

Para um dicionário etimológico do português arcaico

Américo Venâncio Lopes Machado Filho

Universidade Federal da Bahia
Alameda dos Umbuzeiros, 145/401
41820-680 Salvador Bahia Brasil
americovenancio@gmail.com

Abstract

Although an important number of linguistical studies has been devoted to Old Portuguese, up til now no systematical effort has been accomplished to the elaboration of a dictionary of this period of the language. This poster is intended to be a brief notice of the working up of a project that has been developed in Universidade Federal da Bahia, in Brazil, and that aims to organize an etymological dictionary of Old Portuguese, on eletronic databases of medieval documents from the XIIIth, XIVth, XVth and XVIth centuries.

1 Motivação

Embora seja a língua portuguesa registrada pela escrita, pelo menos, desde o século XIII, não dispõe ainda hoje de dicionários relativos a seu período arcaico, senão algumas poucas tentativas de coleção de unidades lexicais, comumente apresentadas em glossários, que não exibem a sistematicidade dicionarística hoje perseguida.

Muito desse material se resume, por vezes, a listas de palavras e frequências, com algumas ocorrências, sem qualquer referência à etimologia ou aos significados que poderia o léxico comportar na língua no período de seu registro, o que, de certa forma, vela sua utilização prática e imediata, como instrumento de consulta.

Verdelho (1995: 138) acredita que

interessa notar que a memória lexical e lexicográfica de uma língua constitui um dos fundos patrimoniais mais interactivos. Garante, por um lado, a ligação com o texto e a mensagem do passado (...). Por outro lado, dá continuidade e assegura, entre os grupos humanos, uma identidade lingüística.

Afinal é no léxico que se espelham não apenas significados e conceitos do mundo material, social e abstrato com que se relaciona e que, ao fim e ao cabo, procura conformar, mas da própria gramática pregressa da língua, já que por mais nevoento ou mortal que se vislumbre o cenário da existência humana a que se atrela qualquer língua, mais sobrevivente tem sido a palavra como suporte a essa história.

Brucker (1988: 105) acredita que

Os dicionários etimológicos representam a última etapa e, por numerosas razões, a mais difícil e a mais

complexa da pesquisa etimológica. A ambição suprema da etimologia e da lexicologia histórica é a de fornecer uma ilustração completa e exaustiva de todo o léxico de uma língua ou de um grupo de línguas, que deveria constituir um tipo de estado de conhecimento e da pesquisa em determinados setores (tradução nossa).

Revela-se, portanto, como fonte de fundamental importância a sistematização e elaboração de dicionários de momentos pretéritos, com vistas a servir de contributo relevante para o conhecimento ainda disperso da história interna e externa do léxico da língua portuguesa, sobretudo os que se voltem ao levantamento de sua etimologia.

É, pois, objetivo principal deste projeto, a elaboração de um Dicionário Etimológico do Português Arcaico, nomeadamente sobre o arco temporal compreendido entre os séculos XIII e meados do XVI.

2 Bases e viabilidade

Pode-se dizer que a “metodologia de dicionário etimológico é ainda um terreno relativamente novo e suscetível a aperfeiçoamentos”, como afirma Brucker (1988:115), antes citado.

Os trabalhos de natureza lexicográfica em língua portuguesa têm por tradição, até o presente, basear-se normalmente em referências arquetípicas de grandes trabalhos precedentes. Essa tradição é, entretanto, bastante restrita, se considerado comparativamente o espólio de outras línguas de cultura de sua família ou mesmo de seu tronco lingüístico. Da Idade Média portuguesa, por exemplo, só sobreviveu à ação causticante do tempo um dicionário de verbos do *latim/português*, provavelmente do século XIV, mas nenhuma outra obra em ordem inversa.

Não obstante, a construção do conhecimento lexicográfico se desenvolve verticalmente a partir do século XVI, com o Renascimento, e reconhece nos séculos posteriores grandes produções, que viriam a servir de base para a construção de dicionários modernos de língua portuguesa.

Na contemporaneidade, entretanto, não se podem mais elaborar dicionários sem uma base de dados eletrônica (cf. Sterkenburg 2003: ix). Por isso a seleção equilibrada dos *corpora* e sua captação em bases digitais são de fundamental importância para a consecução de objetivos de natureza lexicográfica, hoje.

No que tange ao período arcaico do português, todavia, o espólio com que pode contar o pesquisador não permite o mesmo balanceamento em relação a gêneros e tipos de textos, como soeria ocorrer na elaboração de um dicionário geral contemporâneo, sobretudo em função de não existir uma quantidade considerável de edições filologicamente confiáveis, disponibilizadas em meio digital.

Em face dessa realidade, o Projeto Deparc conta, em sua primeira fase, com um *corpus* representativo do período arcaico do português, nomeadamente de textos literários e textos não-literários, cujas edições são integrantes do Projeto BIT-PROHPOR (Banco Informatizado de Textos do Programa para a História do Português), da Universidade Federal da Bahia.

Como as características específicas dos textos medievais têm, de certa forma, obliterado a utilização plena dos recursos de programas informáticos da lexicografia moderna, nomea-

damente no que se relaciona com a variação gráfica dos vocábulos, o projeto tem perseguido desenvolver, junto a analistas, um sistema de informação, denominado *Prodeparc*, para instrumentar a elaboração, a utilização e a evolução do Projeto, sobretudo para construção de um programa informático próprio, contribuindo com isso com o aprimoramento da lexicografia eletrônica. Por ora, tem-se recorrido ao *Oxford WordSmith 4.0*, para elaboração de *wordlists* e concordâncias.

A microestrutura dos verbetes exibem a seguinte configuração:

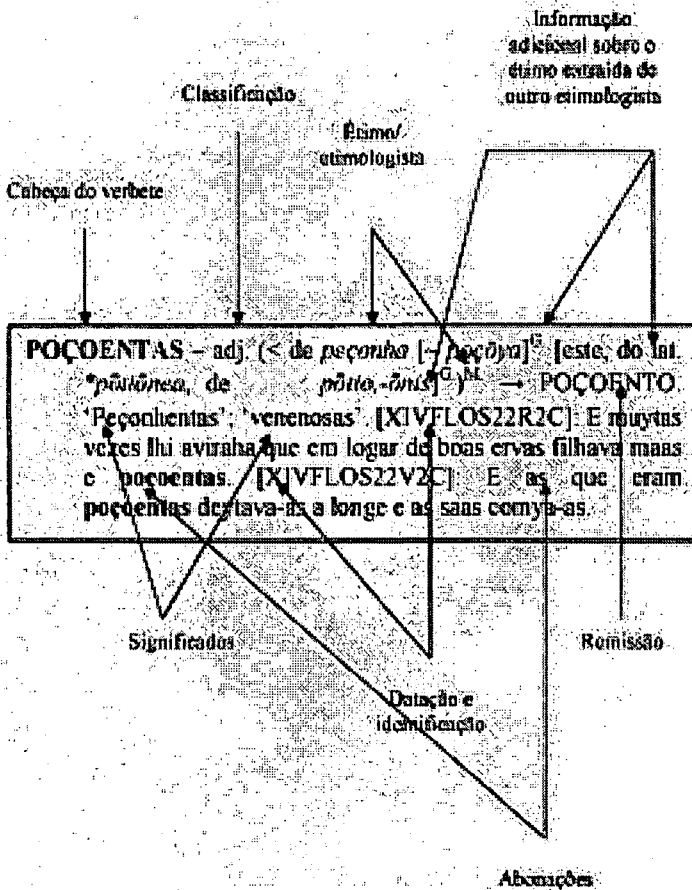


Figura 1 Exemplo de verbete

De macroestrutura estritamente alfabética, o dicionário tem tratado com especial interesse a variação, sobretudo gráfica, que pode revelar comportamentos lingüísticos diatópicos ou diafásicos bastante relevantes para o conhecimento da história da língua, especialmente de natureza fonético-fonológica e morfossintática.

Nesse sentido, o processo de lematização não se define pela forma canônica, mas por

toda e qualquer ocorrência de vocábulos levantados nos *corpora*, a partir das *wordlists* geradas. Os verbetes nesses casos são meramente remissivos à entrada considerada padrão, a partir dos critérios definidos para sua eleição, nomeadamente o que se relaciona com a frequência no *corpus*, e sua abonação é então nesse verbete registrada.

Colocações e fraseologismos são igualmente considerados, constituindo verbete próprio, quando se o justifique.

Por buscar servir sobretudo a estudiosos do passado da língua, os verbos têm tratamento diferenciado na composição de sua microestrutura, exibindo, depois de sua forma canônica no infinitivo, todas as conjugações verbais e as correspondentes variantes detectadas.

Não obstante os resultados práticos de oportunidades de uso, o DEPARC é ainda um projeto de geração de conhecimento, pois trabalhar incessantemente sobre o léxico de uma língua é, como disse Saramago (2002: 61) poder perceber “aquilo que de modo variável vão significando e representando, por imortais, imorredouras, ou eternas”.

Referências

- Brucker, C. (1988), *L'étymologie*. Paris: Presses Universitaires de France.
Saramago, J. (2002), *O homem duplicado*. São Paulo:, Companhia das Letras.
Sterkenburg, P. (ed.) (2003), *A practical guide to lexicography*, Amsterdam, John Benjamins.
Verdelho, T. (1995), *As origens da gramaticografia e da lexicografia latino-portuguesas*, Aveiro, IN-CM.