

Associação Nacional de História – ANPUH
XXIV SIMPÓSIO NACIONAL DE HISTÓRIA - 2007

Uma proposta para o tratamento de fontes nominativas para o estudo da família: o exemplo dos registros paroquiais

Ana Silvia Volpi Scott*

Resumo: Um dos grandes desafios para os pesquisadores interessados em analisar a família numa perspectiva histórica diz respeito ao tratamento das fontes nominativas. Um dos problemas principais está ligado à identificação dos indivíduos e ao cruzamento nominativo das informações provenientes de fontes variadas. Esta comunicação pretende apresentar e discutir alguns procedimentos desenvolvidos ao longo dos últimos anos, com base em pesquisas que se exploram o uso dos registros paroquiais de batizado, casamento e óbito, que estão entre as fontes mais utilizadas nos estudos sobre a história da família. Além da discussão dos procedimentos em si, apresentaremos um programa informatizado desenvolvido para a coleta dos dados oriundos dos registros paroquiais, que está habilitado ainda para efetuar uma reconstituição semi-automática das famílias.

Palavras-Chave: Demografia Histórica; Metodologia e Fontes; Cruzamento Nominativo

Abstract: This paper presents a reflection about nominative record linkage, using luso-brazilian parish records. The procedures were developed considering the experience reached in many research projects, during the last decades. The paper presents a proposal of semi-automatic family reconstitution program.

Keywords: Historical Demography; Methodology and sources; nominative record linkage

A metodologia que deu as bases da Demografia Histórica, a Reconstituição de Famílias, nada mais é do que o cruzamento nominativo dos assentos de batizado/nascimento, casamento e óbito. Entretanto esse processo de cruzamento nominativo revela-se difícil e complexo, especialmente para os estudiosos das populações luso-brasileiras devido a uma série de problemas relativos à falta de regras para transmissão dos nomes de família, a alteração e/ou inversão dos nomes e sobrenomes, a ausência de nomes de famílias para a maioria da população feminina, à concentração na escolha de alguns nomes de batismo - tanto para homens como para mulheres -, e à alta incidência de homônimos.

Com base em estudos realizados ao longo dos últimos anos procurou-se desenvolver alguns procedimentos para facilitar e garantir uma maior eficiência no cruzamento de fontes nominativas, especialmente no que se refere aos registros paroquiais,

* Docente do Programa de Pós-Graduação em História – UNISINOS/ Pesquisadora do CNPq. Doutora em História e Civilização (IUE – Florença-Itália).

analisando-se duas maneiras distintas de se efetuar a coleta e o cruzamento dos dados, e quais são os ganhos e as perdas que cada uma delas pode trazer ao pesquisador.

A primeira maneira de se coletar os dados segue as disposições clássicas de Louis Henry, que propõe a coleta individual por ato e posterior cruzamento das informações; e a segunda que tem por base as proposições de Maria Norberta Amorim, que efetua o cruzamento e a identificação dos indivíduos no ato do levantamento.

Tivemos a oportunidade de trabalhar com os registros paroquiais para efetuar a reconstituição de famílias, tanto utilizando a metodologia clássica proposta por Henry/ Fleury, como a proposta metodológica desenvolvida por Maria Norberta Amorim, denominada Reconstituição de Paróquias, e que deu as bases para o SRP, Sistema de Reconstituição de Paróquias, atualmente em uso no Núcleo de Estudos de População e Sociedade (NEPS), da Universidade do Minho (Portugal).

A tese de doutoramento apresentada ao Instituto Universitário Europeu (Florença-Itália) em 1998, intitulada “Família, Formas de União e Reprodução Social no Noroeste Português” teve por base a paróquia de São Tiago de Ronfe, concelho de Guimarães, reconstituída através da metodologia de Maria Norberta Amorim. A base de dados proveniente do cruzamento dos assentos de nascimento/batizado, casamento e óbito, foi cruzada com um conjunto amplo e variado de outras fontes nominativas, e este processo complexo fez surgir nosso interesse por discutir os problemas enfrentados pelos pesquisadores, os procedimentos que podem facilitar a tarefa, e os desafios enfrentados no momento de identificar os indivíduos em fontes nominativas diversas.

Exatamente por conta da utilização de uma já paróquia reconstituída para a base de nosso estudo, desde logo discutimos com Maria Norberta Amorim as diferenças existentes entre a metodologia clássica e aquela que havia sido desenvolvida tendo por base os registros portugueses, onde é notória a ausência de regras de transmissão do nome de família, o que para muitos, significaria a impossibilidade de usar tais procedimentos para o estudo do passado das populações luso-brasileiras (ainda hoje esta afirmação é veiculada).

Diante disso, propusemos um desafio, quando resolvemos fazer uma análise comparada dos resultados obtidos a partir da utilização de metodologia clássica de Henry e a proposta por Norberta Amorim. Neste sentido, Dario Scott desenvolveu um programa informatizado que permitia a coleta dos dados em fichas de atos (nascimento/batizado/casamento/ óbito) utilizadas para posterior cruzamento automático para a reconstituição das

famílias propriamente ditas. Os resultados seriam, então, comparados àqueles obtidos através dos procedimentos propostos por Amorim. Os resultados desta comparação foram divulgados em artigo publicado em Portugal, e mostraram que ambas as metodologias produziram resultados virtualmente iguais, e que, portanto, a opção por uma ou outra, deveria ser determinada pelo próprio pesquisador, tendo em consideração sua habilidade e familiaridade em trabalhar com cada uma das metodologias, além é claro, dos objetivos que norteariam sua investigação.

Cabe aqui mostrar as diferenças entre os dois procedimentos. Na metodologia clássica, os registros são coletados em fichas individuais por ato, nascimento-batizado/casamento/ óbito. Na seqüência elabora-se um dossiê com todos os registros relacionados a cada casal, recebido em matrimônio, que terá uma *ficha de família* aberta, onde os dados sobre esta família biológica são lançados. Tal procedimento possibilita a elaboração dos cálculos relativos à fecundidade, mortalidade e nupcialidade. Repare-se que a identificação dos registros que correspondem aos indivíduos pertencentes a cada família biológica será feita apenas depois de coletados todos os atos em fichas separadas. Fichas coloridas identificavam os diferentes atos, na metodologia originalmente proposta por Henry..

O programa informatizado desenvolvido entre os anos de 1991 e 1992, “batizado” de NACAOB, permite ao pesquisador ter em mãos uma cópia fiel do documento manuscrito original. Partia dos atos individuais, com uma tela de inserção para os dados do batizado, casamento e óbito, vinculado a cada ato todos os indivíduos que haviam sido arrolados no mesmo, referindo o “papel” desempenhado por cada um e os eventuais laços que os uniam.

Importante é salientar que diversos procedimentos foram desenvolvidos para eliminar ou, pelo menos, a diminuir a incidência dos erros no momento da inserção dos dados. Assim criou-se, por exemplo, um dicionário de nomes, e uma série de tabelas codificadas: uma que definia os diversos tipos de laço entre os indivíduos, uma relativa à naturalidade, residência, ocupação de cada indivíduo, enfim, todas aquelas informações que ao serem codificadas, poupavam tempo no momento da digitação dos dados, e paralelamente reduziam os problemas de erros gerados pela digitação incorreta – imediatamente o computador registrava se o nome Maria, por exemplo, havia sido digitado de maneira errada. Além disso, as informações paralelas ou observações variadas encontradas nos registros sobre cada ato poderiam ser inseridas, pelo pesquisador, na ficha informatizada, relacionando-as a cada campo específico.

A partir do final do levantamento dos assentos, passava-se à fase de reconstituição automática das famílias, gerando-se uma ficha de família específica. Aqui, é que a questão da alteração dos nomes dos indivíduos coloca problemas maiores. Para esse procedimento é necessário que sejam definidos os **parâmetros** para o programa reconstituir os laços entre os diversos indivíduos, nos diversos atos. Se, por ventura, os nomes apresentarem diferenças, ou inversões, o programa não reconhecerá o “link”, provocando, então, a geração de um conjunto de fichas de família maior do que as que de fato existem. Pois se os parâmetros não têm uma coincidência de 100%, o programa não aceitará como válidos os vínculos, e gerará uma ficha de família nova. Assim, é fundamental definir quais serão os elementos fundamentais para identificar os indivíduos nos diversos atos.

Na metodologia desenvolvida por Amorim que, como a de Henry, era feita manualmente, partia do levantamento dos nascimentos/ batizados. Para cada assento deste tipo abre-se uma ficha, com as informações relativas ao batizado e ao casal. Com o levantamento subsequente, todos os batizados atribuídos àquele casal eram anotados na ficha respectiva. Diferentemente de Henry, Amorim usava fichas brancas, não pautadas, e as informações eram coletadas com canetas de cores diversas que indicariam de onde o dado era proveniente (azul – batizado/ verde-casamento/ vermelho-óbito).

O ganho de tempo era grande, uma vez que para cada ficha aberta era necessário acrescentar apenas as informações sobre os nascimentos consecutivos atribuídos àquela família. Ao final do levantamento dos batizados/nascimentos, a ficha de família estaria completa, uma vez que teriam sido identificados todos os indivíduos. A seguir passaria para o cruzamento com os assentos dos casamentos com a informação organizada a partir dos batismos, com todas as informações na cor verde. Se, eventualmente, um casal é identificado nos atos de casamento, que não havia aparecido nos batizados, abre-se uma nova ficha de família, com os dados grafados com caneta verde. A terceira fase é o cruzamento possível com os óbitos, dada a situação comum dos registros de óbito no que se refere à deficiente identificação do defunto. A passagem da reconstituição manual para a informatização da reconstituição de paróquias não mudou a essência do método, nem quando ela se realizava no ambiente DOS, nem hoje, quando ela se realiza através do Sistema de Reconstituição de Paróquias (SPR), em ambiente visual (FARIA & HENRIQUES, 2004).

O grande avanço da reconstituição de paróquias foi o desmembramento posterior da ficha de família, em ficha de indivíduos residentes, em encadeamento genealógico. Ao final do processo, uma paróquia seria considerada reconstituída quando, para cada residente,

temos uma ficha fechada, ou seja, quando caso a caso, dispõe-se de uma data de início de observação e a data de fim de observação (AMORIM, 1991).

Um dos argumentos que sempre foram utilizados por Maria Norberta Amorim, e que tem a sua vantagem de fato, é o ganho de tempo no momento da coleta dos dados, uma vez que não é necessário repetir a transcrição das informações sobre os indivíduos todas as vezes que eles aparecem em cada um dos diversos atos. No máximo, o pesquisador deveria fazer a complementação da informação nova coletada.

Isso significa que, a fase de coleta de dados nos assentos vitais é de fundamental importância, se utilizamos a metodologia desenvolvida por Amorim, hoje conhecida por SRP. Isso porque a **identificação dos indivíduos** se processa exatamente no momento em que o pesquisador está a trabalhar com a fonte. A decisão tem que ser tomada com base nas informações que ele dispõe naquele momento, quando tem o documento à sua frente. Isto quer dizer que a pessoa que faz a coleta dos dados tem uma responsabilidade enorme, pois uma vez efetuada a identificação, a informação será cadastrada, e assumida como correta.

No caso da opção por um levantamento individualizado, por atos, o **processo de cruzamento das informações e a identificação dos indivíduos** como parte desta ou daquela família, é **uma etapa posterior**, e a tomada de decisão se dá tendo em conta o conjunto de informações reunidas nos diversos assentos e ao longo do tempo.

Ora essa diferença metodológica pode não ser tão comprometidora quando trabalhamos individualmente num projeto de pesquisa próprio, para o qual temos o conhecimento do conjunto da documentação, das condições das fontes no momento da coleta, da dificuldade que enfrentamos com a leitura paleográfica (mesmo que se trata de documentos já do século XIX) enfim, com as condições gerais em que a recolha dos dados foi efetuada.

No entanto, se pensamos num estágio de pesquisa em que se pretende estender a abrangência das análises, para uma região ou eventualmente, num projeto ambicioso, com uma abrangência nacional, a situação muda de figura, pois não há mais o domínio do pesquisador sobre o universo de fontes exploradas e da comunidade a ser estudada.

Será necessário que uma equipe seja treinada nos procedimentos específicos, nas variáveis a serem consideradas para a que as decisões sejam tomadas, uma vez que ao identificar um indivíduo, aqueles que vão trabalhar com as bases de dados não terão a documentação original na mão para confrontar a pertinência daquele “link”.

Além disso, embora não haja a necessidade de se copiar “todos os dados, todas as vezes” será necessário despendar algum tempo para realizar a consulta ao banco de dados (considerando-se que estamos a utilizar o SRP), e verificar se devemos abrir uma nova ficha ou, diante das diferentes opções de fichas de família já abertas, inserir o indivíduo nas famílias já existentes. Esse tempo pode ser bastante alargado, dependendo do tamanho do banco de dados que estamos a trabalhar.

No projeto de pesquisa que desenvolvemos atualmente¹, que utiliza os registros paroquiais para uma comunidade portuguesa entre as décadas de 1860 e 1880, tivemos a oportunidade de testar as duas opções metodológicas e fazer diversas experiências com os dados coletados.

Quando iniciamos o projeto, há cerca de dois anos, pouco mais, não tínhamos tido acesso ao SRP², e acabamos por optar por utilizar o programa NACAOB, que havia sido desenvolvido no início dos anos 1990.

Esta decisão foi tomada levando em consideração dois fatores:

- a) a qualidade dos manuscritos era muito boa uma vez que no arquivo paroquial da Lousã pudemos ter acesso aos extratos dos registros paroquiais, que haviam sido compilados para livros específicos no início do século XX;
- b) que os dados seriam inseridos no programa por pessoas sem treinamento específico seja na leitura paleográfica, seja nas questões metodológicas e técnicas envolvidas no cruzamento nominativo de fontes e, desta forma, a utilização dos extratos agilizaria a etapa de coleta de dados, pela facilidade de leitura, o que levaria a um ganho de tempo também.

Assim iniciamos a digitação dos dados no NACAOB, reunindo informações para uma população natural da Vila da Lousã, situada no concelho da Lousã, distrito de Coimbra.

Todos os dados relativos aos nascimentos/ batizados e aos casamentos foram coletados, digitados e inseridos no programa, e somaram 4.547 batizados e 1.040 casamentos. Como esta metodologia pressupõe a inclusão de todos os indivíduos registrados em cada ato,

¹ A imigração invisível: portugueses no Brasil independente, financiado pelo CNPq (Ana Silvia Volpi Scott (coord.) e Oswaldo Mario Serra Truzzi).

² Para uma discussão mais aprofundada sobre os procedimentos desenvolvidos no NEPS nos últimos anos em torno da problemática da reconstituição de paróquias veja-se Faria & Henriques 2004.

ao final da coleta reunimos 37.611 nomes de indivíduos que se revezaram nos diversos papéis: pais, mães, avós, padrinhos, testemunhas, isto é todos os indivíduos que foram arrolados ao lado das crianças batizadas e dos homens e mulheres que se casaram na igreja paroquial da vila da Lousã.

A seguir, depois da inserção dos dados nas fichas de ato, rodamos o programa que gera as famílias através do cruzamento dos registros de batizado e casamento (faria também o cruzamento com os óbitos se eles tivessem sido coletados). O resultado para a Lousã foi um total de 2.736 famílias cruzadas de maneira automática, sem a intervenção do pesquisador/ usuário do programa. O cruzamento foi efetuado tendo como variáveis a serem consideradas o nome da avó materna, a noiva (mãe no batizado) e o noivo (pai no batizado). Ocorre que devido, ao problema conhecido de todos os pesquisadores das populações luso-brasileira, da variação/ inversão dos nomes dos indivíduos, uma quantidade destas famílias reconstituídas automaticamente teriam sido geradas por conta deste problema, sendo na verdade resultado da variação dos nomes encontrados para os mesmos indivíduos, nos diversos atos cruzados.

Vejamos o que ocorreu com a família de Manuel Francisco, casado com Rosa Conceição. Na identificação automática encontramos quatro famílias em que o casal referido aparecia. Manuel Francisco foi registrado no seu casamento como Manuel Francisco Rita. A mãe de Rosa Conceição, mulher de Manoel Francisco, foi identificada com três variações de nome no conjunto dos 13 filhos que o casal gerou na Lousã: foi registrada como Ana Jesus, Ana Conceição, e Ana Maria. Isto é, o cruzamento automático jamais poderia supor que estes nomes dissessem respeito à mesma mulher. Neste caso a variação ocorreu no nome da mãe da noiva, mas houve casos em que mais de um indivíduo teve alterações, criando dessa forma inúmeras famílias, que na verdade diziam respeito a um único casal.

Estas mudanças no nome dos indivíduos geram, portanto, um conjunto de famílias duplicadas, que caberá ao pesquisador, na fase seguinte do processo, identificar e eliminar.

No caso da Lousã, que partiu de 2.736 famílias reconstituídas automaticamente, depois de toda a correção feita, este número havia caído para 1.762. Foram eliminadas 974 famílias.

Esses números mostram que a reconstituição automática acertou em 64,4% dos casos, gerando um re-trabalho em 35,6% das famílias. Esta operação é facilitada porque temos uma série de outras informações que são confrontadas para que a decisão final sobre a identificação seja tomada. A operação se processou com a importação dos dados gerados pelo

NACAOB para uma planilha, na qual pudemos reorganizar a informação sob diversos critérios, e assim eliminar os casos de famílias duplicadas ou triplicadas.

Para se ter uma idéia do problema enfrentado, mais de 55% das mulheres da amostra foram batizadas com o nome de Maria (ou combinações várias com Maria). Além disso, a esmagadora maioria delas não carrega o nome de família, sendo conhecidas através de invocações (da Piedade, do Espírito Santo, de Jesus, do Rosário, da Conceição etc.). Para tornar as coisas mais difíceis, vimos que, com muito freqüência os nomes poderiam ser alterados ou invertidos: Ana de Jesus, que vira Ana da Piedade, Rosaria Serra que se transforma em Rosária de Jesus, assim como as Joanas Marias podem virar Marias Joanas. Isso também pode ocorrer com os homens, embora com menor intensidade. Para eles as situações mais comuns diziam respeito à inclusão ou supressão de expressões como Velho, Novo, Sênior e Júnior, de alcunhas ou nome de família.

A única saída é a utilização de outros atributos identificadores, que se tornam fundamentais para dirimir estas dúvidas. Tais atributos se juntariam aos nomes pelos quais os indivíduos foram sendo reconhecidos ao longo do tempo: naturalidade, local de residência. Até mesmo o intervalo de nascimento entre os filhos dos casais, no nosso caso, tornou-se uma ferramenta de identificação válida, pois o ritmo inter-genésico era de tal maneira regular na população estudada, que muitas vezes auxiliava a identificar famílias duplicadas ou triplicadas, exatamente pela falta de um nascimento entre outros dois anteriormente registrados.

Levando em consideração isso tudo é que nos perguntamos se a utilização de uma metodologia que exige que a identificação seja feita no ato da coleta (o SRP, por exemplo), não pode induzir a julgamentos apressados, especialmente quando quem faz a coleta não tem a sensibilidade e o conhecimento suficiente dos problemas que estão por trás desta etapa fundamental da reconstituição de família/ paróquias.

O NEPS utiliza com sucesso esta metodologia do SRP, e inclusive disponibiliza *on line* as paróquias que já foram reconstituídas. Ocorre que este material foi organizado principalmente de duas maneiras: 1) ou a reconstituição foi feita pelos próprios pesquisadores (comumente alunos do Mestrado em História das Populações; ou 2) a reconstituição é feita pelos funcionários do NEPS já largamente treinados na metodologia manual que foi a base do sistema informatizado que está sendo atualmente utilizado.

Aqui no Brasil os pesquisadores que trabalham nesta área, normalmente o fazem de maneira isolada, sem ter um grupo mais amplo para discutir e desenvolver procedimentos padronizados para a execução desta tarefa. Praticamente, todo estudo que emprega a reconstituição é individualizado e cada um desenvolve procedimentos adaptados a seu caso específico. Cada pesquisador desenvolve a sua “ficha”, “máscara de inserção” de dados em programas como Access ou Excel. Mas a maioria não avança na direção de procedimentos automatizados de reconstituição. Também há aqueles que adaptam programas já desenvolvidos para outras realidades (por exemplo o SYGAP), mas a adaptação nem sempre é fácil.

Na verdade estas reflexões nasceram dos problemas que enfrentamos no projeto de pesquisa que estamos a desenvolver, problemas que não são apenas “nossos”, mas que são compartilhados por todos os utilizadores interessados nestes tipos de metodologias.

A questão fundamental, que queremos discutir, é exatamente o fato de que esta identificação é feita com base no **nome**, uma informação que sabemos é passível de muita modificação. O que isso provoca? Se eu tenho que identificar uma mulher, que aparece como mãe num batizado, que se chama Maria Joaquina, o programa vai me mostrar todas as mulheres com este nome. Uma infinidade delas. Mas eu posso acrescentar outros atributos que me ajudem na identificação, por exemplo, local de nascimento, a legitimidade, data de nascimento ou casamento, ou óbito ou nascimento do primeiro filho. Ou seja, utilizando-se um programa como o SRP, diante do conjunto de Marias Joaquinas, aquele que coleta a informação (pesquisador ou bolsista) deve fazer a sua opção. Além do tempo gasto para identificar, se o trabalho fosse feito por bolsistas, será que o procedimento teria confiabilidade?

Todas estas experiências mostram que, por certo, os procedimentos *totalmente* automatizados não são viáveis para o caso luso brasileiro, mas eles podem constituir uma ferramenta poderosa para auxiliar o pesquisador na difícil tarefa de superar os problemas inerentes ao cruzamento nominativo.

Diante disso é que trazemos estas reflexões para o simpósio temático, para que possamos em conjunto pensar soluções e procedimentos que possam abarcar uma realidade que vá além dos nossos próprios projetos de pesquisa individuais, e que dêem suporte para “vôos” mais ambiciosos.

Referências Citadas: