

# **VERBA MANENT: A MEMÓRIA DA ORALIDADE \***

**Rita Marquilhas**

**Congresso Internacional VERBA VOLANT?**

**Oralidade, escrita e memória**

**Faculdade de Filosofia de Braga  
Universidade Católica Portuguesa (13-15 Nov 2014)**

Chamei à minha intervenção VERBA MANENT não porque queira falar especialmente de oralidade. Quero falar especialmente de história da escrita, que é a minha área de investigação, mas a propósito da escrita, quero lembrar que ela é tão avassaladora na forma como influi, uma vez adquirida, na nossa perceção de fenómenos linguísticos, sociais ou culturais, que nos cega para muitos aspetos do comportamento mais universal das línguas. Cega-nos para a diferença entre língua e cultura, para a diferença entre língua natural e língua artificial, para a diferença entre história das línguas e memória das línguas.

Quero lembrar que podemos dizer, a propósito de uma série de observações, que a **fala permanece**, a minha tradução -- não muito feliz, reconheço-o -- para a expressão VERBA MANENT.

[1a] Em primeiro lugar, podemos observar que a **fala permanece na natureza**. Aqui permanece porque é uma propriedade dos seres humanos.

Por natureza, os humanos falam, tal como as aves voam ou os peixes nadam: desde há 100 mil anos que a fala permanece, se for verdade que é há 100 mil anos que existe o HOMO SAPIENS. É por isto também que as línguas, entendidas como línguas naturais, também permanecem, e só se extinguem se por catástrofe repentina desaparecer todo o seu povo. Caso contrário, mudam, podem até receber outro nome, mas permanecem. Podem também dar origem a outras, em condições de contacto muito radical, como se observa nas grandes conquistas, como a da Romanização da Antiguidade, ou nas línguas de formação rápida da Idade Moderna, conhecidas pelo nome de crioulos. Os indivíduos entendidos como detentores desta fala natural são designados por falantes e são objeto de estudo por parte da linguística, que lhes observa o comportamento gramatical partindo depois para uma série de abstrações, conforme a sub-disciplina e a orientação teórica.

[1b] Em segundo lugar, a **fala permanece** na memória, uma memória coletiva constituída pela tradição oral dos povos. Este é um nível de permanência diferente do anterior porque diz respeito a textos transmitidos geracionalmente, verbalizados de forma mais ou menos fidedigna na celebração de rituais, na narração de mitos, na formulação de crenças. Enquanto a anterior é a fala das línguas naturais, esta é uma fala de língua de cultura, invariavelmente caracterizada por alguma elaboração formal. Não é partilhada igualmente por todos, só pelos

iniciados, e contém léxico e estruturas gramaticais arcaicos e cristalizados. (O pão nosso de cada dia nos dai hoje versus estrutura moderna *Dê-nos hoje o nosso pão de cada dia*). Aqueles que na comunidade possuem o conhecimento destes textos, para os compreendermos bem, devemos descrevê-los enquanto atores sociais e contrastá-los com os outros, cujo diferente percurso os segregou do mesmo conhecimento. Se os quisermos considerar falantes, devemos ignorar-lhes estes textos do conjunto dos seus enunciados relevantes.

[1c] Em terceiro lugar, a **fala permanece na escrita**. Não estou aqui a falar da escrita plenamente desenvolvida, mas de uma outra escrita, a inábil. Como os atores sociais têm, como acabei de referir, diferentes percursos e diferentes graus de integração cultural, alguns, sobretudo nas sociedades pré-industrializadas, ficavam em patamares incipientes da aprendizagem da escrita. Se, por casualidade, puderam deixar documentação gráfica, podemos também nós procurar nos seus textos um testemunho que já pode ser novamente, mas indiretamente, o de falantes de uma língua natural. Encontraremos aí respostas sobre a oralidade do passado, se lhes fizermos as perguntas certas. O que se observa neste tipo de informantes é que eles usam a tecnologia da escrita da mesma forma que a usam as crianças em fase de aprendizagem e a usaram os inventores ou adaptadores de sistemas gráficos (há 5.000 anos na Mesopotâmia, há 2.900 anos na América Central / outros sistemas serão independentes, ou

não? Egito 3400 aC., China 1200 aC, Maias 200 / 300 aC, Gregos 730 aC). Usam-no como instrumento de reflexão e explicitação do seu conhecimento inconsciente da linguagem.

Um exemplo: o vocalismo átono em português. Vêem aqui um exercício de escrita executado por uma criança de 5 anos, idade em que já sabe os valores das letras, sabe assinar o nome, que tem várias consoantes (Beatriz), mas está estacionada no que os psicolinguistas chamam a fase silábica. Põe a hipótese de o alfabeto servir para escrever os núcleos das sílabas do português, ou seja, as suas vogais. Não separa palavras (Jogos Olímpicos); e escreve de maneira quase sistemática o [u], o [6] e o [@] não acentuados. Dá também conta de um fenómeno típico dos falantes de Lisboa que é o da diferenciação do ditongo [ej] no [6j], mais dissimilado. É em torno desse tipo de texto que me tenho especializado. Falarei um pouco de como se pode encontrar, coligir e tratar este património, distinguindo entre o seu significado linguístico e o seu significado cultural.

Antes de prosseguir, deixem-me no entanto fazer um desvio sobre a relevância deste tipo de estudo, para o caso de poderem pensar que são apenas anedotas do passado, curiosidades, que assim se procuram.

O estudo dos enunciados produzidos por informantes pouco escolarizados tem relevância porque permite simular inquéritos a informantes de épocas que já não podemos questionar diretamente por pertencerem ao passado.

Podemos junto deles confirmar a envergadura da mudança linguística que se inscreveu entre o seu tempo e o nosso com muito a segurança de um inquérito sociolinguístico.

Não é sociolinguística do presente, como a que foi fundada por William Labov, é uma sua adaptação ao estudo histórico da língua em sociedade, ou seja, a sociolinguística histórica.

A constatação feita pela sociolinguística laboviana de que a mudança linguística ocorre nas gramáticas grupais das comunidades de fala, liderada por falantes pertencentes à secção central da hierarquia socioeconómica (Labov 2001, 500) constitui um importante desafio para a sociolinguística histórica.

Por um lado, a evidência empírica para uma comunidade de falantes do passado exige investigação arquivística muito direccionada para registos numerosos, bem distribuídos em termos geográficos, cronológicos e socioeconómicos, além de textualmente variados, com destaque para o registo conversacional.

Por outro lado, a captação das hierarquias socioeconómicas do passado e das respectivas secções – altas, centrais ou baixas – não pode ser anacrónica. Exige, por conseguinte, um elevado grau de desenvolvimento da historiografia relativa às comunidades sob escrutínio, atenta à sua composição demográfica e à sua cultura.

Acontece que isso se tem efetivamente verificado no que diz respeito à história social das comunidades ocidentais, como destacam os autores do recente *Handbook of Historical Sociolinguistics*: O interesse dos investigadores de história social pela estrutura dos grupos e das comunidades do passado, juntamente com a reconstrução de estruturas demográficas e socioeconómicas, facilitou muito a tarefa do sociolinguista histórico. (Conde-Silvestre and Hernández-Campoy 2012, 3).

Para Portugal, por exemplo, dificilmente se pode já falar de falta de dados para o estudo das suas comunidades dos séculos XVI a XIX (Idade Moderna). Não só a história da população como a própria história cultural têm sido progressivamente revisitadas e refinadas desde a década de 1980.

No caso da história social de Portugal é possível ter boa noção do que será uma amostra representativa de informantes da etapa Moderna, necessária para captar a pertinência de variáveis históricas independentes (estruturas socioeconómicas, redes de relações interpessoais, comportamentos diferenciados relacionados com o fator sexo, etc) no processo de variação e mudança da língua da época. A ideia dos volumes da população portuguesa, baseada tanto em estimativas como em recenseamentos, é bem concreta, como se pode verificar no quadro X:

Years	Continental Portugal	Islands	Total
1530	1.120.000		1.120.000
1580	1.200.000		1.200.000
1620	1.200.000		1.200.000
1640	1.900.000		1.900.000
1700	2.050.000	125.897	2.175.897
1732	2.143.368	159.921	2.303.289
1768	2.409.698		2.409.698
1801	2.931.930	283.400	3.215.330
1821	3.026.450		3.026.450
1835	3.061.684		3.061.684

Table X Portugal, volumes de população entre os séculos XVI e XIX --  
 excerto de um quadro cronologicamente mais amplo publicado em  
 Rodrigues (2009, 519)

O historiador da língua em sociedade pode combinar este tipo de estatística com o critério, sugerido por Labov, de recolher amostras linguísticas que correspondam a, pelo menos, 0,025% da população de falantes a investigar (Labov 1966, Hernández-Campoy & Schilling 2012: 67 – citar bem). Os números a que se chega, na tentativa de obtenção de uma amostra suficientemente representativa de informantes linguísticos, divididos por períodos razoavelmente estreitos, de 50 anos, não são nada inatingíveis, como se vê pelo quadro XX:

Time interval	Number of necessary
---------------	------------------------

	informants
1551-1600	300
1601-1650	470
1651-1700	500
1701-1750	530
1751-1800	600
1801-1834	780
Total	3,180

Table XX – Amostra ideal em termos de validade para representar o português europeu falado no período Early Modern

Formada esta ideia da amostra ideal, passa a ser preciso ir no encalço daqueles cerca de 3,200 informantes diferentes, fidedignos, com registos informais, tendencialmente conversacionais, que tenham ficado documentados por escrito. E eles efetivamente existem. Como foi investigado entre 2008 e 2010 pelo projeto CARDS, Cartas Desconhecidas (Marquilhas 2012), e agora pelo projeto P.S. Post Scriptum que lhe sucedeu (Marquilhas et al. 2013), há documentos portugueses inéditos e devidamente contextualizados, conservados em número suficiente para concretizar a constituição desta amostra sociolinguística histórica.

O projeto P.S. Post Scriptum, do Centro de Linguística da Universidade de Lisboa, busca, edita e explora documentos da interação privada: são cartas particulares portuguesas e espanholas (mas aqui só nos referiremos ao caso das portuguesas, dado o tema do artigo).

As cartas do *Post Scriptum* correspondem a manuscritos inéditos. São



fontes extraordinariamente ricas para a história das línguas portuguesa e espanhola faladas quotidianamente entre o século XVI e o início do século XIX. Na sua maioria foram dirigidas a familiares, amigos, amantes ou apenas a conhecidos e foram arquivadas por tribunais civis e eclesiásticos dos dois reinos juntamente com documentação que esclarece o seu contexto situacional. Esclarece igualmente os perfis sociológicos de quem as escreveu ou recebeu, uma vez que o que interessava ao tribunal era o valor incriminatório das cartas. O seu discurso provém de estratos sociais muito diversificados: escreveram-nas e receberam-nas homens, mulheres, e até crianças; uns eram ricos, outros pobres, e de todos os lugares da Península Ibérica, bem como dos domínios ultramarinos português e espanhol. Sabemos também, muitas vezes com pormenores tirados do depoimento de réus ou de testemunhas, quem eram mesmo aquelas pessoas.

Pretendemos editar e estudar 7,000 dessas cartas. Para o efeito, estamos a preparar uma *edição crítica digital* dos manuscritos e seu contexto; paralelamente, estamos a converter o conteúdo das cartas em dois *corpora anotados* de um milhão de palavras cada, um com as cartas portuguesas, outro com as espanholas.

Identificada a documentação, torna-se necessário tirar proveito de todas as inovações possíveis em termos de filologia, de humanidades digitais e

de linguística de corpus para a potenciar a captação do seu significado cultural, social e linguístico.

A edição destes manuscritos originais de autor tem de partir, por conseguinte, de uma base de transcrição paleográfica que dê origem, depois, a uma edição crítica genética, reconstituidora do processo de escrita. Tem de adotar o formato mais universal possível em termos de edição digital, ou seja, o XML-TEI. Tem de alimentar uma base de dados biográfica (de autores e destinatários) cuja informação possa ser cruzada com os conteúdos textuais das cartas que escreveram ou receberam e com a situação comunicativa em que participaram. Mediante o uso de *automatic tools*, as word strings da transcrição paleográfica têm de ser fragmentadas em unidades gráficas – tokens – alinhadas uma a uma com palavras ortográficas modernas, com lemas, e com categorias mais abstratas, tanto morfossintáticas, como léxico-semânticas. Desejavelmente, também, as suas unidades sintáticas devem dar origem a uma *tree bank* compatível com as que já existem para outros textos históricos, quer portugueses, quer de outras línguas românicas.

Este é, resumidamente, o programa do projeto P.S. Post Scriptum, em que estão envolvidos linguistas, historiadores e engenheiros de sistemas, e que estará concluído em 2017. Para a perspetiva da sociolinguística histórica, interessa o que no P.S. se encontre sobre factores internos e externos de variação e mudança ao longo de quatro séculos de português.

Vários resultados parciais do projeto estão já acessíveis. Veja-se este primeiro, para demonstração.

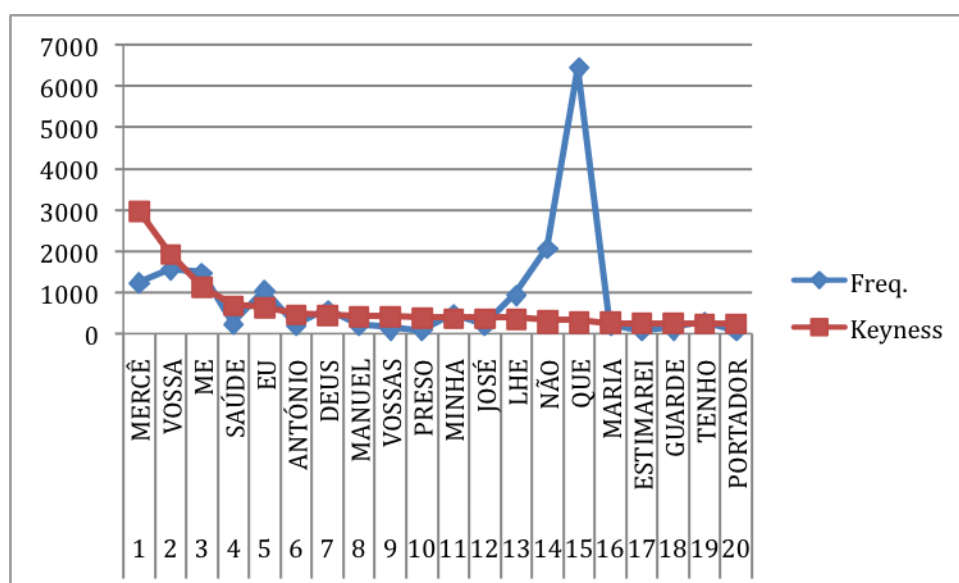
Tomando uma amostra de 402 indivíduos portugueses diferentes, oriundos de todo o reino, três quartos homens e um quarto mulheres, vivendo entre finais do século XVI e inícios do século XIX, percebem-se quais eram as principais fontes de variação fonológica da época: no consonantismo, o *seseo* e o *ceceo*; no vocalismo, a elevação das vogais átonas prétonicas.

O diagnóstico sobre a importância dada pelos falantes da época Early Modern a estes dois fenómenos de variação fonológica consegue-se mediante recurso à ferramenta DICER, *Discovery and Investigation of Character Edit Rules*, desenvolvida por Alistair Baron para a análise estatística de variantes gráficas cruzadas com factores extralinguísticos (Baron 2011).

Um segundo tipo de análise estatística já possível sobre os dados do Post Scriptum é constituído pela avaliação do seu léxico-chave em comparação com corpora de referência. Neste caso, os dados de entrada têm de estar codificados num formato totalmente estandardizado no que diz respeito à variação gráfica. Os vestígios da transcrição paleográfica têm que ser temporariamente apagados porque criariam contrastes muito visíveis e pouco significativos com os dados dos padrões de referência, também eles ortograficamente normalizados.

O recurso a programas de análise lexical como o WordSmith Tools (Scott 2011) permite a busca do léxico chave e clusters chave no corpus Post Scriptum em comparação com dois tipos de padrão: o de falantes da mesma época mas mais cultivados e autores de um discursos não conversacionais; e o de falantes culturalmente e discursivamente comparáveis, mas vivendo no século XX. Neste caso apenas haverá ocasião de exemplificar o primeiro caso, de contraste com a elite literária da mesma época.

O contraste entre uma amostra de cartas do Post Scriptum muito equilibrada em função do género (93 mulheres e 151 homens) vivendo entre o final do XVI e o início do XIX, e o discurso de falantes todos eles cultos, autores literários, dos mesmo séculos, autores coligidos pelo corpus Tycho Brahe já acima referido, apresenta o seguinte léxico-chave para o Post Scriptum:



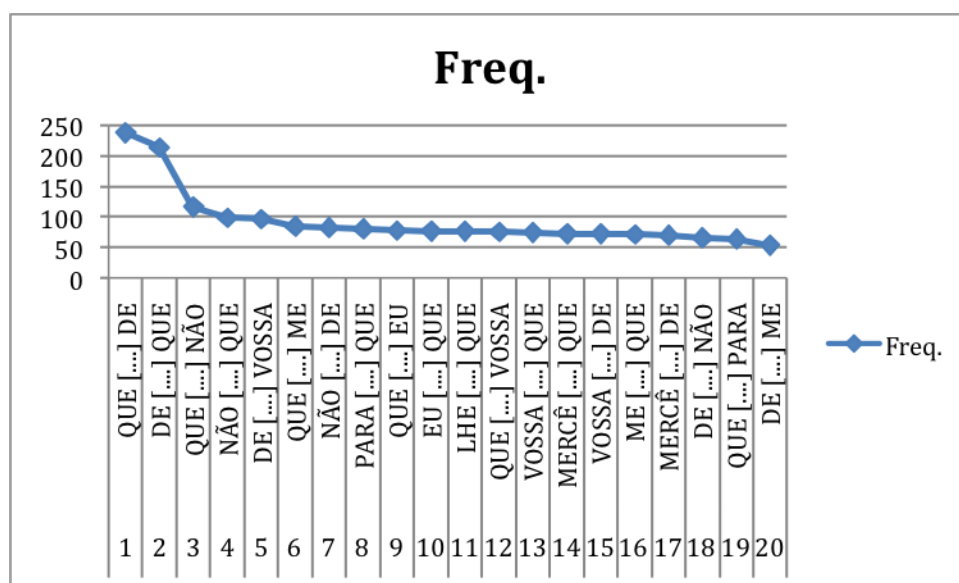
O que se pode concluir que sobressai neste tipo de contraste envolvendo só léxico chave de topo, é que os factores determinantes de variação são, aqui, as marcas de género epistolográfico e de registo conversacional. Seguem-se-lhe as contingências do contexto histórico-situacional, explicitamente lexicalizado na carta.

As palavras mais salientes, consideradas isoladamente, são as do léxico da referência à primeira e à segunda pessoa, quer nominal, quer pronominal, e as dos clichés da *politeness*, com referência à saúde e a Deus.

A relação com o contexto situacional – são tudo cartas, e muitas delas relacionadas com crimes – essa aflora na frequência key das palavras *preso* “prisoner / arrested” e portador.

Só o operador de negação “não” sai do quadro do imediatamente expectável, mas terá sobretudo a ver com as particularidades do registo textual dominante no corpus de referência Tycho Brahe, cujos autores, praticamente todos eles personagens eminentes e estilisticamente *accomplished*, se dirigiam aos leitores para partilhar mais a autoridade das suas concepções do que proposições exprimindo dúvida, medo, ignorância ou ordem, particularmente necessitadas, ao nível sintático, de um operador de negação. E isso já é fenómeno chave da conversação.

Quanto à alta frequência e ao valor chave da palavra ‘que’, esse advém-lhe de dois factores, que se observam melhor olhando para um quadro diferente, com as palavras-chave organizadas em clusters e ordenadas por ordem de frequência:



A forma “que”, por um lado, pode pertencer em português a muitas diferentes categorias funcionais: complementador, conjunção subordinativa, pronome relativo, pronome interrogativo e conector discursivo. Por outro lado, ocorre como operador nas frases clivadas, servindo para focalizar diferentes constituintes sintáticos. Tem pois um papel insubstituível na obtenção de coesão textual, sobretudo a coesão com enunciados ocorridos imediatamente antes, como acontece de forma *pervasive*, mais uma vez, no tão comentado registo conversacional.

Terminarei com duas secções de ordem mais geral, contendo reflexões metodológicas sobre a hesitação e a decisão na recolha deste tipo de dados, típicos daquilo a que se chama a história vista de baixo.

A busca de documentação tendencialmente conversacional e privada para os tempos da Idade Moderna não pode orientar-se pela mesma conceção da divisão entre público e privado que é tomada como válida para as sociedades modernas. O risco de anacronismo é enorme, com explica Pedro Cardim:

No período moderno a divisão entre público e privado ou íntimo era muito diferente da de hoje.

No entanto, nos tempos pré-modernos, anteriormente à formação de uma “esfera pública burguesa” tal como viria a ser teorizada por Jürgen Habermas (1989 [1962]), havia obviamente as suas oposições entre diferentes espaços para a interação em sociedade. Podia-se contar, pelo menos, com toda aquela encenação que é posta em marcha em qualquer sociedade, e em qualquer época, para que a comunicação funcione. Tal encenação, ensinou-nos Erving Goffman, tem dois diferentes espaços, que implicam a assunção de papéis também eles diferentes por parte dos atores sociais (Goffman 1990 [1959]). São os espaços do *palco* e dos *bastidores*. Se o palco é o patamar do fingimento em equipa para benefício de um público que espera convenções muito codificadas, os bastidores são o espaço da representação mais distendida, junto de atores

sociais com os quais se partilha a sensação de confiança advinda da pertença à mesma *equipa*. Com eles pode-se até ser menos *conversador*, pode-se ridicularizar a plateia da sala, podem-se assumir comportamentos espontaneamente emotivos, mais violentos do que cordatos, pode-se conspirar em benefício de representações futuras. Os bastidores são também o espaço onde se podem infiltrar elementos estranhos, vindos da plateia, e que aqui se poderão movimentar se cuidarem bem do seu disfarce.

Todas estas metáforas dramáticas do método proposto por Goffman para se compreenderem as nuances do comportamento comunicativo dos atores sociais se tornam muito úteis na seleção de documentos do passado por parte de um linguista que busque o registo conversacional de uma comunidade linguística antiga.

Interessam-lhe cartas familiares de todo o tipo, sem dúvida, já que o ambiente doméstico é o enquadramento por excelência da representação dos bastidores. Mas o modelo de família, é preciso não esquecer, era muito mais alargado do que o modelo nuclear que triunfou aquando da modernização ocidental. Incluía muito mais indivíduos do que aqueles que partilhavam afinidades de sangue; incluía, por exemplo, também os criados (Ariès and Duby 1986).

Interessam igualmente o registo em que a carta ficou escrita porque certos registos só aparecem nos bastidores. Interessam, por conseguinte, cartas



iradas, cartas amorosas, cartas obscenas, cartas anónimas – extorsionistas ou vingativas –, cartas indiscretas, cartas intriguistas, desde que se estabeleça com alguma segurança que autor e destinatário confiavam suficientemente um no outro para assegurarem em palco, se necessário fosse, uma representação a dois diferente da que era assegurada pela troca da carta. Neste sentido, o facto de se lidar com documentação bastante contextualizada do ponto de vista situacional, como é o caso da documentação judicial, é muito útil.

O facto de terem sido escolhidos textos dos bastidores como fontes para o corpus Cards-Post Scriptum, tal como descrito no apartado supra, tem conduzido à constituição de um grupo de sujeitos relativamente heterogéneo, enquanto punhado de informantes da época portuguesa Early Modern. Mas coloca-se logo outra pergunta: a partir de que momento é que a heterogeneidade relativa de uma amostra histórica se torna em heterogeneidade suficiente? Seriam também elegíveis, todos estes falantes, se pudéssemos fazer um inquérito sociolinguístico viajando até ao passado, ao tempo em que eles viveram? Não, ou pelo menos, nem sempre.

Uma seleção de informantes dependente da atuação de tribunais, ora preocupados com a defesa da ortodoxia católica, como era o caso da Inquisição, ora com a defesa da ordem social, como era o caso da justiça civil, apanha demasiados atores sociais que, no seu tempo, eram

personagens consideradas suspeitas. Entre esses suspeitos, estavam os estrangeiros e os cristãos novos.

Os estrangeiros, obviamente, são maus informantes para quem busque fundamentação empírica para as estruturas de uma língua que para eles foi já só língua segunda. Com imaginação, estes estrangeiros que escreveram cartas portuguesas e que foram alvo de investigação judicial, podem ser usados em estudos sobre aquisição do português na história, mas o estudo da variação sociolinguística precisa de falantes nativos enquanto informantes.

Já os cristãos-novos, esses têm o problema de terem formado um grupo com demasiada coesão dos pontos de vista étnico, religioso e socioeconómico. Seriam as características do enunciado de uma cristã-nova mais influenciadas pelo factor extralinguístico “género”, ou pelo factor “grupo étnico”?

Não nos parece, no entanto, que a documentação estatisticamente ou gramaticalmente menos boa deva ser descartada. Ao mesmo tempo, enquanto conjunto de fontes para a história da composição da sociedade portuguesa, ela é importante. Se os estrangeiros e os cristãos-novos que viviam em Portugal e escreviam em português na época Early Modern eram perseguidos enquanto suspeitos pelos tribunais, era porque estavam mesmo no território e falavam português. São importantes para se conhecer a composição da sociedade, razão pela qual voltam a ser

indispensáveis os seus enunciados em termos de *data* para a perspetiva sociolinguística. Este paradoxo lembra um pouco o aforismo de Labov, sobre fazer “o melhor uso possível de maus dados” nos trabalhos históricos e torna claro que o que pode ser bom ao mau não são os dados, mas a abordagem que a eles é feita por parte do estudioso. A distinção entre micro e macro sociolinguística histórica traz uma ajuda para resolver tal paradoxo. Os dados históricos que não têm qualidade ‘técnica’ suficiente para servirem de base empírica à medição fina de variáveis linguísticas, sejam elas internas ou externas, podem ainda assim servir de base à formação de conceitos de natureza sociocultural sobre a comunidade em observação.

\* Funded by the European Research Council, ERC Advanced Grant 2011, GA 295562.