

EDIÇÕES DIGITAIS NAS LETRAS CLÁSSICAS: ELEMENTOS DE GÊNERO NA PRODUÇÃO E LEITURA DA LÍNGUA GREGA E DO LATIM

DIGITAL EDITIONS IN CLASSICS: ELEMENTS OF GENRE IN THE PRODUCTION AND READING OF THE GREEK AND LATIN LANGUAGES

*Anise Ferreira**

Universidade Estadual Paulista, Araraquara, SP, Brasil

*Christopher Blackwell***

Furman University, Greenville, SC., Estados Unidos

*Chiara Palladino****

Furman University, Greenville, SC., Estados Unidos

Resumo: As edições digitais na área de clássicas têm surgido como aliadas no rompimento de barreiras e no enfrentamento de desafios no ensino das línguas e culturas clássicas. O objetivo deste artigo é destacar algumas características específicas de edições digitais das Letras Clássicas, como possíveis elementos de gênero na sua produção e sua leitura digitais, supondo serem uma forma de contribuir para a mobilização da aprendizagem das línguas clássicas e a formação docente na área. Descrevemos, de um lado, os componentes essenciais no processo de uma edição digital nova do texto grego do *Édipo Rei*, de Sófocles, preparado na Universidade Furman, para um curso de graduação. De outro, apontamos de que modo as anotações de edições digitais, em alinhamentos, *tree-banking* e geoanotações, na qualidade de experiências de aprendizagem, são potenciais instrumentos de desenvolvimento.

Palavras-chave: edições digitais; letras clássicas; artefatos de ensino-aprendizagem; edições digitais anotadas; CITE.

Abstract: *Digital editions in Classics have emerged as allies in breaking down barriers and facing challenges in teaching classical languages and cultures. The aim of this article is to highlight some specific features of digital editions in Classics, as possible genre elements in their digital production and reading, assuming they are a way of contributing to the mobilization of classical language learning and teacher training in the field. First, we describe the essential components in the process of a new digital edition of the Greek text of Oedipus Tyrannus, by Sophocles, prepared at Furman University for an undergraduate course. After that, we point out*

* Universidade Estadual Paulista – UNESP, Araraquara, SP, Brasil. CNPq proc. n. 307431/2019-3. Convênio CAPES/PRINT/UNESP; <http://orcid.org/0000-0001-5755-1434>; anise.ferreira@unesp.br

** The Louis G. Forgiione University Professor Furman University, Greenville, SC, Estados Unidos; <http://orcid.org/0000-0003-2933-3493>; cwblackwell@gmail.com

*** Assistant professor Furman University, Greenville, SC., Estados Unidos; <http://orcid.org/0000-0002-1811-5602>; chiarapalladino1@gmail.com

in what way digital edition annotations, in translation alignments, treebanking and geoannotations, as learning experiences, are potential development tools.

Keywords: *Digital Classics; Teaching-learning Artifacts; Annotated Digital Editions; CITE.*

Introdução

Os avanços na infraestrutura da tecnologia de informação em rede têm sido vistos como aliados no rompimento de barreiras e no enfrentamento de desafios no ensino das letras clássicas, seja no Brasil, seja no mundo (CRANE *et al.*, 2012). As barreiras são conhecidamente o fato de se lidar com línguas históricas restritas à universidade, requerer-se um longo período de formação, competir com outras áreas e disciplinas por alunos e recursos e apontar para uma carreira profissional incerta. Os desafios podem ser de ordem social, tais como estender o acesso do conhecimento à comunidade em geral; de ordem acadêmica, no caso da língua portuguesa, que não conta com traduções e estudos de acesso aberto como em línguas estrangeiras; e de ordem didático-pedagógica, tais como atrair alunos e desenvolver materiais que auxiliem o desenvolvimento e a formação de alunos e o trabalho de professores (FERREIRA; REIS, 2017).

Há algum tempo comentou-se (FERREIRA; MELO, 2016) sobre a análise de gêneros digitais da web à luz do interacionismo sociodiscursivo (ISD) (BRONCKART, 2007) e da análise e transcrição multimodal de Baldry e Thibault (2010), abordando gêneros de textos acadêmicos digitais na área de letras clássicas em contraste com os textos impressos (FERREIRA; TRINDADE, 2015). Nessa comparação, constatou-se, como esperado, que, do ponto de vista do plano global do texto¹, embora os conteúdos e as informações tenham os mesmos componentes das edições impressas, a disposição dos conteúdos segue trajetórias ou trilhas de hipertexto, com seus *clusters* e *subclusters*, sem um padrão fixo de organização temática, seja no âmbito da produção, seja no âmbito da leitura; dependem, em princípio, do design criado para atender a determinados fins comunicativos. Essa

¹ A análise do ISD sobre gêneros textuais é realizada em vários níveis do texto. Outros trabalhos vêm sendo realizados dentro do ISD com textos multimodais, como o de Sumya, 2016.

plasticidade, em tese, tornaria mais complicado elaborar-se um modelo de gênero para a produção de textos dentro de um suposto gênero de produções de letras clássicas digitais. Contudo, qualquer que seja a interface, ela não traz ao leitor um só tipo de informação como as edições tradicionais acompanhadas de notas; uma edição digital pode ser acompanhada simultaneamente de consultas e visualizações diversas, tais como dicionário, morfossintaxe, mapas, etc., como veremos adiante. Algo que somente no meio digital é possível. Os destinatários podem ser os mesmos classicistas leitores de estudos da antiguidade clássica; porém, as edições digitais se abrem para outros leitores; os actantes na infraestrutura textual também são ampliados pela multiplicidade de editores, curadores e anotadores. Na leitura, supomos que a escolha e percurso realizados pelo usuário tornam possível o surgimento de uma voz própria na construção do conhecimento. Quanto aos mecanismos enunciativos, encontram-se múltiplas vozes e modalizações de natureza epistêmica e deontica, em sequências injuntivas e argumentativas nos comentários, glossários, dicionários e gramáticas.

Posteriormente, características comuns foram levantadas em relação aos processos envolvidos na produção de edições digitais nessa área de estudos clássicos (FERREIRA; REIS, 2017). Esses processos foram: o compartilhamento na produção de dados, infraestrutura e saberes; o trabalho em equipe por profissionais com diferentes backgrounds, com estudantes de níveis diferentes de conhecimento, além de colaboradores; e acesso aberto a dados vinculados, a licenças de uso e a linguagens de programação padrão. Digamos que as palavras-chave das edições digitais seriam: o compartilhamento, o trabalho em equipe e o acesso aberto.

É importante ressaltar que o foco na ideia do gênero digital de edições e anotação de corpora digitais no ambiente de ensino surge aqui derivado do conceito de gênero como artefato e potencial instrumento do desenvolvimento, tanto do aprendiz como do profissional do *métier* docente, de forma análoga ao que já expuseram Machado e Lousada (2010). Elas se baseiam em conceitos vigotskianos em que “os artefatos são ‘objetos’ materiais ou simbólicos, sócio-historicamente construídos para mediar a ação do homem sobre o meio ou sobre o outro e para se atingirem determinadas finalidades” (MACHADO; LOUSADA, 2010, p. 625-626) Quando o

artefato é apropriado pelo sujeito, adaptado ou transformado para uso em diferentes situações, conforme as suas necessidades, se torna um instrumento.

Acreditamos, por ora, poder explorar mais detalhadamente alguns dos elementos ou características dos textos ou edições digitais voltadas ao contexto educacional que tenham implicações no trabalho docente e no ensino e aprendizagem da língua grega. Afinal, a aprendizagem e leitura de textos clássicos envolvem um conjunto complexo de textos acadêmicos tradicionais e de suporte – materiais de referência – necessários ou, pelo menos, desejáveis: o texto original com variantes, traduções, léxicos, operações morfossintáticas e histórico-fonológicas, descrição sintática, tabelas de flexão, concordâncias, notas sobre particularidades de uso, expressões e estilos específicos, referências geográficas, literárias, materiais e prosopográficas, etc.

Neste artigo, são levantados alguns elementos que podem caracterizar uma transformação na produção textual acadêmica tradicional para as edições digitais voltadas ao ensino na língua grega ou latina. Trazemos, de um lado, os componentes essenciais no processo de uma edição digital nova do texto grego do *Édipo Rei* de Sófocles, preparado na Universidade Furman para um curso oferecido no período da primavera de 2020. De outro, como as anotações de edições digitais, alinhamentos de tradução e geoanotações, na qualidade de experiências de aprendizagem, são potenciais instrumentos de desenvolvimento.

A edição digital de *Édipo Rei*, de Sófocles, em código fonte²

As novas complexidades de gênero no meio digital e o ambiente computacional podem ser ilustrados pela discussão de uma edição nova do texto grego de Sófocles *Édipo Rei*³ (OT) preparado na Universidade Furman para um curso ministrado na primavera de 2020⁴. O acesso ao aplicativo para os leitores de OT está em <http://folio2.furman.edu/ot/pages/index.html>

² Trata-se da edição cujo código é lido pela máquina por meio de algum aplicativo, como o navegador.

³ Em inglês, *Sophocles, Oedipus Tyrannus*, com abreviação canônica, a partir do latim, OT.

⁴ Pelo professor Christopher W. Blackwell.

Os objetivos dessa ação de edição foram, em geral, a) criar uma edição digital sujeita à citação não ambígua por linha poética, como ocorre na tradição do gênero dramático grego; e b) integrar essa edição às informações lexicais, morfológicas e sintáticas⁵. Tais metas demandaram atos adicionais de edição. De um lado, o alinhamento preciso e automatizado da morfologia e lexicografia requereu *estender* o esquema de citação das linhas da poesia para as linhas poéticas como *containers* de palavras citadas, ou seja, uma *tokenização citável* pela palavra grega. De outro, o alinhamento preciso e automatizado da sintaxe grega requereu que essa tokenização citável incluísse marcas de pontuação individuais. Mas, por que um esquema de citação⁶ é importante e como funciona?

Qualquer estudioso do grego antigo ficaria horrorizado diante da ideia de uma citação canônica de pontuação. Sófocles quase certamente não escreveu com marcas de pontuação e os testemunhos manuscritos mais antigos do drama grego não têm pontuação ou, pelo menos, passam longe de padrões de pontuação do grego antigo que estiveram em uso desde os incunábulo do séc. XVI. Por sinal, Sófocles por certo não separou as palavras individuais com o espaço em branco, uma prática que foi uma inovação da Antiguidade tardia.

Quando leitores humanos⁷ instruídos abordam uma edição de Sófocles, categorizam inconscientemente características do texto como orgânicas ou editoriais, e a numeração por linha parece orgânica, uma vez que, em versos, o texto é organizado por uma linha poética. Convenções modernas de ortografia, divisões de palavra e pontuações são editoriais, adições feitas pelo editor, que as usa como ferramentas de sua autoridade editorial para expressar o sentido do texto.

Quando temos a expectativa de desempenhar processos computacionais sobre o texto, essas adições editoriais necessariamente se tornam orgânicas (ou, pelo menos, indispensáveis). Sófocles pode não ter escrito a vírgula ainda não inventada, mas ele, de fato, escreveu frases apositivas; ele nunca usou aspas, mas suas

⁵ Francesco Mambrini, que realizou a análise lexical, morfológica e sistemática da linguagem, forneceu os dados para essa edição integrada.

⁶ Em inglês, *citation-scheme*.

⁷ Por oposição às leituras realizadas por máquina ou por aplicativos, como navegadores, por exemplo.

personagens citaram outras personagens. No gênero do texto digital com o qual trabalhamos, essas adições editoriais anacrônicas têm função e, de modo algum, arbitrária. Tome-se o exemplo das reticências "...". Em inglês, tal marca é chamada uma *ellipse*. Mas não é, concretamente, uma elipse, só uma marca que indica uma elipse em determinado ponto do texto. Sófocles nunca escreveu tal marca, mas o significado de suas sentenças com frequência envolve elipse e, se quisermos *identificá-la, ligá-la* ou *analisá-la* em um ambiente computacional, precisamos de um símbolo ao qual apontá-la. Em 1990, DeRose *et al.* publicaram um artigo no qual propuseram uma definição de texto para a era computacional. Sua definição era: "um texto é uma hierarquia ordenada de objetos de conteúdo" (DEROSE, 1990, p. 6). Essa definição veio a ser conhecida como o modelo OHCO⁸. Após esse artigo, surgiu imediatamente uma longa e duradoura disputa sobre o que constituiria conteúdo. Marcas são conteúdo? Pontuação é conteúdo?

Em 2009, Smith e Weaver propuseram uma definição alternativa que foi: "Um texto é uma hierarquia ordenada de objetos *citados*" (SMITH; WEAVER, 2009, p. 130). Essa veio a ser conhecida como "OHCO2" e é a base do protocolo de Serviços de Texto Canônico⁹ da Arquitetura CITE¹⁰. Dentro do modelo OHCO2, uma edição de um texto é definida por suas citações. O primeiro elemento de qualquer edição do Édipo Rei de Sófocles é a linha poética rotulada como 1. Ao usar os valores de citação, não só como princípio organizador, mas como princípio de definição, o problema do conteúdo desaparece.

Que um editor afirme que a primeira linha de uma edição de O.T.¹¹ tenha um valor de citação¹² expresso em CTS URN: urn:cts:greekLit:tlg0011.tlg004.fu:1. O número final **1** identifica a passagem, a linha poética; **tlg0011** é o identificador de Sófocles; **tlg004** é identificador da obra O.T., e **fu** indica "Furman University Edition". Com essa citação precisa e rica semanticamente, podemos ser flexíveis com

⁸ Em inglês: *Ordered Hierarchy of Content Objects*.

⁹ Em inglês: *Canonical Text Services* (CTS).

¹⁰ Ver mais em <https://cite-architecture.github.io/about/> e em Blackwell e Smith (2019).

¹¹ Ver nota 1.

¹² Em inglês: *citation-value*.

como tratamos o conteúdo. Como editores, podemos dizer que todas as formas seguintes compartilham uma mesma identidade acadêmica:

urn:cts:greekLit:tlg0011.tlg004.fu:1#ὧ τέκνα, Κάδμου τοῦ πάλαι νέα τροφή¹³,
urn:cts:greekLit:tlg0011.tlg004.fu:1#w)= te/kna, *ka/dmou tou= pa/lai ne/a
trofh/¹⁴,
urn:cts:greekLit:tlg0011.tlg004.fu:1#ωτεκνακαδμουτουπαλαινεατριφη¹⁵

Tal identidade acadêmica foi dada pelos editores que identificaram essas formas como “as mesmas”, atribuindo-lhes a mesma e não ambígua citação URN. No domínio digital, cada uma dessas apresentações de conteúdo de OT1 tem um valor funcional. Se, para propósitos computacionais, quisermos trabalhar no nível de detalhe, não sustentado pelo esquema tradicional de citação, as linhas poéticas numeradas, podemos elaborar um *exemplar analítico*, que *estende o* esquema de citação:

urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.1#ὧ
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.2#τέκνα
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.3#,
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.4#Κάδμου

Note-se que cada um dos quatro *tokens* acima ainda está identificado por :1. Mas o que precede o campo final foi estendido por um campo adicional, :1.n, que enumera o *token* seguido do próprio *token*. Todas essas identificações são baseadas na Edição **urn:cts:greekLit:tlg0011.tlg004.fu:**, mas são parte de um *exemplar*¹⁶ da edição, uma transformação dela, nesse caso, uma *tokenização*.

¹³ Primeiro verso em grego, em português: “Meus filhos, nova geração do antigo Cadmo” (trad. por M.G.Kury).

¹⁴ Transposição do grego em código Beta (cf. <https://www.translatum.gr/converter/beta-code.htm>, *Greek Beta Code Converter*).

¹⁵ Texto grego sem espaçamento entre as palavras.

¹⁶ Em inglês também *exemplar*.

Estudando Sófocles: três textos

Para trazer essas considerações abstratas à terra e defender o valor do modelo OHCO2 de texto, que é a base para os Serviços de Texto Canônico (CTS), podemos considerar três modos pelos quais poderíamos ler OT: 1) leitura humana da peça, visando a uma possível bolsa de estudos, o que requer citação; 2) leitura computacional da linguagem da peça, tal como contar palavras, localizar N-Gramas, de modo que o relatório de resultados requer citação; 3) alinhamento de palavras (e pontuação) na peça com outros dados; isso depende totalmente da citação como um mecanismo de vinculação de dados.

Para tornar essas três leituras mais complicadas, adicionamos a questão: o atributo falante é parte do texto ou não?, já que, em uma edição impressa, os nomes dos interlocutores costumam aparecer seja por extenso, seja de forma abreviada:

Κρέων (Creonte)

91 εἰ τῶνδε χρήσεις πλησιαζόντων κλύειν,

92 ἔτοιμος εἰπεῖν, εἴτε καὶ στείχειν ἔσω.

Οἰδίπους (Édipo)

93 ἐς πάντας αὖδα: τῶνδε γὰρ πλέον φέρω

94 τὸ πένθος ἧ καὶ τῆς ἐμῆς ψυχῆς πέρι.

As palavras Κρέων (Creonte) e Οἰδίπους (Édipo) fazem parte do texto, ou não? A resposta será que depende do tipo de leitura. Pois para a leitura humana (item 1 acima), a resposta é decididamente sim! Os leitores precisam saber qual das personagens disse o quê. Para a leitura pelo computador (item 2 acima), a resposta pode ser talvez; se quisermos contar a frequência de palavras de conteúdo do corpus, ou se quisermos saber quantas vezes o nome da personagem é enunciado na peça, os nomes dos interlocutores vão alterar o resultado se contarmos os nomes da troca de falantes. Por outro lado, para uma análise de discurso, podemos querer isolar as falas de Édipo. No caso de alinhamento de palavras (item 3 acima), a resposta será provavelmente não. Os nomes dos interlocutores na troca de turno dos diálogos não são parte da sintaxe de nenhuma sentença.

A resposta, então, capturada nos dados disponíveis no GitHub,¹⁷ está em elaborar uma *Edição* do texto e dois *Exemplares* derivados. A Edição contém apenas o texto da peça, citada por linha poética:

urn:cts:greekLit:tlg0011.tlg004.fu:1#ὦ τέκνα, Κάδμου τοῦ πάλαι νέα τροφή,
urn:cts:greekLit:tlg0011.tlg004.fu:2#τίνας ποθ' ἔδρας τάσδε μοι θοάζετε
urn:cts:greekLit:tlg0011.tlg004.fu:3#ικτηρίοις κλάδοισιν ἔξεστεμμένοι ;

Um *Exemplar* adiciona o atributo-falante estendendo o esquema de citação tradicional em um nível, com a opção de adicionar o elemento speaker (falante). Essa estratégia funciona mesmo quando uma única linha poética é dividida entre dois falantes:¹⁸

urn:cts:greekLit:tlg0011.tlg004.fu.sp:1522.speaker1#Οἰδίπους
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1522.1#μηδαμῶς ταύτας γ' ἔλη μου .
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1522.speaker2#Κρέων
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1522.2#πάντα μὴ βούλου κρατεῖν :
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1523.1#καὶ γὰρ ἀκράτησας οὐ σοὶ τῷ βίῳ
ξυνέσπετο .
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1524.speaker1#Χορός
urn:cts:greekLit:tlg0011.tlg004.fu.sp:1524.1#ὦ πάτρας Θήβης ἔνοικοι,
λεύσσειτ', Οἰδίπους
ὄδε

A terceira edição, ou segundo exemplar, é uma tokenização por palavra e marca de pontuação. Essa pode ser alinhada a dados lexicais, morfológicos e sintáticos:

urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.1#ὦ
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.2#τέκνα
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.3#
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.4#Κάδμου

¹⁷ <https://github.com/Furman-University-Editions/OedipusTyrannos>

¹⁸ Para quem não está familiarizado com a poesia dramática grega, é comum que a sentença de um verso seja dividida entre dois falantes, o que poderia trazer um problema na identificação das citações.

urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.5#τοῦ
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.6#πάλαι
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.7#νέα
urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.8#τροφή

Todas essas anotações de identificação estão implicitamente alinhadas umas com as outras, uma vez que qualquer máquina ou processo que entenda CTS URNs pode saber que, por exemplo, urn:cts:greekLit:tlg0011.tlg004.fu.tokens:1.7 é um derivado de urn:cts:greekLit:tlg0011.tlg004.fu:1. E todas essas edições podem vir juntas em uma apresentação HTML dessa peça em uma URL, a saber, http://folio2.furman.edu/ot/pages/urn_cts_greekLit_tlg0011_tlg004_1-57.html.

O esquema de citação prova ser, assim, um dos elementos constantes na preparação de uma edição digital no nível de leitura da máquina que permite justamente a flexibilidade de manipulação da informação que poderia ser um elemento de caracterização do gênero digital da área de clássicas em um nível não previsto pelas abordagens de gêneros textuais. Assim, a apresentação por HTML retoma algumas das palavras-chaves do processo: o compartilhamento de dados e saberes, com a transformação de uma edição impressa (Francis Storr, Loeb Classical Library) em um texto legível por máquina (G. Crane, Perseus Digital Library), dotado das anotações lexicais, morfológicas e sintáticas (F. Mambrini); a parceria entre esses e o emprego de linguagens e licenças abertas. Não há uma regra para a disposição na página web, a não ser aquela que combine a usabilidade¹⁹ com os limites de exibição dos aplicativos.

A anotação digital manual de textos antigos

Uma vez que um texto está estabilizado pela identificação de suas partes, com o sistema de citação e *tokens* que o constituem, é possível adicionar mais significados e interpretação a ele por meio de anotações digitais. As anotações realizadas no

¹⁹ Termo usado na Interação humano-computador para referir-se a facilidade de uso de uma interface.

meio eletrônico²⁰ são um meio de sobrepor uma gama de informação adicional na fonte, como resultado de interpretação linguística ou semântica. O meio eletrônico tem oferecido a possibilidade de equipar palavras individuais ou mesmo *tokens* menores como letras e símbolos, com um espaço de significado vertical extra, onde dados adicionais podem ser colocados. No meio impresso, algumas dependências são deixadas implícitas, ou elas são apenas realizadas pela associação na mente do leitor (BAŃSKI; WITT 2019), tais como: estruturas gramaticais ou identificações de pessoas específicas ou lugares com as entidades correspondentes no mundo real. Esse tipo de informação está normalmente espalhado em um número de recursos individuais, como atlas, dicionários ou outras obras de referência. Ou, talvez, fique inteiramente implícito. O meio eletrônico é capaz de combinar diferentes tipos de dados, tornando-os, junto com as associações correspondentes ao texto, explícitos e acessíveis através do próprio texto. Nesse sentido, a anotação é uma inovação do gênero da edição digital: permite investigar o texto em um nível de profundidade e multimodalidade inteiramente novo.

No espaço da sala de aula, a anotação constitui uma experiência de aprendizagem que é tanto ativa como passiva. Os fluxos de trabalho de anotação podem ser usados para adicionar conteúdo ao texto enquanto se faz sua leitura (produção de dados) ou para ter a experiência de um texto já anotado de muitos ângulos diferentes (visualização de dados). Essas experiências são geralmente entrelaçadas, dependendo do contexto de uso: um texto com uma tradução alinhada pode ser lido na classe para enfatizar usos de uma palavra ou o significado de vocábulos específicos, enquanto que um estudante pode alinhar um texto como uma forma de aumentar a capacidade de sua experiência de leitura, integrando-a, por exemplo, com uma anotação morfossintática, de modo a entender melhor a função de palavras específicas em contexto.

Eis alguns exemplos abaixo de ambientes atualmente disponíveis para anotação digital de textos clássicos: alinhamento de tradução, anotação geoespacial e anotação

²⁰ Um exemplo de anotação em edição digital é a marcação XML, também chamada anotação *inline*, porque é armazenada no mesmo documento que está sendo anotado, por oposição às anotações *standoff*, que são armazenadas em lugar diferente daquele do documento anotado (v. Bański e Witt, 2019).

morfossintática via *treebanking*, ou árvore sintática. Com ferramentas como *Ugarit* (fig.1) ou DUCAT²¹, é possível desempenhar analiticamente o alinhamento de texto pelo pareamento manual de um ou mais *tokens* através de múltiplos textos.

O alinhamento de textos em diferentes línguas é uma tarefa bastante complexa, por causa do número ampliado de variáveis envolvidas e porque é virtualmente impossível encontrar perfeitas correspondências entre as línguas que expressam ideias por meio de constructos morfossintáticos e semânticos diferentes. Entretanto, precisamente por causa dessa complexidade, o alinhamento de uma tradução analítica através de anotação digital é uma poderosa experiência pedagógica que estimula o reconhecimento crítico de diferenças culturais e linguísticas entre duas ou três línguas diferentes. Pela realização do alinhamento de texto, um estudante pode ter acesso a muitas formas nas quais conceitos culturais complexos, tais como “honra”, “culpa”, ou “amor”, são recebidos nas línguas modernas e traduzidos por meio de várias estratégias, e investigar a sintaxe e a ordem das palavras no grego antigo em contraste e em discrepância com as traduções de línguas modernas.

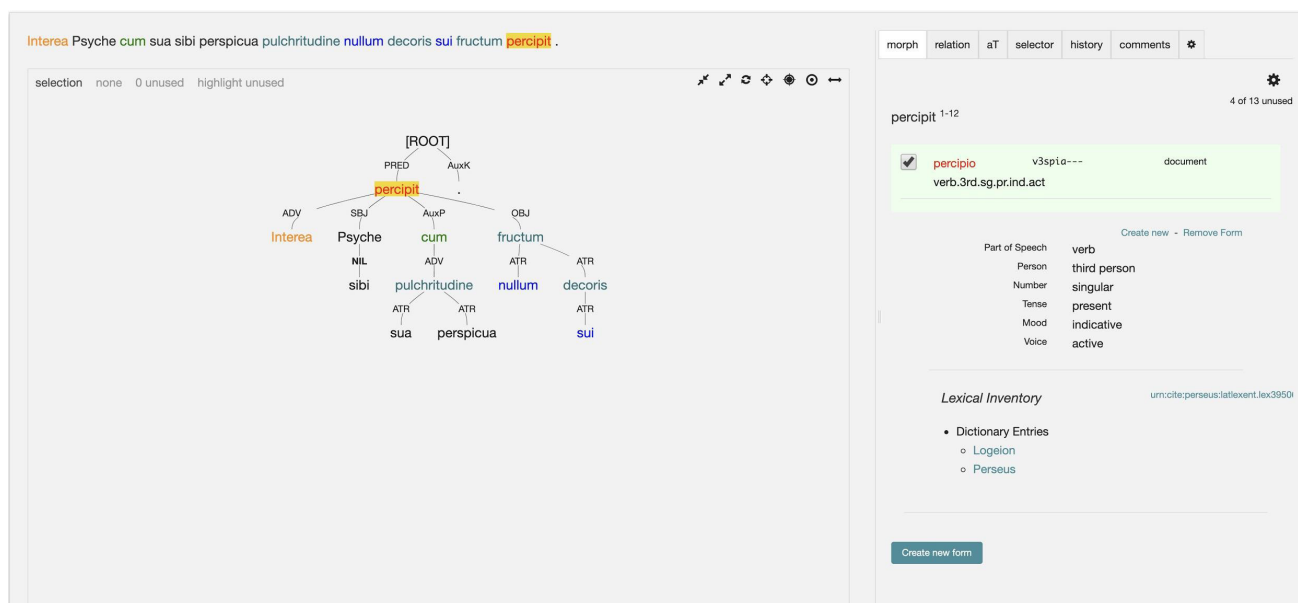
Figura 1: English-Greek text alignment on Ugarit

The screenshot displays the Ugarit text alignment tool interface. At the top, it identifies the source as 'Hesiod, Works and Days, vv. 42-105' by Anna Zhang. Below this, there are two main columns of text. The left column contains the English text, and the right column contains the Greek text. The text is aligned line-by-line, with markers indicating the alignment. The English text is on the left, and the Greek text is on the right. The interface includes a search bar at the top left and a 'Ugarit' logo at the top right. The text is presented in a clear, readable font, with the Greek text in a serif font and the English text in a sans-serif font. The alignment markers are small, dark lines connecting the corresponding lines of text in the two columns.

Fonte: <http://ugarit.ialigner.com/text.php?id=26966>

²¹ Em inglês: *Daughter-of-Ugarit Citation Alignment Tool*, <http://folio2.furman.edu/>, é o aplicativo de alinhamento expandido com esquema de citação e possibilidade de múltiplas línguas.

Figura 2: Um texto em latim anotado em treebank na ferramenta Arethusa na plataforma



Fonte: Elaboração própria na plataforma Perseids.

O *Treebank de Dependência de Grego Antigo e Latim*²² (AGLDT) é um projeto (fig.2) destinado para anotação morfofossintática colaborativa em ambientes digitais através de um aplicativo inovador de gramática de dependência (TESNIÈRE, 1959). As *treebanks*, ou árvores sintáticas, são corpora de sentenças sintaticamente analisadas e emergiram como um recurso valioso não só para a análise linguística computacional, mas também para a pesquisa e ensino tradicionais. A operação de montar árvores ou *treebanking* sobre um texto se dá pela integração da classificação morfológica semi-automática que os usuários confirmam de acordo com seu conhecimento, com a anotação sintática por meio da estrutura de “árvores” (PAL-LADINO, 2015; MAMBRINI, 2016). Por outro lado, árvores anotadas ajudam a visualizar a estrutura global de sentenças complexas com um relance de olhar, realçando as suas dependências, permitindo uma investigação muito refinada das estruturas linguísticas e retóricas de um texto (MAMBRINI, 2011; FERREIRA, 2017; REIS, 2017). Segundo Celano (2019), o AGLDT é o projeto mais antigo

²² Em inglês, no original: *The Ancient Greek and Latin Dependency Treebank*: https://perseusdl.github.io/treebank_data/.

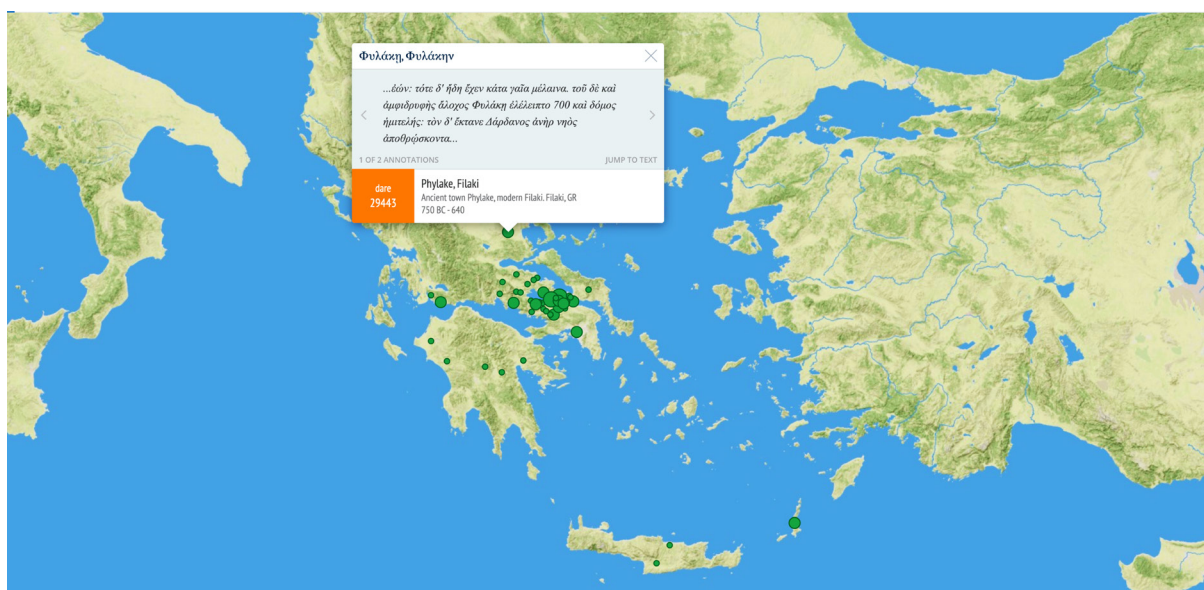
de *treebank* para grego e latim. O aplicativo aberto, utilizado pelos anotadores para sua edição, é a ferramenta Arethusa na plataforma Perseids²³.

Na prática de sala de aula de Palladino (2015; 2020), estudantes de grego antigo e latim de todos os níveis usam uma combinação de alinhamento e *treebanking* durante seus cursos de língua, muitas vezes integrados com exercícios, perguntas e instruções específicas. Os exercícios normalmente incluem alinhamento aos pares e trilingue, *treebanking* de tradução em inglês e original e ensaios e pesquisas guiadas para avaliar sua reflexão sobre o processo. Os alunos examinam a relação entre tradução e original com foco em palavras específicas, avaliam traduções concorrentes, e desenham estratégias de alinhamento baseadas em seus objetivos. A maioria dos exames finais consiste em um conjunto completo de dados de um texto alinhado ou com *treebanking* ou com uma reflexão escrita sobre o processo e seus resultados. Ocasionalmente, os estudantes apresentam seus resultados em conferências de graduação em outros lugares: em 2018, os estudantes da Universidade Furman apresentaram seus projetos de alinhamento integrado e *treebanking* das *Metamorfoses* de Apuleio em dois simpósios de graduação, em duas universidades americanas. Ainda, de acordo com sua experiência, os estudantes de nível avançado tendem a escrever posteriormente suas próprias traduções. A comparação entre diferentes modelos autorais provou ser útil não só para compreender diferentes estratégias de abordagem de problemas linguísticos e semânticos, mas também para ajudar o estudante a encontrar a sua própria voz de tradutor. Alguns estudantes até usaram o alinhamento *trilingue* para ganhar mais proficiência na sua terceira língua que não era latim ou grego: eles usaram o conhecimento de duas línguas para alavancar as questões na compreensão da terceira, reconhecendo padrões sintáticos comuns, morfologias ou semelhanças no vocabulário. No final do curso, a maioria dos alunos desenvolveu um sentido aguçado da limitação de sua compreensão prévia sobre a fonte, a partir da tradução e do contraste com a tradução linguística moderna, investigando os seus vários significados. O reconhecimento das limitações da tradução para transmitir conceitos de forma eficiente em línguas clássicas, ainda que inicialmente cause decepção e crítica, acaba convencendo os estudantes

²³ <http://perseids.org>

da necessidade de pesquisar a língua original para compreender as suas características inerentes e a forma como nela as ideias eram expressadas. Tais experiências são similares às realizadas também em salas de aula de língua grega no Brasil.²⁴

Figura 3: Uma versão anotada e “geo-referenciada” do Catálogo das Naus, de Homero (Ilíada 2.484-2.769)



Fonte: Elaboração própria na plataforma Recogito

A anotação geoespacial (fig. 3 e 4) consiste de duas operações: marcar os nomes, ou *strings*,²⁵ identificados como lugares, e referenciá-los geograficamente, i.e., atribuir-lhes coordenadas geográficas específicas, de modo que as entidades correspondentes possam ser representadas em um mapa. O ambiente de anotação do *Recogito*²⁶ permite a realização dessas duas operações por meio de uma estrutura subjacente de *Linked Open Data*²⁷ (LOD) e anotações semânticas, que também permitem acesso a uma extensa quantidade de informação adicional provida pelos

²⁴ Cf. Ferreira, 2017.

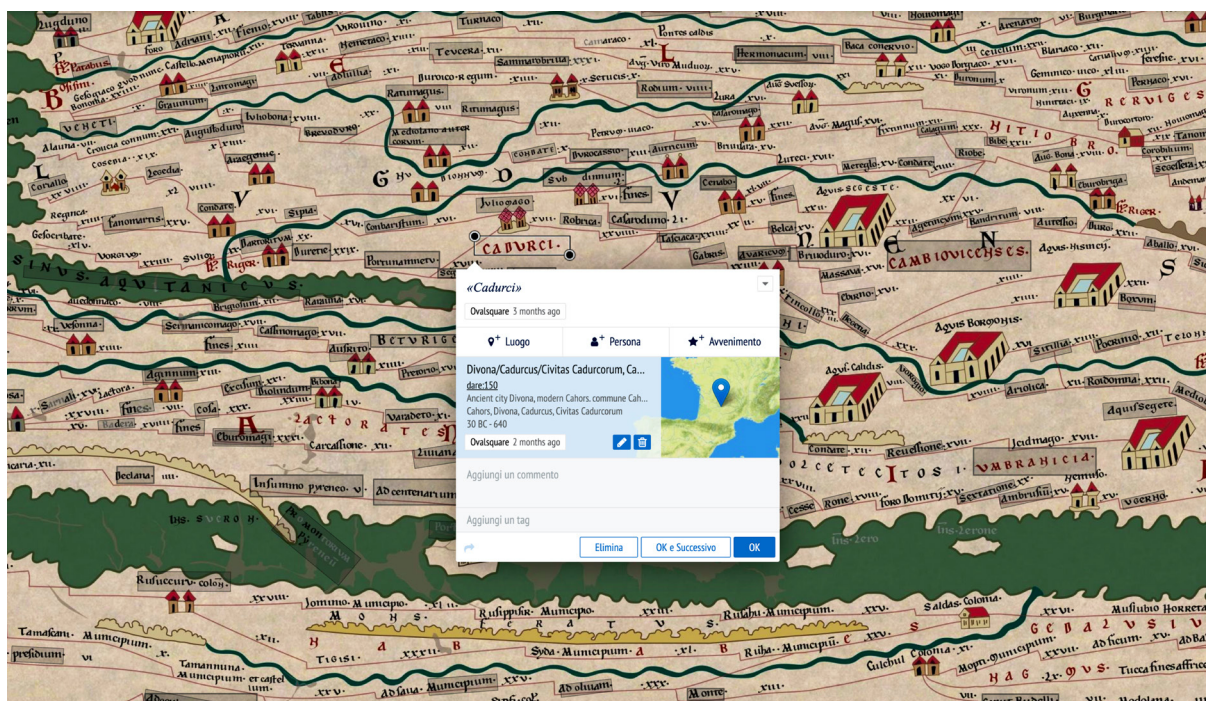
²⁵ Série de caracteres que são processados como uma unidade de informação.

²⁶ <https://recogito.pelagios.org/>

²⁷ Em português: Dados abertos vinculados (ou ligados).

atlas e os chamados *gazetteers* ou dicionários geográficos on-line, como o *Pleiades*²⁸ e o Atlas Digital do Império Romano²⁹.

Figura 4: Uma versão anotada, transcrita e “geo-referenciada” da Tabula Peutingeriana romana.



Fonte: Elaboração própria na plataforma Recogito

Há muitas razões pelas quais as entidades nomeadas (ou de nomes)³⁰ em um texto devam sempre ser anotadas e associadas a identificações correspondentes em referências autorizadas ou confiáveis: primeiramente, para fins pedagógicos, é importante deixar associações explícitas que somente são claras a especialistas, mas podem não ser para estudantes. Por exemplo, a entidade Alexandria pode corresponder a muitas Alexandrias diferentes, cuja identificação depende exclusivamente do contexto. Em segundo lugar, há aspectos de um texto que podem se tornar claros quando o contexto espacial é identificado, permitindo considerações sobre a cultura subjacente ao quadro literário: há textos que simplesmente não fazem

²⁸ <https://pleiades.stoa.org/>.

²⁹ Em inglês: *Digital Atlas of the Roman Empire*, DARE: <https://dh.gu.se/dare/>

³⁰ Em inglês: *named entities*.

sentido sem um contexto espacial, tal como o homérico *Catálogo das Naus*, que, na simples forma textual, apenas lista nomes de pessoas e lugares, mas se plotado em um mapa e marcado de forma apropriada, pode prover uma compreensão muito mais profunda dos espaços e das culturas envolvidas na expedição troiana.

Os textos antigos, portanto, contêm elementos que trazem muita ambiguidade ao leitor, seja no vocabulário comum, seja nos nomes de pessoas ou nos nomes de lugares. O que as anotações contribuem em termos de identificação dos elementos do texto ou em termos de associação de informações, contextualizando-os, constitui um fator de maior aproximação e maior entendimento por parte dos leitores e estudiosos junto ao texto. Uma mesma edição digital pode comportar todos esses possíveis dados de contexto em virtude de sua identificação universal. Os exemplos acima indicam como os alunos-usuários-anotadores podem transformar o objeto digital e talvez transformar um gênero de texto com características tais da edição digital que se tornem um instrumento de aprendizagem da língua e cultura antiga.

Uma nota sobre o projeto do Dicionário Digital Grego-Português (DDGP)³¹

Foi no sentido de obter uma edição digital anotada, ou também chamada enriquecida, que o DDGP foi criado, se encontra em desenvolvimento desde 2017 (FERREIRA; RODRIGUES; DEZOTTI; 2019) e está em sua versão beta 2x. O corpus inicial foi extraído do material originalmente publicado em formato impresso, entre 2006-2010, em cinco volumes, elaborado sob coordenação de D. Malhadas, M. Celeste C. Dezotti e M. Helena M. Neves, da FCLAr/UNESP, com autores das universidades: USP, UFMG, UFRJ e UFF. Sob licença Creative Commons³², a edição digital nunca visou apenas reproduzir o conteúdo lexical do dicionário impresso, mas também a ampliar seu corpus, fornecer ligações com outros dados, tais como entidades nomeadas: nomes próprios de lugares, pessoas e objetos, bem como exemplos de uso associados a identificações em fontes originais e traduções. O fato de edições digitais proporcionarem o acesso a dados linguísticos, de co-texto e contexto da língua antiga aparece, assim, como um artefato digital textual para estudantes

³¹ <http://perseidas.fclar.unesp.br>

³² Atribuição -NãoComercial- Compartilhual 4.0 Internacional.

e estudiosos brasileiros. As modificações realizadas até o momento detiveram-se a estruturar adequadamente os dados, corrigi-los, ampliá-los, facilitar a pesquisa e dispor uma interface que pudesse ser aperfeiçoada em termos de anotação e de estilo a longo prazo, além de contar com uma interface administrativa, em web, restrita aos pesquisadores. Aqui, também, o esquema de citação será aplicado e suas ligações com geo-referências serão feitas como explicado nas seções anteriores deste texto. Trata-se de uma produção que envolve a conjugação de uma multiplicidade de saberes: os recursos necessários de infraestrutura digital e o entendimento da atividade de classicista, quer do aprendiz, quer do especialista. Requer um esforço em equipe. As novas edições digitais evocam formas de leitura pelos alunos e profissionais na sua produção em diferentes níveis de conhecimento, tanto em relação aos textos clássicos, como em relação aos artefatos tecnológicos e computacionais.

Considerações finais

Pudemos notar que há níveis diferentes de edição de um texto digital em virtude do tipo de enriquecimento ao texto, dotado de anotações inexistentes em edições tradicionais e possíveis nas produções textuais das letras clássicas digitais. Um deles é a identificação citável de *tokens* efetuada por meio de programação até a redação deste artigo. À medida que o desenvolvimento da infraestrutura é aperfeiçoado, plataformas com interfaces mais simples permitem que indivíduos inexperientes em computação, mas aprendizes de letras clássicas, possam construir também edições enriquecidas. É o caso das plataformas de geoanotações, dos alinhamentos de tradução e de anotações em árvores sintáticas. O enriquecimento do texto permite ao leitor uma ação similar à que o índice permite: de consultar elementos de um texto e também aqueles que extrapolam ou cruzam variadas dimensões de um texto. Observamos, na produção de textos digitais, seja no caso da edição de Édipo Rei, seja no caso dos alinhamentos e geoanotações, o compartilhamento das fontes de dados, corpus e tecnologias, a sua natureza aberta e a colaboração entre os editores de edições digitais, no plano computacional ou no plano estritamente do conteúdo. Esse compartilhamento mantém-se como marca de reutilização de recursos na produção textual presente nas edições digitais utilizadas no contexto acadêmico,

que servem tanto para pesquisa, como para o ensino-aprendizagem. A arquitetura CITE, em particular, parece comportar-se como um possível elemento do gênero da edição digital, por seguir um caminho estável e necessário para identificação dos *tokens* de um texto; sejam unidades lexicais, sejam indicadores semânticos ou de imagem. Portanto, trata-se de um artefato que se transforma em instrumento para um editor, cada vez que é empregado em um novo projeto; ou para um anotador, ou para quem elabora uma tradução alinhada. Outros meios computacionais de identificação de elementos no texto digital podem ser empregados, como as linguagens de marcação (as *markup*), como a XML ou TEI-XML, que são formas de marcação *on-line* que apresentam restrições quanto à sobreposição de anotações. As marcas de identificação dentro de um corpus, em geral, seguem uma hierarquia e ficam justapostas a cada unidade de conteúdo textual, como vimos no esquema de citação. Essa seria outra característica a ser considerada em uma apropriação de tecnologias como instrumentos visando a produção de edições de textos digitais. Além disso, vimos uma flexibilidade de tal arquitetura de identificação no ato de seu registro, em razão da finalidade proposta.

A edição digital do texto clássico transforma as edições não digitais em dados que precisam ser tratados ou editados com características e conteúdos específicos; dados esses que se reconfiguram em uma nova edição, promovendo uma forma de aprendizagem ou de estudo, poder-se-ia dizer, também baseada em dados, por aproximação ao conceito DDL, i.e., *data-driven learning*³³, comum ao ensino das línguas modernas, porém delimitada ao contexto das letras clássicas. Mas essa seria uma discussão para outro artigo.

As edições de textos digitais têm-se constituído como projetos para atingir um determinado fim, como no caso da edição do Édipo Rei, reaproveitando corpora de textos, de registros de anotações, de modo a ampliar o acesso e o aprendizado da língua e da cultura clássica a comunidades acadêmicas, de estudantes e profissionais, e não acadêmicas, à medida que o aprendizado lhes possibilite contribuir com anotações de corpus. Existe, em crescimento, uma comunidade de classicistas digitais que possuem habilidades de programação construindo plataformas e projetos ao lado de

³³ Ver Boulton (2011).

uma outra comunidade, a que se pode chamar de classicistas da era digital, que são os que utilizam os produtos criados e, ao utilizá-los, ajudam a aperfeiçoar as plataformas de edição ou de enriquecimento de textos e as interfaces de consulta.

É certo que os usuários, sejam profissionais ou alunos, demandam interfaces fáceis de usar para sua consulta e anotação de corpora e não há um modelo fixo ou estável para a visualização e manipulação dos dados, pois esse se altera a cada projeto em função dos avanços tecnológicos, mas o corpus digital enriquecido ou anotado permanece. Trata-se de um desafio que prossegue, uma vez que não é tarefa simples criar equipes de colaboradores com perfis complementares ou com formação suficiente em linguagens apropriadas às necessidades dos classicistas, capazes de estruturar os dados e gerar tais plataformas. Uma tal necessidade poderá, eventualmente, exercer pressão sobre a formação acadêmica dos membros docentes e técnicos.

As interfaces de acesso, por um lado, então, poderão ter formas diversas de apresentação, consulta e anotação de dados, mas, por outro, algumas tecnologias de identificação, anotação, alinhamento, treebank ou geoanotação tendem a ser constantes, com meios de exportação de arquivos de dados acessíveis, no intuito de que esses sejam compartilhados. A adoção dos textos digitais assim configurados propicia gestos didáticos na formação, pois aproxima os estudantes e pesquisadores dos textos de forma mais focalizada e, ao mesmo tempo, abrangente, colocando-os em situação de interação com os dados, ao reunir, em uma única edição, um conjunto de informações necessárias ao estudo da língua e da cultura.

Referências

BALDRY, A.; THIBAUT, P.J. *Multimodal Transcription and Text Analysis*. Portland: Equinox, 2010.

BAŃSKI, P.; WITT, A. Modeling and Annotating Complex Data Structures. In FLANDERS, J.; JANNIDIS, F. (Eds.). *The shape of data in the digital humanities: modeling texts and text-based resources*. London ; New York: Routledge, Taylor & Francis Group, 2019.

FERREIRA, A., BLACKWELL, C. W., PALLADINO, C. Edições digitais nas Clássicas: elementos de gênero na produção e leitura da língua grega e do latim

Todo conteúdo da *Linha D'Água* está sob Licença Creative Commons Attribution-NonCommercial 4.0 International License

BLACKWELL, C.W.; SMITH, D. N. The CITE Architecture: a Conceptual and Practical Overview. In BERTI, Monica. *Digital classical philology: ancient Greek and Latin in the digital Revolution*. 1st edition ed. Boston, MA: De Gruyter Saur, 2019. (Age of access? Grundfragen der Informationsgesellschaft, 10).

BOULTON, A. Data-driven learning: The perpetual enigma. In GOŹDŹ-ROSKOWSKI, S. (ed.), *Explorations across Languages and Corpora*. Frankfurt: PeterLang, 2011, p. 563-580.

BRONCKART, J.-P. *Atividade de linguagem, textos e discursos: por um interacionismo socio-discursivo*. São Paulo: EDUC, 2007.

CELANO, G. A. The Dependency Treebanks for Ancient Greek and Latin. In BERTI, Monica. *Digital classical philology: ancient Greek and Latin in the digital Revolution*. 1st edition ed. Boston, MA: De Gruyter Saur, 2019. (Age of access? Grundfragen der Informationsgesellschaft, 10)

CRANE, G. *et al.* Student researchers, citizen scholars and the trillion word library. In: [THE 12TH ACM/IEEE-CS JOINT CONFERENCE, 2012. Washington, DC, USA: ACM Press, 2012. p. 213. Disponível em: <http://dl.acm.org/citation.cfm?doid=2232817.2232857> Acesso em: 03/01/20.](#)

DEROSE, S.J., DURAND, D.G., MYLONAS, E. *et al.* What is text, really? *J. Comput. High. Educ.* v. 1, n. 2, p. 3–26, 1 dez. 1990. Disponível em: <https://rdcu.be/b1scX> Acesso em: 03/01/20.

FERREIRA, A. D' O. Aprendendo grego no mundo digital do terceiro milênio. *The ESPe-cialist*, v. 38, n. 1, 22 jul. 2017. Disponível em: <https://revistas.pucsp.br/index.php/esp/article/view/32220> Acesso em: 03/01/20.

FERREIRA, A. D'O.; MELO, G. Análise de textos multimodais da Web e o ISD. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, v. 32, n. 1, p. 1–21, abr. 2016. DOI 10.1590/0102-445056772752226428.

FERREIRA, A. D'O.; REIS, M. F. dos. Considerations on digital tools as a means of reshaping the Greek teacher's genre of activity for a prospective curriculum in a Brazilian public university. Digital Infrastructures for Global Philology. GLOBAL PHILOLOGY OPEN CONFERENCE 2017, Leipzig University, Leipzig, 2017, Feb.20-23.

FERREIRA, A. D'O.; RODRIGUES, A.; DEZOTTI, M.C.C. Atualização do Dicionário Digital Grego-Português. O ambiente de desenvolvimento e uma integração de saberes. XXII CONGRESSO DA SBEC 2019, UFJF, Juiz de Fora, MG, 2019 set.2-6.

FERREIRA, A. D'O.; TRINDADE, A.W. Ensino e aprendizagem de grego antigo: novos gêneros de escrita nas letras clássicas digitais? In LOUSADA, E. G. et al. (org.). *Diálogos brasileiros no estudo de gêneros textuais/discursivos*. Araraquara: Letraria, 2016.

MACHADO, A. R.; LOUSADA, E. G. A apropriação de gêneros textuais pelo professor: em direção ao desenvolvimento pessoal e à evolução do “métier”. *Linguagem em (Dis)curso*, Palhoça, SC, v. 10, n. 3, p. 619-633, set./dez. 2010.

MALHADAS, D.; DEZOTTI, M.C.; NEVES, M.H.M. Dicionário Grego-Português. São Paulo: Ateliê, 2006-2010.

MAMBRINI, F. The Ancient Greek Dependency Treebank: Linguistic Annotation in a Teaching Environment.” In: BODARD, G; ROMANELLO, M. *Digital Classics Outside the Echo-Chamber*, Ubiquity Press, 2016, p. 83-99. Disponível em: https://doi.org/10.26530/OAPEN_608306. Acesso em 03/01/20.

_____. L’Ancient Greek Dependency Treebank. Un Nuovo Strumento per Lo Studio Della Lingua Greca. *Lexis* v. 29 n.1 51-70, jan. 2011. Disponível em: http://www.lexisonline.eu/wordpress/wp-content/uploads/2016/07/Lexis29_Mambrini.pdf Acesso em 03/01/20.

PALLADINO, C. Treebanking Ancient Greek in High School: what my students learned, what I learned. *The Stoa: A Blog for Digital Classicists*, 2015. Disponível em: <https://blog.stoa.org/archives/2149/> Acesso em: 26/10/15.

_____. How Can You Work with a Language That You Do Not Know? Paper presented at the panel “Confronting Complexity of Babel in a Global and Digital Age” at the DH2019: DIGITAL HUMANITIES CONFERENCE, University of Utrecht, 2020, Jul. 9-12.

REIS, M. F. *Temas e formas em Héracles: um estudo baseado em keywords e treebank*. 2017. 107f. Dissertação (Mestrado). Universidade Estadual Paulista “Júlio de Mesquita Filho” Faculdade de Ciências e Letras (Campus de Araraquara).

SMITH, D. N.; WEAVER, G.A Applying Domain Knowledge from Structured Citation Formats to Text and Data Mining: Examples Using the CITE Architecture. In HEYER, G.; CONFERENCE ON TEXT MINING SERVICES; LEIPZIGER INFORMATIK-VERBUND (eds.). Text mining services: building and applying text mining based service infrastructures in research and industry; *Proceedings of the Conference on Text Mining Services 2009* at Leipzig University. Leipzig: LIV, 2009.

SUMIYA, A. H. O gênero multimodal tutorial em vídeo e suas contribuições no ensino-aprendizagem de francês como língua estrangeira por adolescentes. 2017. Dissertação (Mestrado) Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2016.

TESNIÈRE, L. *Éléments de syntaxe structurale*. 2e éd. Paris: Klincksieck, 1959. Versão em inglês: *Elements of structural syntax* trad. por Timothy Osborne e Sylvain Kahane. Amsterdam: J. Benjamins, 2015. Disponível em: <https://benjamins.com/catalog/z.185/fulltext/z.185.pdf>. Acesso em 03/01/20.

Recebido: 16/02/2020.

Aprovado: 11/05/2020.