só agora se dispõe de um

balizamento seguro de fatos fonéticos e mórficos datados, da área referida. Trabalho semelhante completaria o painel de dados para o resto do território português.

Ainda em elaboração, mas que será também instrumento básico para os estudos de mudança nessa fase histórica do português se encontra o Vocabulário Histórico-Cronológico do português medieval de A.G. Cunha, conhecido autor do Dicionário Etimológico Nova Fronteira da Língua Portuguesa (1982).

Com o estudo de Maria Clarinda Maia e o dicionário, em elaboração, de Antônio Geraldo Cunha os estudos diacrônicos do português antigo nos níveis fônico, mórfico e léxico já disporão de elementos ordenados para fazer melhor uso dos dados, por natureza "maus", que a documentação medieval remanescente lhes fornece.

3. A TENTATIVA DE CRIAR DADOS "MENOS MAUS" PARA O ESTUDO DIACRÔNICO DA SINTAXE NO PORTUGUÊS ARCAICO

É nessa direção que tenho tentado trabalhar. Despertei para a assistemática dos dados estudados disponíveis para essa fase da história do português quando ainda estudante me iniciei com o professor Nelson Rossi na história da língua portuguesa. Primeiro quando minhas colegas e eu elaboramos sob sua orientação a edição crítica do Livro das Aves (1965).

Depois, ao optar por fazer, como dissertação de Mestrado, em Brasília (1963-1965), a edição crítica do Segundo Livro dos Diálogos de São Gregório (Inédita). Por fim, quando realizei a edição crítica d'A mala antiga versão portuguesa dos Quatro Livros dos Diálogos de São Gregório, que veio a ser a minha tese de Doutorado (USP, 1971, aceita para impressão em 1983 pela IN-CM de Lisboa).

Desde a elaboração da dissertação de Mestrado, ao tentar fazer um estudo lingüístico daquele documento medieval, já consciente das exigências metodológicas do estruturalismo descritivista, vi-me na impossibilidade de realizar o que pretendia por falta de instrumentação, de modelos a que seguir, de tempo. Desisti então da empresa. Por um tempo...

Ao trabalhar, já em Portugal, na edição completa dos Diálogos de Gregório Magno, tive a oportunidade de ter, com o apoio de listagens mecanográficas, elaboradas pelo Centro de Cálculo Científico da Fundação Calouste Gulbenkian, um material exaustivo de levantamento de dados do corpus sobre o qual fizera a edição crítica que me permitia vir a fazer alguns anos depois - entre 1979-1981 - um estudo sistematicamente descritivo de um corpus medieval português que resultou no livro Estruturas trecentistas. Elementos para uma gramática do português arcaico (sob impressão no IN-CM de Lisboa).

Qual o meu objetivo com esse trabalho descritivo de um corpus definido do português da 2a. metade do século XIV?

Dispondo de vasto material listado, segundo uma

codificação por mim previamente preparada, pude, a partir daí, estruturar o que considero uma gramática parcial do português arcaico, daí o título que dei ao estudo depois de pronto. Realizei, de algum modo, o que não conseguira vários anos antes ao elaborar a dissertação de Mestrado. Organizei essa descrição lingüística daquele período trecentista em três grandes partes: análise do sintagma nominal (1a.), análise do sintagma verbal (2a.) e do enunciado simples e complexo (3a.). Segui a metodologia descritiva, estruturalista, não me cingindo em nenhuma das suas escolas.

Parti das formas documentadas, organizando o estudo do sintagma nominal e verbal, considerando a análise paradigmática dos elementos mórficos que os compõem e a combinatória desses constituintes nos sintagmas. Nessas duas partes, tem-se um estudo morfo-sintático descritivo, em que o latim só eventualmente foi acionado, sendo, no entanto, o português atual o ponto de referência básico, já que pensei a descrição para um público que como eu já não transita em profundidade pelo latim.

Na terceira parte, análise do enunciado simples, estabeleci regras de concordância, tipos de sujeito, tipos de frases (aqui entendida como a relação sujeito/predicado) e os diversos tipos de sintagma circunstancial. Quanto ao enunciado complexo, nele tratei dos mecanismos de coordenação, de subordinação e da ordem dos elementos no enunciado. Nesta terceira parte o trabalho valeu-se não só da análise formal, mas

sobretudo de uma análise de caráter funcional.

Todos os elementos submetidos a análise aparecem quantificados na sua freqüência de ocorrência e em percentuais. Com esta pesquisa constitui um alicerce para servir de amarra para futuros estudos de mudanças sintáticas em curso na fase arcaica do português.

Tentel cobrir de forma, é claro, parcial, mas previamente definida e explícita, uma lacuna que, pelo menos para mim, se constitua numa etapa essencial para uma compreensão de conjunto de fatos morfos, morfo-sintáticos e sintáticos e de suas posteriores mudanças nessa fase mais recuada da língua portuguesa, que, assistematicamente, se podia captar em alguns dos seus aspectos nos estudos existentes sobre o português antigo.

O convívio com essa análise de dados e com a bibliografia existente sobre o português arcaico me levantaram várias questões interessantes a serem observadas no percurso da constituição histórica do português.

É nessa nova etapa de estudo que venho dedicando parte do meu trabalho de pesquisa: a observação e a análise de fatos sintáticos em variação na documentação arcaica e, quando possível, determinar o processo de mudança em curso, em alguns casos, o momento de sua conclusão. Tenho batizado esse projeto de Variação e mudança no português arcaico de projeto aberto: por não determinar previamente tudo que pode vir a ser abordado: por não ter a pretensão de investigar todos os fatos passíveis de

investigação, isto é, fatos sintáticos em variação na documentação arcaica e por dele poderem participar aqueles que, no âmbito do Mestrado em Letras da UFBA, além de outros, se interessarem por investigação desse tipo. É antes uma linha de trabalho ou de pesquisa na qual podem ser incluídos projetos mais abrangentes ou mais restritos que se enquadrem nessa temática.

Por exemplo:

Evidenciou-se como muito interessante buscar determinar por que caminho e quando se definiu a oposição ser-estar em estruturas atributivas que não estava definida no português trecentista, mas já indicava o percurso de expansão de estar sobre ser, como procurei mostrar em estudo intitulado "Ser, estar... fazer... andar no português trecentista" (1987). No momento uma mestranda, Maria do Socorro Sepúlveda Neto trabalha sobre o problema em corpore da primeira metade do século XV e da primeira metade do século XVI para verificar quais as respostas às questões que levantei naquele estudo e que outras questões os dados levantavam e respondiam ou não.

Despertou o meu interesse na análise do corpus trecentista a questão da constituição das chamadas conjunções subordinativas e sua interrelação com sintagmas preposicionais. Sobretudo chamava a atenção a variação que havia entre pero e porém, como explicativas ou como adversativas. Examinei isso em uma pequena monografia - "Pero e porém: mudanças em curso na fase arcaica da língua portuguesa" (1984), confrontando os dados trecentistas com dados dos fins do século XIV, da primeira e

segunda metades do século XV e do século XVI. Cheguei a alguma resposta quanto à questão da mudança sintático-semântica que levou pero - do século XIV para o XV - passar do seu valor etimológico explicativo para o adversativo e porém - do século XV para o XVI - também sofrer a mesma mudança, desaparecendo depois pero do inventário das conjunções portuguesas.

Esse problema alertou-me para outros casos de mudança desse tipo no conjunto das conjunções e hoje outra mestranda, Sílvia Rita Magalhães, trabalha sobre o par ca e pois, a primeira originalmente explicativa, a segunda, originalmente temporal e já temporal e explicativa no século XIV. Hoje ca desaparecido do inventário, pois temporal, substituído por depois (que) e pois explicativa cobrindo o campo de ca, desaparecido. O percurso dessa mudança está sendo observado, a partir do corpora trecentista, em confronto com corpora da 1a. metade do século XV, podendo ser necessário ainda observar materiais da 2a. metade daquele século e do XVI.

Além desses, fatos outros tenho estudado como a variação ter/haver nos chamados tempos compostos - "Um aspecto do auxiliar no português arcaico" (1981) e "Ter__ e haver em estruturas de posse" (entregue para publicação), em ambos os dados do português trecentista foram confrontados com dados da 1a. e 2a. metades do século XV. Estes estudos sobre ter/haver pretendo estendê-los a outras estruturas em que tais verbos ocorrem (no existencial, na expressão do futuro, por exemplo) para construir um trabalho mais abrangente sobre a disputa desses

dois verbos na língua portuguesa, aos quais se devem relacionar outros verbos, ser e estar, pelo menos.

Com os dados organizados nas Estruturas trecentistas pude determinar a variação já existente na concordância verbo-nominal - "Contribuição para a leitura crítica de textos medievais portugueses: sintaxe e grafia" (1986). Essa questão deverá ser desenvolvida em confronto com outros documentos da fase arcaica da língua.

Quando cumprirei essas intenções? Quem as cumprirá além de mim? Possivelmente em futuros encontros possa responder a essas perguntas. Por isso considero essa linha de trabalho ou de pesquisa um projeto aberto, que se definirá na medida das possibilidades, contando de certo com um reforço positivo que é o meu prazer pessoal por esse tipo de estudo.

Não quero deixar de frisar, para concluir, que as Estruturas trecentistas. Elementos para uma gramática do português arcaico defino-a como uma tentativa de organizar com sistematicidade dados naturalmente fragmentados e por isso "maus" para o estudo da diacronia no tempo real a fim de, a partir daí, fazer o melhor uso possível deles na busca da compreensão e explicitação das mudanças em curso no português arcaico, reconhecendo, sem dúvida, que tudo isso poderia ter sido realizado com mais arte e engenho.

REFERÊNCIAS

1. CUNHA, A. G. - 1982. Dicionário etimológico Nova Fronteira da língua portuguesa. Rio, Nova Fronteira.
2. HUBER, J. - 1986. Gramática do português antigo. Lisboa, Gulbenkian (tradução do original de 1930 por Manuela Delle).
3. LABOV, W. - 1982. Building on empirical foundations. In: W. Lehmann e Y. Malkiel (eds.). Perspectives on historical linguistics. Amsterdam Philadelphia, J. B. Publishing Company. p. 17-92.
4. LINDLEY CINTRA, L.F. - 1959. A linguagem dos foros de Castelo Rodrigo. Lisboa, C.E.F.
5. LORENZO, R. - 1975. La traducción gallega de la Crónica General y de la Crónica de Castilla. 2 vols. Orense. Instituto "Padre Feijó".
6. LIGHTFOOT, D. - 1979. Principles of diachronic syntax. Cambridge, C.U.P.
7. IDEM - 1984. The language lottery: toward a biology of grammar. Cambridge Mass., MIT Press.
8. MAIA, M. Clarinda - 1986. História do galego-português. Coimbra, I.N.I.C., 1986.
9. MATTOS E SILVA, R.V. - 1981. Um aspecto do auxiliar no português arcaico. Tulane Studies in Romance Languages and Literature. 10. Tulane. p. 93-109.
10. IDEM - 1984. Perç e porém: mudanças em curso na fase arcaica

- da língua portuguesa. Boletim de Filologia, XXIX. Lisboa, p.129-151.
11. IDEM - 1988. Contribuição para a leitura crítica de textos medievais portugueses: sintaxe e grafia. Atas do Colóquio de Crítica Textual Portuguesa. Paris, Gulbenkian. p.85-98.
 12. IDEM - 1987. Ser, estar, lazer, andar no português trecentistas. Arquivos do Centro Cultural Português, XXIII, Lisboa-Paris, p.31-47.
 13. MATTOS E SILVA, R.V. - 1988. Fluxo e refluxo: uma retrospectiva da lingüística histórica no Brasil. Delta, 4 (1). p. 85-113.
 14. VISSER, Th. - 1963-1973. An historical syntax of the English language. V.1-IIIb. Leiden, Brill.
 15. ROSSI, N. - 1985. Livro das Aves. Rio, INL.