

Algumas intersecções disciplinares na recuperação da geografia da *Peregrinação* de Fernão Mendes Pinto

Afonso Xavier Canosa Rodrigues¹

Submetido em 7/12/2016

Aceite em 6/1/2017

RESUMO:

A geografia da Peregrinação de Fernão Mendes Pinto tem sido abordada desde aproximações interdisciplinares com a intenção de reconstruir itinerários e cenários. Nenhuma obra conseguiu, de momento, referenciar toda a geografia da Peregrinação para todas as áreas. Neste artigo descreve-se mais um contributo que combina técnicas do Processamento da Linguagem Natural (PLN) com Sistemas de Informação Geográfica (SIG) para elaborar um novo índice de entidades geográficas mencionadas. A partir da análise comparada de obras especializadas, principalmente no domínio da história e da geografia histórica, o índice oferece uma georeferência exata para todas as localidades que, sendo conhecidas previamente, não apresentam contradição com a descrição dada por Pinto. Configuramos um modelo em que cada entidade é classificada segundo um tipo geográfico físico ou administrativo e ligada com um holónimo na relação Parte_de à qual pertence. A taxonomia resultante é processada numa ontologia e guardada junto de dados adicionais do corpus numa base de dados relacional. Ilustram-se os métodos e resultados com exemplos, descrevem-se os produtos finais e conclui-se ser necessária uma maior análise interdisciplinar para mais desenvolver as georeferências relativas.

PALAVRAS-CHAVE: Fernão Mendes Pinto, georeferenciamento, entidades geográficas mencionadas, geografia histórica, mineração de texto.

ABSTRACT: There have been different interdisciplinary approaches to recreating the routes and locations in Fernão Mendes Pinto's travels. Until now, none has been able to provide a georeference for all the areas involved. In this article we combine techniques from Natural Language Processing (NLP) and Geographic Information Systems (GIS) to produce an index of geographically named entities. After careful comparative analysis of specialized works, mainly from the domain of history and historical geography, our index provides an exact georeference for all locations that were previously known and show no contradiction with Pinto's description. For the rest, we build an initial model that solves a relative georeference where every single entity is assigned a geographical feature and is linked to a holonym as a part to which it belongs. The resulting taxonomy is further processed as an ontology and stored along with additional data from corpus analysis in a relational database. We illustrate methods and results with examples, describe the final products, and conclude more interdisciplinary analysis is required to further develop the relative georeferences.

KEYWORDS: Fernão Mendes Pinto, georeferencing, geographical named entities, historical geography, text mining.

¹ Investigador do CITIUS na Universidade de Santiago de Compostela onde trabalha num projeto de identificação automática de topónimos. Nesta mesma universidade estudou filologia galego-portuguesa e apresentou para defesa uma tese de doutoramento sobre o reconhecimento e referênciação de entidades geográficas mencionadas na *Peregrinação*.

Introdução

A elaboração de um novo índice de entidades geográficas mencionadas na *Peregrinação* serviu para representar cartograficamente as entidades sobre as que há unanimidade na atribuição de umas coordenadas exatas e prover uma georreferência relativa, o mais aproximada possível, para aquelas menos conhecidas ou que geram discrepância nos estudos críticos. Pretende-se que este seja o ponto de partida para um estudo mais amplo, necessariamente interdisciplinar, a contribuir para uma melhoria do conhecimento da geografia da *Peregrinação*. Neste artigo apresenta-se o processo de elaboração do índice, os resultados obtidos e as conclusões que levam a olhar o trabalho futuro como interdisciplinar. Assentado o trabalho dentro dos estudos de geografia histórica sobre a *Peregrinação*, a metodologia segue uma linha incipiente que combina a linguística de *corpus* com a análise geográfica de textos, preferentemente históricos e literários. Para além do próprio índice de entidades geográficas mencionadas, os resultados finais abrangem dois *corpora* anotados e uma base de dados que chamamos de conhecimento prévio que analisa criticamente aquelas entidades cuja situação, em termos de coordenadas, se pode dar como muito provável e serve de ponto de partida para o georreferenciamento relativo das entidades mais complicadas ou desconhecidas. Finalmente, apontam-se alguns exemplos ilustrativos e conclusões a que chegamos, fruto do trabalho sobre o índice. A mais importante, a necessidade de uma nova abordagem interdisciplinar da geografia da *Peregrinação*.

1. Trabalhos sobre a geografia da Peregrinação

A geografia da *Peregrinação* de Fernão Mendes Pinto tem dois grandes contributos exemplo da necessidade de uma abordagem interdisciplinar, se quisermos resgatar o vasto conhecimento dum das maiores e mais originais descrições da Ásia. Por um lado, o trabalho do Visconde da Lagoa, com a sua tentativa de reconstrução dos itinerários (Lagoa, 1949) e publicação do *Glossário Toponímico* (Lagoa, 1950-1953) em que ampliou o trabalho inicial sobre Pinto a um conjunto maior de textos dos descobrimentos Portugueses, obtém resultados ainda por superar ao combinar o conhecimento prático da navegação com o labor de pesquisa histórica. Por outro, as notas e índices dos volumes dirigidos por Alves (2010) constituem o índice de referência mais completo, fruto da suma de análises individuais, principalmente

de tipo histórico e documental, estruturadas, ainda que não de modo explícito, segundo uma compartimentação por áreas geográficas. É necessário citar entre as duas obras o glossário específico da *Peregrinação*, assinado por Reinaldo Varela Gomes (1983), que recolhe as contribuições encontradas nos estudos prévios à sua data de publicação.

Nenhuma obra conseguiu, até ao momento, referenciar toda a geografia da *Peregrinação*, ainda que de modo aproximado, para todas as áreas. Os diversos intentos mostram os limites do indivíduo na pretensão do resgate do conhecimento empírico de Pinto. Mascarenhas, homem de mar, perdeu-se e desistiu da reconstrução dos itinerários interiores, terra adentro. Tampouco reconheceu a influência de interrupções editoriais que pudessem alterar os itinerários, aspecto por primeira vez confirmado no trabalho de documentação do professor Almeida (1989, 2006), nem sempre devidamente reconhecido. A obra editada por Alves (2010), apesar de representar, em conjunto, um grande avanço, mostra ainda exemplos dos preconceitos, fruto de apriorismos -resultado de elaborações históricas e literárias nacionais- que assumem de modo acrítico o contexto da receção da *Peregrinação* e, por outro lado, abordam o texto de modo intencional e subjetivamente crítico. O método aparece repetido em trabalhos referidos a Pinto: o objetivo, mais que reconstruir os itinerários e feitos históricos, parece ser achar algum erro no homem de Montemor-o-Velho. Toda a aproximação deve ter uma perspetiva crítica e compreendemos ser mesmo necessário o exercício de nada ter por certo, sem se demonstrar. Exigimos o mesmo tratamento tanto para quem exercer a crítica como para as suas fontes (em que e por que se aceitam como válidas?). A exegese incapaz de questionar a natureza apriorística daquela receção a assumir a calúnia, sem questão nenhuma, e fundamentar nela o seu interesse analítico, achamos não só errada, mas desfasada: teve sentido, se calhar, uns séculos atrás, na emergência de rivalidades nacionais e de blocos, no contexto europeu e mediterrâneo, global até, e, possivelmente, haja que buscar aí a sua origem e causa. Perigo, ou apriorismos interpretativos, de que, desde logo, quem este artigo escreve não se considera livre, por muito que o pretenda. Resulta que é mais fácil ver o erro no outro que o defeito em si próprio. Dito isto em favor de quem estudou a geografia de Pinto, ainda que de um ponto de partida errado, e com a admiração e o reconhecimento do conjunto de ambas as duas obras (Lagoa, 1950-53; Alves 2010), pilares fundamentais em que se sustenta qualquer aproximação contemporânea à toponímia da *Peregrinação*.

2. Análise geográfica de textos

A combinatória de técnicas da linguística de *corpus* e SIG aparece como uma nova área com aplicações nas humanidades, especialmente por permitir a visualização das geografias dum texto (Gregory & Baron, 2013; Alves & Queiroz, 2015; Cooper, Gregory, Hardie & Rayson, 2015; DeLozier, Wing, Baldrige & Nesbit, 2016). Num procedimento tipo de análise geográfica (fig. 1), prepara-se um *corpus* em que se anotam as entidades geográficas mencionadas, quer manualmente, quer mediante a aplicação de sistemas de reconhecimento automático (nesta altura do seu desenvolvimento, frequentemente necessitados de uma revisão posterior, particularmente quando se pretender criar um padrão). O *corpus* pode levar mais anotações de tipo semântico para os termos reveladores de uma temática. Dos textos georreferenciados extraem-se dados para um SIG, permitindo a visualização cartográfica de entidades e relações derivadas da combinatória dos resultados provenientes das ocorrências no *corpus* com a análise espacial da geografia.

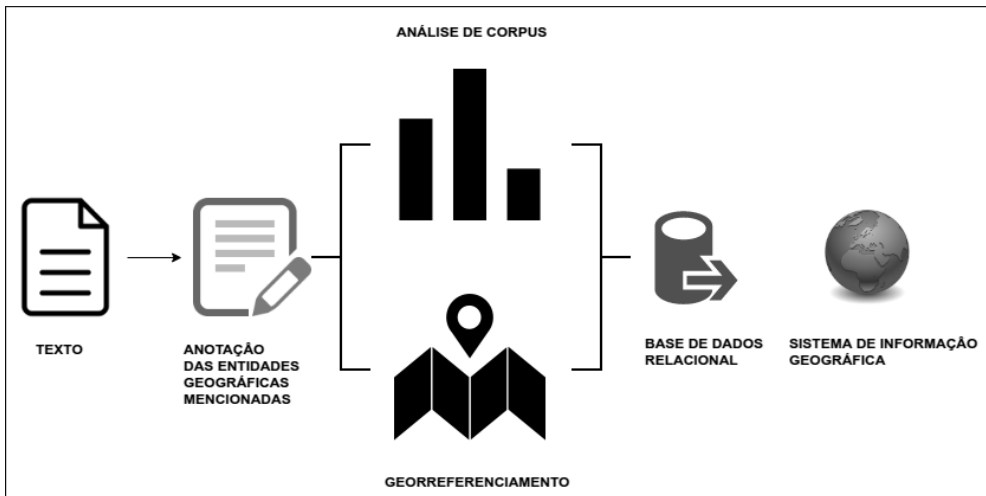


Figura 1 - Esquema do procedimento seguido para a análise das entidades geográficas mencionadas na Peregrinação.

Como exemplo de projeto da análise geográfica de textos históricos, desde a preparação e extração de dados até o seu processamento num SIG, Gregory e Hardie (2011) usaram um *corpus* reduzido de notícias de jornais de Londres em 1653-1654, anotado semanticamente, e analisaram as colocações das entidades geográficas

mencionadas para pesquisarem termos reveladores de uma temática, combinando os dados de *corpus* com análise espacial. Alves e Queiroz (2013) georreferenciam as entidades geográficas mencionadas em um corpus de 35 obras publicadas entre 1854 e 2009 para estudar a evolução do espaço urbano de Lisboa. Bruggmann e Fabrikant (2014) identificam as entidades geográficas mencionadas num *corpus* histórico da Suíça por meio de uma lista específica e analisam as frequências das suas coocorrências, assumindo que representam uma relação semântica, para visualizar graficamente e sobre um mapa a evolução histórica da relação entre topónimos. No contexto do georreferenciamento de um *corpus* literário do Lake district na Grande Bretanha (Gregory, Baron, Cooper, Hardie, Murrieta-Flores & Rayson, 2014), Donaldson, Bushell e Gregory (2016) revisam aplicações para a criação de mapas digitais nas humanidades e dos SIG, no trabalho com textos literários.

Uma ferramenta que combina o processo de anotação e representação geográfica numa só aplicação é o Edinburgh Geoparser. Tem sido aplicado para a georreferenciação de coleções de textos históricos (Grover et al., 2010) e permite a integração de listas de topónimos definidas pelo usuário. Alex, Byrne, Grover e Tobin (2015) mostram três exemplos do seu uso nas humanidades, no primeiro caso para o estudo do comércio no Império Britânico no século XIX, no segundo para o georreferenciamento de textos clássicos, e, no terceiro, para textos históricos locais. Nestes casos, a resolução das entidades geográficas faz-se com listas de entidades geográficas especializadas, o Geoparser permite a sua identificação no *corpus* e visualização cartográfica. Os resultados podem ser posteriormente reelaborados por meio de técnicas de *corpus* e análise espacial (Gregory et al., 2014). Mais específico para o Português, o Rembrandt² foi uma ferramenta especialmente concebida para o trabalho com entidades geográficas (Cardoso, 2012), de *software* aberto mas sem atualização, vai para meia década. Uma análise do desempenho de ferramentas disponíveis para o reconhecimento de entidades em Português (pessoas, locais, organizações as mais comuns) é apresentado por Amaral, Fonseca, Lopes e Oliveira (2014).

A análise textual e geográfica pode ser também aplicada numa perspetiva histórica dos acontecimentos. Numa tentativa de ligar fontes de tradições diversas, face a uma história global, Cheong, Nanetti e Philippov (2016) anotam entidades para explicarem factos históricos, usando ferramentas de extração do conhecimento, de forma a permitirem a visualização cartográfica das narrativas. Tomaszewski e

2 <http://xldb.di.fc.ul.pt/Rembrandt/> Acedido em 17 novembro de 2016.

MacEachren (2012) apresentam o funcionamento de um sistema completo para anotar entidades e conceitos num *corpus* de notícias e produzir geovisualizações com aplicação em situações de crises humanitárias. Neste tipo de análises, a entidade geográfica é mais um elemento anotado, junto com outro tipo de entidades (pessoas, organizações), considerando nas suas relações a dimensão temporal.

3. O caso da *Peregrinação* de Mendes Pinto

A análise positivista sobre a hipótese da *Peregrinação* ter sido concebida como um relatório de matéria geográfica continua a ser revista e renovada (Canosa-Rodrigues, 2013, 2014, 2015, 2016). Recolhendo estes trabalhos e aproveitando os glossários anteriores da *Peregrinação* (Gomes 1983; Alves 2010) e das Navegações Transoceânicas (Lagoa 1950-1953; Albuquerque, 1994) pareceu conveniente elaborar um novo índice de entidades geográficas mencionadas. A sua criação implicou a revisão e análise crítica de uma base ampla, de cerca de cem obras, de geografia histórica e da historiografia portuguesa (ao conjunto chamamos base documental), priorizando aquelas que têm em conta, ainda que de modo muito parcial e limitado apenas uma entidade, a geografia de Pinto.

A presente contribuição tem a particularidade de incluir os gentílicos como mais uma variante do topónimo. Também inclui os edifícios e construções, quando estas foram mencionadas de modo explícito por um nome próprio, como mais um elemento da microtoponímia. A nível metodológico, combinam-se técnicas de análise de *corpus*, PLN e SIG para agilizar os processos de anotação e compararmos criticamente as descrições oferecidas por Pinto com as geovisualizações dos objetos geográficos e georreferências achadas na base documental.

3.1 Materiais e ferramentas

Como transcrição da primeira edição da *Peregrinação* usamos um documento digital adquirido a Edições Vercial³. Advertida a Editora da falta de fólios inteiros no documento facultado e confirmadas as dificuldades para a sua reparação, procedeu-se à correção de gralhas, transcrição de quatro fólios completos a partir do original e emenda de substituições e deslocação indevida de parágrafos. Refira-se em favor do documento digital inicial o preço de tão vasta obra ter sido praticamente simbólico e a transcrição válida, não obstante os casos citados. Para a

3 <http://alfarrabio.di.uminho.pt/vercial/evercial/index.html>

comparação com a edição de 1614 usou-se o texto revisto por Elisa Lopes da Costa (Alves, 2010) e o fac-símile disponível na Biblioteca Digital Nacional de Portugal⁴.

O processamento de dados foi feito com ferramentas de software livre. No ambiente de trabalho XAMPP⁵, aplicaram-se *scripts* de elaboração própria para a anotação do *corpus*, documentação do próprio projeto, consulta e extração de resultados em diversos formatos. A análise de dados e testes foram feitos com o software estatístico R⁶. Para a criação de uma base de dados num formato padrão SIG, usou-se o QGIS⁷. A geovisualização foi obtida, basicamente, por aplicações de acesso público: o GoogleEarth⁸ (imagem satélite e geovisualização 3D), GoogleMaps⁹ (anotação sobre base cartográfica) e GeoNames¹⁰ (recuperação de toponímia e identificativo para a atribuição de coordenadas para um objeto geográfico contemporâneo).

3.2 Procedimento

3.2.1 Elaboração de um corpus da *Peregrinação*

A partir da transcrição revista do texto original da primeira edição (Pinto, 1614), anotaram-se as entidades geográficas mencionadas em todas as suas variantes, considerando tanto os topónimos quanto os gentílicos formas de um mesmo le-xema. O desenho do *corpus* foi condicionado pela sua finalidade, a extração de concordâncias para a análise crítica e obtenção de dados para trabalhos de PLN, as atividades mais comuns. No assinalado, usaram-se formatos compatíveis com padrões XML. A partir do *corpus* extraiu-se uma primeira listagem das entidades ordenadas e o texto estruturado em unidades: capítulos e orações.

3.2.2 Estudo crítico e geovisualização dos objetos geográficos

A partir da análise crítica de cerca de uma centena de obras do tipo glossários de geografia histórica, estudos das Navegações e Descobrimentos, geografia descritiva e histórica, anotaram-se criticamente as entidades geográficas da listagem ordenada obtida do *corpus*. Criou-se um documento em formato KML (notação compatível XML para a visualização sobre capas cartográficas e 3D) em que o estudo crítico é adicionado aos dados geográficos do objeto anotado.

4 <http://purl.pt>

5 https://www.apachefriends.org/pt_br/index.html

6 <https://www.r-project.org/>

7 http://www.qgis.org/pt_PT/site/

8 <http://www.google.com/intl/pt-PT/earth/index.html>

9 <https://maps.google.pt>

10 <http://www.geonames.org/>

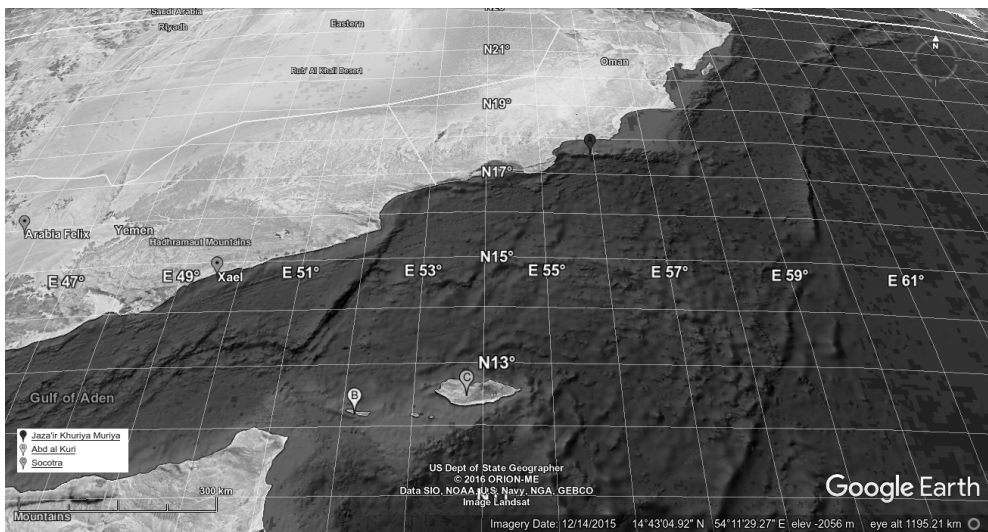


Figura 2 - Geovisualização editada no GoogleEarth para a pesquisa das entidades *Curia*, *Muria* e *Abedalcuria*. Destacados o arquipélago Khûryâ Muriya (A), a ilha de Abd al Kuri (B) seguida por duas ilhas menores e Socotra (C).

Uma vez analisadas todas as entidades, aplicamos uma distinção epistemológica: 40%, por serem conhecidas previamente (devido ao facto de serem resolvidas diretamente a partir da base documental) damos o nome de *entidades de conhecimento prévio*; 60% restantes necessitam de maior elaboração, e designamos como *entidades conhecidas por descrição*¹¹. Para as entidades georreferenciadas por conhecimento prévio, fornecemos umas coordenadas geográficas em termos de latitude e longitude, com a máxima probabilidade: existe unanimidade na atribuição das coordenadas na base documental e não achamos na geovisualização contradição com a descrição oferecida por Pinto. O resto das entidades geográficas dependem de uma descrição: têm de ser ainda mais processadas para obtermos uma referência (mais ou menos provável e mais ou menos definida).

A figura 2 mostra um exemplo de como a geovisualização do objeto favorece a descrição de Pinto para as ilhas de *Curia*, *Muria* e *Abedalcuria* como três unidades

11 Os termos baseam-se na distinção feita por B. Russell (1905, p. 479) entre *acquaintance* e *knowledge about*: “For example, we know that the centre of mass of the Solar System at a definite instant is some definite point, and we can affirm a number of propositions about it; but we have no immediate acquaintance with this point, which is only known to us by description. The distinction between acquaintance and knowledge about is the distinction between the things we have presentations of, and the things we only reach by means of denoting phrases. It often happens that we know that a certain phrase denotes unambiguously, although we have no acquaintance with what it denotes; this occurs in the above case of the centre of mass”.

geográficas independentes, a primeira Abd al-Kûrî em lat. 12.20, long. 52.22 , com mais duas ilhas em direção a Socotra (lat. 12.50, long. 54.00). Porém, na base documental, a referência apontada são as ilhas de Khûryâ Muriya em lat. 17.50, long. 56.00. Nestes casos de contradição, ainda quando houver coordenadas disponíveis, optamos por deixar a entidade como não conhecida previamente.

3.2.3 Georreferenciamento

Definimos um modelo semântico de tipo referencial, em que a solução da georreferência vem dada pela ligação entre a expressão (o topónimo) e o objeto geográfico (o referente). Na sua implementação atual usaram-se apenas duas relações, hiponímia e meronímia, para formular um conceito num esquema inspirado em modelos cognitivos (Jackendoff, 2010), com a diferença muito notável, nesta versão, de o referente ser um objeto no mundo real, com existência independente da sua conceção na mente humana. O esquema conceitual resolve uma definição que categoriza a entidade como membro de um tipo geográfico e estabelece uma relação a respeito de uma outra entidade geográfica a que a primeira pertence. O resultado final é uma definição simples, do tipo: *Bardees é barra de Goa*, em que uma entidade mencionada, *Bardees*, é classificada dentro do tipo geográfico *barra* (por sua vez integrado numa taxonomia que o inclui como acidente hidrológico) pertencente a (é_Parte_de) *Goa*, esta última resolvida por conhecimento prévio com coordenadas lat. 15.33, long. 74.08.

O *corpus* serviu de padrão para testar procedimentos automáticos de identificação, de extração de terminologia e relações, pelo momento limitados aos elementos usados na definição do conceito. No entanto, na análise durante o estudo crítico, anotamos também relações do tipo *distância* (x,y , *medida de distância*), *área próxima* (x,y), que descrevem a entidade de modo relativo. Todos os dados espaciais (coordenadas de conhecimento prévio e relações para georreferenciamentos relativos) foram integrados numa base de dados relacional.

3.3 Resultados

3.3.1 Corpus anotado de entidades geográficas mencionadas da *Peregrinação*

A anotação das entidades permite uma leitura mais atenta e ágil das concordâncias da entidade mencionada a georreferenciar. A base de dados relacional facilita o contraste entre descrição de Pinto e os dados recolhidos na base documental. A figura 3

mostra uma captura de ecrã do painel de pesquisas em que as concordâncias de cada entidade são apresentadas junto com dados do estudo crítico da base documental.

A anotação do corpus foi auxiliada por procedimentos semiautomáticos, os dados obtidos do estudo crítico e geovisualização serviram de padrão para medir a eficácia de métodos de extração de termos do domínio geográfico e aprendizagem de máquina para a extração de relações de meronímia e hiponímia. Em todos os trabalhos se obtiveram resultados significativos. Os testes tiveram a utilidade adicional de reverem e melhorarem o *corpus*. Consequentemente, uma segunda aplicação do *corpus* é servir de padrão dourado para a análise e desenvolvimento de aplicações de Processamento da Linguagem Natural, particularmente na adaptação e medida de desempenho de métodos e ferramentas sobre um texto a divergir do padrão contemporâneo.

The screenshot shows a search interface for the entity 'Bitonto'. At the top, the title 'Bitonto' is displayed in a grey box. Below it, the section '(1) Entidade nomeada' is shown. Under this section, the text 'Bitonto nos capítulos: | 4 |' and 'Nas orações: 41.' is visible. A single search result is listed, starting with '1. [41](Cap. 4) E aquelle mesmo dia fomos dormir a hum Mosteyro de officinas nobres & ricas que se dizia Satilgão, & como ao outro dia foy menham, caminhamos ao longo de hum rio mais cinco legoas, até hum lugar que se chamaua Bitonto, no qual nos agasalhamos aquella noite em hum bom Mosteyro de Religiosos que se chamaua Sao Miguel, com muyta festa & gasalhado do Prior & Sacerdotes que nelle estauão, onde nos veyo ver hum filho do Barnagais Governador deste imperio de Etyhopia, moço de idade de dezassete annos, & muyto bem desposto, acompanhado de trinta de mulas, & elle somente vinha em hum cauallo ajaezado à Portuguesa, com hum arreyo de veludo roxo franjado douro, que da India lhe mandara o Governador Nuno da Cunha auia dous annos, por hum Lopo Chanoca, que depois foy catiuo no Cayro, ao qual este Principe mandaua resgatar por hum mercador Iudeu natural de Azebibe, porem quádo este lâ chegou, o achou ja morto, de que dizem que mostrou muyto sentimento, & nos affirmou o Vasco Martins, que aly naquelle Mosteyro de São Miguel lhe mandara fazer o mais honrado saymento que elle nunca viria em sua vida, no qual se ajuntarão quatro mil Sacerdotes, a fora outra mór copia de nouiços aque elles chamão Santileus.'

Below the text, the section '(2) Objecto geográfico interpretado' is shown, followed by 'Conhecimento: D'. A small note in parentheses explains the source: '(16i) Pinto situa aqui um grande mosteiro e refere chegar desde Satilgão depois de caminhar ao longo de um rio por cinco léguas. Segundo Gianfrancesco Lusini (Encyclopaedia Aethiopica, s.v. Dabra Bizan) o topónimo a preceder a criação do mosteiro é Bizan. Identificamos como provável Däbra Bizän por descrição, contexto geográfico e homofonia. Provável: Däbra Bizän em 15° 20' N, 39° 05' E.'

At the bottom of the interface, there are two buttons: 'Ir ao indice de entidades geográficas' and 'Ir a qualquer capítulo (1 até 226):' followed by a text input field and a 'Pesquisa' button.

Figura 3 - Captura do painel de pesquisas com resultado de concordâncias obtidas do *corpus* e anotações do estudo crítico para a entidade geográfica mencionada *Bitonto*.

3.3.2 *Corpus* paralelo da Tartária

Um *corpus* paralelo alinha segmentos de texto, para o nosso caso, as orações dos capítulos relativos à Tartária da *Peregrinação*, na sua primeira edição em português (Pinto, 1614) e inglês (Pinto, 1653), em princípio com um fim contrastivo, orientado face à análise crítica. A sua importância como fonte para sucessos

centrais numa história global (assédio a Pequim, relações diplomáticas na Ásia), o facto de recolher expressões e mesmo orações para um língua pouco estudada (área linguística do Mongol) e, finalmente, o contraste entre a língua de produção do autor (português) mais a tradução para uma língua relativamente próxima (inglês), fazem do *corpus* um recurso particularmente relevante para o estudo da história da Ásia e da receção da *Peregrinação* (como são interpretadas grafias e escolha de termos na tradução das distintas línguas). A tabela 1 mostra um exemplo da sua aplicação com este fim, ao recolher as concordâncias para as formas *bada* e *badas*, animal de tiro e gado dos Tártaros (na Ásia Central há ainda hoje em dia várias espécies pouco comuns para os europeus usadas como recurso quer alimentício, quer como meio de transporte). Achamos Pinto constante no uso da mesma expressão, como acostuma fazer noutros casos em que mantém o termo original da Ásia em vez de procurar uma tradução. Em absoluto procede, portanto, uma identificação com o rinoceronte, surpreendente e injustificadamente (no que respeita a Pinto), o eleito para a tradução.

Concordâncias da forma *bada* e alinhamentos no *corpus* paralelo

1614	1653
que por terra eraõ vindos da cidade de Lançame, & de Famstir, & de Mecuy, dõde partiraõ com oitenta mil <i>badas</i> em que vinha o mantimento & toda a bagage,	which were come by land from the Cities of Luançama, Famstir, and Mecuy, with fourscore thousand Rhinocerots, that draw the waggons, wherein was all the Bagage of the Army;
Entre os hom ^{rs} que então acompanhauão o Mitaquer, estaua hum por nome Bonquinadau, homem ja de dias, & dos principais senhores do reyno, & que aly era Capitão da gente estrangeyra, & das <i>badas</i> da guarda do câpo,	Amongst those which accompanied the Mitaquer at that time, there was one named Bonquinuda, a man in years, and of the principalest Lords of the Kingdom, who in this Army commanded over the strangers and Rhinocerots, that served for the Guard of the Camp;
& elle se partio por terra com sós trezentos mil de cauallo, & vinte mil <i>badas</i> .	he himself went away by Land with three hundred thousand Horse, and twenty thousand Rhinocerots:
de que a mayor parte morrera de doença, & trezentos mil caualllos, & sessenta mil <i>badas</i> , que se comerão em dous meses & meyo que tiuerão de esterilidade.	the most of whom died of sickness, as also an hundred thousand Horses, and threescore thousand Rhinocerots, which were eaten in the space of two months and an half, wherein they wanted victual,
a qual lhes mãdou dar todas as caualgaduras que ouerão mister para sy & para os seus, & oito <i>badas</i> para leuarem o seu fato.	who for that purpose caused them to be furnished with horses for themselves, & their people, as also with eight Rhinocerots for the transportation of their bagage.
Apos estes sacerdotes, mais atras hum pequeno espaço hião quarenta carros com duas <i>badas</i> em cada carro,	After them followed forty other Chariots, each of them being drawn by two Rhinocerots,

Tabela 1 - Concordâncias para o lexema BADA e alinhamentos no *corpus* paralelo da Tartária

O *corpus* paralelo foi ainda mais usado para comparar o desempenho de sistemas de Reconhecimento de Entidades Mencionadas (REM) em modalidades estatísticas e de regras. Como produto final obtivemos um *corpus* menor em inglês com as entidades geográficas anotadas, reutilizável para trabalhos de PLN, extração de concordâncias e comparação dos topónimos e gentílicos na edição portuguesa e a sua tradução inglesa.

3.3.3 SIG de entidades de conhecimento prévio

Fruto da classificação das entidades geográficas em conhecimento prévio e descrito, e da restrição da primeira categoria apenas a aquelas entidades sobre as que achamos consenso na atribuição de coordenadas, obtemos um índice de entidades georreferenciadas que extraímos para um mapa vetorial num formato padrão, conforme a um SIG histórico (Gregory & Ell, 2007; Bol 2013). Todas as entidades, independentemente de se corresponderem com pontos, linhas ou áreas, são representadas como pontos. Esta solução também tem sido adotada no SIG histórico da China (Bol & Ge, 2005), pela dificuldade de definir umas áreas para as entidades de tipo administrativo.

Os dados da nossa base podem ser combinados e visualizados sobre capas que contribuem à sua compreensão. Na figura 4 mostramos uma representação cartográfica das entidades da área do Brama e Pegù (área central e sul da atual Birmânia) sobre uma capa vetorial de rios e oceanos facilitada por Natural Earth¹². O objetivo principal, porém, é servirem como referência no georreferenciamento das entidades do *corpus*, resolvidas no índice em termos relativos.

¹² <http://www.naturalearthdata.com/>

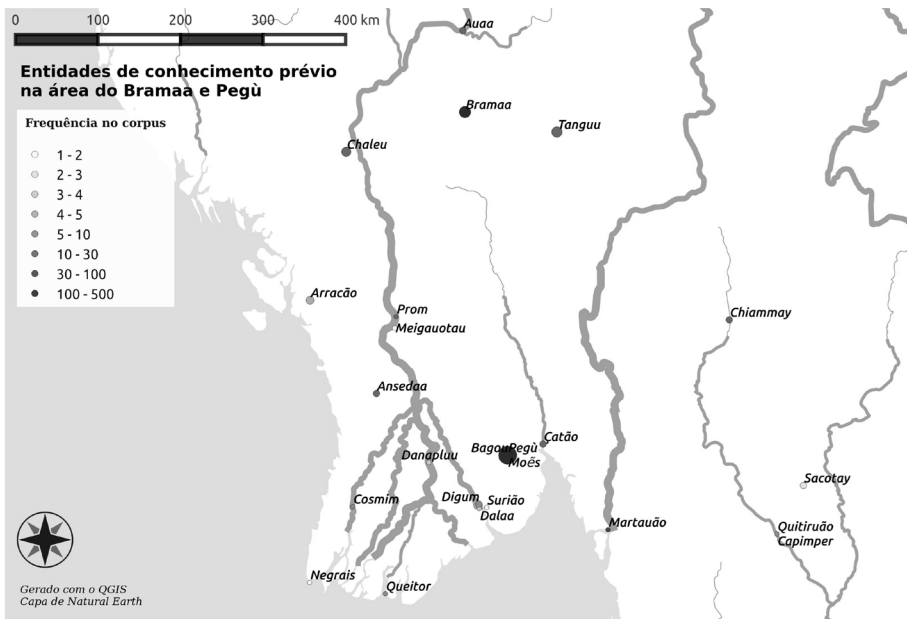


Figura 4 - Mapa de entidades geográficas mencionadas no *corpus*, dadas como de conhecimento prévio para a área do Bramaã e Pegù.

3.3.4 Índice de entidades geográficas mencionadas

O mais completo índice de entidades geográficas da *Peregrinação* de que tenhamos conhecimento é o publicado no trabalho coletivo editado por Alves (2010). Nesta obra, cada entidade aparece representada pelo seu nome padrão contemporâneo, com uma referência a um tipo e uma localização numa área (semelhante à definição usada neste texto). O nosso índice amplia o trabalho anterior porquanto: 1) organiza todas as entidades, incluindo topónimos e gentílicos, a partir da expressão do *corpus*; 2) referência as ocorrências de modo a serem recuperáveis em qualquer edição da obra (por capítulos); 3) inclui o nome padrão contemporâneo e as coordenadas geográficas para as entidades para as que há consenso no seu georreferenciamento; e 4) adiciona relações de distância e proximidade para a georreferência relativa. Neste estágio de desenvolvimento, as entidades e relações da definição foram ordenadas numa ontologia. Os objetos são organizados conforme a uma taxonomia de tipos geográficos, em que incluímos as unidades convencionais da toponímia (cidade, rio, ilha,) e mais as construções, quando tiverem um nome

próprio expresso de modo explícito no *corpus*, por ex. *Singrachirau* é a expressão usada por Pinto para se referir à Grande Muralha¹³ da China, como tal é incluído com tipo *muro* pertencente à classe *construções*. A ontologia estrutura as entidades e classes com base em duas relações básicas e as suas inversas: hiponímia (pertença a classe) e hiperonímia (a classe contém o tipo e a instância); a segunda a meronímia (uma entidade é parte de outra) e holonímia (uma entidade contém outra). Os dados obtidos do ordenamento das entidades na ontologia são integrados com os referidos ao *corpus* e relações espaciais adicionais na base de dados relacional. Para a elaboração do índice recolhemos:

- * Entrada e variantes: todas as variantes associadas a um lexema. Como forma representativa do lexema (entrada) escolhemos o topónimo com maior índice de frequência.
- * Ocorrências: Número de ocorrências para cada capítulo em que aparece a variante.
- * Relações da ontologia: o tipo geográfico e a entidade maior em que incluímos a entidade.
- * Outras relações: anotações sistematizadas que contribuem para a georreferenciação da entidade em termos relativos a uma outra.
- * Objeto contemporâneo: quando a entidade seja dada como de conhecimento prévio, incluímos dados recuperados de GeoNames para o nome em inglês, entidade administrativa superior, se a houver, continente e coordenadas.

13 Na língua falada atualmente na Mongólia, próxima, mas não necessariamente a mesma que a transcrita por Pinto, a Muralha da China é chamada hoje em dia цагаан хэрэм, onde *tsagaan* é branco, e *kherem* /xerem/ muro. O *ch* em *Singrachirau* deve ser lido como uma velar fricativa, fonema inexistente no português. Consideramos que Pinto transliterou na grafia mais próxima no Português, a oclusiva velar surda /k/, neste caso o uso do *ch* possivelmente justificado pelo modelo do Grego e Latim. A terminação *-au* precisa de uma maior atenção e deixamos aqui apenas proposta como hipótese. Nos topónimos e antropónimos (cf. Mongol *khan* e as formas *Abicau*, *Tuymicão* e *Moncamicau* em Pinto), ao não ter um referente acústico sobre o que homologar a terminação, a intervenção editorial não é homogénea. Para o seu estudo mais atento convém revisar como são transcritas estas formas na versão em inglês (Cf. *Tuymicão* vs. *Tuymicoa* e *Tuymican* que interpretamos como “o khan dos *tuymed*”), comparar mais casos em que a forma na língua asiática é recuperável (ex. *Abicau nilancor*, traduzido por Pinto como *Deus da saúde dos reis*) e considerar a padronização editorial sobre o manuscrito original (ex. homologação de terminações *-ANU*, *-ANE*, *-ONE*, *-ANT*) com explicações internas à língua, tais como a solução nasal destas terminações ter distribuição dialetal no português continental, a sua representação e pronúncia comum atual fruto da normatização. Como hipóteses externas, a fonética da perceção, fonemas distintos nas línguas asiáticas serem percebidos por assimilação a um outro, por analogia acústica, dentro do esquema fonológico do recetor. Agradecemos ao professor José António Souto Cabo a conversa sobre estas hipóteses da perspetiva histórica do Português e temos de adicionar como mais uma hipótese, proposta pelo professor, um possível erro de interpretação do editor da grafia *n* por *u*.

Exemplo de entidades mencionadas no índice

<p>Queitor. <i>Queitor</i> 156(1), 157(1), 158(1). Rio. Parte de: Índico Oriental. Outras relações: #área#@Auaa@; #área#@Prom@; #área#@Guampanoo@.</p> <p>Irrawaddy River, Myanmar [Burma] (AS). Lat. 15.83333, long. 95.1.</p>
<p>Meleitay. <i>Meleitais</i> 195(1). <i>Meleitay</i> 150(1), 155(1), 156(5), 157(1), 194(1). <i>Meleytay</i> 128(1), 156(1), 157(2). Cidade (fortaleza, rio). Parte de: Chaleu. Outras relações: #18 léguas#@Prom@; #área#@Pumfileu@.</p>

Tabela 2 - Exemplo de duas entidades mencionadas no índice.

Na tabela 2 mostramos um exemplo de entradas recolhidas no índice. A aplicação do mapa da figura 4 mostra a área para a sua georreferência relativa. A primeira entidade, *Queitor*, é o rio que flui por *Auaa* e *Prom*, o qual permite a sua localização sem ambiguidade: na atribuição de coordenadas, o rio é situado como um ponto na sua boca (todas as entidades, ainda que sendo linhas ou polígonos, têm valor inicial de ponto no SIG). O segundo exemplo, *Meleitay*, aparece no índice com uma relação que o situa a 18 léguas de *Prom*: a distância provê uma georreferência aproximada (o dado da direção, neste caso rio acima, não recolhido no índice, é recuperável nas concordâncias do *corpus*), não obstante, neste estado do nosso trabalho, não resolvemos em termos de coordenadas, a não ser quando houver mais elementos (na toponímia, geografia física e documentais) a afastar a ambiguidade, com alta probabilidade acerca de um único objeto geográfico.

4. Para além da geografia e o PLN

O resultados anteriores mostram como a combinação de técnicas provenientes do PLN com a análise em SIG podem contribuir para uma atualização e melhor compreensão da geografia da *Peregrinação*. Supõe apenas um ponto de partida. Na elaboração do índice e georreferenciamento adotou-se um critério conservador, oferecendo coordenadas apenas para aquelas entidades cujo conhecimento aparece como óbvio e não oferece contradição nos estudos prévios. Não obstante, o estudo crítico permite desenvolver as georreferências inicialmente relativas, ao ponto de obter coordenadas geográficas em termos de alta probabilidade e reduzir o conjunto de entidades denotadas como potenciais referentes, dentro da área de influência de uma outra entidade cujas coordenadas são bem conhecidas. Embora, para chegarmos a este nível de resolução para o conjunto, convinha a interdisciplinaridade ser ampliada.

Particularmente relevantes aparecem os conhecimentos náuticos na compreensão das rotas e direção mais provável na navegação, velocidade das distintas embarcações, marés e navegabilidade em áreas fluviais, recursos todos usados nas descrições de localidades na *Peregrinação*. Também a topografia, quando tivermos duas entidades relacionadas e quisermos resolver pontos intermédios ou um destino, dada uma direção e ponto de partida. A astronomia mesmo. Sirva de exemplo o caso prático da inspeção dum ponto obtido a partir de uma georreferência relativa dada em léguas (caso de *Meleitay* na tabela 2 acima). Para obtermos uma distância de referência precisamos primeiro calcular o valor da légua. Resolvemos a partir da descrição feita por Pinto em que determina a posição em graus e minutos pela altura do Sol¹⁴: considerando a légua ser 17.5 intervalos do grau (Albuquerque, 1987, p. 49), a longitude aproximada da circunferência da Terra 40,000,000 m, 360° a circunferência, temos $40,000,000 \text{ m} / 360^\circ = 111,111.1$, a dividir pelo valor da légua $111,111.1 / 17.5 = 6,349 \text{ m}$ tem uma légua. A resolução do problema passa agora pela toponímia. Quando tivermos uma distância concreta numa área muito reduzida, necessitamos de um topónimo para recuperar a georreferência (*reverse geocoding*), nem sempre transcrito a caracteres ocidentais nem transparente (dificuldades gráficas e fonéticas), nem disponível na cartografia e recursos de geovisualização mesmo (particularmente na microtoponímia e construções). O conhecimento da toponímia e línguas locais são a chave neste caso. A combinação de fontes documentais, preferentemente com uma orientação para a história global, é também um modo de estabelecermos ligações quando acharmos novas evidências sobre os espaços e personagens para a solução definitiva de gentílicos e topónimos, até à data apenas conhecidos pela descrição dada por Pinto.

Finalmente, a cartografia histórica ficou como a grande área sem tratar no nosso trabalho. Nas aproximações limitadas e generalistas que tivemos, contribuímos para ilustrar as áreas e pontos de referência como evidência para explicar o porquê do desconhecimento dos topónimos mais difíceis da *Peregrinação*. Com presenças e ausências. Assim os topónimos representados nas zonas costeiras tendem a confirmar os dados de Pinto. Nas ausências, a visualização de apenas os topóni-

14 Na *Peregrinação*, Cap. 224: “& atrauessando daquy com ventos de moução tendente, continuamos nosso caminho por espaço de sete dias, no fim dos quais o tempo com a conjunção da lúa noua nos saltou ao Sul, & ameaçandonos com chuueyros & mostras de inuerno, veyo em tamanho crescimento, que nos foy forçado arribar em fim de roda com a proa ao rumo de Nornordeste por mar incognito, & nunca nauegado de nação nenhũa, sem sabermos por onde hiamos, entregues de todo ao arbitrio da fortuna & do tpo, com hũa tão braua & tão excessiua tormenta, qual os homs nunca imaginaraõ, que nos durou cinco dias: & como em todos elles nunca vimos o Sol para o piloto saber porque altura caminhaua, só pela sua fraca estimatiua, sem conta de graos nem de minutos, pouco mais ou menos foy demandar a paragem das ilhas dos Papuaas, Selebres, & Mindanous que distauaõ daly seiscentas legoas” (Pinto, 1614, fól. 284 reto).

mos de conhecimento prévio com uma ferramenta SIG mostra os pontos de ilhas e portos na Insulíndia como mais uma evidência a favor das teses que explicam uma ausência intencional da Austrália na cartografia da época (Cortesão 1981, pp. 1282-1307)¹⁵. O desconhecimento das zonas interiores da Ásia e aumento na resolução das zonas mais orientais (Marques, 1996), conforme avançamos no período vital do autor de Montemor-o-Velho, explicam, em parte, a grande consideração que teve Pinto como fonte de conhecimento geográfico de volta a Portugal.

Considerações finais

A combinatória de disciplinas facilita o estudo geográfico da *Peregrinação*. No processo de elaboração do *corpus* aplicamos uma série de métodos, resultado da convergência de campos na compreensão e engenharia da linguagem natural, em que incluímos a linguística, a lógica e trabalhos de Processamento da Linguagem Natural, ramo este atendido na utilização de dados e Inteligência Artificial. O georreferenciamento implica, principalmente, a Geografia.

A análise espacial da Geografia, em ambiente SIG, combinada com técnicas de *corpus* sobre documentos, particularmente relevantes para nós, os históricos e literários, está a facilitar uma nova aproximação aos textos para a caracterização geográfica de tópicos e georreferenciamento dos sucessos e entidades. O ponto de partida foram, porém, os estudos de geografia histórica, atlas e glossários toponímicos.

Com a combinação da análise crítica da base documental e a geovisualização dos objetos conseguimos georreferenciar 40% das entidades em função de umas coordenadas a que lhe damos a maior probabilidade. Deixamos 60% restantes descritos em termos de uma georreferência relativa a partir de um tipo geográfico e uma outra entidade com que se relaciona.

Anotamos ainda mais outras relações espaciais, sistematizadas para prover distâncias e pontos de proximidade. É possível aumentar o nível de resolução (redução da área de pesquisa e número de candidatos) mediante a análise crítica da representação cartográfica destas relações, do *corpus* e da base documental.

Entidades geográficas para as quais sim temos conhecimento prévio foram deixadas apenas como descritas (sem atribuição de coordenadas) quando houver algum tipo de contradição nos resultados da análise documental. Uma solução para a

¹⁵ Cf. (Albuquerque, 1990, pp. 87-103) para uma revisão crítica da hipótese com os dados apresentados na altura.

atribuição de uma georreferência exata para estas entidades passa pelo uso duma escala de probabilidades. Em todos os casos, achamos que é possível aumentar o nível de probabilidade por acumulação de evidências. Neste labor, os recursos por nós usados deveram ser ampliados a mais disciplinas. A solução última de mais um topónimo resolvido da *Peregrinação* é um fruto de um trabalho interdisciplinar.

Referências bibliográficas

- Albuquerque, L. (Dir.). (1994). *Dicionário de História dos Descobrimentos Portugueses* (2 vols.). Lisboa: Caminho.
- Albuquerque, L. (1990). *Dúvidas e certezas na História dos Descobrimentos Portugueses* (2ª ed.). Lisboa: Vega.
- Albuquerque, L. (1987). *As navegações e a sua projecção na ciência e na cultura*. Lisboa: Gradiva.
- Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh Geoparser for Historical Georeferencing. *International Journal of Humanities and Arts Computing*, 9(1), 15-35. Acedido em 9 janeiro de 2017, em <http://www.eupublishing.com/doi/full/10.3366/ijhac.2015.0136>
- Almeida, F. A. (2006). *Fernão Mendes Pinto. Um aventureiro Português no Extremo Oriente*. Almada: Câmara Municipal da Almada.
- Almeida, F. A. (1989). Alguma biografia de Fernão Mendes Pinto. In F. M. Pinto, *Peregrinação & Cartas* (vol. 2, pp. 1075-1090). Lisboa: Fernando Ribeiro de Mello / Edições Afrodite.
- Alves, D., & Queiroz, A. I. (2015). Exploring literary landscapes: From texts to spatiotemporal analysis through collaborative work and GIS. *International Journal of Humanities and Arts Computing*, 9(1), 57-73. Disponível em: <http://www.eupublishing.com/doi/full/10.3366/ijhac.2015.0138>
- Alves, D., & Queiroz, A. I. (2013). Studying urban space and literary representations using GIS: Lisbon, Portugal, 1852-2009. *Social Science History*, 37(04), 457-481.
- Alves, J. S. (Dir.). (2010). *Fernão Mendes Pinto and the Peregrinação* (4 vols.). Lisbon: Fundação Oriente.

- Amaral, D. O., Fonseca, E. B., Lopes, L., & Vieira, R. (2014). Comparative Analysis of Portuguese Named Entities Recognition Tools. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014, Reykjavik* (pp. 2554-2558). European Language Resources Association (ELRA). Acedido em 9 janeiro de 2017, em http://www.lrec-conf.org/proceedings/lrec2014/pdf/513_Paper.pdf
- Bol, P. K. (2013). On the Cyberinfrastructure for GIS-Enabled Historiography: Space-Time Integration in Geography and GIScience. *Annals of the Association of American Geographers*, 103(5), 1087-1092. Acedido em 9 janeiro de 2017, em http://dash.harvard.edu/bitstream/handle/1/13366679/on_the_cyberinfrastructure_for_gis-enabled_historiography.pdf?sequence=1
- Bol, P., & Ge, J. (2005). China Historical GIS. *Historical Geography*, 33, 150-152.
- Bruggmann, A., & Fabrikant, S. I. (2014). Spatializing time in a history text corpus. In *Proceedings of the 8th International Conference on Geographic Information Science, GIScience (extended abstracts)*. Acedido em 9 janeiro de 2017, em http://www.geo.uzh.ch/~sara/pubs/bruggmann_fabs_gisci14.pdf
- Canosa-Rodrigues, A. X. (2016). *Asia in the 16th century*. GIS database. Acedido em 28 de outubro de 2016, em <http://worldmap.harvard.edu/maps/8478>
- Canosa-Rodrigues, A. X. (2015). Estudo, seleção e classificação de entidades geográficas para um mapa global da “Peregrinação”. *Boletim da Academia Galega da Língua Portuguesa*, 8.
- Canosa-Rodrigues, A. X. (2014). *The geography of early transoceanic navigations. A positivist approach to place names in The Travels of Mendes Pinto. Slides for MIU Colloquium Series, Fall 2014. Ulaanbaatar (Mongolia)*. Documento digital. Acedido em 17 novembro de 2016, em https://www.academia.edu/29936526/The_geography_of_early_transoceanic_navigations_A_positivist_approach_to_place_names_in_The_Travels_of_Mendes_Pinto_Slides_from_a_presentation_at_MIU_Colloquim_Series
- Canosa-Rodrigues, A. X. (2013). Notas biográficas e estudo das referências documentais de Fernão Mendes Pinto. *Veredas*, 20, 9-34. Acedido em 9 janeiro de 2017, em <http://ojs.lusitanistasail.org/index.php/Veredas/article/view/2>
- Cardoso, N. (2012). Rembrandt - a named-entity recognition framework. In *LREC 2012, Eighth International Conference on Language Resources and Evaluation* (pp. 1240-1243). Acedido em 9 janeiro de 2017, em http://www.lrec-conf.org/proceedings/lrec2012/pdf/409_Paper.pdf

- Cheong, S. A., Nanetti, A., & Philippov, M. (2016). Digital Maps and Automatic Narratives for the Interactive Global Histories. *The Asian review of World Histories*, 4(1), 83-123. Acedido em 9 janeiro de 2017, em http://www.thearwh.org/journal/ARWH_4_1_Articles/arwh_4-1_cheong,%20nanetti,%20philippov.pdf
- Cooper, D.C., Gregory, I.N., Hardie, A., & Rayson, P. (2015). Spatializing and Analyzing Digital Texts: Corpora, GIS, and Places. In D. J. Bodenhamer, J. Corrigan & T. M. Harris (Eds.), *Deep Maps and Spatial Narratives*. Bloomington, Ind.: Indiana University Press. Acedido em 9 janeiro de 2017, em <http://espace.mmu.ac.uk/579357/2/Spatializing%20and%20Analyzing%20Digital%20Texts.pdf>
- Cortesão, J. (1981). *Os Descobrimentos Portugueses* (3ª ed., 6 vols.). Lisboa: Livros Horizonte.
- DeLozier, G., Wing, B., Baldrige, J., & Nesbit, S. (2016). Creating a Novel Geolocation Corpus from Historical Texts. *LAW X*, 188. Acedido em 9 janeiro de 2017, em <https://www.aclweb.org/anthology/W/W16/W16-17.pdf#page=200>
- Donaldson, C. E., Bushell, S. C., Gregory, I. N., Rayson, P. E., & Taylor, J. E. (2016). Digital literary geography and the difficulties of locating ‘Redgauntlet Country’. *Studies in Scottish Literature*, 42(2), 174-183. Acedido em 9 janeiro de 2017, em <http://scholarcommons.sc.edu/ssl/vol42/iss2/5/>
- Gomes, R. V. (1983). Roteiro Geográfico através da Peregrinação. In A. M. Flores, R. Varela Gomes & R. H. Pereira de Sousa, *Fernão Mendes Pinto: subsídios para a sua bibliografia* (pp.81-167). Almada: Câmara Municipal da Almada.
- Gregory, I. N., & Baron, A. (2013). Geographical Text Analysis Mapping and spatially analysing corpora. *Corpus Linguistics*, 104.
- Gregory, I. N., Baron, A., Cooper, D. C., Hardie, A., Murrieta-Flores, P., & Rayson, P. (2014). Crossing Boundaries: Using GIS in Literary Studies, History and Beyond. *Collections électroniques de l'INHA. Actes de colloques et livres en ligne de l'Institut national d'histoire de l'art*. Acedido em 9 janeiro de 2017, em <https://inha.revues.org/4931?lang=fr>
- Gregory, I. N., & Ell, P. S. (2007). *Historical GIS: technologies, methodologies, and scholarship* (Vol. 39). Cambridge: Cambridge University.
- Gregory, I. N., & Hardie, A. (2011). Visual GISing: bringing together corpus linguistics and Geographical Information Systems. *Literary and linguistic computing*, 26(3), 297-314.
- Grover, C., Tobin, R., Byrne, K., Woollard, M., Reid, J., Dunn, S., & Ball, J. (2010). Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1925), 3875-3889. Acedido em 9 janeiro de 2017, em <http://rsta.royalsocietypublishing.org/content/368/1925/3875.short>

- Jackendoff, R. (2010). *Meaning and the Lexicon*. Oxford: Oxford University.
- Lagoa, V. (1950-53). *Glossário Toponímico da Antiga Historiografia Portuguesa Ultramarina* (4 Vols.). Lisboa: Junta de Investigações Coloniais.
- Marques, A. P. (1996). *A Cartografia Portuguesa do Japão. The Portuguese Cartography of Japan*. Lisboa: Imprensa Nacional - Casa da Moeda.
- Pinto, Fernam Mendez. (1614). *Peregrinaçam*. Lisboa: Pedro Crasbeek. Acedido em 9 janeiro de 2017, em <http://purl.pt/82>
- Pinto, Fernão Mendes. (1653). *The voyages and adventures, of Fernand Mendez Pinto, a Portugal ... done into english by H. C. Gent*. London : printed by J. Macock, for Henry Cripps, and Lodowick Lloyd. Acedido em 9 janeiro de 2017, em <http://purl.pt/16425>
- Russel, B. (1905). On denoting. *Mind*, 14, 479-493.
- Tomaszewski, B., & MacEachren, A. M. (2012). Geovisual analytics to support crisis management: Information foraging for geo-historical context. *Information Visualization*, 11(4), 339-359. Acedido em 9 janeiro de 2017, em <https://pdfs.semanticscholar.org/0421/27b361ec4a19758611a12e07eb3d3f206b3f.pdf>