

TRANSPONDO PALAVRAS PARA UMA PLANILHA: A CONSTRUÇÃO DE UM BANCO DE DADOS INTENCIONALMENTE SUBJETIVO**TRANSPOSING WORDS INTO A SPREADSHEET: THE CONSTRUCTION OF AN INTENTIONALLY SUBJECTIVE DATABASE**Vinícius Sodré Maluly¹

Resumo: A construção de bancos de dados é um processo bastante naturalizado no meio da informática e das humanidades digitais. É pouco discutida tanto em termos de sua concepção quanto em termos de sua implementação. Os obstáculos presentes nas análises dos dados também pouco figuram em trabalhos, independentemente da área do conhecimento. Dessa forma, não apenas a concepção/construção do banco de dados é irrefletida, como também temos a impressão de que as análises são meros produtos intuitivos e inevitáveis de uma formulação anterior. Este artigo visa a questionar essa causalidade rígida, demonstrando não apenas determinadas minúcias na construção de um banco de dados específico (sobre as viagens do botânico francês Auguste de Saint-Hilaire no Brasil), como as suas falhas, os desvios de rota, as análises mal elaboradas e as estratégias utilizadas para contornar dificuldades. Não entramos no mérito qualitativo deste banco de dados em si, mas apenas no que há por trás de sua formulação, construção e aplicação. Exemplos cartográficos são trazidos para demonstrar a aplicabilidade dos dados, principalmente tendo em vista os *softwares* utilizados e as intenções que baseiam a pesquisa, mas buscamos, finalmente, dar ênfase ao desafio existente na conversão entre o qualitativo e o quantitativo neste processo.

Palavras-chave: Bancos de dados. Análise de dados. Cartografia digital. Humanidades digitais.

Abstract: Database construction is a rather naturalized process in the computer science and digital humanities scene. It is little discussed either in terms of its conception or in terms of its implementation. The obstacles present in data analysis are also scarcely mentioned in works, regardless of the area of knowledge. Thus, not only is the design/construction of the database thoughtless, but we also have the impression that the analyses are merely intuitive and inevitable products of a previous formulation. This paper aims to question this rigid causality, demonstrating not only certain details in the construction of a specific database (on the travels of the French botanist Auguste de Saint-Hilaire in Brazil), but also its flaws, the detours, the poorly elaborated analyses, and the strategies used to circumvent difficulties. We do not go into the qualitative merit of this database itself, but only what lies behind its formulation, construction and application. Cartographic examples are brought to demonstrate the applicability of the data, especially in view of the software used and the intentions behind the research, but we seek, finally, to emphasize the challenge existing in the conversion between the qualitative and the quantitative in this process.

Keywords: Databases. Data analytics. Digital cartography. Digital humanities.

Introdução

¹Graduado e mestre em Geografia pela Universidade de Brasília, atualmente cursa doutorado em Geografia pela École des Hautes Études en Sciences Sociales em Paris-França sob a orientação da Prof. Dra. Cláudia Damasceno Fonseca e co-orientação do Prof. Dr. Laurent Vidal (Université de La Rochelle). E-mail: vmaluly@gmail.com – Orcid iD: <https://orcid.org/0000-0002-3989-7842>

Este texto visa a expor a construção de um banco de dados que teve por objetivo quantificar palavras densamente tecidas em uma narrativa de ampla divulgação e ciência. Trata-se da obra de maior difusão do botânico francês Auguste de Saint-Hilaire. Este, nascido e falecido em Orléans, em 1779 e em 1853, respectivamente, foi um estudioso da flora brasileira, que esteve no Brasil entre os anos de 1816 e 1822. Produziu larga literatura, principalmente sobre os aspectos naturais do Brasil, porém Saint-Hilaire provavelmente seja mais conhecido por ter publicado uma obra colossal intitulada “Viagens ao interior do Brasil” na qual ele relata todas as suas jornadas nesses seis anos de estadia. Saint-Hilaire fez uma série de itinerários de extensões continentais, partindo do Rio de Janeiro e passando por diversas capitânicas e comarcas, como as de Minas Gerais, Goiás, Espírito Santo, São Paulo, Santa Catarina, Rio Grande do Sul, entre outras, descritos em mais de 3.500 páginas ao longo de quatro partes, cada uma com dois tomos, publicados entre 1830 e 1851, com uma obra suplementar, póstuma, de 1887. Neste texto, no entanto, não buscaremos entrar no mérito das viagens supracitadas². Tentaremos nos debruçar sobre as escolhas técnicas e metodológicas que estiveram por trás da construção do banco de dados que guiou e continua a guiar a nossa pesquisa sobre este naturalista³.

Dividir para conquistar

O texto do botânico francês é, de antemão, extremamente desafiador para qualquer leitor, entre os mais habituados a este tipo de documentação até os mais desavisados. De construção narrativa densa e com bastante vagar, Saint-Hilaire não nos poupa de dar detalhes aprofundados sobre cada elemento pelo qual ele passou ou pelo qual seus olhos repousaram. Não à toa, o poeta Carlos Drummond de Andrade

² Para uma análise mais aprofundada do texto do viajante com relação a alguns territórios de povos indígenas no extremo leste da capitania de Minas Gerais, ver MALULY, Vinicius Sodré, Auguste de Saint-Hilaire e os territórios de exceção (Minas Gerais, 1816-1817), *PatryTer*, v. 3, n. 6, p. 266-280, 2020.

³ Com respeito aos momentos iniciais de construção do banco de dados e ao interesse pelo qual iniciamos a pesquisa sobre o naturalista francês, ver MALULY, Vinicius, **Voyages et formation territoriale : une approche exploratoire du récit d’Auguste de Saint-Hilaire (1816-1817)**, Mémoire de master - territoires, espaces, sociétés, École des Hautes Études en Sciences Sociales, Paris, 2019.

escreveu, em outubro de 1979, no *Jornal do Brasil*, que fazia “Turismo na toca” ao ler os escritos do francês e ainda complementou: “haverá maior comodidade?”⁴. A seguir destacamos um excerto retirado em aleatório de um de seus livros apenas enquanto ilustração da narrativa citada:

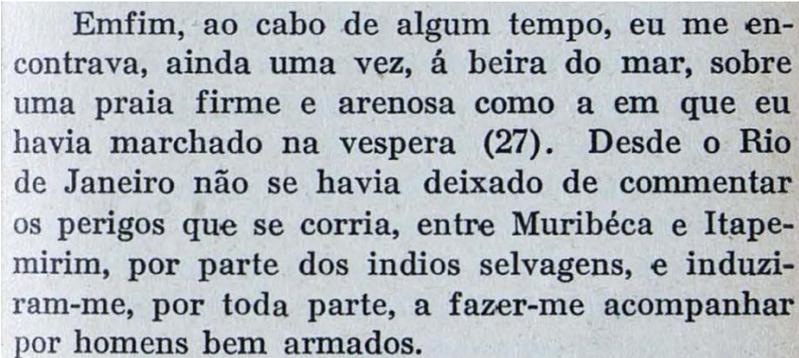
The image shows a rectangular box containing a text excerpt in a serif font. The text is centered and reads: "Emfim, ao cabo de algum tempo, eu me encontrava, ainda uma vez, á beira do mar, sobre uma praia firme e arenosa como a em que eu havia marchado na vespera (27). Desde o Rio de Janeiro não se havia deixado de commentar os perigos que se corria, entre Muribéca e Itape-mirim, por parte dos indios selvagens, e induziram-me, por toda parte, a fazer-me acompanhar por homens bem armados." The background of the box is a light, textured grey.

Figura 1 – Excerto de “Segunda viagem ao interior do Brasil: Espírito Santo” de Saint-Hilaire⁵

Podemos notar, superficialmente, a partir da figura 1, o teor do documento a ser trabalhado, tendo como característica preponderante fornecer, junto à descrição material da viagem, comentários mais gerais com respeito às capitâneas e ao próprio Brasil pré-republicano. Repetidas são as vezes em que Saint-Hilaire tece comentários aprofundados sobre a política e a economia dos territórios pelos quais ele passa, realizando comparações com outras já conhecidas por ele e fornecendo exemplos europeus a depender do assunto tratado. O ritmo da leitura dessa obra é equiparável, em termos, com o ritmo de suas próprias viagens, dada a minúcia das descrições e reflexões do botânico. A esse respeito, Cláudia Damasceno defende a tese de que a própria recepção indesejada desses escritos no continente europeu, ao longo dos 21 anos de publicação após seu retorno à França, poderia surtir do fato de que a leitura promovida por Saint-Hilaire não era condizente com o que o público-alvo desejava⁶. As

⁴ ANDRADE, Carlos Drummond, Meu amigo Saint-Hilaire, *Jornal do Brasil*, p. 5, 1979.

⁵ SAINT-HILAIRE, Auguste, *Segunda viagem ao interior do Brasil: Espírito Santo*, São Paulo: Companhia Editora Nacional, 1936, p. 41.

⁶ DAMASCENO FONSECA, Cláudia, *Voyages dans l'intérieur du Brésil : les observations historiques et géographiques d'Auguste de Saint-Hilaire*, in: LAMY, Denis *et al* (Orgs.), *Un botaniste français au Brésil*, Paris: Muséum National d'Histoire Naturelle, 2016, p. 217.

leituras de histórias de sobrevivência, como a viagem de Charles-Marie de La Condamine pelo rio Amazonas⁷, atraíam muito mais a atenção dos leitores, ávidos em explorar, por meio das palavras desses personagens, as terras pouco conhecidas do Novo Mundo⁸.

Dada essa breve contextualização, a construção de um banco de dados tornou-se inevitável se quiséssemos compreender, em larga escala e de forma organizada, as considerações do francês sobre o Brasil do início do século XIX. O nosso problema inicial era o de buscar organizar os comentários do naturalista francês sobre os territórios pelos quais ele passou e, para isso, tivemos de selecionar os documentos tratados, já que não podíamos, naquela altura, trabalhar com a totalidade de sua obra – isto é, ao redor de 3.500 páginas divididas em oito tomos. Assim, foram selecionados os seguintes livros: Parte 1 – Tomo 1 e Parte 3 – Tomos 1 e 2, totalizando uma amostra de mais de 1.000 páginas. O primeiro tomo da Parte 1 refere-se à chegada do francês no Brasil e, portanto, às suas impressões e ideias iniciais carregadas consigo desde o outro lado do Atlântico. Já os dois tomos da Parte 3 foram escolhidos por fazerem parte das viagens de Saint-Hilaire desde o oeste de Minas Gerais até a capitania de Goiás. O interesse em pesquisar a história do território goiano pode ser mais bem esmiuçada com base em outras pesquisas nossas⁹. Finalmente, realizamos o “desmonte” da fonte, como sugerido por Tiago Gil na figura 2, e passamos, então, a construir o “material ou metafontes”.

⁷ LA CONDAMINE, Charles-Marie de, **Viagem na América Meridional descendo o Rio das Amazonas**, Brasília: Senado Federal, 2000.

⁸ Sobre estas discussões, são incontornáveis as leituras de PRATT, Mary Louise, **Imperial Eyes: travel writing and transculturation**, London & New York: Routledge, 2003; e SAFIER, Neil, **Measuring the New World: Enlightenment Science and South America**, Chicago: The University of Chicago Press, 2008.

⁹ MALULY, Vinicius, **Como se fossem para o cabo do mundo: geohistória e cartografias sobre os caminhos e os descaminhos de Goyaz (1725-1752)**, Dissertação (Mestrado em Geografia), Universidade de Brasília, Brasília, 2017.



Figura 2 – “Esquema das etapas da produção do conhecimento histórico” extraído do livro
 “Como se faz um banco de dados (em história)” de Tiago Luís Gil¹⁰

Segundo Ângelo Carrara, Carlos Valencia e Massimiliano Grava, devemos a Jean-Philippe Genet a primeira referência à ideia de metafonte aplicada à informática, na qual se busca construir uma estrutura que difira necessariamente do texto documental¹¹. Esses autores, por sua vez, complementam a ideia ao afirmar que nem todo objeto digital é uma metafonte. Esta deve ser composta “pelo cruzamento de informação que geram outras novas. Esse cruzamento emprega meios informáticos, já que processa milhões de dados de forma simultânea. Desse cruzamento, desse conjunto de fontes, provém a geração de informação que não era diretamente observável na fonte tradicional.”¹². Para finalizar, Emmanuel Le Roy Ladurie concorda com o colocado anteriormente, afirmando que

na história, como em outros campos, o que conta não é a máquina, mas o problema. A máquina só tem interesse na medida em que nos permite abordar novas questões, originais em seus métodos, conteúdos e, sobretudo, em seu escopo. Uma vez aceito este ponto, muitas direções frutíferas estão abertas à pesquisa de historiadores assistidos por computadores. Uma das direções mais claras é a análise de vastos *corpus* documentais, cujos dados eram

¹⁰ GIL, Tiago Luís, **Como se faz um banco de dados (em história): metodologia de pesquisa e informática**, Porto Alegre: Ladeira Livros, 2015, p. 16.

¹¹ GENET, Jean-Phillipe, *Histoire sociale et ordinateur*, in: **Informatique et histoire médiévale**, Istituto di Storia Medievale de la Università di Pisa/École Française de Rome, Italia.

¹² “*por el cruzamiento de informaciones que generan otras nuevas. Ese cruzamiento emplea medios informáticos ya que procesa millones de datos de forma simultánea. De ese cruzamiento, de ese ensamble de fuentes, proviene la generación de información que no era observable directamente en la fuente tradicional.*” Ver CARRARA, Ângelo; VALENCIA, Carlos; GRAVA, Massimiliano, *Metafuente y el uso de los sistemas de información geográfica en historia económica*, **América Latina en la Historia Económica**, v. 25, n. 3, p. 40-70, 2018, p. 48.

cruciais, mas cujas dimensões até agora tinham desafiado os esforços dos pesquisadores¹³.

Posta a questão da metafonte enquanto construção informática necessária de uma nova fonte, baseada e construída na fonte originária, tivemos de ordenar as densas palavras do botânico francês. Para cumprir esse objetivo, dividimos as temáticas por ele abordadas em 12 temas: 1) administração pública; 2) caminhos; 3) demografia; 4) escravidão; 5) tributação; 6) fundiário; 7) paisagem; 8) povos indígenas; 9) produção; 10) território; 11) toponímia; 12) urbano. Algumas considerações devem ser feitas com respeito a essa divisão.

Primeiramente, ela é claramente subjetiva. Se qualquer outro pesquisador buscasse temáticas presentes no texto de Saint-Hilaire, ele certamente faria essa tarefa de forma diferente. Poderia haver semelhanças ou uma total discrepância. Nenhuma das opções seria mais ou menos viável, como mostram Anne-Kathrin Reuschel e Lorenz Hurni ao discutir as incertezas na produção da cartografia literária. As diferenças entre nós, os leitores da obra a ser cartografada, estaria calcada em ao menos duas influências principais, segundo os autores: 1) “a imaginação pessoal, o conhecimento sobre o espaço, os detalhes adicionais advindos de uma literatura secundária e a generalização das configurações”; 2) “o mapa base em geral (seja imagem de satélite ou mapa), a escala de entrada e as ferramentas digitalizadoras disponíveis”¹⁴.

Uma outra questão é notória: a nossa pesquisa é feita no âmbito de uma geografia histórica e seu autor é geógrafo. Assim, determinadas questões saltam mais aos nossos olhos do que aos de um historiador, por exemplo, ou de um antropólogo. Talvez o tema “Povos Indígenas” fosse subdividido. Talvez não existisse “Toponímia”,

¹³ “en histoire, comme ailleurs, ce qui compte, ce n'est pas la machine, mais le problème. La machine n'a d'intérêt que dans la mesure où elle permet d'aborder des questions neuves, originales par les méthodes, les contenus et surtout l'ampleur. Ce point admis, bien des directions fécondes s'ouvrent aux recherches d'historiens assistés par l'informatique. L'une des orientations les plus claires, c'est l'analyse de vastes corpus de documents, dont les données étaient capitales, mais dont les dimensions avaient jusqu'ici défié les efforts des chercheurs.” Ver LADURIE, Emmanuel Le Roy, L'historien et l'ordinateur, in: **Le territoire de l'historien**, Paris: Gallimard, 1973, v. 1, p. 11.

¹⁴ “the personal imagination, the knowledge about the space, additional details in secondary literature and the generalisation of settings. (...) the base map in general (satellite image or map), the input scale and available digitizing tools.” Ver REUSCHEL, Anne-Kathrin; HURNI, Lorenz, Mapping Literature: visualisation of spatial uncertainty in fiction, **The Cartographic Journal**, v. 48, n. 4, p. 293-308, 2011, p. 299.

inclusa em “Urbano”. Talvez “Produção” não fosse separada de “Fundário”. Infinitas seriam as possibilidades. Uma outra questão é a natureza imprecisa de todos os temas, visto que, quando falamos de “Demografia”, falamos de certa forma da subdivisão eclesiástica do Brasil, visto que as coisas não estavam separadas¹⁵. Ou, então, tratar de “Escravidão” é, segundo o relato de Saint-Hilaire, necessariamente de escravidão africana. Isto não quer dizer que não houvesse escravidão de indígenas na história do Brasil. No entanto, neste relato, a escravidão referida é finalmente africana. “Território” é outra categoria bastante imprecisa, talvez a mais imprecisa de todas. O que a distingue e por que ela existe? Saint-Hilaire faz várias considerações de cunho “territorial”, isto é, de referência à organização territorial do Brasil, porém não no intuito da administração pública em si (como cargos ou normas) e também não necessariamente atrelado às cidades. Portanto, temos a divisão dos temas, mas admitindo a subjetividade, a imprecisão e a porosidade deles. O que fazer agora com essas delimitações do texto?

Por uma necessária simplificação do discurso

A figura 3 é uma captura de imagem da interface criada por nós para construir o banco de dados desejado. O trabalho foi todo realizado utilizando o *software* livre QuantumGis, já que este nos permite dispor de todas as informações de maneira georreferenciada e com formas simplificadas de exportá-las para uso posterior em outros *softwares*, como Microsoft Excel, por exemplo. O caminho inverso também poderia ser feito, inserindo as informações no Excel e depois exportando-as em arquivo formato .csv para o QGIS, mas a montagem do banco de dados intra-QGIS tem, ao menos, duas vantagens: 1) a forma didática pela qual podemos montar as fichas, como o exemplo da figura 3; 2) a visualização das informações inseridas por nós já em um ambiente georreferenciado, o que faz com que consigamos ver, *pari pasu*, a

¹⁵ DAMASCENO FONSECA, Cláudia; VENÂNCIO, Renato Pinto, Vila Rica : prospérité et déclin urbain dans le Minas Gerais (XVIIIe-XXe siècles), in: VIDAL, Laurent (Org.), **La Ville au Brésil (XVIIIe-XXe siècles) : naissances, renaissances**, Paris: Les Indes Savantes, 2008.

especialização dos dados. Aqui, faz-se necessário reiterar a importância de o próprio pesquisador construir o seu banco de dados, os seus mapas, as suas visualizações. Sabemos que nem todos são afeitos a isso, mas a terceirização desse tipo de serviço certamente provoca a perda de possíveis alternativas que se mostram visíveis apenas para aquele que está imbuído no ato de processamento da informação. Quando terceirizamos uma visualização, na maior parte das vezes, estamos buscando apenas a reprodução material de algo que já temos estabelecido e construído em nossas mentes. Uma das vantagens, por assim dizer, do saber-fazer no campo das humanidades digitais é de justamente poder tirar proveito da democratização das ferramentas informáticas, colocando-as nas mãos do próprio investigador.

The image shows a web-based data entry form with the following sections and fields:

- Générale**
 - id: 416
 - Chapitre: P3C25
 - Date: 06/08/1819
- Lieu**
 - Lieu: Vila Boa de Goiás
 - Lieu obs: Capital da capitania
 - Page: 223
 - Commentair: Notas sobre os montes e a vegetação que cercam a cidade
 - Qté mots: 70
- Sujet**
 - Sujet: Paysage
 - Échelle: Locale
- Sentiments**
 - Sentiments: Joie
 - Pos ou né: Positif
- Num**

Adm	NULL	Che	NULL	Dém	NULL
Esc	NULL	Évé	NULL	Fisc	NULL
Fonc	NULL	Org	NULL	Pay	1
Pop	NULL	Prod	NULL	Topo	NULL
Urb	NULL	Pos	1	Nég	NULL

Buttons: Cancel, OK

Figura 3 – Exemplo de preenchimento de ficha do banco de dados sobre Auguste de Saint-Hilaire

Esmiuçemos, então, a ficha presente na figura 3¹⁶. O primeiro campo é referente ao “id”, identificador de cada feição adicionada ao projeto. Ou seja, cada ficha é numerada. Neste caso, seria a ficha de número 416. A seguir temos mais dois campos, o de “Capítulo” e de “Data”. No nosso exemplo, estaríamos na Parte 3 – Capítulo 25, no dia de 06 de agosto de 1819. Os três campos formariam o cabeçalho da ficha. Em seguida, temos a seção dedicada às informações sobre o lugar.

Primeiramente preenchemos o campo “Lugar”, sendo que nem sempre é possível saber com precisão onde está o viajante. No nosso exemplo, estaríamos em Vila Boa de Goiás, mas, em casos de dúvida, vale o esforço de preencher com “nas redondezas de...”. Adicionamos um campo de observação com respeito ao local, neste caso, capital da capitania, mas poderíamos também inserir outras informações, como, por exemplo, “a 25 léguas de...”. Em seguida, a página da fonte na qual estamos naquele exato momento. Pode também ser inserido um intervalo, como, por exemplo “223-225”. Isto depende inteiramente da fonte consultada, ainda que queiramos, ao máximo, diminuir a quantidade e o tamanho dos intervalos em nosso banco de dados. A seguir temos os campos de “Comentário” e de “Quantidade de palavras”. O primeiro é o mais importante de toda a ficha, no qual inserimos um resumo da observação do viajante que é pertinente à nossa investigação, e o segundo é, simplesmente, a quantidade de palavras que ele levou para fazer toda a descrição. Neste caso, temos um resumo sobre as observações relacionadas à paisagem no entorno de Vila Boa de Goiás, mas, na fonte em si, vemos que ele utilizou 70 palavras em toda a descrição. Isto quer dizer que, na ficha proposta na figura 3, não podemos inserir de fato o comentário por completo do viajante. Temos um limite a obedecer, neste caso, de 254 caracteres, que provavelmente é inferior ao número de 70 palavras indicado. Porém, utilizaremos a ficha para não apenas classificar e resumir os comentários selecionados. Esta é somente uma de suas funções, como veremos mais adiante.

A próxima parte da ficha é a de “Tema”. Nesta seção podemos ver que não há campos de preenchimento “string”, ou seja, de inserção de sequências de caracteres.

¹⁶ Nos desculpamos com o leitor por parte da ficha estar em língua francesa. Isto é em decorrência da pesquisa ter sido realizada no seio da École des Hautes Études en Sciences Sociales, Paris-França.

São listas nas quais podemos escolher apenas uma opção. Na figura 3, podemos selecionar apenas um tema dentre os 12 indicados anteriormente e apenas uma “Escala” dentre duas – a saber, local ou geral. No nosso exemplo, escolhemos o tema paisagem e a escala local. A eleição de adotar também a classificação de escala nos pareceu crucial para filtrar o caráter de determinados comentários e observar determinados padrões de comportamento. Por exemplo, a maior parte das observações referentes ao tema “Demografia” são em escala “Local”, mas, em termos de “Território”, a maioria é “Geral”. Outros cruzamentos são potencialmente interessantes, como o de saber quantas palavras são usadas, em média, nos comentários em escala local e quantas para a geral.

A seguir, temos a seção referente aos “Sentimentos”. Também só há opções a escolher dentre uma lista restrita e refere-se a uma pesquisa ainda em andamento, mas que quer acrescentar mais uma camada de variáveis à investigação, desta vez buscando, por um lado, sentimentos específicos que podem estar presentes no discurso do viajante (alegria, medo, tristeza, raiva, nojo e surpresa) e, de outro, sentimentos amplos, classificados em “Positivo” ou “Negativo”. Nem sempre é possível indicar um dos 6 sentimentos listados, mas é sempre aconselhado inserir ora “Positivo” ora “Negativo” para manter registro das flutuações do discurso do viajante a depender dos territórios pelos quais este passa. É um exercício naturalmente limitador, visto que são categorias binomiais e a realidade do discurso é muito mais profunda do que isso, mas toda pesquisa científica é, de uma maneira ou de outra, um esforço de redução e de simplificação da realidade. Cabe a nós ressaltar isso para o leitor.

Finalmente, encerramos a ficha da figura 3 com a seção “Num” que diz respeito à repetição de algumas informações indicadas anteriormente, desta vez em caráter numeral. Isto tem um propósito específico, relacionado aos mapas por contagem de ponto em polígono que veremos mais adiante (figura 6). Neste momento, basta dizer que são campos que classificam cada temática abordada e os campos gerais relativos aos sentimentos. No nosso exemplo, preenchemos apenas os campos referentes à paisagem e ao sentimento geral positivo com o número 1, deixando todos os outros automaticamente com resultado 0.

Tendo preenchido a nossa ficha, seguimos então para o próximo passo. Como identificar as espacializações do discurso do viajante de acordo com os parâmetros indicados?

Mapas não falam, mas podem dizer

Tendo criado, segundo as necessidades de nossa pesquisa, o banco de dados diretamente no *software* QGIS, passamos agora ao momento de visualizar cartograficamente tais informações. Primeiramente, começaremos com a visualização mais intuitiva: a localização dos dados divididos por temas (figura 4).

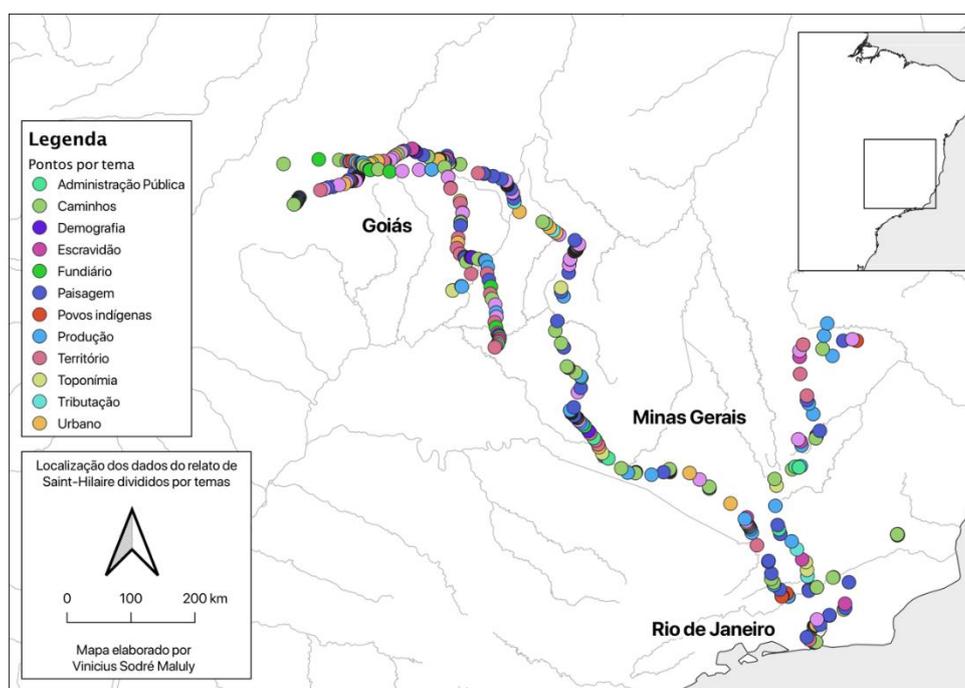


Figura 4 – Localização dos dados do relato de Saint-Hilaire divididos por temas

A figura 4 fornece a localização de todos os dados coletados ao longo do relato de Saint-Hilaire, segundo a seleção por nós aclarada anteriormente. Podemos observar que temos dois percursos distintos: ambos saindo do Rio de Janeiro, mas um rumo a leste de Minas Gerais e outro rumo a Goiás. O primeiro é referente à Parte 1 – Tomo 1 e o segundo à Parte 3 – Tomos 1 e 2. Porém, o que nos interessa aqui é a visualização

proposta em temas. Como indicamos acima, temos 12 ao total e a figura 4 é uma tentativa não muito bem-sucedida de localizá-los por meio do uso de cores. Temos uma profusão de diferentes pontos, mas a sobreposição deles em decorrência da escala de visualização, além da quantidade elevada de cores utilizada (sortida aleatoriamente), não nos permite chegar a nenhuma conclusão contundente. Se dividíssemos a análise por seções ou amostras, talvez chegássemos a mínimas conclusões de como que o viajante fez uso de cada temática para tratar deste ou daquele território, mas, efetivamente, não há nada de substancial. A escala 1:200km também fornece suas próprias dificuldades, visto que o *software* não consegue arranjar todos os elementos de forma simultânea. Há uma necessária sobreposição a ser realizada e esta não depende de nós. A figura 5 explicita o processo.

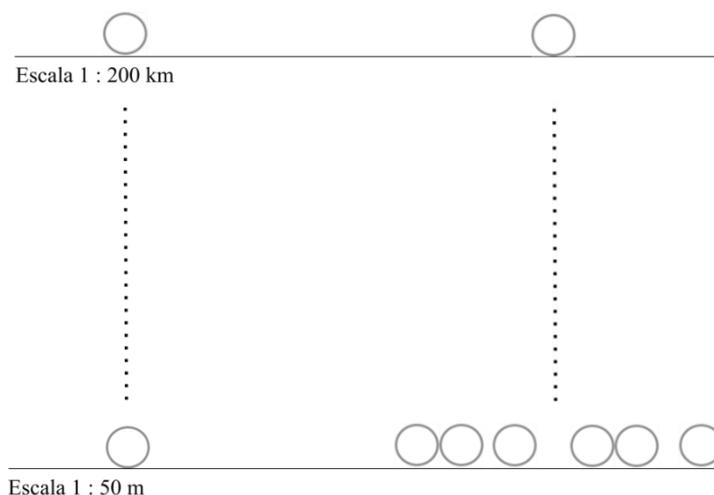


Figura 5 – Sobreposição de dados a partir da variação de escala¹⁷

A partir da figura 5, podemos visualizar o problema da escala proposto anteriormente. Dada a extensão continental do percurso analisado, a figura 4 apresenta necessárias limitações técnicas, visto que não é possível representar

¹⁷ Esta figura é uma adaptação de “Figure 8 – Schéma concernant le problème de l'échelle” de MALULY, *Voyages et formation territoriale: une approche exploratoire du récit d'Auguste de Saint-Hilaire (1816-1817)*, p. 37.

cartograficamente todas as feições presentes no banco de dados, ao menos em forma de pontos. Se estivéssemos em escala 1:50m, como sugere a figura 5, poderíamos ver as feições com muito mais propriedade e precisão. Porém, necessitaríamos de uma quantidade quase que infinita de mapas para dar conta de todo o percurso do botânico francês. Também devemos notar que, segundo a figura 5, um ponto em escala grande (1:50m) é representado enquanto um ponto em escala pequena (1:200km). Já uma coletânea de pontos muito próximos em escala grande é igualmente representada enquanto um ponto em escala pequena. Ou seja, as conclusões que podemos chegar analisando as cores da figura 4 podem ser profundamente ilusórias. Podemos achar que estamos tratando de um ponto, mas estamos, em verdade, tratando de diversos.

A variação de cores, portanto, não nos forneceu as devidas respostas que buscávamos, principalmente para dar sentido ao banco de dados elaborado inicialmente. Não conseguiremos distinguir os temas se não pudermos, de antemão, compreender a distribuição dos pontos em si. À primeira vista, eles parecem ordenados de forma igualitária, obedecendo a um percurso bem distribuído, porém a figura 6 oferece uma outra compreensão a esse respeito.

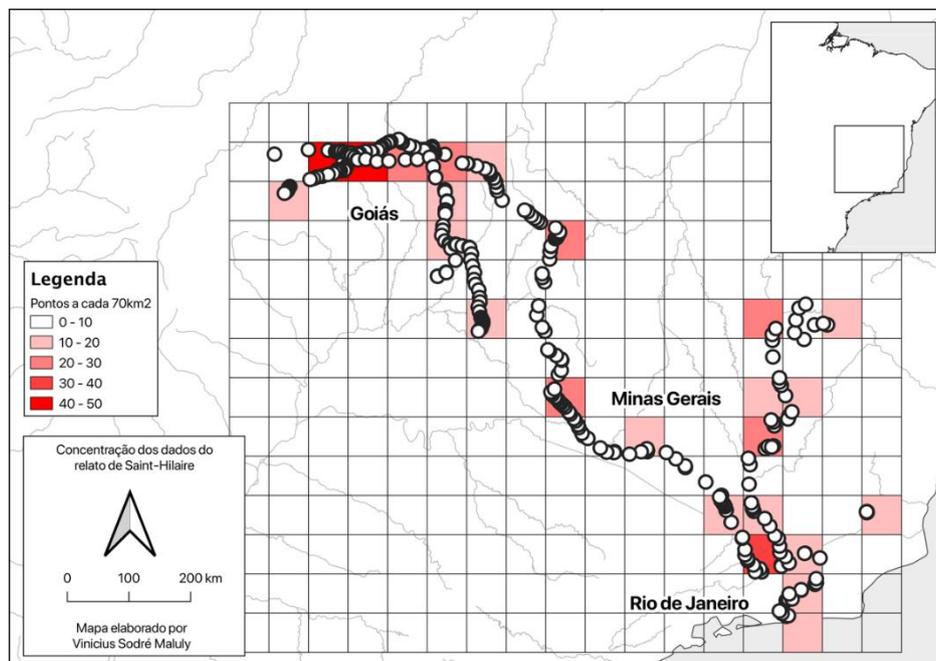


Figura 6 – Concentração dos dados do relato de Saint-Hilaire

Se, destarte, temos a impressão de que os pontos estão distribuídos de forma igualitária ao longo do espaço euclidiano, podemos ver, na figura 6, que isso é uma simplificação promovida pelo *software* de georreferenciamento. De fato, há inúmeras sobreposições e, se quisermos alcançar a espacialização dos temas, temos de antes compreender a forma pela qual os dados se distribuem. Neste caso, tivemos de utilizar um recurso chamado “Contagem de pontos em polígono”. Criamos uma malha abstrata de vetores quadrados sobre um plano e, então, implementamos a análise a partir da contagem desses pontos. Cada polígono possui uma área de 70km² e há uma variação de intensidade da cor vermelha indicando a quantidade de pontos localizados dentro do dito quadrado. A partir da legenda podemos identificar que há 5 níveis de concentração de pontos: 0-10, 10-20, 20-30, 30-40 e, finalmente, 40-50 pontos. Assim, conseguimos, de certa forma, driblar as limitações da visualização na escala 1:200km, mantendo as informações alimentadas no banco de dados.

Como em qualquer outro mapa, a figura 6 foi resultado de uma série de escolhas arbitrárias, respeitando 1) os interesses da pesquisa; 2) a estrutura do banco de dados. Ela poderia ser, assim como colocamos com respeito ao banco de dados, extremamente diferente se fosse realizada por outro pesquisador com outros fins. Neste caso, a malha de polígonos que cobre a área analisada tem quadrados com área de 70km², mas poderiam ser de maior ou menor área. Entretanto, fizemos os respectivos testes e, na escala 1:200km, essa foi a distribuição mais generosa com o arranjo dos próprios dados e com os prováveis leitores do material. Quadrados menores ou maiores forneceriam conclusões similares, tendo apenas leituras menos convenientes. É de se ressaltar, nesta parte, a questão dos campos numéricos indicados no preenchimento da ficha do banco de dados (figura 3). A seção “Num” foi criada justamente porque a ferramenta “Contagem de pontos em polígonos” não aceita campos “string” (texto). Assim, tivemos de elaborar a parte inferior da ficha apenas para indicar a concentração dos pontos de forma numérica, distribuindo-os em seus devidos temas. A figura 7 é a decomposição dos resultados da figura 6, respeitando as divisões da fonte consultada.

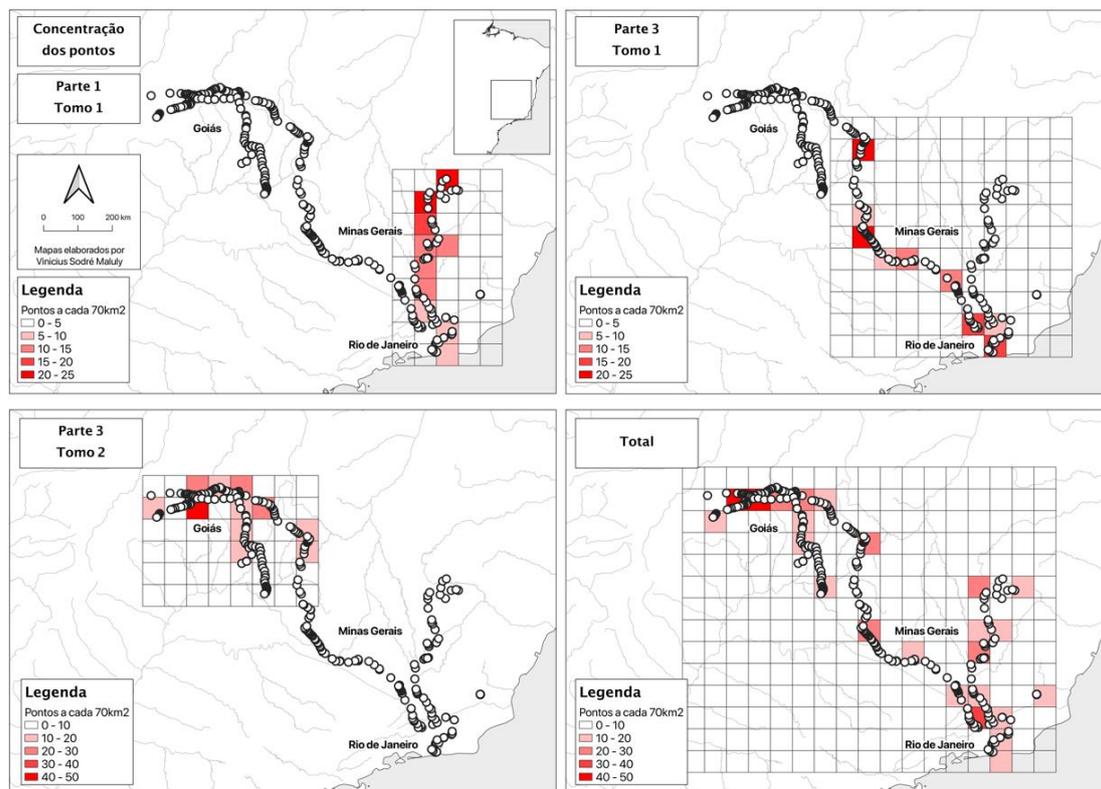


Figura 7 – Concentração dos dados do relato de Saint-Hilaire por Partes/Tomos

O recurso da “Contagem de pontos por polígonos” não necessita ser aplicado na integralidade do banco de dados. A figura 7 dá o exemplo da aplicação da mesma técnica, mas respeitando as divisões da própria fonte consultada. Como já dissemos anteriormente, não entraremos no mérito da questão, mas apenas ressaltaremos as particularidades técnico-metodológicas do processo de pesquisa executado até o momento. Distribuindo cartograficamente os pontos segundo as partes do texto do botânico francês, notamos sub-concentrações muito particulares e, ao mesmo tempo, vemos como que a contagem por etapas é variável a depender da malha analisada. A Parte 1 – Tomo 1 é a mais didática nesse sentido, dado que o acúmulo dos pontos é progressivo, partindo do litoral em direção ao interior. Já as demais partes não dão esse entendimento, fornecendo, portanto, outras interpretações do motivo pelo qual o viajante francês se deteve mais em determinados pontos do que em outros. Porém, é de se considerar igualmente que os intervalos na legenda diferem. A quantidade de pontos

presentes na Parte 1 – Tomo 1 e Parte 3 – Tomo 1 são menores do que na Parte 3 – Tomo 2 e no Total. Dessa forma, a divisão é distinta, com intervalos progressivos de até 5 pontos e não 10, como na figura 6.

Por último, antes de tentarmos novamente nos referir cartograficamente aos temas, devemos desfazer uma aparente causalidade provocada pela leitura irrefletida dos mapas. Todas as figuras até aqui apresentadas indicam que não há uma relação direta entre a aparente quantidade de pontos e a quantidade deles em si. A figura 6 representa isso com clareza. Em partes mais alongadas, nas quais pensaríamos haver uma maior densidade de pontos, isso não ocorre. Efetivamente, é nas proximidades do Rio de Janeiro e em Goiás que Saint-Hilaire se detém para tratar dos assuntos referidos e sintetizados no banco de dados. Porém, é notório que ele passou mais tempo percorrendo a capitania de Minas Gerais. Ou seja, o esforço da produção da metafonte, ainda que possa ser debatida a própria natureza de metafonte neste caso, tendo em vista que ainda estamos focados na mesma fonte, sem cruzá-la com outras, visa a corroborar ou a destronar determinadas ideias/concepções. Temos de estar abertos ao resultado e à interpretação dos dados, sem buscar encerrá-los em nossas expectativas. Reaparece a importância do pesquisador em tratar seus próprios dados, já que o mais notório não é o banco de dados pronto ou a visualização encerrada e, sim, as várias tentativas, desistências e acertos que compõem a jornada.

Expandindo variáveis

Para encerrar a demonstração técnica e analítica do banco de dados construído, retornemos aos 12 temas. O discurso do botânico francês Auguste de Saint-Hilaire foi sub-categorizado por nós, segundo os interesses da pesquisa. Tentamos espacializá-los na figura 4 com o recurso das cores, mas não obtivemos muito sucesso. Apresentamos, em seguida, a abordagem da contagem de pontos sobre malhas abstratas de polígonos e, agora, tentaremos nos apropriar dessa técnica para visualizar os ditos temas. Começamos pela figura 8 que evidencia o tema “Paisagem”.

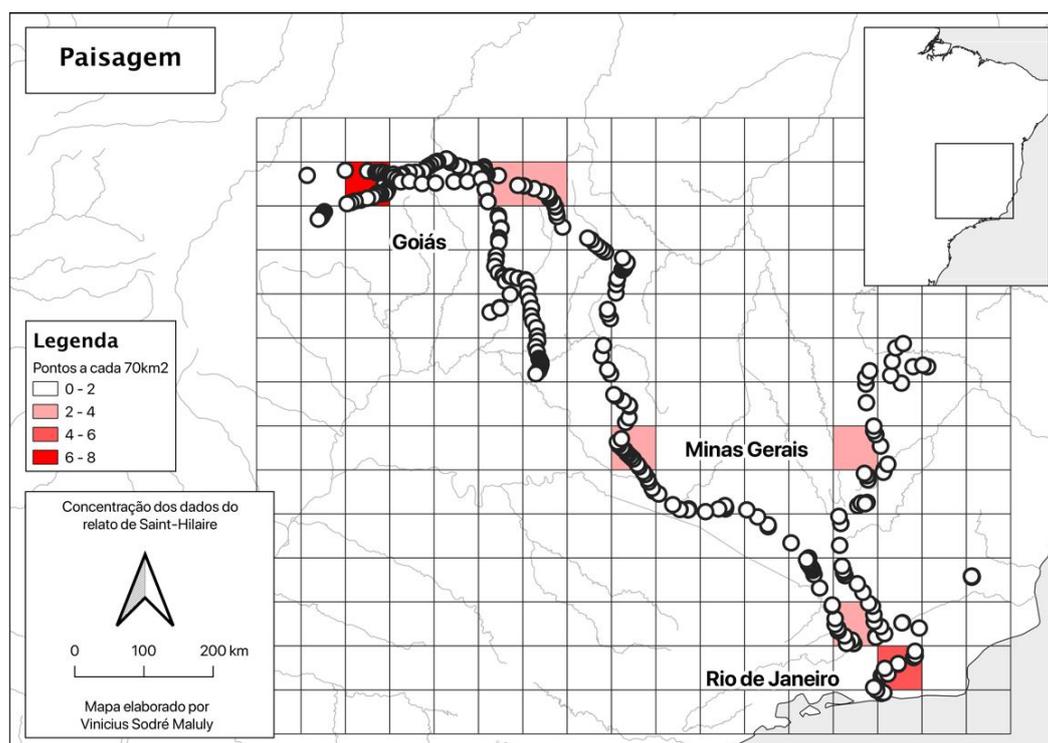


Figura 8 – Concentração dos dados do relato de Saint-Hilaire – Paisagem

A menção à paisagem é recorrente na obra de Saint-Hilaire. Sempre atento às suas redondezas, o francês frequentemente tomava nota do que via por onde passava. Não obstante, há lugares que chamaram mais a sua atenção nesse sentido do que outros. A figura 8 demonstra isso. Em sua chegada no Rio de Janeiro e em sua chegada a Goiás, Saint-Hilaire debruça-se com mais ênfase nessa temática. Deixa para trás outras colocações para aprofundar-se no que, de fato, conseguia ver em seus pontos de partida/chegada. Porém, também há notas referentes à paisagem no meio do caminho e isto se deve, em termos, à ausência de outros temas. Na falta do que dizer, Saint-Hilaire recorria ao que lhe rodeava: “Não é de se estranhar que eu me detenha tão longamente sobre vilarejos simples. Sente-se que eles devem ter importância em terras nas quais se pode viajar muitos dias sem encontrar um único e meses sem ver a menor cidade¹⁸.” A figura 9 adiciona mais 4 temas à discussão.

¹⁸ *On ne doit pas s'étonner si je m'étends aussi longuement sur de simples villages. On sent qu'ils doivent avoir de l'importance dans une contrée où l'on peut voyager bien des jours sans en rencontrer un seul, et des mois sans apercevoir la plus petite ville.* SAINT-HILAIRE, Auguste de, *Voyages dans les provinces de Rio de Janeiro et*

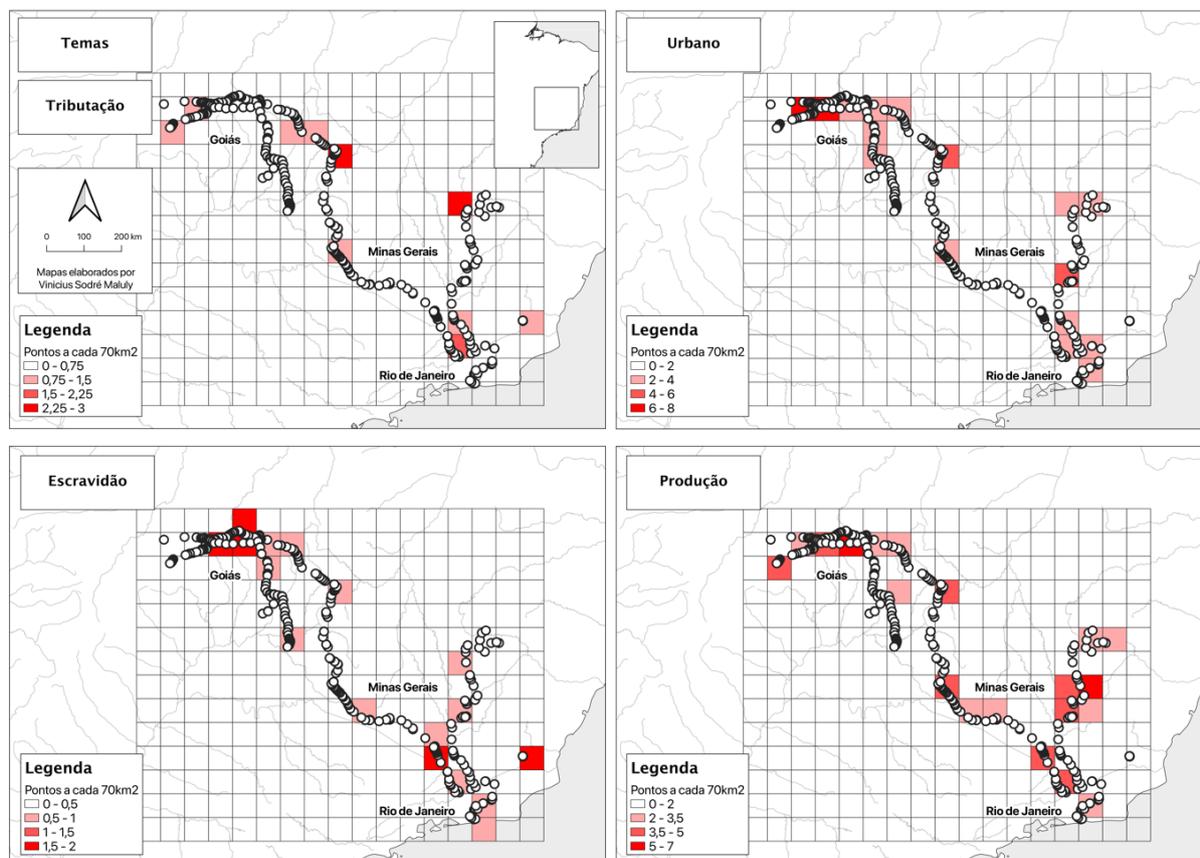


Figura 9 – Concentração dos dados do relato de Saint-Hilaire – Tributação, Urbano, Escravidão e Produção

As temáticas tratadas por Saint-Hilaire são variadas e também o é a quantidade de vezes às quais ele se refere a elas ao longo de seus textos. Isto se reflete nas legendas da figura 9. Todas apresentam quatro intervalos, todos distintos entre si. Tributação vai de 0 a 3 menções. Urbano, por sua vez, de 6 a 8. Escravidão de 0 a 2 e Produção de 0 a 7. Esses intervalos, por mais distintos que possam parecer, não invalidam a visualização comparada entre os mapas, visto que eles mantêm a proporcionalidade. A forma de distribuição é a de “Intervalo igual”, definida pelo QGIS a partir da quantidade de vezes que apareceu tal tema no banco de dados. Assim, podemos ver como que, em termos absolutos, há bem menos menções ao tema

Minas Geraes, in: *Voyages dans l'intérieur du Brésil*, Tome premier. Paris: Grimbart et Dorez, Libraires, 1830, p. 296.

“Escravidão” do que ao de “Produção”, mas isto não incorre em uma impropriedade matemática. Ao contrário, o *software* promove para nós a equiparação possível entre os temas. Há outras formas de fazê-lo, mas o importante, para nossa discussão, é reter o fato de que devemos sempre aplicar a mesma forma de divisão dos dados no momento de atomizá-los. Não podemos dar tratamentos diferenciados ou iremos, dessa forma, incorrer em equívocos algorítmicos.

Por último, trazemos a figura 10 como outra forma de visualizar o relato do viajante francês. A partir das palavras despendidas pelo mesmo, podemos ver *onde mais ele falou e onde ele falou mais vezes*. Isto dá contornos interessantes quando adentramos o nível do qualitativo, trazendo esses dados à luz das discussões da pesquisa. Porém, nos deteremos aqui. Basta dizer que são praticamente infinitas as possibilidades de expansão e de cruzamento das variáveis.

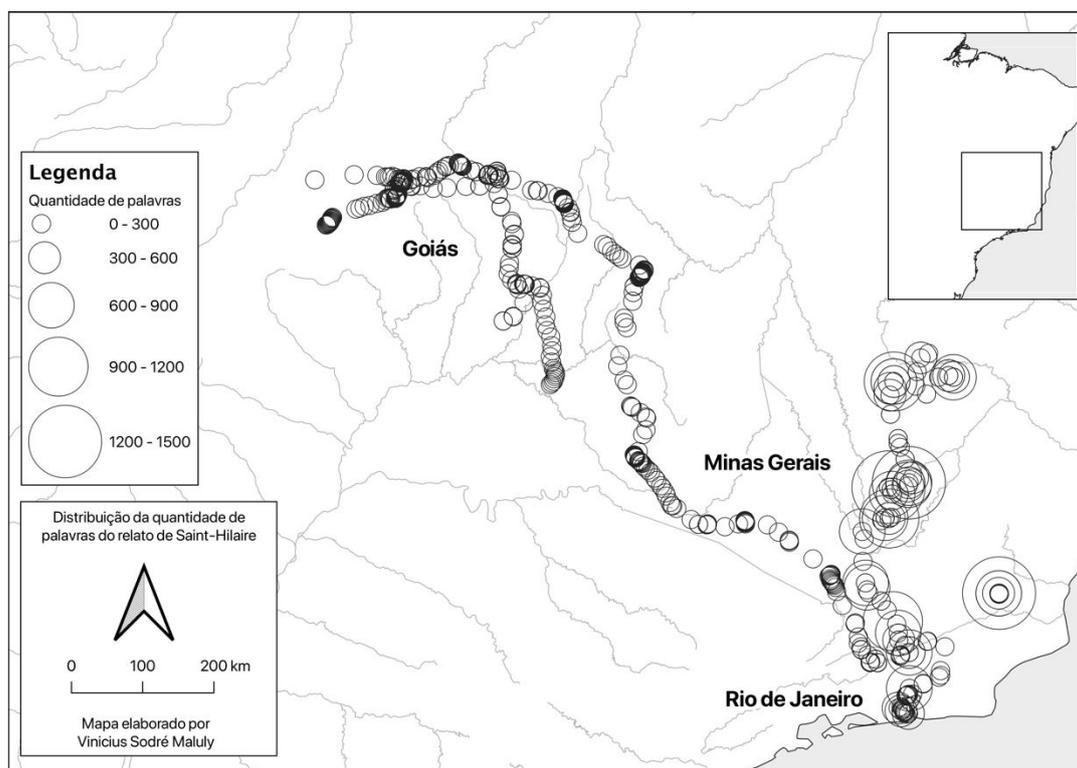


Figura 10 – Distribuição da quantidade de palavras do relato de Saint-Hilaire

Considerações finais

Este artigo tentou demonstrar a forma pela qual foi construído um banco de dados intencionalmente subjetivo. Isto é, por mais que estejamos lidando com ferramentas quantitativas de análise, é sobre o discurso e o território que se detém esta pesquisa. Portanto, o banco de dados teve de refletir isso. Não há fórmulas prontas na elaboração desses constructos sintetizadores da informação, mas há corta-caminhos e padrões que podem ser úteis a outros pesquisadores em suas diferentes colaborações. Desta maneira, tentamos elucidar os entremeios da nossa construção, destacando o motivo pelo qual determinados campos foram criados, a estratégia ora utilizada para contornar certos problemas, a relevância de determinadas análises em detrimento de outras etc. Buscamos nos aprofundar também na confecção e na análise de mapas visto que estes são desdobramentos do próprio banco de dados. Isto quer dizer que, se obedecermos a determinados princípios e tivermos alguns objetivos claros em mente no início da construção do banco de dados, as análises virão com muito mais fluidez e facilidade em etapas posteriores. Por último, é importante destacar: muitas serão as versões criadas e apagadas ao longo do percurso. Em determinados momentos, pode ser bastante frustrante não conseguir transpor, acertadamente, palavras para uma planilha. Porém, esse é o maior esforço que há. A conversão e o diálogo estreito entre o qualitativo e o quantitativo são os maiores desafios dos pesquisadores em humanidades digitais. Devemos ter coragem de abrir e mostrar os fundamentos e as ferramentas de nossas pesquisas para dar vez à colaboração, visto que o conhecimento deve ser sempre construído coletivamente, apesar de muitas vezes sermos induzidos a pensar o oposto.

Referências

ANDRADE, Carlos Drummond de. Meu amigo Saint-Hilaire. *Jornal do Brasil*, p. 5, 1979.

CARRARA, Ângelo; VALENCIA, Carlos; GRAVA, Massimiliano. Metafuente y el uso de los sistemas de información geográfica en historia económica. *América Latina en la Historia Económica*, v. 25, n. 3, p. 40-70, 2018.

DAMASCENO FONSECA, Cláudia. Voyages dans l'intérieur du Brésil : les observations historiques et géographiques d'Auguste de Saint-Hilaire. In: LAMY, Denis; PIGNAL, Marc; SARTHOU, Corinne; *et al* (Orgs.). *Un botaniste français au Brésil*. Paris: Muséum National d'Histoire Naturelle, 2016, p. 207–235. (Publications Scientifiques).

DAMASCENO FONSECA, Cláudia; VENÂNCIO, Renato Pinto. Vila Rica : prospérité et déclin urbain dans le Minas Gerais (XVIIIe-XXe siècles). In: VIDAL, Laurent (Org.). *La Ville au Brésil (XVIIIe-XXe siècles) : naissances, renaissances*. Paris: Les Indes Savantes, 2008.

GENET, Jean-Phillipe. Histoire sociale et ordinateur. In: *Informatique et histoire médiévale*. Istituto di Storia Medievale de la Università di Pisa/École Française de Rome, Italia, 1977.

GIL, Tiago Luís. *Como se faz um banco de dados (em história): metodologia de pesquisa e informática*. Porto Alegre: Ladeira Livros, 2015.

LA CONDAMINE, Charles-Marie de. *Viagem na América Meridional descendo o Rio das Amazonas*. Brasília: Senado Federal, 2000.

LADURIE, Emmanuel Le Roy. L'historien et l'ordinateur. In: *Le territoire de l'historien*. Paris: Gallimard, 1973, v. 1, p. 11–14.

MALULY, Vinicius. *Como se fossem para o cabo do mundo: geohistória e cartografias sobre os caminhos e os descaminhos de Goyaz (1725-1752)*. Dissertação (Mestrado em Geografia), Universidade de Brasília, Brasília, 2017.

MALULY, Vinicius. *Voyages et formation territoriale : une approche exploratoire du récit d'Auguste de Saint-Hilaire (1816-1817)*. Mémoire de master - territoires, espaces, sociétés, École des Hautes Études en Sciences Sociales, Paris, 2019.

MALULY, Vinicius. Auguste de Saint-Hilaire e os territórios de exceção (Minas Gerais, 1816-1817). *PatryTer*, v. 3, n. 6, p. 266–280, 2020.

PRATT, Mary Louise. *Imperial Eyes: travel writing and transculturation*. London & New York: Routledge, 2003.

REUSCHEL, Anne-Kathrin; HURNI, Lorenz. Mapping Literature: visualisation of spatial uncertainty in fiction. *The Cartographic Journal*, v. 48, n. 4, p. 293–308, 2011.

SAFIER, Neil. *Measuring the New World: Enlightenment Science and South America*. Chicago: The University of Chicago Press, 2008.

SAINT-HILAIRE, Auguste. *Segunda viagem ao interior do Brasil: Espírito Santo*. São Paulo: Companhia Editora Nacional, 1936.

SAINT-HILAIRE, Auguste de. Voyages dans les provinces de Rio de Janeiro et Minas Geraes. In: *Voyages dans l'intérieur du Brésil*. Tome premier. Paris: Grimbert et Dorez, Libraires, 1830, p. 458.

Submetido em 25.10.2021 – Aceito 01.12.2021