



A Musa Digital e o ‘Distant Reading’ ou os *Mistérios de Lisboa* como ‘Big Data’

PETER HANENBERG

Centro de Estudos de Comunicação e Cultura (CECC)

Universidade Católica Portuguesa

A prática de leitura sempre dependeu dos suportes em que ‘o legível’ estava à disposição. Antes dos suportes externos, as histórias só se encontravam na memória humana – frágil na sua materialidade e estável na sua essência. O manuscrito aliviou a memória humana da necessidade de “gravar” as suas narrativas (e outros textos) numa memória, cujas relações com o presente e o futuro se faziam exclusivamente pela virtualidade da comunicação circunstancial – sem garantias materiais para a sua subsistência. O manuscrito convidava a uma prática de leitura que se caracterizava pela intensidade e por uma certa exclusividade motivada pela escassez da sua disponibilidade. Portadores desta prática eram, por isso, membros de uma determinada elite tida como sábios e normalmente associada às condições da vida monástica. Os manuscritos (e antes de mais os textos sagrados) foram lidos e relidos inúmeras vezes. O ato de leitura pugnava pela intensidade e pela repetição, dando um estatuto especial a quem sabia entrar na profundidade dos textos.

A tipografia e a invenção do livro impresso vieram alterar esta prática e substituir a leitura intensa por uma leitura extensa, principalmente através da massificação e industrialização na produção do suporte de leituras. Numa época em que romances e jornais vieram a substituir a leitura intensa dos textos de uma tradição filosófica e religiosa, a leitura pugnava pela extensão e pela novidade – alargando, desta forma, tanto a quantidade de textos lidos como o número de leitores envolvidos na leitura.

É no contexto desta industrialização do livro e da massificação da leitura que surgiu uma nova necessidade de recuperar o “saber da leitura”, antigamente restrito a um número reduzido de textos sob a forma de uma leitura intensa e exegética. Olhando para o aparecimento de novos e mais textos de origens e significados variadíssimos, de verdade cada vez menos constantes e de uma singularidade interminável não bastava estender as leituras, mas transformá-las num conceito novo. Foi

assim que nasceram os Estudos Literários, herdeiros de uma tradição de leitura intensa e exegetica e portadores de um saber menos eclesiástico e escolástico e, por isso, entendido como ‘científico’. Como a Filosofia deu lugar à Filologia, a leitura intensa não só se transformava numa leitura extensa como mantinha uma pretensão de proximidade que, mais tarde, se chamou ‘Close Reading’. Os Estudos Literários posicionam-se face à leitura extensa como os monges em relação ao saber religioso: como portadores do verdadeiro sentido do texto. Os Estudos Literários compensavam, assim, a fugacidade de uma leitura sem fim.

Um dos desafios que os Estudos Literários enfrentavam desde o seu início deu-se a conhecer sob duas formas. De um lado, todos os membros da classe têm consciência (crescente) que não seriam nunca capazes de ler tudo o que seria importante ler.

Quando decidimos (sob a orientação do nosso mentor Wulf Segebrecht; 2000) oferecer aos estudantes do Instituto de Literatura Alemã na Universidade de Bamberg uma lista de leituras recomendáveis, fizemo-lo na convicção que nenhum de nós tinha, de facto, lido todos os títulos da lista – e poucos intencionavam fazê-lo, sabendo das suas limitações temporais (afinal, a vida acaba antes da lista). De outro lado, a escolha que fizemos era discutível a muitos níveis, retomando, assim, um clássico das preocupações nos Estudos Literários, o cânone. Quais são, afinal, os livros dignos de um ‘Close Reading’?

Enquanto os Estudos Literários ainda tentavam responder à questão do cânone, o desenvolvimento tecnológico proporcionou uma nova forma de leitura, baseada na comunicação digital e nos chamados ‘Big Data’. Com uma rapidez incrível e com consequências ainda sob avaliação, os media digitais propuseram-se modificar e, porventura, substituir as formas materiais, incluindo os pilares do património cultural como o museu e o livro. Transformando o museu em museu virtual e o livro em *e-book*, oferecem-se práticas de utilização e de leitura inimagináveis sem estes novos suportes. Estes suportes não só alteram novamente as formas e condições materiais da leitura, como permitem outras questões e abordagens. É, neste sentido, que Franco Moretti (2013) surgiu o termo ‘Distant Reading’, uma forma de leitura que se baseia nos textos em formatos digitais e – em última instância – vistos como ‘Big Data’.

Para melhor entender esta nova prática de leitura, as suas vantagens e os seus limites, gostaria de apresentar aqui uma leitura à distância de

um romance de Camilo Castelo Branco. Não sendo especialista em Literatura Portuguesa e ainda menos no género do romance em questão, esta minha insipiência pode ser um bom indicador para as potencialidades de uma leitura que – à partida – rejeita ou ignora qualquer tradição de ‘Close Reading’ ou abordagem histórico-contextualizante. Para demonstrar como um ‘Distant Reading’ pode provocar novas questões e responder a perguntas que não estavam ao alcance do ‘Close Reading’, vou apresentar três exemplos para os quais os meios necessários são publica e gratuitamente disponíveis.

Mistérios e cidades

Uma primeira pergunta tem a ver com o título do romance de Camilo Castelo Branco, isto é, *Mistérios de Lisboa* de 1854. O título desta obra relaciona-se com uma certa moda na sequência do sucesso dos *Mystères de Paris* de 1843 de Eugène Sue que encontrou seguidores e adaptações um pouco por toda a Europa. O que é que terá motivado este interesse? Será que os leitores estavam mais interessados em questões de “mistérios” ou em questões de “cidades”? Para tentar responder a esta pergunta, utilizo o Google Ngram-Viewer que demonstra a frequência de palavras ou frases num determinado corpus de livros¹. Comparando a frequência das palavras “city” e “mysteries” entre os anos de 1750 e de 1914, o Ngram-Viewer apresenta o seguinte resultado:

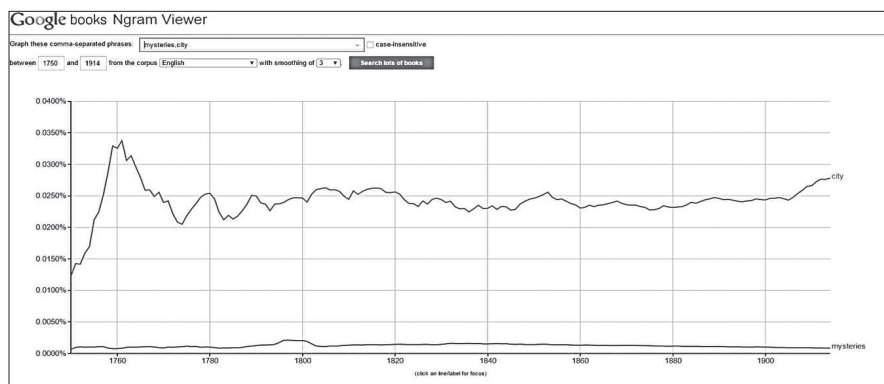


Fig. 1

¹ Vd.: <https://books.google.com/ngrams/info>

É muito óbvio que o interesse é bastante maior em questões de “cidade” do que em questões de “mistérios”. Mas o resultado também provoca uma dúvida: Será um resultado só válido para a língua inglesa? No entanto, uma segunda pesquisa, desta vez no corpus em francês, confirma o resultado, acentuando ainda a subida de frequência na palavra “ville” precisamente nos anos da publicação dos *Mistérios*:

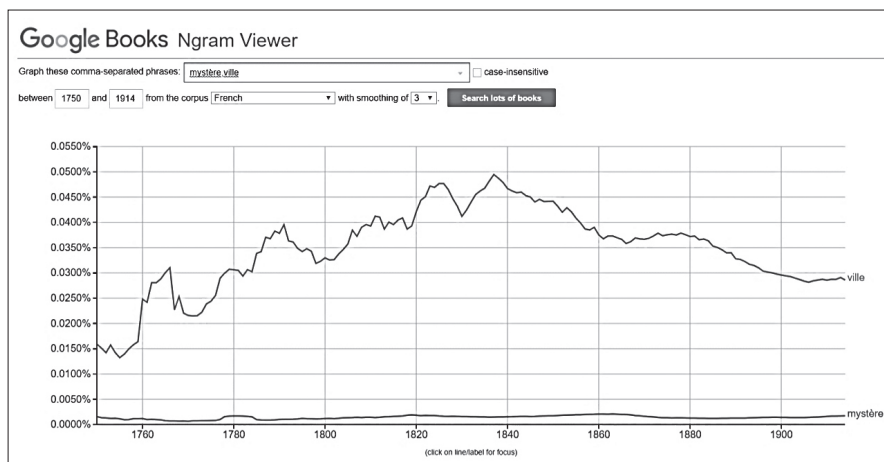


Fig. 2

Infelizmente, não é possível fazer uma pesquisa comparada com livros em português, uma vez que o *Ngram Viewer* só permite pesquisar em *corpora* de língua inglesa, francesa, alemã, espanhola, chinesa, russa, italiana e hebraica. Em todos os casos, a base destas pesquisas são os livros digitalmente disponíveis no *Google Books* – o que significa mais uma restrição à sua representatividade. De qualquer forma, o resultado é claro, mesmo se reduzimos a pesquisa aos livros ficcionais – uma diferenciação que o *Ngram Viewer* curiosamente oferece. Aqui o gráfico demonstra uma maior oscilação dos resultados:

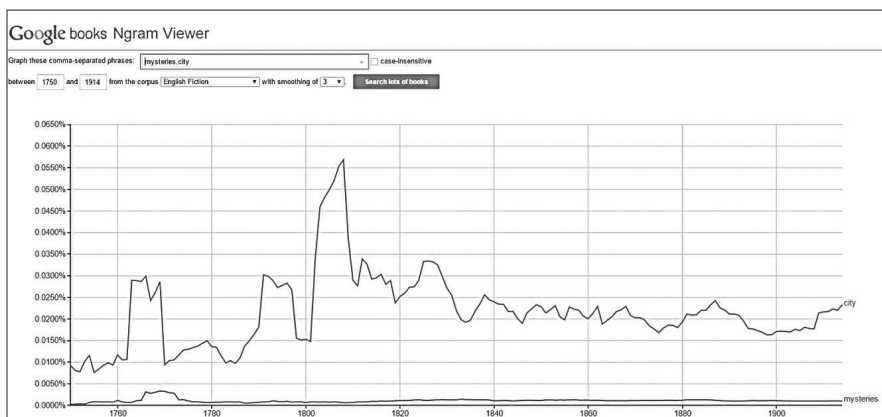


Fig. 3

Uma conclusão simples destas observações iria destacar a importância do conceito da 'cidade' para o século XIX, um conceito certamente mais relevante do que o conceito de 'mistério'. Quer-me parecer que se trata de uma conclusão que vai ao encontro do que se sabe sobre as grandes transformações sociais naquela época.

Na mesma linha, podemos agora perguntar qual a cidade que mais atenção atraiu nessa altura. Entre Londres, Paris e Lisboa não é surpreendente que Londres seja claramente a cidade mais mencionada no corpus de língua inglesa:

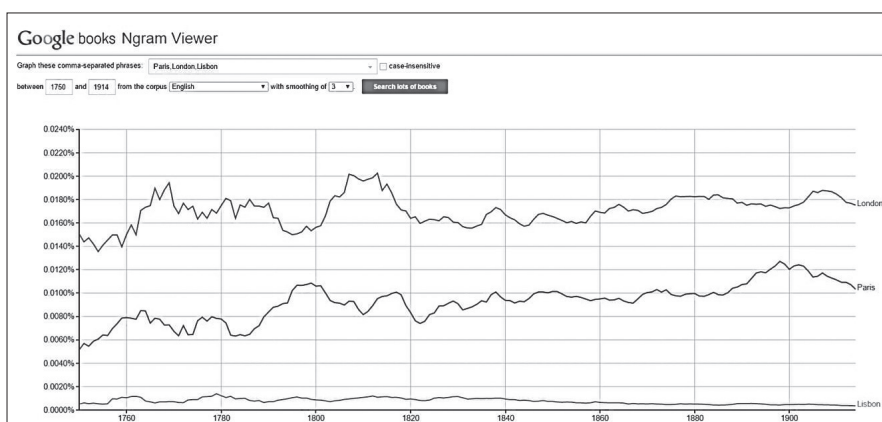


Fig. 4

Igualmente, não surpreende o resultado para o *corpus* francês, destacando – como era de esperar mas de uma forma ainda mais acentuada – a cidade de Paris:

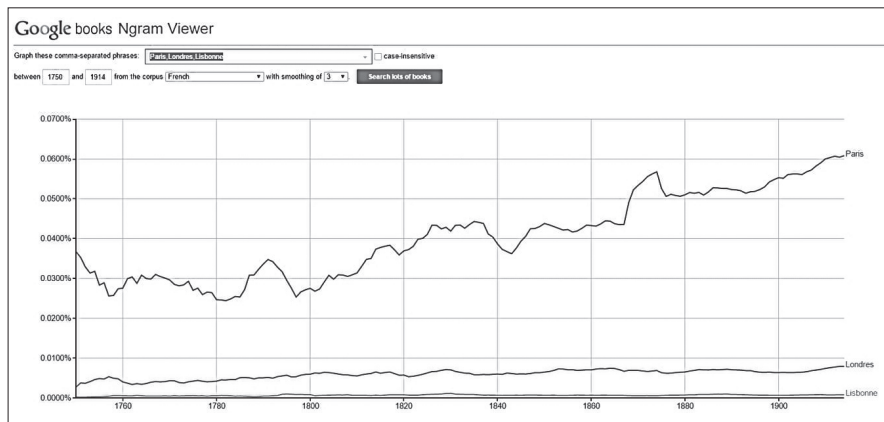


Fig. 5

De um interesse especial é, neste caso, a comparação com o *corpus* em língua alemã que confirma um interesse maior em relação a Paris comparado com Londres e Lisboa. Além disso, o gráfico testemunha como a atenção para Berlim só conseguiu ultrapassar o interesse em Paris na segunda metade do século XIX – uma ilustração simples, por certo, do nacionalismo crescente.

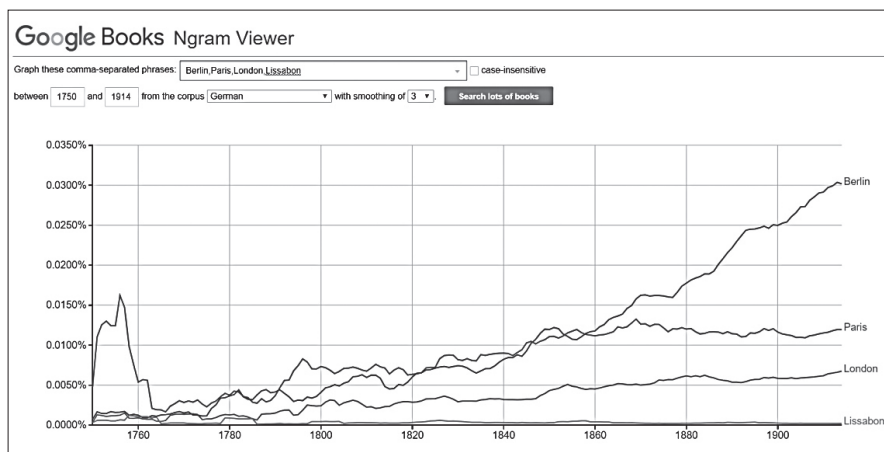


Fig. 6

É de destacar neste como nos casos anteriores um certo “pico” de interesse por Lisboa nos finais dos anos 50 do século XVII, pico este que corresponde ao debate sobre o terramoto de Lisboa.

Estas observações valem o que valem. São pura e simplesmente o espelho da frequência de uma determinada palavra num determinado *corpus*. Neste caso, a musa digital não leva a descobertas muito surpreendentes. Mesmo assim, esta pesquisa permite ‘contextualizar’ um caso como o romance *Mistérios de Lisboa* de uma forma que seria impossível sem o recurso aos ‘Big Data’. Os contextos são aqui o que se poderia chamar os com-textos, os textos que utilizam uma palavra-chave ‘espalhada’ sob um grande número de livros contemporâneos. ‘Distant Reading’ significa, neste caso, a consideração de uma quantidade muito elevada de textos sob um determinado parâmetro de comparação.

Uma análise de ‘Close Reading’ proporcionaria, provavelmente, a noção que os leitores estariam interessados em primeiro lugar nos ‘mistérios’ e a seguir na definição destes mesmos mistérios. Pelo contrário, as consultas realizadas sob a conduta da musa digital permitem concluir que a motivação com-textualizante por detrás dos romances antes é o interesse pela cidade do que o interesse pelos mistérios. Para uma sociologia da leitura no século XIX, esta observação pode ter um peso relevante. Neste sentido, ‘Distant Reading’ permite descobrir e entender o que realmente foi escrito, publicado e lido numa determinada altura – não limitando a análise a exemplos escolhidos seguindo critérios da tradição, da fama ou do cânone, muitas vezes não suficientemente assinalados como tais. Sem dúvida, os dados têm o seu charme *sui generis* – mesmo em questões de leitura.

Números e nuvens

Um outro exemplo de pesquisa sob a égide da musa digital vai-nos aproximar ao texto dos *Mistérios de Lisboa* em si – se bem que o método aplicado é muito semelhante aos exemplos anteriores, uma vez que também aqui se trata de trabalhar a frequência de palavras. São muito conhecidas as chamadas ‘nuvens de palavras’ que programas como o *Wordle* permitem produzir². Tendo obtido o texto do romance

² Vd. <http://www.wordle.net/>

Já num primeiro olhar, a nuvem de palavras identifica as personagens e os tópicos mais salientes num texto – e a comparação indica aqui, à partida, que os dois romances se situam em ambientes bem distintos. As profissões de “padre” e de “Chorineur” levam a mistérios bem distintos!

Entretanto, também salta à vista o destaque que recebe a palavra “disse” em português, correspondendo aqui à frequência da palavra “dit” em francês. Ambos os romances parecem realçar o “dizer”, evidenciam privilegiar a fala sobre a ação – o que pode já ser uma pista relevante para a sua interpretação. Por detrás de uma nuvem de palavras está simplesmente uma lista das suas frequências no texto. Estas listas, por sua vez, permitem análises interessantes como os exemplos a seguir podem comprovar.

Como a seguinte tabela demonstra, a palavra “dizer” nas suas várias conjugações aparece mais do que 1400 vezes ao longo do romance.

disse	593	0.27%	diria	13	0.01%	dissessem	5	0.00%
dizer	253	0.12%	diziam	12	0.01%	dissestes	5	0.00%
dizia	94	0.04%	ditos	12	0.01%	digas	4	0.00%
diz	93	0.04%	disser	12	0.01%	ditas	4	0.00%
diga	84	0.04%	disseste	11	0.01%	dizerem	3	0.00%
digo	54	0.02%	dissera	10	0.00%	dissemos	3	0.00%
disseram	44	0.02%	dizes	8	0.00%	diriam	3	0.00%
dito	38	0.02%	desdizer	6	0.00%	diremos	3	0.00%
dizendo	24	0.01%	dizeis	6	0.00%			
dissesse	22	0.01%	dita	6	0.00%	1469	0,66%	
dizem	20	0.01%	disserem	6	0.00%			
direi	13	0.01%	digam	5	0.00%			

Fig. 9

Se olharmos para outros verbos, a sua frequência é muito inferior – e tendo em consideração que alguns destes verbos pertencem semanticamente ao campo da fala, reforçam a noção de uma certa predominância de conversação ao longo do romance.

Ser	499	0.23%	estava	242	0.11%	parece	133	0.06%
tinha	321	0.15%	ver	204	0.09%	perguntou	123	0.06%
tem	287	0.13%	fazer	151	0.07%	respondeu	123	0.06%
tenho	267	0.12%	quero	135	0.06%			
ter	195	0.09%	dar	134	0.06%			

Fig. 10

Estes resultados necessitam obviamente de uma interpretação mais cuidadosa, cujos princípios gostaria de sugerir no fim deste artigo. De qualquer forma, e mesmo antes de qualquer interpretação, estes resultados têm uma objetividade numérica que dificilmente se deixa ignorar. Neste sentido, não deixa de ser surpreendente que a palavra “mistério” em todas as variantes – incluindo “misteriosamente” (2) – apareça só 54 vezes ao longo do texto – comparado com as cerca de 800 ocorrências da palavra “padre”.

Por outro lado, a segunda parte do título do romance indica a cidade mais mencionada ao longo do texto. Pode, entretanto, surpreender o facto de que outras cidades do país apareçam menos frequentemente do que as principais capitais da Europa, confirmando uma preferência já constatada anteriormente para livros franceses e ingleses:

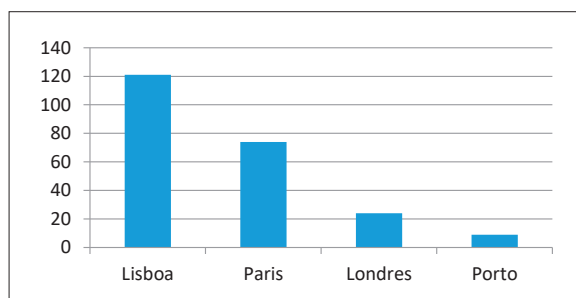


Fig. 11. Cidade nos *Mistérios de Lisboa*

Mas será este resultado realmente um espelho dos espaços mais frequentados no romance de Camilo? Será a cidade mesmo o lugar em que os Mistérios ocorrem? Uma comparação com outros espaços indicados no romance parecem motivar uma certa reserva em relação a esta conclusão:

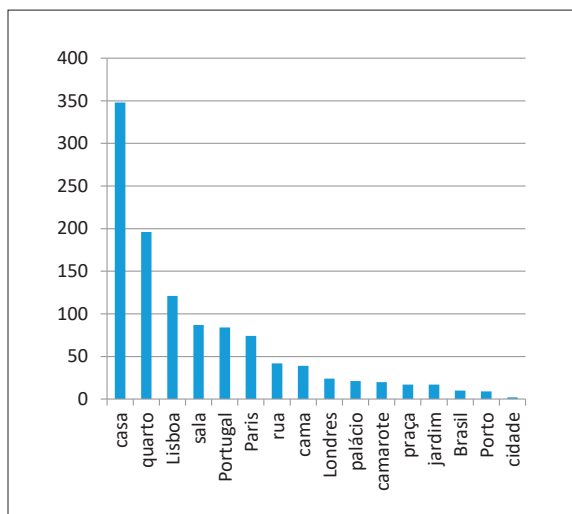


Fig. 12. Os espaços de Camilo

Espaços privados parecem ser privilegiados neste romance – mais uma vez uma observação que necessita de uma interpretação mais abrangente que cabe aos especialistas nesta matéria.

Como estes exemplos demonstram, facilmente se chega aos limites do 'Distant Reading'. Para efetuar uma interpretação conclusiva, não basta olhar para os 'Big Data'; é preciso depois penetrar estes dados com outros critérios e conceitos. Julgo que, mesmo assim, ficou claro em que sentido os dados aqui apresentados abrem pistas para uma análise mais aprofundada capaz de reconciliar distância e proximidade. Estas pistas são tanto mais válidas quanto se baseiam em metodologias e bases de dados transparentes e seguras. O que urge, neste sentido, é a constituição de uma base sólida de textos disponíveis e trabalhados em formato digital. O CRPC – *Corpus de Referência do Português Contemporâneo* desenvolvido na Universidade de Lisboa é, neste sentido, um início muito importante, principalmente porque oferece uma parametrização dos textos sem o qual uma análise diferenciada não é possível⁴. Uma dimensão mais histórica está disponível no *Corpus do Português*, desenvolvido por

⁴ Vd. <http://www.clul.ulisboa.pt/pt/23-investigacao/714-crpc-corpus-de-referencia-do-portugues-contemporaneo#cqj>

Michael Ferreira (Universidade de Georgetown) e Mark Davies (Universidade Brigham Young)⁵.

Por isso, em toda a Europa e no mundo estão, neste momento, a ser feitos esforços para oferecer à musa digital instrumentos necessários para um trabalho relevante. É de destacar aqui a *Digital Research Infrastructure for the Arts and Humanities* (DARIAH) que procura desenvolver as humanidades digitais desde a Numismática até aos Estudos Literários, oferecendo teorias, ferramentas e formações⁶. Portugal é um dos países parceiros nesta iniciativa, contando, entre outros, com a colaboração da FCT, da Fundação Calouste Gulbenkian, da Cinemateca, da Direção Geral do Património Cultural e do Teatro Nacional D. Maria. Uma reflexão mais concentrada nos Estudos Literários foi desenvolvida em 2015 numa conferência do Programa de Doutoramento «Estudos Avançados em Materialidades da Literatura», criado pela Faculdade de Letras da Universidade de Coimbra e pelo Centro de Literatura Portuguesa⁷.

O campo dos Estudos Literários Digitais ainda está em desenvolvimento. Ou talvez seja mais correto constatar que a musa digital, isto é, a utilização de ferramentas digitais e de ‘Big Data’ começa a ser uma realidade incontornável para e nos Estudos Literários no século XXI, sem isso ter necessariamente de significar uma viragem epistemológica radical. Pelo contrário, parece que a utilização destas ferramentas torna-se cada vez mais normal como um processo complementar aos interesses e questões dos Estudos Literários hoje em dia. Três exemplos breves podem ilustrar esta ‘nova normalidade’ em curso.

São João e Thomas Mann

O primeiro exemplo de uma aplicação da musa digital ao estudo literário data de um livro que foi publicado já em 1968 e que conheci no âmbito de um seminário do meu Professor Dieter Wuttke sobre “Matemática e Literatura”, já nos anos oitenta⁸. Trata-se de um livro hoje quase esquecido sob o título *Nach allen Regeln der Kunst* (*Seguindo todas*

⁵ Vd. <https://www.corpusdoportugues.org>. Para uma lista dos *corpora* em português, veja-se <http://comet.fflch.usp.br/corporaportugues>.

⁶ Vd. <https://www.dariah.eu/>

⁷ O programa da conferência está disponível em: <https://eld2015.wordpress.com/programme/>

⁸ Para este contexto, veja-se Wuttke (2002).

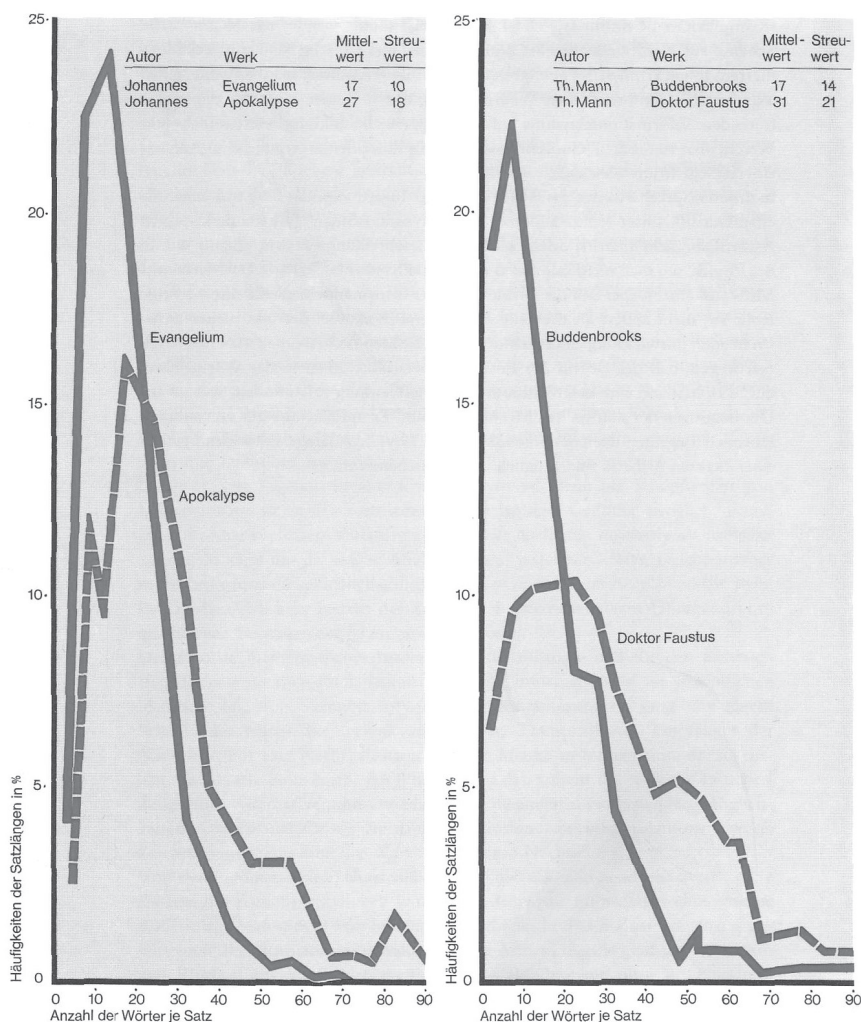


Fig. 13

as regras da arte) do físico Wilhelm Fucks. Os dois gráficos (fig. 13) resultam de uma contagem do número de palavras por frase em quatro textos.

Do lado esquerdo é uma comparação entre o *Evangelho de S. João* e o *Livro do Apocalipse* e do lado direito dos dois romances *Doktor Faustus* e *Buddenbrooks* de Thomas Mann. Os gráficos permitem – sem entrar em debates muito elaborados – duas conclusões. A primeira aborda a problemática se o *Evangelho* e o *Apocalipse* serão ambos escritos por S.

João ou não. A diferença no estilo parece indicar que não. Mas a comparação com os dois romances que são, de facto, do mesmo autor Thomas Mann, indica que um autor – pelos vistos – pode mudar de estilo, neste caso, aplicando estilos ainda mais radicalmente diferentes do que nos dois textos bíblicos. Uma segunda conclusão poderia sugerir que as frases tendem a ficar mais longas quando se distancia da Boa Nova e que se aproxima – de facto – do fim.

Hamlet e Horatio

Um outro exemplo já clássico do ‘Distant Reading’ é a aplicação da Teoria Ator-Rede à análise literária, sugerida por Franco Moretti ainda antes de dispor de ferramentas e bases digitais muito sofisticadas.

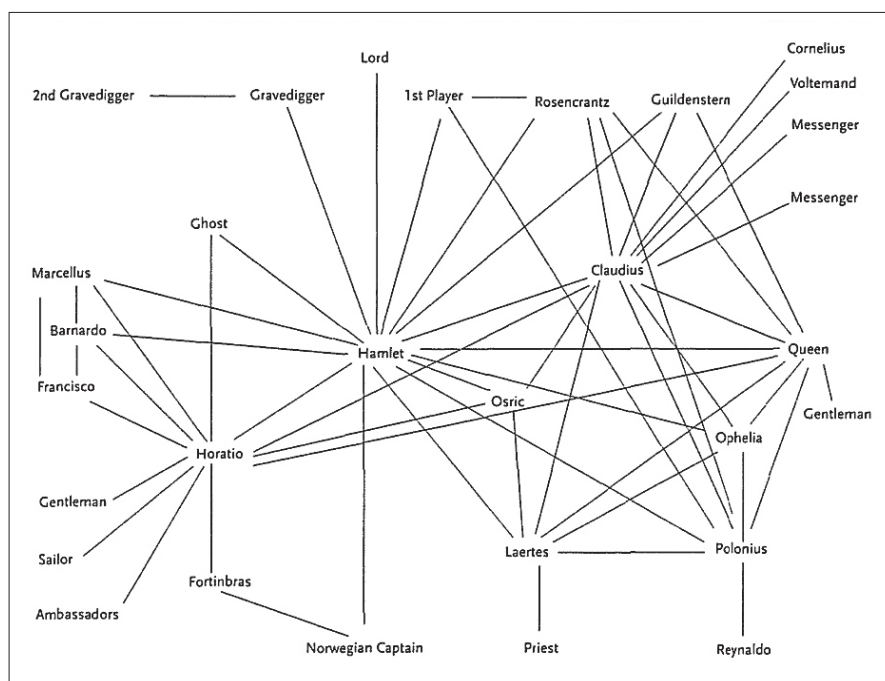


Fig. 14

A fig. 14 apresenta um mapeamento de todas as relações entre as personagens de *Hamlet* de William Shakespeare. Uma fala entre personagens motiva uma linha. Hamlet fala com Barnardo e, por isso, estão

ligados na rede, mas Hamlet não fala com Francisco – não havendo, por isso, uma linha que os conecta. Mas Francisco faz parte da rede porque Barnardo e Marcellus falam com ele.

A relevância sugestiva de uma abordagem tão distante das personagens do drama ainda se torna mais óbvia quando se olha, por exemplo, para a “rede de contactos” que Hamlet e Claudius partilham diretamente – isto é, as personagens com as quais ambos falam.

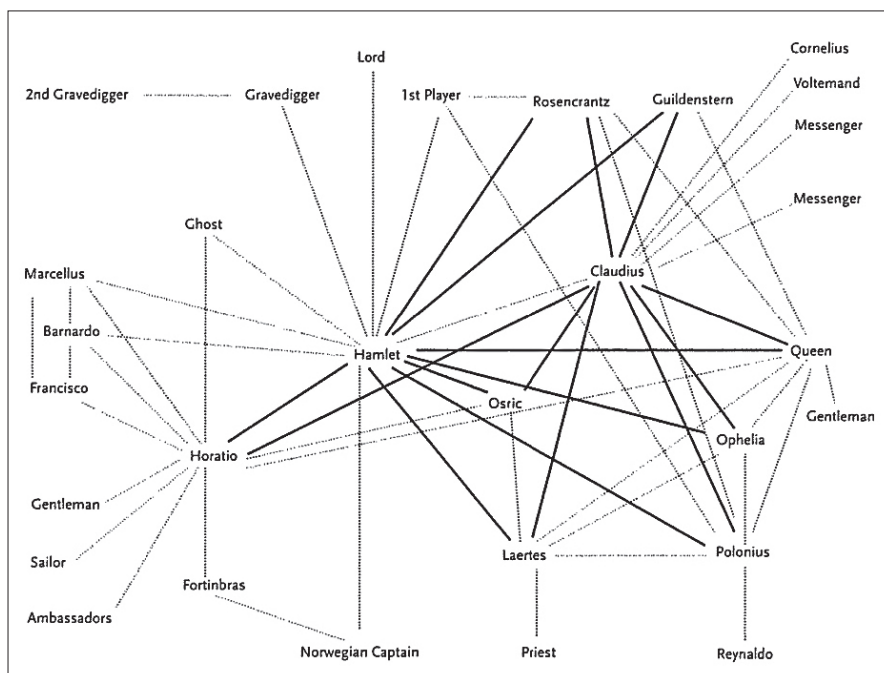


Fig. 15

Moretti chama esta rede entre Claudius e Hamlet “a região da morte”, porque, de facto, e com a exceção de Osric e Horatio, todos os protagonistas que falam tanto com Hamlet como com Claudius morrem ao longo do drama. Moretti continua a sua análise com experiências em relação à centralidade da figura de Horatio em termos estruturais: sem o Horatio o drama e a rede não podem existir, mesmo verificando que o número de falas dele é inferior às do Claudius. Horatio é tão importante para a rede porque mantém contactos exclusivos – ao contrário de Claudius cuja rede pertence a um “cluster” de relações múltiplas que poderia

existir, como tal, mesmo sem ele. O estudo de Franco Moretti é um bom exemplo para a capacidade do ‘Distant Reading’ para esclarecer questões tão difíceis como a centralidade de uma personagem supostamente secundária – um ponto crucial na poética shakespeariana e, ao mesmo tempo, numa sociedade em que a centralidade da corte começa a dar lugar a uma nova centralidade do Estado.

Mistérios e Teorias

Por fim, gostaria de voltar aos *Mistérios de Lisboa* e de propor uma abordagem que combina a distância de um modelo analítico com a proximidade ao texto. Noutro contexto, sugeri os chamados modelos culturais, estabelecidos por Bradd Shore (1998), como um instrumento útil para descrever e entender uma determinada cultura (Hanenberg 2018: 89). Por isso, tentei olhar para o início do romance de Camilo, focando, num primeiro passo, a atenção exclusivamente nos modelos culturais aplicados:

Era eu um rapaz de catorze anos, e não sabia quem era...

Vivia na companhia de um padre e de uma senhora que diziam ser irmã do padre, e de vinte rapazes, que eram meus condiscípulos. Destes, algum mais cultivado em conhecimentos do mundo perguntava-me se eu era filho do padre. E eu não sabia responder-lhe. Ora este padre parecia um homem muito virtuoso; mas nem por isso seria extraordinário eu ser seu filho.

Não o ouvira eu nunca salmear na harpa cantares de contrição; mas é rigorosamente lógico que não haja David sem harpa?! Muitas vezes senti o atrevido ímpeto de dizer-lhe: “Mestre!, perguntam-se se sois meu pai; deverei responder que não, para me deixarem?”

Nunca, porém, fiz isto, porque entendi que não me era uma das primeiras necessidades da vida saber de quem era filho.

Propenso para pensamentos elevados, erguendo os olhos ao céu, via eu, muitas vezes, voar um passarinho. E dizia comigo:

“Perguntem lá àquela criatura de Deus quem é seu pai? Como ela corta por tão alto um espaço que é todo dela! Que liberdade, e que independência! O meu espírito é como aquela andorinha! Eu tenho um mundo tão amplo para voejar como ele! Se eu puder subir, subir, subir até Deus, não terei encontrado meu pai? Isto da terra parece-me uma coisa tão pequena!...”

Seria isto uma frioleira de criança: mas eu pensava assim, e não gostava que me acordassem neste meu berço, em que eu próprio me embalava, como se assim quisesse indenizar-me de carinhos, que nunca recebera ao pé do berço da minha infância.

Quem mais vezes me inquietava nestas ociosas ilusões era o padre. Eu aborrecia o latim e a lógica e os livros e a ciência. A andorinha era o meu modelo, e a andorinha não sabia latim. “Isto de que serve”, dizia eu folheando, aborrecido, o Tito Lívio, “será necessário devorar meia existência, consumi-la num luxo de palavões estéréis, para no fim de tudo ficar o mesmo homem, sem ao menos

ter descoberto o sexto sentido do corpo humano?"

Não afirmo que fosse textualmente assim o meu raciocínio; mas, afora as palavras que a sociedade me ensinou, e que eu lhe não agradeço, a ideia era aquela.

Mas a ideia do padre era outra. Constrangia-me a estudar e especializava-me entre os meus condiscípulos. Se o carinho fosse sintoma de paternidade, nunca eu devera inspirar suspeitas de ser filho do mestre. Eu não tinha férias, nem passeios, nem prémios, nem elogios. Era um pária, um bastardo de pai, de mestre, de todo o mundo.

E, contudo, dizia-me a pobre irmã do padre, que eu era o discípulo amado do seu irmão. Explicava, ao seu modo, aquela teoria de amar, e chegava à triunfal conclusão de que, sendo a ciência o meu património, quanto

mais cultivado o recebesse das mãos do mestre, mais sagrados títulos recebia para a minha gratidão.

Custava-me a perceber isto; mas, sem grande esforço de inteligência, compreendia que era pobre.

Não me apaixonava por isso. A andorinha passava nua nas campinas do céu; e adormentava à tarde, sem granjear o alimento da manhã seguinte.

Estas razões, dadas assim àquela boa D. Antônia, faziam-na chorar. A sensível mulher chorava com qualquer coisa, e mais não conhecia ainda o mundo... ou parecia não conhecê-lo.

Mas a andorinha não remediava todas as minhas ânsias de curiosidade

Recorrendo aos modelos propostos por Shore tentei agrupar os conceitos da seguinte forma:

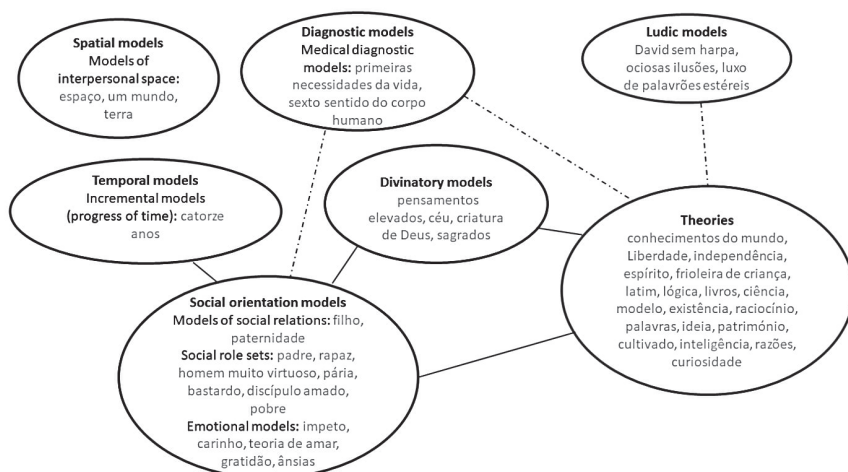


Fig. 16

O que este agrupamento demonstra à partida é uma certa predominância de teorias que vão desde a “curiosidade” até ao “conhecimento do mundo”.

Ao lado de modelos lúdicos, classificatórios, ritualísticos ou dramáticos (os últimos três ausentes nesta parte do texto), teorias fazem parte dos modelos conceituais e expressivos. “Theories”, escreve Bradd Shore, “are important cultural models that provide communities with a conceptual picture of a complex state of affairs” (Shore 1998: 1194). As múltiplas teorias a que o texto alude, agrupam-se semanticamente da forma seguinte:

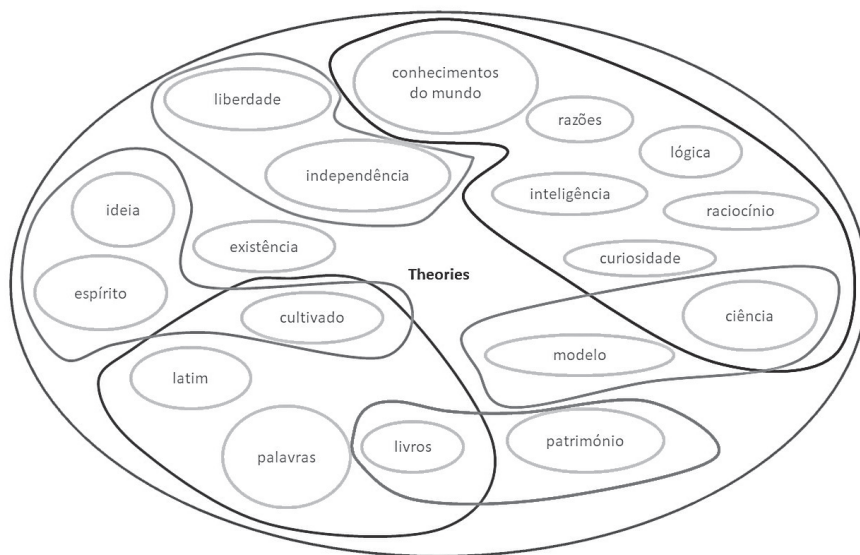


Fig. 17

Teorias ligadas a instrumentos de saber (como conhecimentos do mundo, lógica, ciência, raciocínio, inteligência, razões e curiosidade) formam um grupo, teorias de “alta cultura” (latim, livros, palavras, cultivado) outro. O espírito junta-se às ideias, a independência à liberdade, a ciência ao modelo, os livros ao património – só a existência parece ficar condenada a ser uma teoria isolada. Ou será a existência a teoria sob a qual todas as outras se subsomem? Julgo que a fig. 17 ‘traduz’ num esquema intuitivo o que serão os desafios centrais do romance de Camilo: a existência como questão.

Num segundo passo, olhei para os verbos neste início do texto que “circulam” à volta dos modelos culturais – ou que fazem circular estes modelos culturais:

Era eu um rapaz de catorze anos, e não sabia quem era...

Vivia na companhia de um padre e de uma senhora que diziam ser irmã do padre, e de vinte rapazes, que eram meus condiscípulos. Destes, algum mais cultivado em conhecimentos do mundo perguntava-me se eu era filho do padre. E eu não sabia responder-lhe. Ora este padre parecia um homem muito virtuoso; mas nem por isso seria extraordinário eu ser seu filho.

Não o ouvira eu nunca salmejar na harpa cantares de contrição; mas é rigorosamente lógico que não haja David sem harpa?! Muitas vezes senti o atrevido ímpeto de dizer-lhe: "Mestre!, perguntam-se se sois meu pai; deverei responder que não, para me deixarem?"

Nunca, porém, fiz isto, porque entendi que não me era uma das primeiras necessidades da vida saber de quem era filho.

Propenso para pensamentos elevados, erguendo os olhos ao céu, via eu, muitas vezes, voar um passarinho. E dizia comigo: "Perguntem lá àquela criatura de Deus quem é seu pai? Como ela corta por tão alto um espaço que é todo dela! Que liberdade, e que independência! O meu espírito é como aquela andorinha! Eu tenho um mundo tão amplo para voejar como ele! Se eu puder subir, subir, subir até Deus, não terei encontrado meu pai? Isto da terra parece-me uma coisa tão pequena!..."

Seria isto uma frioleira de criança: mas eu pensava assim, e não gostava que me acorrassem neste meu berço, em que eu próprio me embalava, como se assim quisesse indenizar-me de carinhos, que nunca recebera ao pé do berço da minha infância.

Quem mais vezes me inquietava nestas ociosas ilusões era o padre. Eu aborrecia o latim e a lógica e os livros e a ciência. A andorinha era o meu modelo, e a andorinha não sabia

latim. "Isto de que serve", dizia eu folheando, aborrecido, o Tito Lívio, "será necessário devorar meia existência, consumi-la num luxo de palavões estereis, para no fim de tudo ficar o mesmo homem, sem ao menos ter descoberto o sexto sentido do corpo humano?"

Não afirmo que fosse textualmente assim o meu raciocínio; mas, afora as palavras que a sociedade me ensinou, e que eu lhe não agradeço, a ideia era aquela.

Mas a ideia do padre era outra. Constrangia-me a estudar e especializava-me entre os meus condiscípulos. Se o carinho fosse sintoma de paternidade, nunca eu devera inspirar suspeitas de ser filho do mestre. Eu não tinha férias, nem passeios, nem prémios, nem elogios. Era um pária, um bastardo de pai, de mestre, de todo o mundo.

E, contudo, dizia-me a pobre irmã do padre, que eu era o discípulo amado do seu irmão. Explicava, ao seu modo, aquela teoria de amar, e chegava à triunfal conclusão de que, sendo a ciência o meu patrimônio, quanto mais cultivado o recebesse das mãos do mestre, mais sagrados títulos recebia para a minha gratidão.

Custava-me a perceber isto; mas, sem grande esforço de inteligência, compreendia que era pobre.

Não me apaixonava por isso. A andorinha passava nua nas campinas do céu; e ador-mecia à tarde, sem granjear o alimento da manhã seguinte.

Estas razões, dadas assim àquela boa D. Antônia, faziam-na chorar. A sensível mulher chorava com qualquer coisa, e mais não conhecia ainda o mundo... ou parecia não conhecê-lo.

Mas a andorinha não remediava todas as minhas ânsias de curiosidade.

Semanticamente, os verbos deixam-se agrupar da forma seguinte:

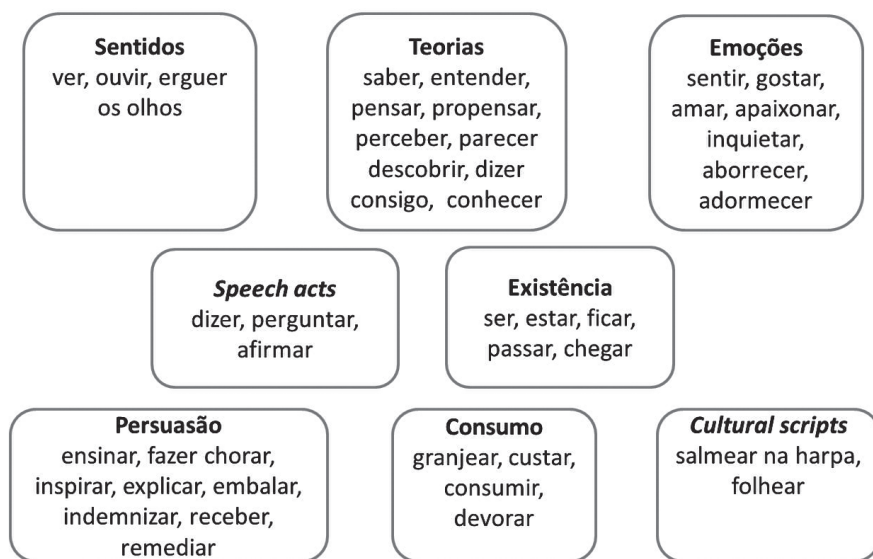


Fig. 18

Também aqui dominam palavras relacionadas com teorias para entender o mundo e outros verbos relacionados com o próprio ato de falar e da persuasão. Nesta análise olhámos para os conceitos e os verbos utilizados no início do romance de Camilo simplesmente como se fossem dados – dados num sentido abstrato. Ignorámos os contextos culturais, autorais ou pragmáticos, como se as palavras estivessem num dicionário ou numa base de dados. Julgo que os resultados que esta análise produz não deixam de ser relevantes e válidos para os Estudos Literários. O nosso ‘Distant Reading’ do romance permite-nos afirmar com uma certa confiança empírica: Os *Mistérios de Lisboa* são feitos de uma forma em que prevalece uma supremacia teórica num enredo, onde a fala domina a ação.

Contudo, esta leitura distante permite-nos reformular as questões iniciais e retomar a competição epistemológica entre as promessas da distância e as exigências da proximidade. Será que dos mistérios só ficaram os ‘Data’? Ou serão os ‘Big Data’ os verdadeiros mistérios? Afinal, o objetivo do presente texto não era a produção de um saber novo sobre o romance de Camilo de Castelo Branco. Talvez até tenha sido meramente

um pretexto para falar das transformações que a musa digital trouxe ao estudo do património literário. Mas se foi, à partida, só um pretexto, também é um facto que os dados em si são um pilar de qualquer património. O património representa uma prática humana que transforma os dados em algo significativo que merece o nosso interesse.

Bibliografia

- Fucks, Wilhelm. 1968. *Nach allen Regeln der Kunst. Diagnosen über Literatur, Musik, bildende Kunst – die Werke, ihre Autoren und Schöpfer*. Stuttgart: DVA.
- Hanenberg, Peter. 2018. *Cognitive Culture Studies*. Lisboa: Universidade Católica Editora.
- Moretti, Franco. 2013. *Distant Reading*. London, New York: Verso.
- Segebrecht, Wulf. 2000. *Was sollen Germanisten lesen?* 2., überarbeitete Aufl. Berlin: Erich Schmidt.
- Shore, Bradd. 1998. *Culture in Mind: Cognition, Culture, and the Problem of Meaning*. Oxford: Oxford University Press.
- Wuttke, Dieter. 2002. *Para uma visão holística das ciências e das artes*. Viseu: passagem.

Sítiografia:

- [Corpora em Português]: <http://comet.fflch.usp.br/corporaportugues>
- Corpus do Português: <https://www.corpusdoportugues.org>
- CRPC – Corpus de Referência do Português Contemporâneo: <http://www.clul.ulisboa.pt/pt/23-investigacao/714-crpc-corpus-de-referencia-do-portugues-contemporaneo#cqp>
- DARIAH. Digital Research Infrastructure for the Arts and Humanities: <https://www.dariah.eu/>
- Digital Literary Studies. Conferência internacional. Coimbra 2015: <https://eld2015.wordpress.com/programme/>
- Ngram Viewer: <https://books.google.com/ngrams/info>
- Poeteiro Editor Digital; São Paulo -2014: www.poeteiro.com
- Wordle: <http://www.wordle.net/>