

06

CIDEHUS' *e-Working papers*

LER UMA NARRATIVA MEDIEVAL COM OUTROS OLHOS:

EXERCÍCIOS EXPLORATÓRIOS DE ANÁLISE DE
TEXTO EM TORNO DO MANUSCRITO L DA
CRÓNICA DE 1344

José Manuel Simões



UNIVERSIDADE DE ÉVORA

Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora

CIDEHUS

06 - CIDEHUS' e-Working Papers

Título/Title: Ler uma narrativa medieval com outros olhos: exercícios exploratórios de análise de texto em torno do manuscrito *L* da *Crónica de 1344*

Autor/Author: José Manuel Simões

Data/Date: 2019

Assessor para a Série/Series Adviser: João Paulo Salvado

Comissão Editorial/Editorial Board:

Mafalda Soares da Cunha
Ana Cardoso de Matos
Maria Filomena Gonçalves
João Paulo Salvado

Por favor, citar como / Please, quote as: Simões, José Manuel – Ler uma narrativa medieval com outros olhos: exercícios exploratórios de análise de texto em torno do manuscrito *L* da *Crónica de 1344*. Évora: CIDEHUS' e-Working Papers, n.6, 2019

ISSN: 2183-8003



Este trabalho está licenciado sob a Licença Creative Commons Atribuição-SemDerivações 4.0 Internacional.

**Ler uma narrativa medieval com outros olhos:
exercícios exploratórios de análise de texto em torno do manuscrito L da
Crónica de 1344**

José Manuel Simões

PIUDHist/Universidade de Évora
zemanuelsimoes@gmail.com

Resumo

Apresentam-se aqui alguns exercícios de exploração das potencialidades de um software de análise de texto, o *Voyant Tools*, com vista ao desenvolvimento de procedimentos de interpretação quantitativa e qualitativa de uma crónica medieval portuguesa. No caso, utilizámos o manuscrito L da *Crónica de 1344*, a chamada versão *duarteana* desse texto. O objetivo deste trabalho será simultaneamente explorar de forma breve as possibilidades de leitura que os softwares deste tipo podem introduzir na análise de textos medievais, assim como olhar criticamente para a complementaridade entre os métodos quantitativos e qualitativos, no quadro de uma análise mais ampla, realizada com recurso às humanidades digitais e dedicada à construção da memória dos espaços na historiografia medieval portuguesa.

Palavras-chave: análise de texto, *Voyant Tools*, Manuscrito L da Crónica de 1344, humanidades digitais.

Abstract

Here are presented some exercises of exploration of the potentialities of a text analysis software, *Voyant Tools*, in order to develop procedures for the quantitative and qualitative interpretation of a medieval Portuguese chronicle. In this case, we used the manuscript L of the *Chronicle of 1344*, the so-called *duarteana* version of this text. The objective of this work will be simultaneously

to explore in a short way the reading possibilities that a software of this type can introduce in the analysis of medieval texts, as well as to look critically at the complementarity between quantitative and qualitative methods in the framework of a broader analysis, using digital humanities and dedicated to the construction of memory of spaces in Portuguese medieval historiography.

Keywords: text analysis, Voyant Tools, Manuscript L of Chronicle of 1344, digital humanities.

Neste trabalho apresentam-se alguns dos exercícios de exploração das potencialidades de um software de análise de texto, o *Voyant Tools* (uma aplicação *web-based* disponível em acesso aberto, que permite a leitura e interpretação de textos no âmbito das Humanidades Digitais), com vista ao desenvolvimento de procedimentos de interpretação quantitativa e qualitativa de uma crónica medieval portuguesa. No caso, foi utilizado o manuscrito *L da Crónica de 1344*, a chamada versão *duarteana* desse texto (ms. 1 Azul da Academia das Ciências de Lisboa¹), por apresentar algumas particularidades que levam a crer que terá sido copiado e alterado durante o reinado de Duarte I de Portugal. Com efeito, e embora seja basicamente uma cópia do texto português da segunda redação dessa crónica, diverge no relato do reinado de Afonso VI de Castela e Leão, omite por completo a narrativa dos reis de Portugal e substitui a parte final da *Crónica de 1344* original por uma tradução incompleta da *Crónica de Alfonso X* escrita por Fernan Sánchez de Valladolid².

Recorrendo a uma das principais, e mais imediatas, funcionalidades do software obtivemos uma lista ordenada das 59 palavras mais frequentes ao longo do texto. Para chegarmos a esse conjunto foi necessário definir restrições ao nível da contagem, nomeadamente através da identificação e

¹ Fizemos aqui uso da transcrição proposta por Marta Pedrosa, “*Reconstituição do ms. L da Crónica Geral de Espanha de 1344*”, Relatório de Estágio de Mestrado, Universidade de Lisboa, Lisboa, 2012.

² Sobre a questão ver Filipe Alves Moreira e Arthur Askins, “A Crónica de 1344 para além de Pedro de Barcelos: perspetivas recentes e novidades”, *e-Humanista*, 31 (2015), pp. 64-79.

sinalização de *stopwords* (ou *palavras vazias*), isto é, de palavras desprovidas de significado que são normalmente as mais comuns numa determinada língua, como por exemplo em português são *o, a, em, no, com, se*, entre outras. O facto de o texto se encontrar num português «arcaico», segundo a proposta de periodização de José Leite de Vasconcelos, que poderíamos também designar como português «antigo», seguindo Luís Filipe Lindley Cintra, obrigou-nos a realizar esse processo manualmente³. Assim, foram selecionadas sucessivamente para a tabela de *stopwords* aquelas palavras que surgiam entre as 60 mais frequentes e que considerámos serem vazias de significado. A lista das 166 palavras identificadas como tal encontra-se no Anexo 1 deste documento.

Uma vez findo esse processo foi-nos possível observar o conjunto das 59 palavras mais frequentes, com significado, no texto em análise. A relação destas palavras encontra-se no Anexo 2 deste mesmo documento. Não deixa de ser importante olhar para algumas das particularidades da listagem obtida, assim como para algumas virtualidades dos resultados em razão da sua pertinência para o estudo da *Crónica de 1344*. Desde logo, a palavra que será a mais comum, *rei*, surge no texto 1785 vezes, praticamente o triplo das vezes que ocorre a segunda palavra mais comum, *grande* (651 vezes), que a palavra *terra* (que surge 588 vezes), e que a terceira mais comum, *filho* (num total de 284 ocorrências). Seguem-se os termos *godos* (279 ocorrências), *mouros* (262), *senhor* (261), *cidade* (239), *Espanha* (237) e *logar* (repetida 235 vezes).

Estes dados apontam só por si, nos seus valores absolutos, em duas direções que nos parecem relativamente claras: por um lado, abundam as referências às figuras régias, que funcionarão como referencial mais significativo do texto, mas também aos sucessores destes e dos senhores dos *godos* e dos *mouros*, que ocupam um lugar bastante destacado na economia da narrativa; por outro lado, a repetição de palavras que se reportam às terras que estes governaram e à forma de as adjetivar favoravelmente como *grande* ou *grandes*, às cidades que conquistaram, perderam ou pelas quais

³ No que diz respeito à periodização e definição da língua ver Ivo Castro, *Introdução à História do Português* (Lisboa: Colibri, 2011), pp. 73-77.

batalharam, na tentativa de conquistar a *Espanha* (leia-se *Hispânia*). Se a primeira ordem de palavras nos remete para o tema que qualitativamente conseguíamos perceber como aquele que será a temática geral da crónica, a história da realeza hispânica (incluindo nesse sentido a recorrência da palavra *filho* e *filhos*, que decorrerá certamente da linearidade da diegese, isto é, de uma procura pela narração continuada e não meramente circunstancial ou enunciativa), aquele último indicador, mais focado no espaço, permite-nos compreender que também o *lugar onde*, o espaço diegético, parece ocupar um lugar central no discurso.

De facto, se atentarmos nas ligações entre as expressões mais frequentes (Figura 1), obtidas através da opção *Links* apresentada no software, podemos verificar como a *terra* parece ser a ponte entre os diferentes episódios: será pela terra que os exércitos cristãos lutam contra os *mouros*, esse inimigo plural que não aparece aqui personificado por nenhum soberano (califa, rei de Taifa ou outro) e será também na terra – a *Espanha* – que esses reis governam, herdeiros que eram dos *godos* e dos últimos reis da monarquia visigótica (*Rodrigo*, principalmente, por ter sido o último a governar uma península unificada, mas também *Ramiro*, *Ordonho* e os vários *Afonso* da monarquia asturo-leonesa).

Figura 1. Ligações entre alguns dos termos mais frequentes no manuscrito *L da Crónica de 1344*



Mas se olharmos de forma mais atenta para as correlações mais significativas da palavra *terra* com outros termos, através da ferramenta *Correlations* do *Voyant Tools* que nos permite analisar as variações da frequência entre termos através do coeficiente de correlações de Pearson (sendo identificados como valores significativos os iguais ou inferiores a 0.05), verificamos que essa expressão estabelece ligações muito fortes com outros termos mais específicos e menos frequentes (Tabela 1). É o caso das *gêtes* (significativa ao nível de $p=0.0002$) e das *obras* ($p=0.001$), simbolizando a presença humana e as edificações no espaço, mas também com as suas ações, daí a correlação com expressões como *poboou* ($p=0.001$), caso de *Espam*, antepassado mitológico do povo hispânico, que *despois que repairou a terra e as fortellezas e poboou os portos do mar con gentes que lhe veeron de Grecia e de Hercolles que poboou Tracona e outros logares*, mas também os outros heróis e povos, como os *Almonizes* (uma tribo Celta) que *pobrarom* ($p=0.001$) *muytas vyllas, asiinadamente Tolledo e fezerõna cabeça do regno e ainda Panpolona e Segonça e Cordova*. E igualmente a relação com adjetivos que caracterizam esse mesmo espaço, qualificando-o: a terra que será *comprida* ($p=0.001$), *avondada* ($p=0.001$), mas também de origens remotas ou *antigas* ($p=0.001$), como Tarragona que *he hũa das villas antigas em que ficaron fundamentos velhos e acham hy rastro das cousas antigas muy maravyllhosas* ou Beja, *hũa das antigas cidades que ha em Espanha*. Para aceder ao contexto em que estas palavras surgem foi de grande utilidade a ferramenta *Contexts* no referido software, que nos permitiu ver qualitativamente a proximidade entre palavras.

Tabela 1. Correlações mais significativas de outros termos com *terra*.

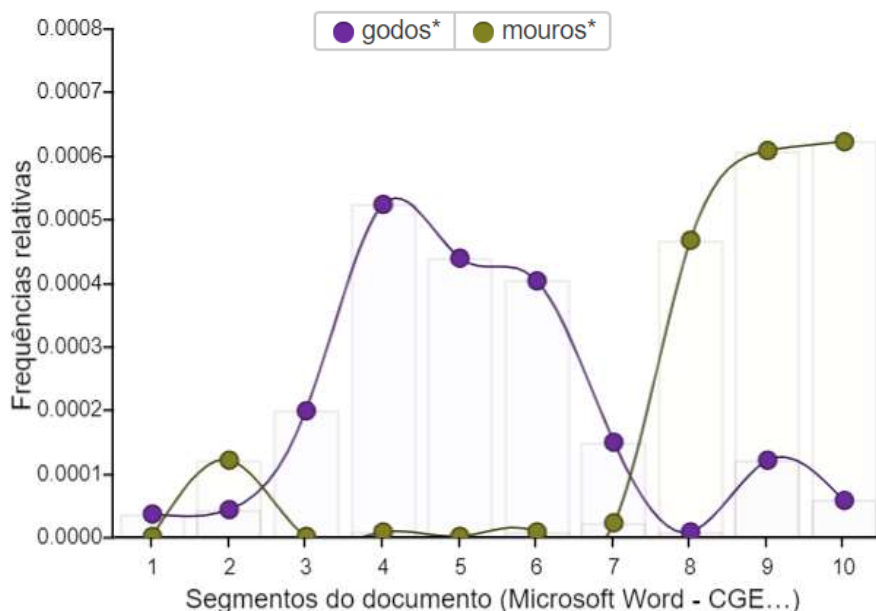
Termo 1	Termo 2	Correlação (<i>r</i>)	Significância (<i>p</i>)*
<i>marmor</i>	<i>terra</i>	0.92	0.0001
<i>gêtes</i>	<i>terra</i>	0.91	0.0002
<i>terceira</i>	<i>terra</i>	0.87	0.0009
<i>primeira</i>	<i>terra</i>	0.87	0.001
<i>obras</i>	<i>terra</i>	0.87	0.001
<i>antigas</i>	<i>terra</i>	0.87	0.001
<i>comprida</i>	<i>terra</i>	0.87	0.001
<i>avondada</i>	<i>terra</i>	0.86	0.001
<i>mete</i>	<i>terra</i>	0.86	0.001
<i>poboou</i>	<i>terra</i>	0.86	0.001
<i>corre</i>	<i>terra</i>	0.86	0.001
<i>pobrarom</i>	<i>terra</i>	0.86	0.001
<i>caliz</i>	<i>terra</i>	0.86	0.001
<i>aar</i>	<i>terra</i>	0.86	0.001
<i>acanhas</i>	<i>terra</i>	0.86	0.001
<i>alçaçova</i>	<i>terra</i>	0.86	0.001
<i>aragon</i>	<i>terra</i>	0.86	0.001
<i>covedos</i>	<i>terra</i>	0.86	0.001
<i>fala</i>	<i>terra</i>	0.86	0.001
<i>jazen</i>	<i>terra</i>	0.86	0.001

* $p \leq 0.05$

Com recurso à produção de indicadores visuais deste software gerámos ainda dois gráficos para compreendermos a simultaneidade das frequências dos termos *godos* e *mouros*, uma vez que grande parte dos capítulos nos dão conta do processo comumente designado como *Reconquista*, pese embora todas as discussões e dificuldades que este levanta, mas também entre os termos *deus*, *terra* e *batalha*, pelos mesmos motivos. Como podemos ver no primeiro dos gráficos, as frequências relativas da expressão *godos* opõe-se praticamente à expressão *mouros*, como se estas coexistissem no texto e se cruzassem poucas vezes. Com efeito, na divisão em dez partes que o gráfico

nos oferece o cruzamento entre estes dois termos (ou poderíamos antes dizer entre estes dois grupos) ocorre no momento número 7, correspondente à entrada dos *mouros* no ano de 711 na Península Ibérica e subsequente queda da monarquia visigótica, iniciando-se aí o período de domínio muçulmano. Vemos como a linha que diz respeito a esse grupo assume a partir daí uma tendência crescente, mantida até ao fim, ao passo que as referências ao passado visigótico continuarão a existir, sabemo-lo, sobretudo como discurso de recuperação da unidade hispânica e de um tempo de hegemonia cristã.

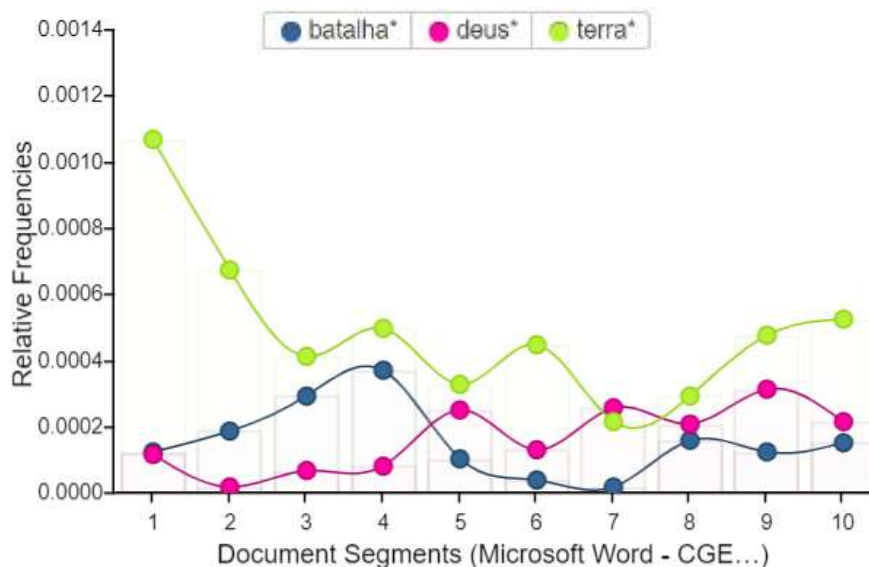
Gráfico 1. Comparação entre as frequências relativas dos termos *godos* e *mouros*.



Por outro lado, quando contrastamos as expressões *deus*, *terra* e *batalha* verificamos como as ocorrências da expressão *batalha* parecem acompanhar a evolução da expressão *terra* num primeiro momento (entre os pontos 1 e 4), abrandar num momento entre os pontos 5 e 7, para voltar depois a acompanhar a evolução desse termo juntamente com a expressão *deus*. Simbolicamente estas frequências e evoluções similares indicam-nos que, em primeiro lugar, a *terra*, como já havíamos visto parece ser omnipresente no discurso, ou seja, não será um termo acessório, mas que também apresenta variações ao longo da crónica. Além disso, quando a frequência das

ocorrências de aproxima, parece marcar a passagem de um momento de conquista e domínio da terra desde o período pré-romano até ao domínio visigótico (entre os pontos 1 e 4), altura em que as batalhas diminuem e se assiste a uma proliferação do culto cristão na Península, e depois, a partir do ponto 7, aquando da conquista muçulmana e único momento em que a expressão *deus* ultrapassa toda as outras (talvez por ser esse momento de «perda» para o Islão aquele em que os cronistas pensaram que os seus antepassados mais teriam falhado ao divino). A partir daí, a evolução das linhas *batalha* e *deus* aproxima-se, simbolizando aquele ideal de *Reconquista* de que falámos antes, uma guerra santa pelo domínio de um espaço que se supunha pertencer ao poder cristão por direito próprio.

Gráfico 2. Comparação entre as frequências relativas dos termos *batalha(s)*, *deus(es)* e *terra(s)*.



Conclusões

Podemos olhar para as potencialidades deste software com alguma satisfação pelas possibilidades de leitura que abre, mas também com outras tantas reservas. Desde logo, não dispensa uma leitura atenta do texto e mesmo uma análise qualitativa do mesmo. A sensibilidade do investigador, embora falível, poderá conduzi-lo melhor na leitura dos dados. Na verdade, os softwares deste tipo limitam-se a fornecer dados com base em quantificações numéricas que pouco ou nada nos dizem se não forem interpretadas com base na leitura cuidada das narrativas. Os chamados *mixed-methods*, que combinam dados qualitativos e quantitativos, apontam precisamente nesse sentido. E, por isso mesmo, dão-nos esse outro contraponto que se baseia em informações pragmáticas que levariam imenso tempo a contabilizar manualmente e nos permitem confirmar apontamentos intuitivos e de difícil demonstração. Por exemplo, ao preparar uma tese sobre a importância do espaço na *Crónica de 1344*, só com base na leitura e na sensibilidade podia apontar que o espaço era de facto um elemento importante e com uma presença significativa. Com a análise quantitativa que um software deste tipo nos fornece temos agora dados sólidos sobre a presença constante de indicadores que apontam nesse mesmo sentido.

Uma última nota em relação às correlações que utilizámos aqui. Estas são ferramentas tão poderosas quanto difíceis de controlar. Com efeito, a tentativa de interpretação que ensaiámos aqui trata-se somente disso mesmo, embora tenhamos tido o cuidado de utilizar um critério de verosimilhança dos dados. Muitas vezes as correlações, por serem o número de vezes que dois termos surgem em conjunto e variam também conjuntamente, sugerem apenas a proximidade textual e pouco simbolizam. É necessário aí fazer a distinção entre os termos que são efetivamente importantes e aqueles que são apenas acessórios.

Bibliografia

Castro, Ivo. *Introdução à História do Português* (Lisboa: Colibri, 2011).

Moreira, Filipe Alves, e Arthur Askins. “A Crónica de 1344 para além de Pedro de Barcelos: perspetivas recentes e novidades”. *e-Humanista*, 31 (2015), pp. 64-79.

Pedrosa, Marta. “Reconstituição do ms. L da Crónica Geral de Espanha de 1344”. Relatório de Estágio de Mestrado, Universidade de Lisboa, Lisboa, 2012.

Anexo 1

Lista de *stopwords* definida para o manuscrito *L da Crónica de 1344*.

a	cada	do	fora	muy
aa	capitulo	dom	forom	muytas
ally	com	dos	foron	muyto
anno	começou	dous	fosse	muytos
annos	como	e	foy	na
ante	con	el	guisa	nem
antre	contra	ella	guysa	no
ao	cousa	elle	ha	nom
aos	cousas	elles	he	nome
aquella	cõ	em	hi	nos
aquelle	d	en	ho	nosso
aquelles	da	enno	hua	nõ
as	das	enton	hy	nũca
assi	de	era	hũa	nẽ
assy	del	eram	hũu	o
ataa	della	esta	hũus	os
aver	delle	estas	ja	ou
avia	delles	este	lhe	outra
avya	depois	estes	lhes	outras
avyam	des	esto	logo	outro
bem	despois	eu	mais	outros
ben	deste	fazer	mal	outrossy
boa	desto	feito	mandou	ouve
boas	disse	fez	mas	parte
ca	dizer	foi	me	per

pera	quatro	seus	tanto	vos
pero	que	sobre	te	vyo
pode	se	soube	toda	ãno
por	seer	sua	todallas	ãnos
quaaes	segundo	suas	todo	ẽ
qual	seis	sy	todollos	ẽna
quando	sete	tal	todos	
quanto	seu	tam	tres	

Anexo 2

59 palavras mais frequentes no *corpus* e número de ocorrências no manuscrito.

rei (1785)	termho (154)	bispos (96)
grande (651)	reino (151)	afonso (95)
terra (588)	mar (148)	conde (95)
Espanha (526)	homões (145)	emperador (95)
filho (284)	homen (144)	reis (91)
godos (279)	jaz (137)	dō (90)
mouros (262)	villa (130)	hercolles (90)
senhor (261)	castello (127)	bispado (89)
cidade (239)	cristãaos (126)	bamba (86)
logar (235)	papa (124)	bispo (86)
deus (226)	ryo (124)	igreja (85)
reinado (203)	onde (123)	
romãaos (202)	poder (121)	
tolledo (177)	morto (116)	
morreo (176)	jhesu (112)	
batalha (174)	cristo (110)	
tempo (172)	rodrigo (110)	
grandes (169)	arcebispo (109)	
reinou (169)	terras (107)	
hoste (168)	cesar (104)	
roma (166)	gentes (104)	
padre (156)	filhos (103)	
morte (155)	estorya (102)	
cordova (154)	força (102)	