

Construção de Software para o Estudo da Língua Grega Clássica: o Sistema Nominal do Dialeto Ático na Visão Temática

Edenis Gois Cavalcanti¹

¹ Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo (USP)

edenis@usp.br

Abstract. *The aim of this research is to study the nominal flexion of the Attican dialect in a digital environment, so as to fill a gap in studies of ancient languages with the use of digital technologies. The main objectives are: 1) the construction of a computer program that will make nominal flexion possible, providing the analysis of terms with, namely: the nominal theme, the case, gender, number, a table with all the combinations proposed by the program and the dictionary entry; 2) creation of a corpus of Attican dialect, with the generation of a database lexicon, in order to verify the existence of forms created by the program and provide subsidies for linguistic analysis, 3) creation of a direct search mechanism in the Thesaurus Linguae Graecae (TLG), from the Research Center of the University of California, Irvine, founded in 1972, with 99 million words - from Homer to the sixteenth century. In the generation of the program, we have used the Pascal Object Oriented Language, developed by Borland Delphi, for the Enterprise Software Corporation. For the development of the database - lexicon, table of desinencies and vocabulary - the Manager Database SQL Server 2005 Express, Microsoft Corporation was employed. This project offers two main contributions to Greek language studies: one - of a strictly linguistic character - presenting desinencies briefly, without the usual lengthy descriptions, facilitating viewing and understanding for language students. Another contribution - interdisciplinary - is the convergence of new digital technologies with the study of Ancient Greek, since morphological analysis is made available in the digital environment, fully and simultaneously. The information is provided in a very short time and with a negligible margin of error.*

Key-words: nominal flexion; Greek language; Attican dialect; new digital technologies; computer programming; linguistics; morphological calculus

Resumo. *Esta pesquisa tem por objeto de estudo a flexão nominal do dialeto ático em ambiente digital. A justificativa central relaciona-se ao preenchimento de uma lacuna existente nos estudos das línguas clássicas, isto é, ao uso das tecnologias digitais nos estudos das línguas antigas. Os objetivos principais são: 1) a construção de programa de computador que viabilize a flexão nominal, fornecendo a análise da palavra digitada, ou seja: o tema*

nominal, o caso, o gênero, o número, tabela com todas as combinações propostas pelo programa e entrada no dicionário; 2) criação de corpus do dialeto ático que permita a geração de um léxico em banco de dados, visando verificar a existência das formas criadas pelo programa e fornecer subsídios para eventual análise lingüística; 3) criação de mecanismo de busca direta no corpus do TLG, Thesaurus Linguae Graecae (Centro de Pesquisa da Universidade da Califórnia, Irvine, fundado em 1972), com 99 milhões de palavras – de Homero até o século XVI. Na geração do programa, empregamos a Linguagem Pascal Orientada a Objetos, desenvolvida pela Borland Delphi, da Enterprise Software Corporation, e na constituição da base de dados – léxico, tabela de desinências e vocabulário - o Gerenciador de Banco de Dados SQL Server 2005 Express, da Microsoft Corporation. São duas as contribuições ao estudo da língua grega: uma – estritamente lingüística –, quando apresentamos as desinências nominais sinteticamente, sem as longas descrições tradicionais, facilitando a visualização e compreensão, para o estudioso da língua. Outra – interdisciplinar, na convergência das novas tecnologias digitais com as letras clássicas –, quando disponibilizamos a análise morfológica em ambiente digital de forma integral e simultânea. As informações são fornecidas em curto espaço de tempo e com margem de erro desprezível.

Palavras-chave: flexão nominal; língua grega; dialeto ático; novas tecnologias digitais; programa; lingüística; cálculo morfológico.

1. Objeto de estudo, justificativa e objetivos

Este trabalho tem por objeto de estudo a flexão nominal do dialeto ático – uma das variantes dialetais do grego antigo, juntamente com o jônico, o dórico e o eólico. O interesse em processar o dialeto ático em ambiente digital surgiu em 2001, nas aulas de grego clássico do professor Dr. Henrique Graciano Murachco – Departamento de Letras Clássicas –, nos cursos de extensão da Universidade de São Paulo.

Pelo fato de o método do professor Henrique não privilegiar o modelo das gramáticas tradicionais, considerado por ele descritivo, impositivo e prescritivo, mas *pensar* as formas lingüísticas – nominais e verbais – levando em consideração suas partes fundamentais – a saber, o tema (parte fixa, sede do significado virtual, absoluto ou abstrato) e as desinências (que estabelecem a função, a pessoa, o número, a voz, o modo e o aspecto) – e os processos fonéticos que integram a sua *história*, criou-nos a convicção de sua coerência metodológica, visto desaparecerem as irregularidades morfológicas, e, portanto, a idéia de que era possível flexionar as formas em um programa de computador. Bastaria, para isso, criar um banco de dados com todas as desinências nominais e verbais, extrair a desinência da palavra a ser pesquisada e submetê-la a um processo comparativo com as desinências do banco e, desse modo, gerar as combinações possíveis. Criou-se então um protótipo digital da flexão ática e, para este trabalho, decidimos desenvolver o módulo relativo à flexão nominal.

Embora concordando com as gramáticas gregas são impositivas, não podemos negar o papel de importância indiscutível que exerceram no ensino da língua grega. Recorremos, no nosso estudo, inúmeras vezes a elas, ora coletando informações, ora confrontando-as em pontos duvidosos ou lacunosos, mas *sempre as consultando*. São *relas* que, bem ou mal, encontramos o *depósito das formas*. Ainda que façamos uso de programas de computador em nossas pesquisas, nossa intenção não é produzir um discurso de refutação, de desconsideração ou negação da interferência do outro, mas de complementação. Acreditamos que o saber é coletivo, interdiscursivo, dialógico, feito por várias mãos e povoado por várias vozes. Reconhecemos as limitações, mas, também, o valor histórico das gramáticas.

O avanço das novas tecnologias digitais em todos os ramos da vida humana coloca, entretanto, desafios intransferíveis às pesquisas lingüísticas. Não obstante as gramáticas tradicionais tenham desempenhado papel relevante na história do ensino, é imprescindível, hoje, que o conhecimento lingüístico esteja integrado ao conhecimento digital.

Há um *déficit lingüístico* muito grande em estudos de língua grega no Brasil, que utilizam os recursos computacionais de forma integrada. Não há ferramenta informática – grego-português – que forneça *simultaneamente* diversas informações lingüísticas sobre as formas nominais. Por causa disso, entendemos que a justificativa inicial de nosso trabalho esteja relacionada ao preenchimento da lacuna nos estudos das línguas clássicas. A nossa contribuição relaciona-se, portanto, à criação de um programa de computador que processe informações sobre as formas nominais, isto é, realize a flexão nominal, estabelecendo, *simultaneamente*, o tema, o caso, o gênero, o número, *o conjunto das combinações* derivadas da *análise e do cálculo morfológicos*, quadros paradigmáticos de flexão, entrada no dicionário e busca rápida em *corpus extenso* com margem de erro desprezível, que possibilite o confronto com as formas criadas pelo computador, de forma a apontar para a validade ou não das combinações. Isso realizado, diríamos que à ferramenta caberia o processamento de todas as informações e ao estudioso a tarefa de compreender os conceitos e interpretar os resultados.

Outro motivo relaciona-se à eficácia fornecida pelas tecnologias digitais. Todos nós sabemos como é penoso "decifrar" o nominativo singular dos nomes da terceira declinação (nomes com temas em consoante, soante, semivogal) ou das formas contratas, no grego antigo. Por causa dos acidentes fonéticos – que provocam as aparentes irregularidades morfológicas –, a busca da entrada nos dicionários torna-se, muitas vezes, desanimadora. É necessário que se tenha pleno domínio dos processos fonéticos, para efetuar a decomposição das formas e, assim, encontrar o nominativo singular e, portanto, a entrada no dicionário. Isso implica em investimento de *muito* tempo em exercícios de tradução para se adquirir o *know-how* satisfatório, o que provoca a desistência, não raras vezes, do estudante da língua grega. Por outro lado, se armazenássemos em um banco de dados todas as desinências nominais, com seus respectivos processos fonéticos, e transformássemos a "rotina" humana de decodificação nominal em um programa de computador, necessariamente o tempo da análise diminuiria absurdamente.

Postulamos, portanto, inequivocamente, o uso das novas tecnologias digitais nos estudos lingüísticos, não como algo *estranho* que se agrega, mas como elemento constitutivo

do processo de produção do conhecimento lingüístico. Essa postulação está filiada, como vimos, a uma tendência universal irreversível de informatização generalizada das rotinas humanas que exigem repetição, precisão, volume e rapidez no processamento; tendência, aliás, presente nos processos científicos e acadêmicos de produção do conhecimento dos grandes centros de excelência.

Em coerência com o exposto acima, objetivamos, com este trabalho:

1) construção de programa de computador que viabilize a flexão nominal, fornecendo a análise da palavra digitada, isto é: o tema nominal – em vogal, consoante, soante, semivogal; o caso da forma digitada pelo usuário – nominativo, vocativo, acusativo, genitivo, dativo, locativo e instrumental; o gênero – masculino, feminino, neutro; o número – singular e plural; o participípio – *infectum*, *aoristo* e *perfectum*, em suas vozes ativa, média e passiva; tabela com todas as combinações propostas pelo programa, com seus respectivos quadros de flexão; entrada no dicionário – nominativo singular; quadros paradigmáticos para eventual confronto com as combinações propostas pelo programa.

2) Criação de *corpus* do dialeto ático que permita a geração de um léxico em banco de dados, com indexação diferenciada dos itens lexicais, indicando o autor, a obra e a linha, visando verificar a existência das formas criadas pelo programa – busca indireta – e fornecer subsídios para eventual análise lingüística, através do cálculo dos valores lexicais em cada autor e em cada item lexical. Para isso, usaremos a noção *desvio reduzido* da Estatística Paramétrica, explicitada no método matemático-estatístico-computacional de análise de textos de André Camlong, para a confecção das T.V.Ls. (Tabelas de Valores Lexicais).

3) Criação de mecanismo de *busca direta* no *corpus* do TLG, *Thesaurus Linguae Graecae* (Centro de Pesquisa da Universidade da Califórnia, *Irvine*, fundado em 1972, que digitalizou textos literários escritos em grego, desde Homero – século VIII a.C. – até a queda do Império Bizantino em 1453, totalizando 3800 autores, 12000 obras, 99 milhões de palavras). A busca direta nos textos tem por finalidade única verificar, de forma rápida e ampla, a existência das formas propostas pelo programa. A diferença entre as duas buscas – a direta e a indireta – reside no tamanho do universo lingüístico pesquisado e na forma da organização dos dados: a indireta realiza a pesquisa *no léxico de 11 autores*, dentro de um *banco de dados*, e a direta busca a forma *nos textos dos 3800 autores*. A intenção de alargar o universo de busca é pôr em prova, através de um *corpus* "gigante", as combinações das formas propostas. Se houver convergência entre as combinações propostas e a busca realizada no *corpus*, isto é, se a busca estornar resultado positivo, isso indica que as formas propostas pelo programa procedem.

4) Reorganização das desinências nominais em *conjuntos de desinências*, diferentemente da organização apresentada pelas gramáticas;

5) Criação das tabelas de flexão dos participípios. As gramáticas apresentam apenas o nominativo masculino, feminino, neutro singular, em cada aspecto verbal, deixando de desenvolver os demais casos;

6) Apresentação do processo fonético que deu origem à determinada forma nominal.

2. Desenvolvimento: ambiente e banco de dados

A escolha do ambiente de programação implicou algumas dificuldades. A primeira delas diz respeito à grande variedade de linguagens de programação existentes no mercado. Precisávamos de uma linguagem que fosse produtiva, confiável e que tivesse acesso a vários bancos de dados. A segunda, à não-existência de códigos abertos (*open source*) ou projetos disponíveis semelhantes para que pudéssemos ter um ponto de partida na construção de um projeto de flexão nominal do dialeto ático. A terceira, ao suporte técnico à linguagem escolhida

Dentre as linguagens do mercado, optamos em trabalhar com as que eram orientadas a objetos. Dentre elas, destacam-se o C++ e o Delphi, em sua versão 7, que tem, como linguagem nativa, o Pascal orientado a objetos. Decidimos pelo Delphi, pelo fato de ser um ambiente de fácil manipulação e de alta produtividade, além de ter suporte técnico disponível, fornecido por várias empresas, no Brasil (filiamo-nos ao Clube do Delphi, sediado na cidade de Avaré, interior do estado de São Paulo).

A escolha do banco de dados também não foi simples. Deveria ser robusto, confiável, com excelente performance e gratuito. Os que se encaixavam nessas características eram: SQL Server 2005, PostGresSQL, FireBird e MySQL. Para efetuarmos a escolha, realizamos um teste de performance com a tabela do léxico do dialeto ático, com 200689 registros. A busca realizou-se sobre o último registro da tabela. O resultado foi o seguinte:

200689 registros							
Busca do último registro							
Banco de Dados (Gratis)	Capacidade	Componente	Tempo em ms	Performance		Problemas de cache	Nativo do Delphi 7
SQL Server 2005	4 GB por banco	ADOQuery	140	21,34%	1º	N	S
PostGresSQL	sem limites	ZQuery	172	26,22%	2º	N	N
FireBird	sem limites	SQLQuery	188	28,66%	3º	S	S
MySQL	sem limites	ZQuery	656	100,00%	4º	S	N

Tabela1: performance dos Bancos de Dados

A escolha, a partir do teste de performance, ficou clara: SQL Server 2005. Embora sua capacidade esteja limitada a 4 GB por banco, é extremamente rápido, não dá problemas de *cache* local e pode ser acessado por componente nativo do Delphi – o componente ADO, *Active Data Objects*.

Na atual fase do projeto, utilizamos o SQL Server 2005 e o FireBird. Na etapa final, migraremos todos os dados para um único banco, isto é, o SQL Server 2005.

3. Constituição do *Corpus* Ático: fontes, obras, autores, extrato dos textos e léxico

O *corpus* que deu origem ao léxico foi formado a partir da coleta de textos na *Internet*, nos *sites* de domínio público, que disponibilizam obras de autores antigos. Basicamente, foram duas as fontes: a) *Perseus Digital Library*, em <http://www.perseus.tufts.edu/>; e b) *Thesaurus Linguae Graecae*, na seção aberta ao público, em <http://www.tlg.uci.edu/demo.html>.

Os textos extraídos da *Internet* foram processados e submetidos a uma indexação diferenciada das encontradas comumente. Para tal, criou-se uma *ferramenta digital*, de nome *Corpus*, que gerou o Léxico com: a) as frequências de uso dos itens lexicais nas obras processadas, no todo e nas partes – autor por autor; b) o endereço de cada item lexical, contendo: nome do autor, obra e linha em que se deu a ocorrência da palavra; c) o peso (valor) lexical nas obras do *Corpus*.

Tendo em vista o objeto deste trabalho, optamos por selecionar um conjunto de autores que escreveram em grego ático. A obra completa de cada autor está em formato texto (*.txt), num mesmo arquivo. Isso facilita a busca morfológica e o reconhecimento por qualquer plataforma, como Windows e Linux.

As obras dos autores que escreveram em grego ático, no nosso *corpus*, somam um total de 316.

4. Conclusão

Enfatizamos que a pesquisa está em curso e, por isso, muitos aspectos podem sofrer alterações. Acreditamos, entretanto, que a idéia geral do projeto, a saber, a construção de um programa de computador que forneça um *conjunto de informações* sobre a flexão nominal do dialeto ático, está, a cada passo da nossa pesquisa, concretizando-se.

A orientação principal de se constituir um banco de dados para armazenar informações para tratamento computacional da língua grega já é uma realidade neste trabalho. Inúmeras vantagens poderiam ser listadas, mas ficamos apenas com a relacionada à eficácia e alta produção lingüísticas. A análise e o cálculo morfológicos efetuados pelo computador – que possibilita a flexão nominal (e futuramente a flexão verbal) – e a busca das formas nos textos dos autores, num espaço de tempo ínfimo – num *corpus de aproximadamente 99 milhões de palavras* – são vantagens indiscutíveis do uso das Novas Tecnologias Digitais nas pesquisas lingüísticas.

Referências

BAILLY, Anatole. *Dictionnaire grec français*. 26e ed. Paris: Hachette, 1963.

CAMLONG, André. *Lexicométrie I - La corrélation simple*. O teste t de Student-Fisher ou a regressão linear simples? Toulouse: Université de Toulouse Le-Mirail, 2003.

_____. *Lexicométrie II - La corrélation simple et la corrélation multiple*. Toulouse: Université de Toulouse Le-Mirail, 2003.

_____. *Lexicométrie III ou La lexicométrie avec Stablex*. Toulouse: Université de Toulouse Le-Mirail, 2003.

_____. *Méthode d'analyse lexicale textuelle et discursive*. Paris: C.R.I.C. & OPHRYS, 1996.

_____. *Stablex pratique*. Toulouse: Teknea, 1991.

_____; BELTRAN, Thierry. *Stablex. Manuel d'utilisation*. Toulouse: A.P.I, 1991.

_____; _____. ZAPPAROLI, Zilda Maria. *A propósito do Stablex - versão PC*. São Paulo: Pirus Tecnologia, 2004.

CANTÙ, Marcos. *Dominando o Delphi 2*. Trad. de Edimilson Kazwyoshi Miyasaki. São Paulo: Makron Books, 1996.

MURACHCO, Henrique. *Língua grega: visão semântica, lógica, orgânica e funcional*. São Paulo: Discurso Editorial / Vozes, 2001.

ZAPPAROLI, Zilda Maria; CAMLONG, André. *Do léxico ao discurso pela Informática*. São Paulo: EDUSP/FAPESP, 2002.