

Abordando palavras-chaves e árvores sintáticas em textos de grego antigo: o caso de “Héracles”, de Eurípides

Approaching keywords and syntactic trees in Ancient Greek texts: the case of “Heracles”, by Euripides

Michel Ferreira dos Reis¹

Universidade Estadual Paulista, Brasil

RESUMO

O presente artigo, resultante da pesquisa de mestrado, apresenta a extração automatizada de palavras-chaves da tragédia grega *Héracles*, de Eurípides, associada à anotação morfossintática em árvore (*treebanking*) com a finalidade de determinar temas e aspectos formais da obra. Dessa forma, primeiramente *Héracles* fora comparada estatisticamente com as outras dezoito obras de Eurípides, sem lematização. Posteriormente, foram selecionados trechos contendo as palavras-chaves com a finalidade de analisá-las morfossintaticamente em árvore. Desse modo, os resultados evidenciaram quatro temas centrais na tragédia como os personagens, os laços familiares, a caracterização de *Héracles* por meio de suas armas e trabalhos, e, por fim, sua loucura.

PALAVRAS-CHAVE:

Keywords. Treebank. Língua grega antiga. Eurípides. Ensino e aprendizagem de línguas.

ABSTRACT

This article, which is the result of the master's research, presents the keyword extraction of Euripides' Greek tragedy called *Heracles*, associated with the treebanking in order to determine themes and formal aspects of the work. This, first *Heracles* was compared statistically with the other eighteen works by Euripides, without lemmatization. Subsequently, sections containing the keywords were selected for the purpose of analyzing them morphosyntactically in treebank. The results showed four central themes in the tragedy as the characters, the family ties, the characterization of Heracles through their weapons and labors, and his madness.

KEYWORDS:

Keywords. Treebank. Ancient Greek language. Euripides. Language learning and teaching.

Recebido em: 30/07/2019
Aceito em: 10/04/2020

¹ E-mail: michelfereis@gmail.com | ORCID: 0000-0003-2018-4188.

1. Introdução²

Os textos de línguas históricas como grego e latim carregam em si uma longa jornada de sobrevivência e difusão. Essa questão recai dentre outros elementos na diversidade de materiais nos quais os registros históricos e literários foram originalmente grafados, como tabuinhas de argilas, estelas e papiros, e em seus suportes posteriores nos quais as obras foram copiadas, replicadas e até modificadas pelas mãos dos copistas e dos editores na época da imprensa. O fato é que novos tempos tendem a proporcionar novos suportes para os textos e novas abordagens de estudos, e isso não seria diferente com os textos históricos que encontram mais recentemente, no campo das Ciências Computacionais em consonância com as Ciências Humanas, novos meios e formas de preservação e manipulação por intermédio da exploração e criação de corpora digitais.

Nessa direção, entre os anos 70 e 80 dois grandes projetos ligados a textos greco-latinos e corpora digitais se originam. Nos anos 70, o projeto *Thesaurus Linguae Graeca* (TLG) se forma com a finalidade de criar um escopo de textos e políticas de digitalização para a criação de um *lexikón* de língua grega, e anos mais tarde passa a ser uma coletânea de textos, que inicialmente era distribuída por fitas magnéticas e CD-ROMs. Em 2001, todo o cânone de textos gregos é disponibilizado para consulta online, havendo a possibilidade de análise vocabular como frequência no ambiente virtual, dentre outros recursos. Atualmente, o TLG online tem um corpus de mais de 110 milhões de palavras, no qual o seu uso só exige um registro online, mas o acesso a todo o corpus só é possível por meio de uma assinatura.

O segundo projeto é a *Biblioteca Digital Perseu*³ (BDP), fundado nos anos 80, surgiu com objetivo de investigar o processo de passagem das bibliotecas físicas para o ambiente digital. Hoje, a *Biblioteca* mantém um grande acervo de acesso aberto de obras em línguas históricas como o grego antigo, latim, árabe e outras, contendo uma coleção com mais de 13,5 milhões de palavras em grego e mais de 10,5 milhões em latim. O acervo ainda permite o acesso a traduções dessas obras em língua inglesa, consulta a dicionários, análise morfológica de palavras e de frequência. Além disso, a *BDP* apoia, financia e trabalha colaborativamente com outros projetos que, de alguma maneira ou outra, se baseiam nos mesmos dados digitais com a finalidade de proporcionar novas ferramentas e análises para que possam ser reutilizadas em contextos diversos de pesquisa.

² O presente artigo é um trabalho derivado da dissertação de mestrado *Temas e formas em Hércules: um estudo baseado em keywords e treebank* (2017), com o apoio da FAPESP – Processo nº 2015/00266-6. Atualmente, o autor é bolsista de doutorado pelo CNPq Processo nº 141660/2017-1.

³ Disponível em <<http://www.perseus.tufts.edu/hopper/>> Acesso em: 10 dez 2018.

Nessa direção, o atual ambiente colaborativo de produção de dados digitais abertos chamado *Perseids Project*⁴, que trabalha em colaboração com a BDP, disponibiliza em sua plataforma ferramentas de alinhamento de traduções, editor de *treebanking*, consulta a dicionários, análise morfológica, anotação de epígrafias entre outros. A priori, as ferramentas surgiram como um auxílio no processo de ensino e aprendizagem de línguas históricas, mas permitindo também a criação de mais dados digitais, visto que o uso da plataforma e suas ferramentas possibilitam a criação de dados novos que podem ser compartilhados e estudados.

Assim exposto, acredita-se que primeiramente a exploração de corpora de línguas históricas combinada à criação de dados linguísticos digitais em atividades de ensino/aprendizagem é uma prática que deva ser mais explorada e investigada, embora já haja experiências de tecnologias digitais e corpora abordadas em situações reais de ensino de língua grega antiga (FERREIRA, 2004; LEE *et al*, 2013; HARRINGTON, 2013, 2016), mas ainda de forma escassa. Desse modo, a questão neste trabalho concerne à aplicação de métodos da Linguística de Corpus, mais especificamente o de palavras-chaves por meio da ferramenta Antconc em consonância com a *Greek Vocabulary Tool*, da BDP, visando extrair temas da tragédia e aspectos formais da peça teatral *Héraclès*, de Eurípides, e posteriormente realizar uma análise desses trechos por meio de anotação linguística. Para tal iniciativa, *Héraclès* foi o corpus de estudo e as outras dezoito obras de Eurípides serviram como corpus de referência.

2. Corpora e palavras-chaves

A Linguística de Corpus é uma área que se dedica ao estudo de corpora que são os conjuntos de textos compilados e sistematizados, a fim de serem analisados linguisticamente com o auxílio de programas computacionais (BERBER SARDINHA, 2004; SÁNCHEZ, 1996). Embora o uso de corpora mais recentemente esteja relacionado diretamente ao uso de computador, ressalta-se que a ideia de corpus é anterior ao surgimento de ferramentas computacionais, tendo sido compilados anteriormente de forma manual. O desenvolvimento de áreas da informática e computação ajudou a desenvolver ferramentas e métodos para a compilação e análise de conjuntos de textos facilitando o manuseio e anotação de corpora cada vez maiores.

⁴ Projeto Perseidas em português. Disponível em <<https://www.perseids.org/>> Acesso em: 10 dez 2018.

Epistemologicamente, a Linguística de Corpus se baseia em dados empíricos por meio da observação e experiência e compreende a linguagem a partir de uma concepção na qual o sistema linguístico é visto como um sistema de probabilidades, tal como o linguista Michael Halliday (1992), rejeitando o racionalismo da teoria chomskiana na qual a linguagem é um sistema de possibilidades. Na perspectiva *hallidayana*, as palavras não aparecem ou se manifestam aleatoriamente em um contexto linguístico, mas sua frequência evidencia uma relação entre a escolha de um vocábulo e o número de ocorrências de outros em seu entorno (BERBER SARDINHA, 2004; SINCLAIR, 1987).

Dessa forma, a combinação de palavras pode revelar padrões de uso uma vez que há uma relação essencial entre o léxico e a sintaxe, acrescidas a frequência de ocorrências e suas associações a um determinado sentido (HUNSTON; FRANCIS, 1998). Tais padrões podem se manifestar e ser formalizados em alguns conceitos caros à Linguística de Corpus como, por exemplo, a colocação, que consiste na “associação entre itens lexicais [...] e campos semânticos” (BERBER SARDINHA, 2004, p. 40). Essa relação é exposta também no processo de extração de *keywords* (palavras-chaves), pois segundo Stubbs (2010) elas são

palavras que alegam ter um *status* especial, seja porque expressam significados sociais de avaliação importantes, ou porque desempenham um especial papel em um texto ou tipo de texto. **De um ponto de vista linguístico, elas contribuem para a longa “procura por unidades de significado” (Sinclair, 1996).** De um ponto sociológico de vista, são parte de “um vocabulário da cultura e sociedade” (Williams 1976/1983). No trabalho com palavras-chaves, análise semântica e social são inseparáveis.⁵ (p. 21, tradução e grifo nosso)

Conforme a afirmação de Stubbs (2010), palavras-chaves contribuem para a compreensão semântica, ponto de vista defendido neste trabalho. Além disso, o viés adotado para a extração de *keywords* é uma abordagem quantitativa, na qual os resultados se dão por cálculos estatísticos, uma vez que são consideradas “palavras significativamente mais frequentes em uma amostra de texto do que seria esperado, dada a frequência delas em um grande corpus de referência geral” (STUBBS, 2010 p. 25). A diferença de frequência se relaciona diretamente a uma característica

⁵ Original: “Keywords are words which are claimed to have a special status, either because they express important evaluative social meanings, or because they play a special role in a text or text-type. From a linguistic point of view, they contribute to the long “search for units of meaning” (Sinclair 1996). From a sociological point of view, they are part of “a vocabulary of culture and society” (Williams 1976/1983). In work on keywords, semantic and social analysis are inseparable.”

textual das palavras que é chamada de chavidade visto que ela “é uma qualidade que palavras podem ter em um dado texto ou conjunto de textos, sugerindo que elas são importantes, refletem sobre o que o texto é realmente, evitando futilidade e detalhes insignificantes” (SCOTT; TRIBBLE, 2006, p. 55-56). Isso significa que as palavras-chaves podem apontar temas de um ou conjunto de textos (SCOTT; TRIBBLE, 2006; BONDI, 2010), além de trazer questões de estilo de um texto (SCOTT; TRIBBLE, 2006). Embora neste trabalho, tenham sido obtidas tanto palavras de classes gramaticais fechadas que representam elementos gramaticais e indicariam o estilo, quanto palavras de classes abertas que carregam em si significado lexical, o foco principal foi esta última uma vez que pretendia-se determinar os temas da obra.

3. Corpora e *treebanking*

Treebank é um termo cunhado nos anos 80 pelo linguista Geoffrey Leech e se refere a um corpus linguístico anotado cuja representação visual se configura em formato de árvore⁶, incluindo “análises gramaticais além do nível das classes de palavras” (NIVRE, 2008) no qual as sentenças passam por uma interpretação dos indivíduos de forma explícita, elemento a elemento, fornecendo dados para a área computacional e Linguística de Corpus (BAMMAN; CRANE, 2011), em outras palavras, *treebanks* (ou banco de árvores) são corpora anotados linguisticamente desde o nível morfológico até o semântico, mostrando as relações sintáticas dos elementos da sentença. As informações linguísticas a serem anotadas podem variar conforme a teoria gramatical adotada. Inicialmente, a anotação manual das sentenças pode contribuir na construção e alimentação de bancos de dados a fim de fornecer mais informações linguísticas e digitais, em um processo que auxiliaria, em uma etapa posterior, na indução automática de e por ferramentas computacionais.

Com o avanço das ciências da computação e áreas correlatas, os *treebanks* ganham força entre os anos de 1990 e 2000, período em que a maioria deles foram criados. Primeiro, surgem os de línguas modernas como o *The Penn Treebank* (MARCUS; MARCINKIEWICZ; SANTORINI, 1993) de língua inglês, *The Prague Dependency Treebank* (HAJIC, 1998) de tcheco, o *TIGER Treebank* (BRANTS *et al*, 2002) de alemão e mais recentes de línguas históricas como o de português medieval, o de latim e o de grego antigo. Quanto a suas configurações, contrapõem-se

⁶ Originalmente, o nome se refere a esse tipo de visualização, mas atualmente os *treebanks* não apresentam mais exclusivamente o formato em árvore.

principalmente dois sistemas gramaticais: de um lado, a gramática de constituintes de Chomsky; do outro, a gramática de dependência baseada nos trabalhos de Tesnière, Mal'cuk e Sgal. O *Trebank de dependência de grego antigo* (BAMMAN; CRANE, 2008), baseado no de língua tcheca, foi formulado conforme a gramática de dependência na qual o verbo é o elemento central de uma sentença, de modo que as palavras são ligadas diretamente uma a outra, não podendo ser ligadas a mais de um elemento linguístico, mas todos dependem direta ou indiretamente do verbo.⁷

A escolha da gramática de dependência para o grego antigo se justifica, pois ela “funciona bem especialmente com línguas que envolvem ordem relativamente livre das palavras (o que em uma gramática transformacional envolveria, de outro modo, um alto grau de confusão⁸)” (BAMMAN; CRANE, 2011, p. 3, tradução minha). Ainda segundo os autores, *treebanks* de línguas históricas são diferentes dos de línguas modernas uma vez que não há falantes nativos para a anotação dos dados linguísticos, somada também a alta estilização, característica própria desses textos, o que dificulta a compreensão deles e exige mais tempo e rigor. Entretanto, uma vez que os textos históricos são estudados há séculos, a leitura e entendimento deles ganham o auxílio dos trabalhos científicos dos acadêmicos.

Assim posto, um questionamento a se fazer é em que medida os *treebanks* se encaixam no estudo de línguas históricas e computação? Quais as contribuições? Primeiramente, a anotação manual de bancos linguísticos proporcionaria mais conteúdo para a alimentação de ferramentas computacionais que permitem a extração automática de dados de corpora de línguas históricas, além de fornecer mais dados para a automatização de anotação linguística, contribuindo, por exemplo, na consulta de dicionários online abertos como aqueles que estão disponíveis na página da Biblioteca Digital Perseu. Em segundo lugar, os *treebanks* podem apontar similaridades textuais fornecendo informações sobre reuso de textos. (CRANE; BAMANN, 2011, 2009).

Além disso, os *treebanks* têm sido usados como ferramenta pedagógica. Bamman e Crane (2010) afirmam que eles podem ser usados na aprendizagem de “fenômenos gramaticais complexos” (p.546) desde o nível sintático até o semântico, revelando habilidades linguísticas diferentes entre os aprendizes e pontos gramaticais deficientes. Harrington (2013), a partir de suas aulas de latim, discute que as anotações possibilitam a discussão sobre as suas escolhas

⁷ Para saber mais, consultar: Ágel & Fischer (2009); Volk & Marek (2011).

⁸ O termo em inglês é *scrambling* (embaralhamento), dessa forma é possível entender que a representação da conexão da dependência entre os elementos pode interferir na relação entre outros elementos, causando confusão na compreensão.

linguísticas e além de serem vias possíveis para estudo de estilo dos textos e de seus autores, e simultaneamente, os aprendizes estariam criando mais dados linguísticos por meio de trabalhos interativos e colaborativos em ambiente virtual. (HARRINGTON, 2016)

No mesmo sentido, Lee *et al* (2013), tendo desenvolvido uma experiência com a anotação de chinês clássico na Universidade de Hong Kong, apontaram questões semelhantes como a percepção de pontos gramaticais deficientes dos aprendizes, possibilidades de leitura e anotação não pensadas anteriormente, e a contribuição dos dados para trabalhos de outros pesquisadores. Esses são alguns aspectos levantados em estudos utilizando as anotações sintáticas (*treebanking*) em situações de ensino e aprendizagem de línguas históricas, porém é uma temática ainda um pouco recente e que deve ser investida e investigada em sala de aula, não só como uma ferramenta pedagógica, mas também como um recurso de criação de dados digitais que pode consequentemente ser utilizado em pesquisas computacionais, linguísticas, pedagógicas e ademais. É nesse contexto que o presente artigo se configura. Assim exposto, na próxima seção serão descritos os processos metodológicos do trabalho.

4. Procedimentos metodológicos

A pesquisa optou por uma abordagem mista, visto que a coleta de dados e a extração de palavras-chaves ocorrem por meio de cálculo estatístico, podendo ser quantificados e recebendo valores de chavidade⁹. Dessa forma, os resultados obtidos passam por uma seleção e análise dos temas depreendidos a partir da lista de vocábulos, buscando compreender padrões de linguagem e especificidades morfossintáticas e lexicais da obra associando extração de palavras-chaves e anotação em árvore.

4.1. O corpus de estudo e de referência

A extração de palavras-chaves ocorre por meio da comparação de dois corpora linguísticos: o de estudo, aquele que contém o(s) texto(s) a ser(em) analisado(s), e o de referência, que serve como controle de comparação entre textos. Na tabela 1, listamos as 19 obras de Eurípides retiradas da *Biblioteca Digital Perseu* mostrando a) o corpus de estudo ou de referência; b) os

⁹ Conceito abordado na seção 2.

tokens: o número total de vocábulos que aparecem em cada obra e c) e os *types*: número total de vocábulos, excluindo-se a repetição delas.

Tabela 1: Obras de Eurípides retiradas da *Biblioteca Digital Perseu*

Tipo de corpus	Obras	Tokens	Types
Estudo	<i>Héraclès</i>	8.243	2.988
Referência	<i>Alceste</i>	6.982	2.456
Referência	<i>Andrômaca</i>	7.645	2.711
Referência	<i>As Bacantes</i>	7.984	3.016
Referência	<i>As Fenícias</i>	10.335	3.370
Referência	<i>Ciclope</i>	4.389	2.183
Referência	<i>Electra</i>	8.087	2.861
Referência	<i>Hécuba</i>	7.570	2.702
Referência	<i>Helena</i>	10.435	3.186
Referência	<i>Hipólito</i>	8.500	2.892
Referência	<i>Íon</i>	9.834	3.090
Referência	<i>Ifigênia em Áulis</i>	9.879	2.971
Referência	<i>Ifigênia em Táuris</i>	8.838	2.990
Referência	<i>Medeia</i>	8.323	2.698
Referência	<i>Orestes</i>	10.592	3.119
Referência	<i>Os Heráclidas</i>	6.489	2.155
Referência	<i>Reso</i>	5.648	2.394
Referência	<i>Suplicantes</i>	7.344	2.678
Referência	<i>Troianas</i>	7.434	2.797

Fonte: Adaptado de REIS, 2017, p. 35

4.2. Os instrumentos de coleta e de análise de dados

Para a coleta dos dados, foram utilizadas três ferramentas: 1) *Greek Vocabulary Tool*, disponível na *Biblioteca Digital Perseu*, para o processo de extração de palavras-chaves, que é configurada a mostrar pelo *Key Term Score*¹⁰. Nesse caso, ela passa por lematização e não há controle do corpus de referência; b) *AntConc 3.4.3* para a extração de palavras-chaves sem lematização e com controle do corpus de referência (as 18 peças citadas na tabela 1); e c) o editor de *treebank* Aretusa, disponível na plataforma do projeto *Perseids*, para a anotação morfosintática das sentenças.

¹⁰ Disponível em: <<http://www.perseus.tufts.edu/hopper/help/vocab>> Acesso em 11 nov. 2018.

4.3. O procedimento de coleta de dados

O primeiro procedimento de coleta consistiu na seleção manual do corpus de estudo e de referência, ou seja, as dezenove obras de Eurípides do site da Biblioteca Digital Perseu. Posteriormente, foram realizadas a limpeza e formatação do corpus, conforme os seguintes passos: a) eliminação de numeração de versos; b) junção de palavras que originalmente estão separadas no final dos versos por questões da composição poética e métrica do grego antigo; c) inserção de chaves { } no nome das personagens, quando representava um turno de fala, visto que são tragédias e se não fizesse esse passo haveria uma mudança significativa nas palavras-chaves, porque os nomes dos personagens apareceriam com maior frequência na obra de estudo; d) formatação em extensão *txt*, requerido pelo Antconc, com decodificação em Unicode (UFT-8) a fim de evitar a desconfiguração do grego politônico¹¹.

No segundo procedimento, foram extraídas as palavras-chaves com o Antconc, sem lematização, tendo *Héraclès* como corpus de estudo e as outras 18 obras de Eurípides como corpus de referência. A prova estatística para determinar os valores de chavicidade foi o Log-Likelihood (a probabilidade logarítmica) e o ponto de corte, ou seja, o limiar crítico de 10,83. Isso representa que a probabilidade de as palavras aparecerem aleatoriamente nessa lista é de $p < 0.001$, ou seja, menos de 0,01%. Posteriormente, foi executada a extração de *keywords* pela *Greek Vocabulary Tool*. Esse método se difere em partes do processo com o Antconc, visto que há lematização dos vocábulos e não há a possibilidade de controle do corpus de referência. Desse modo, foram selecionados os 300 primeiros resultados, o que representa 10% do número de *types* em *Héraclès*, para serem comparados com a lista do Antconc.

4.4. A análise dos dados

Após as palavras-chaves serem extraídas, elas passaram por uma análise de concordâncias e colocados no programa Antconc. A partir disso, as sentenças nas quais constavam as palavras-chaves foram separadas para a anotação morfossintática em árvore na ferramenta Arethusa, pois se pretendia entender melhor os vocábulos em torno dessas, mas tendo só o auxílio do Antconc

¹¹ O grego moderno adota ortografia monotônica, na qual há a redução de sinais diacríticos em relação ao grego antigo, restando somente os sinais agudo (´) e diérese (¨).

não seria possível, visto que precisavam ser analisadas e anotadas sistematicamente o que é proporcionado pela Arethusa.

5. Sobre o que dizem as palavras-chaves?

A partir da delimitação metodológica explicitada na seção anterior, foram obtidas duas listas de palavras-chaves. A primeira lista do programa Antconc tinha sessenta e quatro palavras-chaves sem lematização. (Tabela 2)

Tabela 2: Palavras-chaves sem lematização

Rank	Freq.	Palavra-chave ¹²	Rank	Freq.	Palavra-chave	Rank	Freq.	Palavra-chave
1	12	γέροντες [idosos]	23	6	σῖγα [em silêncio]	44	2	πενία [pobreza]
2	11	ἥρακλέους [Héracles]	24	4	ἔτλην [sofrer]	45	2	περιβόλαι [involucros]
3	12	εὐρυσθέως [Euristeu]	25	3	αἰδοῦς [aedo]	46	2	πτερωτός [alado]
4	17	τέκν' [filho]	26	3	ὑδραν [hidra]	47	2	σκαίων [tolo]
5	12	ἄιδου [Hades]	27	4	ἥρακλείους [Héracles]	48	2	σκευάζεται [fazer; preparar]
6	5	ἥρακλῆς [Héracles]	28	7	ἦλθεν [ir; vir]	49	2	ταφίων [táfios]
7	4	λύκος [Lico]	29	6	ἥρας [Hera]	50	2	τεκόμενος [gerar]
8	4	ἀμφιτρώων [Anfitrião]	30	2	βακχεύσει [entrar em frenesi]	51	2	τρίκρανον [três cabeças]
9	19	τέκνων [crianças]	31	2	βλέπει [ver]	52	2	ἐναίρων [matar]
10	4	κύνα [cão]	32	2	γοργωπούς [gorgôneos]	53	2	ἐπισφάζων [sacrificar]
11	4	τόξ' [arco]	33	2	δάιοι [hostis]	54	2	ἔφανε [dizer]
12	8	πρέσβυ [velho]	34	2	δειλίας [covardia]	55	2	ὄμαρτεῖτ' [olhar]
13	3	θηβῶν [Tebas]	35	2	εἰσέβαινε [observar]	56	2	ὄπλον [armas]
14	3	στάσει [sedição]	36	2	εὐνήτωρ [esposa]	57	2	ὦφθης [ver]
15	17	τέκνα [crianças]	37	2	κάμνουσι [sofrer; cansar-se]	58	6	δράσας [fazer]
16	16	φίλων [amigos]	38	2	κίονος [coluna de pedra]	59	6	κτανών [matar]

¹² Embora as palavras estejam declinadas, optou-se por não as especificar na tradução visto que algumas formas podem ser iguais para casos diferentes. Em relação aos verbos, optou-se por traduzi-los pelo infinitivo e não por sua forma conjugada equivalente ao grego.

17	13	πατέρα [pai]	39	2	λυγραι [deplorável]	60	4	σκῆπτρα [cetro]
18	15	παῖδας [filhos]	40	2	λύσσα [Lissa]	61	48	ὁ [o/a]
19	4	τάν	41	2	μεγάρα [Mégara]	62	3	δεδραμένων [fazer]
20	5	νέρθεν [de baixo]	42	2	μεταλλαγαί [mudanças]	63	3	πηγάς [lágrima]
21	5	τέκεα [crianças]	43	2	παιδοκτόνους [filicida]	64	3	ἡγοῦμαι [levar; conduzir]
22	6	άνόσιον [profano]						

Fonte: Adaptado de REIS, 2017, p. 47

A segunda lista, fornecida pela pontuação de *Key Term* da ferramenta de vocabulário grego da *Biblioteca Digital Perseu*, resultou em dezenove palavras (Tabela 3). Nesse processo, foram elencadas todas as palavras em ordem decrescente de *Key Term* e selecionadas as trezentas primeiras a serem comparadas à lista do Antconc.

Tabela 3: Palavras-chaves lematizadas

Rank	Palavra-chave	Rank	Palavra-chave	Rank	Palavra-chave
31	σιῖγα [silenciosamente]	121	σιγάω [ficar em silêncio]	231	παῖς [filho]
32	σιγάς [silêncio]	162	δράω [fazer, acompanhar]	258	τόξον [arco]
37	τέκνον [criança]	179	βακχεύω [entrar em frenesi báquico]	267	κύων [cão]
72	δάιος [hostil]	187	πηγή [água corrente, lágrima]	277	ἔνερθε [abaixo]
98	φίλος [amigo]	219	γέρων [idoso]	283	λυσσάς [ira]
100	σιγάζω [estar em silêncio]	223	τέκος [jovem]	289	ῥῆδρα [hidra]
102	σιγή [silêncio]				

Fonte: Adaptado de REIS, 2017, p. 48

Como esperado, o número de palavras-chaves foi menor do que aquele do Antconc visto que a ausência de controle do corpus de referência e a inserção do próprio corpus de estudo (a peça *Héacles*) no corpus de referência modifica o valor de *key term* e a classificação das palavras-chaves. Embora essa lista seja menor do que a primeira, ela reforça a relevância da ocorrência de algumas de palavras não lematizadas que serão analisadas e discutidas a seguir.

5.1. Os quatro temas

Antes de abordar os temas, é necessário contextualizar o enredo da peça *Hércules*. A tragédia *Hércules* foi escrita por Eurípides (480 - 406 a.C.). Nela, o herói que dá nome à peça começa ausente, na morada de Hades, para realizar o último dos 12 trabalhos a mando de Euristeu. Após seu último feito, dirige-se a Tebas, onde o tirano Lico está prestes a matar seu pai Anfitrão, sua esposa Mégara e seus filhos a fim de tomar o poder da pólis. Inesperadamente, Hércules interrompe os planos de Lico e o mata, salvaguardando seus familiares. Todavia, quando tudo parece finalizado e resolvido, a deusa Hera, que seria uma das maquinadoras dos doze trabalhos, lança ao herói a loucura e a ira, essa personificada pela deusa Lissa. Enfurecido, Hércules confunde seus filhos com os de Euristeu e os mata, porém, quando está na iminência de tirar a vida de Anfitrão, a deusa Atena intercede enviando-lhe um sono profundo. Depois de despertar, Anfitrão revela o acontecido e Hércules incrédulo deseja a si próprio a morte, porém Teseu surge e intercede pela vida do amigo recomendando que o acompanhe até a cidade de Atenas, local em que poderia se redimir de sua falta cometida.

Em suma, esse é o enredo da tragédia, e uma primeira leitura permite afirmar que a temática na atmosfera teatral é a loucura e a violência de Hércules contra seus familiares. Entretanto as questões que se colocam aqui são *quais marcas linguísticas extraídas automaticamente revelam o plano do conteúdo? E quais os papéis temáticos dessas palavras?* Os resultados do Antconc pela extração de palavras-chaves não lematizadas trazem à tona, concomitantemente, temas abarcados na obra e fornecem indícios da função sintática das palavras, visto que na língua grega as palavras aparecem declinadas com suas desinências de casos. Então, a partir da extração de palavras-chaves foi possível identificar quatro grupos temáticos maiores em *Hércules* (quadro 1): os personagens da tragédia, laços familiares e de amizade, as armas e os trabalhos do herói e, por fim, a loucura de Hércules.

Quadro 1: Temas e palavras-chaves

TEMAS	Personagens	Laços familiares e de amizade	Armas e trabalhos de Héracles	Loucura de Héracles
PALAVRAS-CHAVES	Ἅιδου [Hades]	γέροντες [idosos]	κύνα [cão]	ἀνόσιον [profano]
	Ἀμφιτρύων [Anfitrião]	εὐνήτωρ [marido]	ὄπλον [armas]	βακχεύσει [entrar em frenesi]
	Εὐρυσθέως [Euristeu]	πατέρα [pai]	τρίκρανον [três cabeças]	γοργωπούς [gorgôneos]
	Ἡρακλείους [Héracles]	πρέσβυ [velho]	ῥδραν [hidra]	δάιοι [hostis]
	Ἡρακλέους [Héracles]	τέκνα [crianças]		ἐπισφάζων [sacrificar]
	Ἡρακλῆς [Héracles]	τέκνων [crianças]		Ἥρας [Hera]
	Ἥρας [Hera]	φίλων [amigos]		κίονος [coluna]
	Λύκος [Lico]			Λύσσα [Lissa]
	Λύσσα [Lissa]			παιδοκτόνους [filicida]
	Μεγάρα [Mégara]			
	γέροντες [idosos]			

Fonte: Adaptado de REIS, 2017

O **primeiro tema** é composto pelos personagens presentes na obra, contabilizando o total de onze palavras, das quais dez são nomes próprios e uma, que representa um coletivo de pessoas, o coro. Esse conjunto ainda poderia ser classificado em três categorias conforme seu aparecimento na peça:

a) personagens presentes e atuantes: as palavras Ἀμφιτρύων [Anfitrião], Ἡρακλείους, Ἡρακλέους, Ἡρακλῆς (três formas lexicais/declinadas do nome Héracles), Λύκος [Lico], Λύσσα [Lissa] que também é parte do vocabulário da loucura, e Μεγάρα [Mégara] nomeiam os indivíduos que participam efetivamente da encenação da peça. Acrescenta-se aqui também γέροντες [idosos], a única palavra desse grupo que não é um nome próprio, mas identifica um conjunto de participantes que formam o coro da tragédia.

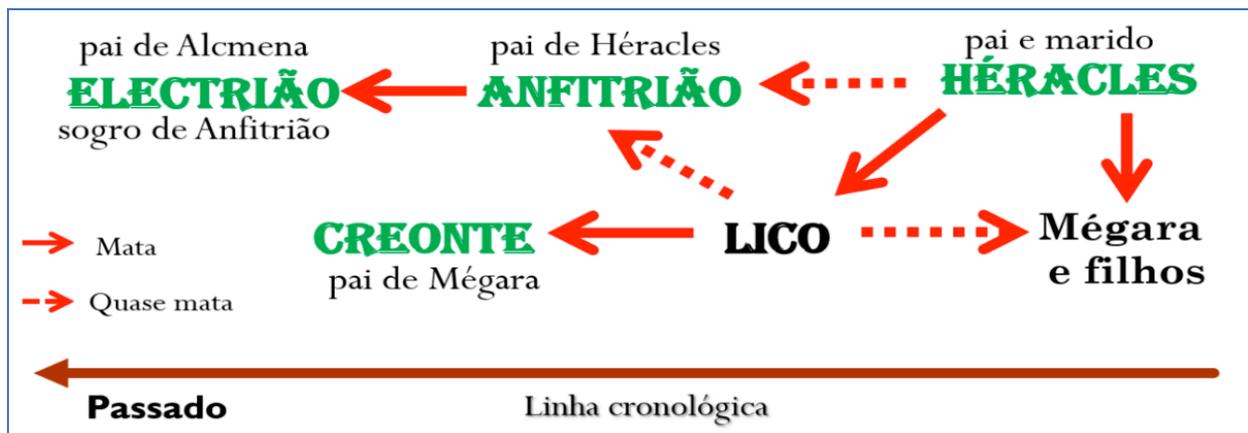
b) personagens não presentes e atuantes: a palavra Εὐρυσθέως [Euristeu] e Ἥρας [deusa Hera] apresentam figuras ausentes da encenação, mas representam um papel atuante no enredo. Euristeu é o mandatário de Héracles aos seus doze trabalhos, explicitado no início da tragédia, e durante o assassinato de seus próprios filhos Héracles confunde-os com os de Euristeu, e revela de

tal modo o ódio que tinha dele. Além disso, a loucura recai sobre o herói, porque fora tramada por ação da deusa Hera.

c) personagem-espaco: a palavra Ἅιδου, forma declinada do lema Ἅιδης [Hades], está no genitivo masculino singular, ocorrendo doze vezes na peça. Em sua maioria, ela é usada logo após a preposição εἰς [para, em direção a] – indicando movimento de entrada, ou depois de ἐκ [de] – no sentido de origem, identificando Hércules como aquele que “adentra a morada do Hades” ou “sai dos domínios de Hades”. A passagem pelo mundo dos mortos marca uma saída vitoriosa do herói perante a mortalidade dos humanos, no caso, Hércules se imortaliza por sua passagem no Hades. (ROSA, 2012)

Há de se indagar que a obtenção de nomes de personagens seja um fenômeno óbvio, uma vez que forneceria o nome dos protagonistas. Entretanto, há dois contra-argumentos que demonstram a relevância desses resultados. Em primeiro lugar, o texto analisado é uma peça de teatro, constituída por diálogos e turnos de fala, o que acarretaria a indicação do nome dos personagens no início de suas falas, porém o Antconc foi programado para suprimir o aparecimento desses nomes. O segundo contra-argumento é que o aparecimento de nomes próprios na peça deveria aparecer primordialmente no caso vocativo do grego, quando houvesse turnos de falas entre os personagens por ser característico de situações dialogais, entretanto só uma pequena parcela de ocorrências assim se apresentam, sendo Ἀμφιτύων (3x) e Λύσσα (1x). Isso representa somente 6,45 % em comparação ao total de 62 ocorrências dos nomes de personagens.

O **segundo tema** é a atmosfera de relações familiares entre os personagens. O enredo se desenrola entre sujeitos de proximidade de laço afetivo, das quais percebe-se quatro funções temáticas distintas. O primeiro grupo representa aqueles que cometeram algum tipo de violência, sendo marcado pelas palavras-chaves εὐνήτωρ [marido] e πατέρα [pai]. As treze ocorrências da forma πατέρα [pai] citam quatro figuras paternas distintas: Anfitrião (9 vezes), pai de Hércules; Electrião (1 vez), pai de Alcmena, a mãe de Hércules; Creonte (1 vez), pai de Mégara, a esposa de Hércules; e o próprio Hércules nos papéis de pai e marido.

Quadro 2: As figuras paternas na peça *Héacles*

Fonte: Autoria própria

Inicialmente, o perigo se instaura com a presença do tirano Lico que buscava dizimar toda a família de Héacles, porém já tinha iniciado seu plano ao revelar seu primeiro ato na cidade de Tebas: “Minha atitude não é ausente de vergonha, velho,/ mas cautelosa: pois sei que, como matei Creonte,/ **pai** desta [Mégara], possuo seu trono” (vv. 165-167)¹³. Após assassinar Creonte, o pai de Mégara, Lico quer tirar a vida do pai de Héacles, o ancião Anfitrião, de Mégara e de seus filhos, mas não consegue, pois o herói retorna do Hades e os salva ao matar o tirano Lico. Entretanto o assassinato é concretizado pelas mãos de quem os salvou anteriormente: Héacles aniquila seus filhos e sua esposa, e só não tira a vida de seu pai por intervenção da deusa Atena. Nota-se que quase se realiza uma sequência de matanças entre os homens de uma família (Quadro 2), pois Héacles salva seus entes eliminando o perigo, mas se torna culpado ao matar sua própria família, além de quase acabar com a vida de seu pai Anfitrião, que também tinha sangue de família em suas mãos de acordo com a própria fala de Héacles:

πρῶτον μὲν ἐκ τοῦδ’ ἐγενόμην, ὅστις κτανῶν
μητρὸς γεραιὸν πατέρα προστρόπαιος ὦν
ἔγημε τὴν τεκοῦσαν Ἀλκμήνην ἐμέ.
ὅταν δὲ κρηπίς μὴ καταβληθῆ γένους
ὀρθῶς, ἀνάγκη δυστυχεῖν τοὺς ἐκγόνους.

Primeiro, nasci deste [Anfitrião] que tendo matado
o velho pai de minha mãe e, sendo maculado,
casou-se com aquela que me deu à luz, Alcmena.
Quando o alicerce da raça não é estabelecido bem
é inevitável que os descendentes tenham má sorte.
(vv.1258-1262)

O segundo grupo é formado pelos indivíduos que sofrem violência, isto é, os filhos de Héacles que são evidenciados pelas palavras-chaves τέκν’, τέκνων e τέκνα, todas formas lexicais

¹³ Todas as traduções de *Héacles* apresentadas no artigo são de minha autoria.

do lema τέκνον [criança], παῖδας forma declinada do lema παῖς [filho] e τέκεα do lema τέκος [jovem/filho]. Essas formas lexicais marcadas pelas desinências de caso podem informar suas funções sintáticas nas sentenças, porém algumas formas coincidem mesmo com função sintática diferente, fato que exige uma análise mais atenta do contexto da palavra, com o auxílio, por exemplo do treebank de grego visto a flexibilização sintática dos termos na sentença grega. As únicas palavras das quais é possível determinar e afirmar categoricamente a morfologia são τέκνων [das crianças], genitivo neutro plural, e παῖδας [filhos], no caso acusativo masculino plural. Se forem analisadas as circunstâncias das ações nas quais as palavras-chaves aparecem, é revelada uma atmosfera ruim para os filhos, ligada à morte majoritariamente. Pensando numa dicotomia de uma violência ou não, pode-se definir o ambiente no qual os filhos estão em três tipos de processos: a primeira determinada pela negatividade das ações, marcada pela violência, ou seja, os filhos têm a vida ameaçada e é justamente nessa situação disfórica que a maioria da ocorrência dessas palavras se encontra; a segunda seria um polo positivo no qual revela-se a positividade das ações, como por exemplo, a tentativa de salvar os filhos, e uma terceira que não significa nenhuma das anteriores, mas uma neutralidade desse ponto de vista. Analogamente, as ocorrências tanto de τέκν' quanto de τέκνα seguem essa mesma perspectiva negativa, além de τέκνων, παῖδας e τέκος.

O terceiro grupo representado pelas palavras-chaves γέροντες (lema γέρων - idoso) e πρέσβυ (lema πρέσβυς - velho) aponta os indivíduos observadores e incapazes de impedir a violência. Seja no papel de coro formado por idosos [γέροντες], seja no papel do velho Anfitrião [πρέσβυς] fica evidente a impotência da velhice face a juventude de Lico e Hércules no decorrer da peça.

O quarto grupo se caracteriza por aquele que ajuda, aquele que é amado, o amigo, representado pela palavra φίλων (lema φίλος - amigo). Aqui, a representação da amizade é o herói Teseu que desempenhará esse papel mais ao final da peça, logo após a loucura de Hércules. É Teseu, que fora outrora resgatado do Hades por Hércules, que exerce o papel de salvador, evitando que o protagonista se mate.

O **terceiro tema** é a caracterização e representação de Hércules por meio de suas armas e seus deveres impostos por Euristeu. De um lado, κύνα [cão], τρίκρανον [três cabeças] e ὕδραν [hidra] indicam dois dos trabalhos de Hércules: a) a busca do guardião da morada de Hades, conhecido como Cérebro e caracterizado como o cão de três cabeças; e b) o trabalho anterior que

é a morte da criatura conhecida como Hidra de Lerna. A fim de tais feitos, o herói carregava consigo suas armas [ὄπλον] e seus arcs [τόξ’], porém as mesmas armas que foram símbolo de vitória e grandeza se tornam vergonha e desgraça ao terem sido usadas no assassinato dos próprios filhos de Hércules.

O **quarto tema** demonstra a presença da loucura de Hércules, desvelando três características do evento: os causadores – Hera [Ἥρα] e Lissa [Λύσσα], os sintomas – gorgôneos [γοργωπούς] e frenesi báquico [βακχεύσει] e o contexto das mortes. Como já abordado anteriormente, Lissa, a personificação da loucura, é enviada sob o comando de Hera para confundir a mente de Hércules e fazê-lo perder a razão ao ponto de matar seus filhos. Em relação aos sintomas, Lissa caracteriza o início da loucura do herói, pois “vê ainda como ele agita a cabeça desde a chegada/ e, em silêncio, revira retorcidos olhos gorgôneos” (vv. 867-868). Os olhos de Hércules tornam-se gorgôneos [γοργωπούς] fenômeno que permite uma aproximação simbólica da figura do herói com a deusa Lissa, visto que ela é descrita como uma Górgona da Noite [Νυκτὸς Γοργῶν] na própria peça. Além disso, estes olhos evidenciam mais uma imagem na peça que é a presença de serpentes de mesmo olhar enviadas por Hera, quando Hércules era criança, com a finalidade de matá-lo. Desse modo, uma vez que não consegue destruir seu inimigo, Hera o transforma em um dos seus através dos olhos enfurecidos despertados por Lissa, como se fosse uma vitória simbólica.

Por fim, as cenas de morte dos filhos configuram um ambiente sacrificial evidenciado pelas palavras ἀνόσιον [profano], κίονος [coluna do altar de Zeus], ἐπισφάξων [sacrificar] e παιδοκτόνους [filicida]. Leitura corroborada por Papadoulou (2005) que afirma que a aparição de vestimentas fúnebres e o canto do coro, que descreve as cenas violentas, permitem essa compreensão. Da mesma forma, Franciscato (2003) compreende que Hércules deva passar pela purificação da morte de Lico, mas para isso seu sacrifício tem de ser o próprio sangue ao ponto de tornar-se um filicida [παιδοκτόνους].

A extração de palavras-chaves permitiu evidenciar e elencar quatro temas maiores na tragédia Hércules, de Eurípides, quando comparada com demais obras do autor. A aparição de personagens participantes ativamente ou não nas cenas, os laços familiares e a amizade, as armas e trabalhos do herói e a loucura são alguns temas abordados não só aqui, mas também reforçados pela crítica literária.

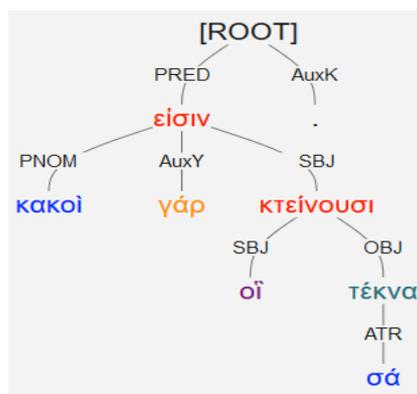
5. Dois exemplos de árvores anotadas

A seção anterior foi dedicada à apresentação e discussão das palavras-chaves obtidas na pesquisa, já nesta parte serão trazidos dois exemplos de anotação morfossintática em árvore selecionados a partir das sentenças que contêm as palavras-chaves a fim de demonstrar o funcionamento de treebank de dependência de grego antigo e a contribuição de seu uso. O objetivo não é realizar uma análise exaustiva das árvores sintáticas, apontando pontos polêmicos de anotação, mas sim ser uma pequena amostra de como o procedimento funciona e trazer alguns questionamentos.

A ferramenta de anotação é chamada Arethusa e está disponível na plataforma do Perseids Project que é um ambiente online aberto para a produção colaborativa de dados a partir de textos históricos, no qual é possível fazer alinhamento de traduções de textos, anotação morfossintática e semântica etc. A anotação morfossintática segue os parâmetros previstos pelo Guia de Anotação Sintática do Treebank de Dependência do Grego Antigo (2008) contendo um conjunto específico de etiquetas para anotação sintática.

O primeiro exemplo, retirado da fala de Mégara, é o verso *κακοὶ γάρ εἰσιν οἱ τέκνα κτείνουσι σά* [“de fato, maus são aqueles que matam teus filhos”] (v.496). Aqui, temos a ocorrência da palavra-chave *τέκνα* [filhos].

Figura 1: anotação de treebank do verso 496



Fonte: REIS, 2017, p. 67

A sentença acima (Figura 1) é um exemplo de uma sentença simples do ponto de vista morfossintático, constituída de um verbo conjugado e principal (PRED) *εἰσιν* [ser/estar] – o verbo é uma forma de 3ª pessoa do plural do verbo *εἰμί* [ser/estar] –, mais um sujeito (SBJ) e um

predicativo do sujeito (PNOM) – κακοὶ [maus] –, ambos dependentes do verbo principal. A peculiaridade recai sobre a forma finita de κτείνουσι [matar], ou seja, uma forma conjugada, que constitui outra oração dentro desse verso οἱ τέκνα κτείνουσι σά [“aqueles que matam teus filhos”]. Como visto, οἱ exerce a função do pronome demonstrativo “aqueles” e de sujeito em relação a κτείνουσι, que conseqüentemente depende do PRED - εἶσιν, essa construção poderia ser reduzida a “os assassinos”. Ainda dependente de κτείνουσι há o objeto “filhos” [τέκνα] acompanhado de um atributo “teus” [σά], ambos separados por κτείνουσι na oração por causa da flexibilidade da ordem sintática da língua grega.

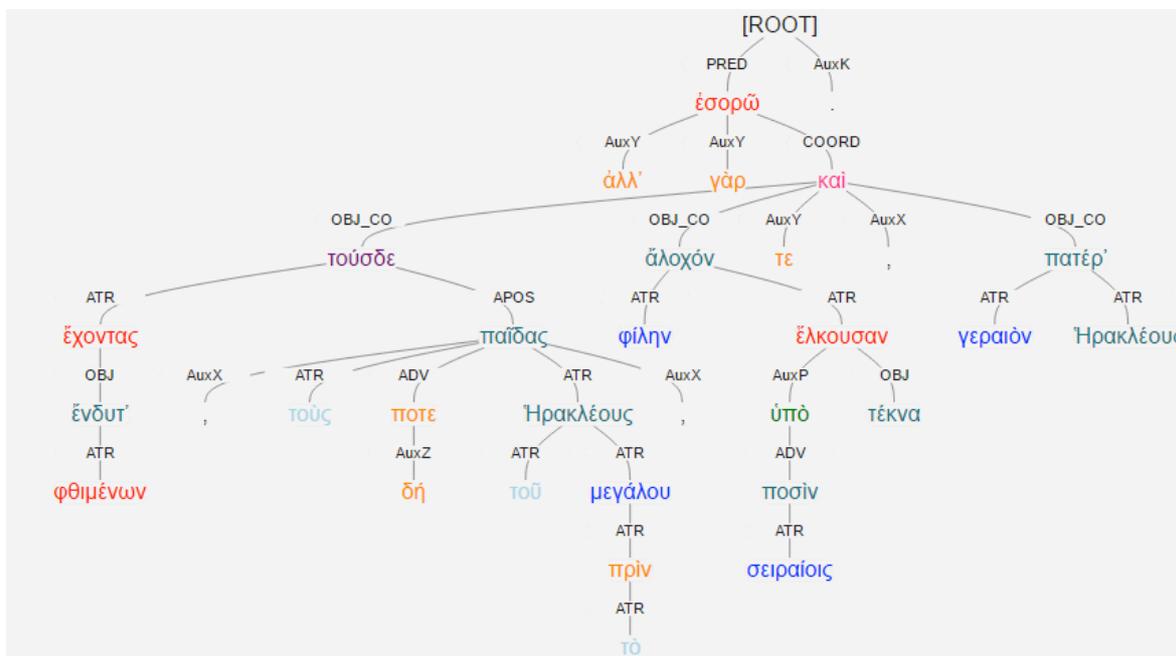
O segundo exemplo de anotação é o trecho da fala do coro entre os versos 442 e 447 que traz a ocorrência das palavras-chaves παῖδας, τέκνα e Ἡρακλέους (duas vezes):

ἀλλ' ἔσορῶ γὰρ τούσδε φθιμένων
 ἔνδυτ' ἔχοντας, τοὺς τοῦ μεγάλου
 δὴ ποτε παῖδας τὸ πρὶν Ἡρακλέους,
 ἄλοχόν τε φίλην ὑπὸ σειραίοις
 ποσὶν ἔλκουσαν τέκνα, καὶ γεραιὸν
 πατέρ' Ἡρακλέους. [...]

Mas observo estes, que vestem
 indumentárias de mortos, os filhos
 do outrora grande Hércules,
 e a amada esposa que arrasta
 as crianças sob pés atados e o velho
 pai de Hércules. [...] (vv.442-447)

Os versos anteriores são compostos praticamente por uma oração principal mais três objetos coordenados, acompanhados, em alguns casos, de orações atributivas (Figura 2). Assim, ἔσορῶ [observar] é o verbo principal (PRED), logo está conectado à raiz da sentença, o sujeito está oculto – “eu” – os dois advérbios oracionais ἀλλ' [mas] e γὰρ [de fato] dependem do PRED, e os três objetos de ἔσορῶ [observar] sofrem uma coordenação de καὶ [e] e por isso estão anotados como OBJ_CO, ou seja, objetos coordenados. Ainda dependente de καὶ [e] são a partícula τε [e] e a vírgula, ambas com a mesma função coordenativa de καὶ, entretanto recebem outra etiqueta, visto que segundo as regras de anotação somente o último elemento coordenativo deve ser anotado como COORD. O primeiro objeto coordenado é τούσδε [estes] que tem dois elementos dependentes: o verbo ἔχοντας [ter] na função de oração atributiva e o nome παῖδας [filhos] como um aposto. Dependente de παῖδας [filhos] há o atributo Ἡρακλέους [de Hércules]. O segundo objeto é ἄλοχόν [esposa] e tem ἔλκουσαν [arrastar] como um dependente anotado como ATR.

Figura 2: Treebank dos versos 442-447



Fonte: REIS, 2017, p. 74

Essas duas árvores sintáticas são uma pequena amostra das que foram analisadas e anotadas durante a pesquisa, mas servem para levantar alguns apontamentos. Em primeiro lugar, a escolha de trechos baseado na extração de palavras-chaves serve como um critério de delimitação de sentenças para o treinamento e uso de treebank no qual o usuário possa praticar a anotação morfossintática, seja com períodos mais curtos ou mais longos, seja com períodos simples ou mais complexos. Em segundo lugar, as árvores servem como exemplos de relações sintáticas que permitiriam uma compreensão de tipicalidades nas sentenças, mas para que isso possa ser estudado é necessária a anotação de mais dados digitais, logo é preciso que mais pessoas anotem corpora de forma manual para que seja possível ter dados em grande escala e, dessa forma, as sentenças sejam comparadas e analisadas. O último ponto é que elas também podem ser utilizadas em estudos sobre tradução, pois a anotação dos elementos na sentença passa pela leitura e interpretação dos sujeitos, revelando assim compreensões diferentes de um mesmo texto ou trecho.

Considerações finais

Baseado nas questões teóricas e metodológicas da Linguística de Corpus, o presente

trabalho alcançou seu objetivo inicial que foi o de extrair palavras-chaves da tragédia *Hércules*, de Eurípides, de forma automática com o auxílio da ferramenta Antconc quando comparada com as demais obras do autor grego. Num segundo momento, os trechos contendo as palavras-chaves foram selecionados para uma anotação morfosintática, com o auxílio da Arethusa, com o objetivo de compreender as relações das demais palavras nesses contextos, e servindo como exemplo de anotação morfosintática para criação de dados linguísticos digitais. Os resultados advindos da extração das palavras-chaves permitiram determinar quatro grandes temas em *Hércules*: a) os personagens da peça, sendo participantes diretos nas ações ou indiretamente; b) a relação afetiva familiar entre pais e filhos que perpassa a tragédia; c) os elementos caracterizadores do herói como suas armas e suas tarefas; e d) a loucura de *Hércules*. Desse modo, são postas duas perguntas quanto a replicação do processo no estudo das demais peças de Eurípides: Quais temas seriam evidenciados pelas palavras-chaves nas demais tragédias? Que elementos linguísticos e temáticos caracterizá-las-iam no conjunto de obras do autor? Essas questões ainda em aberto podem ser orientadoras para novos trabalhos.

Em relação aos recursos e métodos usados na pesquisa como lista de frequência e extração de palavras-chaves, eles se demonstram como mais uma possibilidade aos estudiosos de línguas históricas para que realizem análise linguística em diálogo com a crítica literária por meio da exploração de corpus. Como visto, as palavras extraídas trouxeram elementos temáticos já abordados pela crítica literária de modo a reforçá-las (e até mesmo podendo rebatê-las) e possibilitando uma via de aproximação entre ambos campos de estudo. Para tal empreitada, os recursos e dados digitais deram suporte para se pensar em novos meios e abordagens de estudos e até mesmo de ensino/aprendizagem. Nesse sentido, tanto o corpus e a lista de palavras-chaves quanto o treebank que foram elaborados e/ou utilizados configuram-se como uma amostra em potencial para novas análises linguístico-literárias e para atividades didáticas em aulas de grego antigo, mas não exclusivamente. Esse tipo de abordagem com línguas modernas não é novidade, porém em relação ao grego antigo é possível pontuar situações particulares (FERREIRA, 2004; HARRINGTON, 2013, 2016; LEE et al, 2013), mas ainda escassas, principalmente no Brasil. Além disso, o uso de treebank anteriormente exemplificado proporcionaria a criação e surgimento de mais dados digitais e linguísticos que poderiam ser compartilhados e usados em contextos diversos de pesquisa, ensino e divulgação científica numa interface entre Humanidades e Ciências Computacionais.

Referências

BAMMAN, D.; CRANE, G. The Ancient Greek and Latin Dependency Treebanks. In: SPORLEDER, C.; VAN DEN BOSCH, A.; ZERVANOU, K. (Ed.) *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series*, Berlin Heidelberg: Springer-Verlag, 2011, p. 79-98.

_____. Corpus Linguistics, Treebanks and the Reinvention of Philology. In: *Informatik 2010*, p. 542-551, 2010-01 Disponível em: <<http://subs.emis.de/LNI/Proceedings/Proceedings176/558.pdf>> Acesso em: 30 dez. 2018.

_____. Structured Knowledge for Low-Resource Languages: The Latin and Ancient Greek Dependency Treebanks. In: *Text Mining Services 2009*, Leipzig, Germany: Springer Verlag, 03-2009.

_____. Guidelines for the syntactic annotation of Ancient Greek treebanks, version 1.1. Tech. rep., Tufts Digital Library, Medford, 2008. Disponível em: <<http://nlp.perseus.tufts.edu/syntax/treebank/greekguidelines.pdf>> Acesso em: 30 dez. 2014.

BERBER SARDINHA, T. B. *Linguística de Corpus*. Barueri: Manole, 2004. 410 p.

BRANTS, S. et al. The TIGER Treebank. In: SIMOV, K. (Ed.) *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, Bulgaria, 2002. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.6241&rep=rep1&type=pdf>>. Acesso em: 22 nov. 2018.

BONDI, M. Perspectives on keywords and keyness: an introduction In: BONDI, M.; SCOTT, M. (Eds) *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins, 2010, p. 1-18.

FERREIRA, A. A. G. D. Oneirocritica: ancient desires and modern concepts of metaphor. In: *Tercer Coloquio Internacional. Ética y Estética. De Grecia a la Modernidad*, 2004, La Plata, Argentina. Tercer Coloquio Internacional. Ética y Estética. De Grecia a la Modernidad. La Plata: Universidad Nacional de La Plata, 2004. v. 1. p. 194-21.

FRANCISCATO, C.R. Introdução. In EURÍPIDES. *Héacles*. Introdução, tradução e notas de Cristina Rodrigues Franciscato. São Paulo: Palas Athena, 2003.

HALLIDAY, M. A. K. 1992. Language as system and language as instance: The corpus as a theoretical construct. In SVARTVIK, J. (Ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*. Mouton de Gruyter, Berlin, p. 61-77.

HARRINGTON, J. M. Iterative Morfossyntax: treebanking pedagogy as an application of the experimental method from 1st semester to graduate-level at Tufts University. 2016. In: FOWLER, R.; GRUBER, G.; HARRINGTON, J. M.; MORRELL, K. *Teaching Classics in the digital era: pedagogical and logistical issues*. Disponível em <<https://www.youtube.com/watch?v=8K7kVrfdRXo>> Acesso em: 30 nov. 2018.

_____. *Meaningful Distinctions: The Utility of Perseus Project Latin*

Treebanking Tools for Latin Research-Based Pedagogy. 2013. Disponível em: <https://camws.org/meeting/2013/abstract_list.php> Acesso em: 30 nov. 2018.

HAJIC, J. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In: HAJICOVA, E. (Ed.) *Issues of Valency and Meaning*. Studies in Honor of Jarmila Panevová. Prague, Charles University Press. Karolinum, Praha, 1998. p.106-132. Disponível em: <https://ufal.mff.cuni.cz/pdt/Corpora/PDT_1.0/References/pdtiovam.pdf>. Acesso em: 22 mai. 2016.

HUNSTON, S.; FRANCIS, G. verbs observed: a corpus-driven pedagogic grammar. In: *Applied Linguistics*, Oxford, Oxford University Press, v. 19, n. 1, 1998, p. 45-72.

LEE, J.; HUI, Y.C.; KONG, Y.H. Treebanking for Data-driven Research in the Classroom In: *Proceedings of the Fourth Workshop on Teaching Natural Language Processing*. Sofia, Bulgaria: Association for Computational Linguistics, 2013. p. 56-60. Disponível em: <<http://www.aclweb.org/anthology/W13-3409>>. Acesso em: 20 set. 2015

MARCUS, M. P., MARCINKIEWICZ, M. A.; SANTORINI, B. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. In: *Computational Linguistics* v. 19, n. 2, p. 313-330. Disponível em < <http://aclweb.org/anthology/J93-2004>>. Acesso em 10 nov. 2018.

NIVRE, J. Treebanks. In: KYTÖ, M; LÜDELING, A. (Ed.) *Corpus Linguistics: an international handbook*. Berlin: Mouton de Gruyter, 2008, p. 225-241

REIS, M.F. Temas e formas em Hércules: um estudo baseado em keywords e treebank. 2017. 107 f. Dissertação (Mestrado em Linguística e Língua Portuguesa) — Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências e Letras (Campus Araraquara), Araraquara, 2017.

ROSA, E. B. (2012). *Conflicto, resistencia e philía no Hércules furioso, de Eurípides*. 6º Coloquio Internacional, 19 al 22 de junio de 2012, La Plata, Argentina. Agón: Competencia y Cooperación. De la antigua Grecia a la Actualidad. Homenaje a Ana María González de Tobia. EN: Graciela N. Hamamé y María Cecilia Schamun, eds.. Actas. La Plata: UNLP. FAHCE. Centro de Estudios Helénicos. Disponível em <http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.4015/ev.4015.pdf> Acesso em 10 nov 2018.

SÁNCHEZ, A.; CANTOS, P. *Cumbre - Curso de Español*. Madri: SGEL, 1996.

SCOTT, M.; TRIBBLE, C. *Textual Patterns: key words and corpus analysis in language*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006. 203 p.

SINCLAIR. J.M.H. *Looking up: an account of the Cobuild project in lexical computing and the development of the Collins Cobuild English language dictionary*. Londres: Collins, 1987. 182 p.

STUBBS, M. Three concepts of keywords. In BONDI, M.; SCOTT, M. (Eds) *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins, 2010, p. 21-42.