

A Pesquisa Diacrônica e o Problema do *Corpus*

César Nardelli Cambraia

(UFMG)

1. Introdução

Este trabalho tem como objetivo chamar atenção para algumas das dificuldades com quais um pesquisador interessado em realizar estudo diacrônico do Português necessariamente terá que lidar ao selecionar os textos que constituirão o *corpus* para sua análise. Serão abordadas aqui três dificuldades em especial: (a) a falta de uma tipologia de textos bem definida; (b) a inexistência de um mesmo tipo de texto em todas as fases da história do Português; e (c) a dificuldade de localização de edições críticas e fidedignas de textos antigos.

2. Problemas na escolha do *corpus*

2.1. A falta de uma tipologia bem definida de textos

Antes de passar para a análise dessa questão relacionada ao estudo diacrônico do português especificamente, será necessário falar a respeito do trabalho de Susan Romaine, intitulado *Linguística Socio-histórica – seu status e metodologia*, pois é nele que uma proposta metodológica de utilização de diferentes tipos de texto na pesquisa diacrônica é colocada de uma forma mais sistemática

2.1.1. O trabalho de ROMAINE (1982)

Susan Romaine utiliza a expressão *Linguística Socio-histórica* (o título do seu livro) para nomear a abordagem de análise linguística que propõe. Logo no prefácio de seu livro, a autora esclarece o que seria exatamente a *Linguística Socio-histórica*:

Este livro procura desenvolver um quadro teórico e metodológico para um campo de pesquisa que tratarei como “linguística socio-histórica”. O objetivo principal de uma disciplina como essa seria o de investigar e analisar as formas/usos em que a variação pode

manifestar-se em uma dada comunidade através do tempo, e como determinadas funções, usos e tipos de variação desenvolvem-se em determinadas línguas, comunidades de fala, grupos sociais, redes e indivíduos. (ROMAINE (1982: X)) (Tradução de minha autoria)

Essa proposta de pesquisa lingüística sócio-histórica é discutida pela autora através de uma análise dos marcadores de orações relativas (*relative clause markers*) no Escocês médio e no moderno. Na realização desse estudo, a autora adotou, como metodologia, diferentes tipos de textos para controlar variáveis externas, ou seja, extralingüísticas.

Nessa sua pesquisa, ROMAINE (1982:114) distribuiu os diferentes tipos de textos analisados nas seguintes categorias:

- a) Prosa: - prosa oficial e legal;
 - prosa literária (narrativa);
 - prosa epistolar.
- b) Verso: - Verso sério ou relativo à corte;
 - Verso religioso ou moralizante;
 - Verso cômico.

A divisão dos textos nessas categorias tinha como objetivo permitir que se analisasse o uso dos marcadores de orações relativas no Escocês médio, levando-se em conta o fator estilo. Uma das conclusões a que a autora chegou foi a de que realmente o estilo é um fator importante no uso das relativas (ROMAINE (1982:194)).

2.1.2. A pesquisa diacrônica com textos em língua portuguesa

Se, por acaso, algum pesquisador resolver adotar o método de pesquisa proposto por ROMAINE (1982), terá que enfrentar um grande problema: a questão da falta de uma tipologia bem definida de textos.

Se se tratar, por exemplo, de uma análise que levará em conta o fator estilo, não parece, a princípio, haver nenhum inconveniente em se selecionarem os textos, baseando-se nas categorias utilizadas por ROMAINE (1982). O problema é que, uma vez que as categorias que essa autora utilizou em seu trabalho não foram explicitamente definidas, a tarefa de escolha dos textos torna-se bem complicada. Para ilustrar essa complicação e suas

consequências, serão analisados a seguir dois problemas: o da oposição texto literário x não-literário e o da classificação de textos por assunto.

ROMAINE (1982) fez uma distinção entre prosa literária narrativa e prosa epistolar na divisão dos textos que utilizou em categorias. Uma vez que o adjetivo *literária* aparece somente no caso da prosa narrativa, fica a impressão de que textos em prosa epistolar não seriam considerados literários. Além disso, o adjetivo *narrativa* da categoria prosa literária narrativa aparece entre parênteses, o que reforça, então, a idéia de que a oposição é prosa literária x epistolar e não prosa narrativa x epistolar.

O problema da falta de definição fica, na verdade, mais evidente quando se tenta encontrar um texto que seria um representante da categoria em questão no caso, prosa epistolar não-literária. Seriam, por exemplo, as *Cartas dos Primeiros Jesuítas no Brasil* (LEITE (1954)) um exemplo de prosa epistolar não-literária? E *Cartas* do Padre Antônio Vieira (VIEIRA (s/d))? Seriam esses dois textos representantes de uma mesma categoria ou de categorias diferentes?

Ainda em relação à diferença entre prosa epistolar literária e não-literária, parece pertinente citar aqui as palavras de FIGUEIREDO (1960: 286):

Bom será não confundir a epístola, gênero poético e de intenção crítica, com a curta peça de correspondência comum, em prosa e de ampla liberdade de composição.

Como se vê, esse autor faz uma distinção entre a *epístola*, que seria um gênero poético e, portanto, de caráter literário, e a *correspondência comum* – ao que parece, um texto não-literário. Talvez partindo dessa definição, seria possível dizer, então, que as *Cartas de Antônio Vieira* seriam um exemplo de *epístola*, e as *Cartas dos Primeiros Jesuítas. no Brasil* seriam um exemplo de *correspondência comum*.

A questão, no entanto, não é tão simples assim. Há ainda o problema da difícil definição das fronteiras entre o literário e o não-literário. Um trabalho que parece colocar em pauta esse problema é o de WRIGHT (1989), intitulado *Linguagem privada trazida a público: a linguagem das cartas como literatura* (infelizmente ainda não foi possível ter acesso ao texto de Wright, mas o título provocativo já antecipa claramente tratar-se

de uma visão segundo a qual a correspondência privada é considerada texto literário).

O mesmo problema da oposição entre o que é literário e o que é não-literário aparece também no trabalho de CINTRA (1951). Em seu trabalho, são classificados apenas os textos literários medievais portugueses publicados. A autora trabalha com as seguintes categorias:

- a) Poesia: - profana e - religiosa;
- b) Poética;
- c) Novelística;
- d) Literatura religiosa, crônicas e regras monásticas;
- e) História;
- f) Viagens / epístolas;
- g) Prosa moralística;
- h) Fábulas;
- i) Textos jurídicos;
- j) Tratados técnicos.

Ao se analisarem as categorias acima, pode-se perceber que a autora parece trabalhar com um conceito de texto literário bem mais amplo. Ela considerou todos os textos analisados como literários (“*Procuro apontar na seguinte bibliografia todos os textos medievais portugueses de caráter literário (séc. XII a XV) até hoje publicados*” (CINTRA (1951:60)). Seria realmente um texto literário o *Tratado de cozinha* que a autora cita ao tratar da categoria tratados técnicos? Será que são igualmente literários as *Cantigas de Martim Codax* e o *Livro de como se fazem as cores*, ambos os textos presentes em sua classificação? São questões difíceis de se responder, já que lamentavelmente a autora não define o que entende por textos “*de caráter literário*”.

Para desfazer essa confusão entre o literário e o não-literário, seria necessário estabelecer explicitamente as características típicas de cada tipo, a fim de permitir a qualquer pesquisador, baseando-se em parâmetros bem definidos, decidir se determinado texto pertence a uma ou outra categoria.

Uma proposta nesse sentido parece ser a de PLATÃO & FIORIN (1990). Segundo esses autores, o texto literário apresenta os seguintes traços: *plurissignificação, desautomatização, conotação, relevância no plano da expressão e intangibilidade da organização lingüística*” (PLATÃO &

FIORIN (1990:353). Tal caracterização pode não ser perfeita, mas tem a vantagem de ser explícita. Teria, no entanto, que ser revista, se se quiser utilizá-la na classificação de textos antigos, pois algumas das características citadas parecem ser perceptíveis apenas quando se trata de texto da própria época do falante: a questão da desautomatização, por exemplo. PLATÃO & FIORIN (1990:351) entendem por *desautomatização* o processo em que se procura “*criar novas relações entre as palavras, estabelecer associações inesperados e estranhas*”. Seria um falante de Português do século XX capaz de perceber, em um texto do século XIV, casos de novas relações entre as palavras?

Um outro problema relacionado à questão da tipologia de textos é a questão da classificação por assunto. CINTRA (1951) faz a distinção entre *novelística e historiografia* (também sem definir claramente o que caracterizaria cada uma dessas categorias), mas reconhece que há casos em que é difícil decidir se o texto pertenceria a uma ou outra categoria. Isto fica claro no seguinte trecho:

Uma classificação sistemática como a que procuro fazer levanta inevitavelmente dúvidas. A mais importante das que se me apresentaram diz respeito ao lugar a atribuir à Vida de Júlio César e à Crónica Troyana, obras que oscilam entre a historiografia e a novelística. (CINTRA (1951:61))

A sua solução para esse caso específico revela, de certa forma, que a sua categorização parece ser mais intuitiva do que racional. Ela diz o seguinte:

Preferi introduzi-las um pouco arbitrariamente neste último capítulo [ou seja, na novelística] a uni-las às Crónicas e aos Livros de Linhagens incluídos no primeiro [isto é, na historiografia], junto dos quais me parece que ficariam muito deslocadas. (CINTRA (1951: 61))

2.2. A inexistência de um mesmo tipo de texto em todas as fases da história do português

Um segundo problema com que o pesquisador se depara é o fato de que nem sempre é possível encontrar um mesmo tipo de texto em todas as fases da história do Português.

Um estudioso que se proponha a analisar a língua portuguesa através de textos teatrais só poderá fazê-lo se for analisar o Português do século XVI em diante: os primeiros textos teatrais escritos em Português de que há notícia são os de Gil Vicente.

Ainda em relação aos textos teatrais, há a questão da oposição prosa x verso. No caso de Gil Vicente, as suas peças são escritas em verso. Já as peças de teatro contemporâneas muito dificilmente são escritas em verso – o que cria um certo impasse. Um estudo comparando o Português do século XVI com o do século XX através de peças teatrais teria o inconveniente de misturar textos de diferente natureza: a utilização da língua em um texto em verso obedece restrições (como rima e ritmo, por exemplo) que não se verificam em um texto em prosa.

2.3. A dificuldade de localização de edições críticas e fidedignas de textos antigos

Um terceiro problema é o da localização de edições críticas e fidedignas dos textos antigos. Apesar de haver alguns trabalhos em que são listadas as edições críticas de textos antigos em língua portuguesa que já foram publicadas (cf. o trabalho de CINTRA (1951), já citado anteriormente, e o capítulo II, intitulado *Publicação de textos*, do *Manual de filologia portuguesa* de SILVA NETO (1957)), esses trabalhos são antigos e não dão conta do que já foi publicado até os dias de hoje.

Para ilustrar essa questão da localização, será dado aqui um exemplo pessoal. Precisava de selecionar textos dos séculos XVI e XVII para utilizar como *corpus* em minha pesquisa de mestrado e resolvi escolher textos narrativos em prosa. Corno já havia achado um texto narrativo do século XVI, a *Peregrinação de Fernão Mendes Pinto* (PINTO (1952)), tentei encontrar um texto da mesma categoria – ou seja, relato de viagem – que pertencesse ao século XVII. Uma vez que os trabalhos de CINTRA (1951) e SILVA NETO (1957) concentram-se em texto do período medieval (até o século XV), foi necessário recorrer aos compêndios de literatura portuguesa para ter notícia de possíveis textos da categoria de em questão no século XVI. Nesses compêndios (como, por exemplo, na *História literária de Portugal* (FIGUEIREDO (1962))), não consegui localizar nenhum texto que se adequasse aos meus propósitos. Foi, no

entanto, consultando as estantes da biblioteca do Centro de Estudos Portugueses (na Faculdade de Letras da UFMG) que descobri o texto *Itinerário do Pe. Jerônimo Lobo* (LOBO (1971)). Esse texto é um relato que o autor faz de suas viagens pelo mundo, exatamente o mesmo tipo de tema de Fernão Mendes Pinto.

Tendo encontrado esse texto, procurei obter alguma informação sobre ele nos compêndios de literatura portuguesa: ao autores dos compêndios não citam sequer a existência do texto. Ao se comparar a data dos compêndios consultados (FIGUEIREDO (1962), CIDADE (1959) e MOISÉS (1970)) com a data de publicação do texto o mistério se desfez: a publicação do texto do Pe. Jerônimo é posterior à data dos compêndios e, portanto, não poderia constar neles.

3. Para concluir

Como conclusão das breves considerações feitas acima, farei alguns comentários sobre propostas de solução para algumas das questões levantadas:

Em primeiro lugar, é necessário desenvolver uma classificação de textos, ou seja, uma tipologia, que seja explícita: não basta apenas adotar uma nomenclatura (como foi feito por CINTRA (1951), por exemplo), é preciso explicitar quais características um texto deve apresentar para que seja incluído em uma ou outra categoria. Uma classificação explícita permitiria inclusive a qualquer pesquisador classificar aqueles textos que, por ventura, não constem nos compêndios.

Além disso, uma classificação explícita talvez permitirá o reconhecimento de categorias semelhantes – algo que justificaria a adoção de um texto de uma outra categoria com o objetivo de preencher lacunas resultantes da inexistência de determinado tipo de texto em todas as fases da história do Português. CYRINO (1993), por exemplo, utiliza textos teatrais como *corpus*, em seu trabalho em preparação. Utiliza, no entanto, as poesias de Gregório de Matos para preencher a lacuna decorrente da não-identificação de um texto teatral brasileiro do século XVII (e talvez nem exista).

Por fim, há a questão da dificuldade de localização dos textos que poderia ser solucionada através de uma espécie de catálogo em que se registrariam as publicações de edições críticas – o importante é que esse catálogo esteja sempre sendo atualizado.

Referências Bibliográficas

CIDADE, Hernani. *Lições de cultura e literatura portuguesas*. 4. ed. Coimbra: Coimbra Editora, 1959.

CINTRA, Maria Adelaide Valle. Bibliografia de textos medievais portugueses publicados. *Boletim de filologia*. Lisboa, 12: 60-100, 1951.

CYRINO, S. M. L. Observações sobre a mudança diacrônica no português do Brasil: objeto nulo e clíticos. In: ROBERTS, I. & KATO, M. A. (Orgs.). *Português brasileiro; uma viagem diacrônica*. Campinas: Editora da Unicamp, 1993.

FIGUEIREDO, Fidelino de. *História literária de Portugal*. 2. ed. Rio de Janeiro: Fundo de Cultura, 1960.

LEITE, Serafim S. I. *Cartas dos primeiros jesuítas no Brasil I*. São Paulo: [s.n.], 1954.

LOBO, Pe. Jerônimo. *Itinerário e outros escritos inéditos*. Barcelos: Livraria Civilização, 1971. Ed. crítica pelo Pe. M. Gonçalves da Costa. (Biblioteca Histórica - Série Ultramarina)

MATTOS E SILVA, Rosa Virgínia. *Estruturas trecentistas*, Elementos para uma gramática do português arcaico. Lisboa: Imprensa Nacional-Casa da Moeda, 1989.

MOISÉS, Massaud. *A literatura portuguesa*. 8. ed. São Paulo: Cultrix, 1970.

PINTO, Fernão Mendes. *Peregrinação*; seguida de suas cartas, Lisboa/Rio de Janeiro: Sociedade de Intercâmbio Cultural Luso-Brasileira/ Casa do Estudante do Brasil, 1952. Vol. I e II.

PLATÃO, Francisco & FIORIN, José Luiz. *Para entender o texto*; leitura e redação. São Paulo: Ática, 1990.

ROMAINE, Suzanne. *Socio-historical linguistics; its status and methodology*, Cambridge: Cambridge University Press, 1982.

SILVA NETO, Serafim da. *Manual de filologia portuguesa*. 2. ed. Rio de Janeiro: Acadêmica, 1957.

VIEIRA, Pe. Antônio. *Cartas do Padre Antônio Vieira*. Lisboa: Empreza Litteraria Fluminense, s.d.

WRIGHT, Susan. Private Language Made Public: The Language of Letters as Literature. *Poetics: Journal for Empirical Research on Literature, the Media and the Arts* 1989 dec. v. 18(6) p.549-578.