

**“AO VIVO” E “LEGENDADO”:  
DA TRANSCRIÇÃO À DIGITALIZAÇÃO DE DADOS  
DO ATLAS LINGÜÍSTICO SONORO DO PARÁ**

*Orlando Cassique*  
Universidade Federal do Pará

- **RESUMO:** *Este trabalho apresenta, a partir da experiência de elaboração do Atlas Lingüístico Sonoro do Pará, a discussão sobre a possibilidade de evitar a intermediação representada pela transcrição e de oferecer à análise do lingüista o material coletado tanto mais “ao vivo” quanto possível, através de sistemas informatizados, ou, o que poderia ser melhor, colocar à disposição dos interessados esse mesmo material simultaneamente à transcrição.*
- **PALAVRAS-CHAVE:** *Geo-Sociolingüística; Atlas Sonoro; Sonorização de Dados; Transcrição.*
- **ABSTRACT:** *This paper discusses, based on the experience of Atlas Lingüístico Sonoro do Pará composing, the arguing about possibility of avoiding the mediating represented by transcription and of giving to the linguist's analysis the collected material as alive as possible, through computing systems or, what would be better, make the material available at same time of transcription.*
- **KEY WORDS:** *Geo-sociolinguistics; Sonorous Atlas; Sets of Sonorization; Transcription.*

## **1 INTRODUÇÃO**

Não há consideração, por mais breve que seja, a respeito da lingüística que deixe de se referir à língua falada como centro de atenção:

(...) a maioria dos atuais lingüistas modernos aceita axiomáticamente que o falar é fundamental e que a linguagem es-

crita é secundária e derivada[...]o som (e, mais especificamente, a gama de sons que podem ser produzidos pelos chamados 'órgãos vocais') é o meio a que a linguagem está incorporada [...] as línguas escritas resultam da transferência da fala para o meio secundário, visual. (...) (Lyons (1976))

Essa tomada de consciência no sentido da preeminência da língua falada começa, nunca é demais lembrar, em Saussure (1975). Com efeito, foi o mestre de Genebra quem, ao afirmar a necessidade do estudo sincrônico do sistema lingüístico a par do estudo diacrônico, oportunizou, para o século XX, o posicionamento da língua falada em primeiro plano:

A sincronia conhece somente uma perspectiva, a das pessoas que falam (grifou-se), e todo o seu método consiste em recolher-lhes o testemunho.

Contudo, a tentativa (sempre quimérica) de mergulho obrigatório na manifestação das castiças variações lingüísticas fica a cargo das inspirações culturalistas do final do século XIX, cujo intento de examinar os fatores exteriores condicionantes das línguas acabou por desaguar no caudaloso rio dos estudos dialetológicos e no principal afluente destes, a geografia lingüística.

Ora, esse mergulho no empírico favoreceu de fato uma preocupação decisiva com a língua falada, cuja principal consequência é, certamente, a enfática preocupação com a complexidade do fenômeno lingüístico, a estabelecer a jurisdição do diferente, e ainda:

- a) a melhor definição do fenômeno (língua falada);
- b) o interesse pelo exame de fenômenos até então pouco ou não estudados, tais como os marcadores conversacionais, as estratégias do diálogo, as motivações discursivas de usos soldados à estrutura lingüística (gramati-

- calização) com o decorrer do tempo (conformações pancrônicas do fenômeno lingüístico)<sup>1</sup>;
- c) a gestação de novas metodologias do ensino de línguas de natureza funcional-comunicativas;
- d) Maior reflexão sobre a apreensão dos dados lingüísticos, seja do ponto de vista da coleta, seja do ponto de vista da transcrição.

É a esse último aspecto que este escrito se refere. O chamado "apagamento do falante" no decurso de transcrições de dados, bem como as demais consequências da interferência da subjetividade do transcritor serão abordados na perspectiva de se colocar os dados diretamente à disposição dos estudiosos através de meios ligados à informática.

## LIMITES DA TRANSCRIÇÃO

Urbano (1993), tratando sobre o que ele mesmo denomina transcrição ortográfica, utiliza as expressões "fixação de dados", "transferência de registro sonoro para o registro gráfico", "expressão gráfica dos sons e dos elementos prosódicos", "passagem do sonoro para o visual... do som", "desenho da fala", "fixação visual (gráfica) e convencional das realidades sonoras da língua oral".

<sup>1</sup> Votre & Naro (1989) referem como um exemplo agora clássico o estudo de Gilian Sankoff, de 1982, sobre o processo de clitização do pronome sujeito em Tok Pisin (pidgin falado na Nova Guiné), exposto em "*Variability and explanation in language and culture*". As análises de dados feitas em quatro recortes sincrônicos entre 1885 e 1971 mostraram que o pronome sujeito cumpriu uma deriva caracterizada pela existência de motivações discursivas originais que avançaram para uma situação (atestada em 1971) de subordinação a uma série de regras morfofonêmicas.

Ensina também que a transcrição ortográfica não deve ser confundida com a transcrição fonética, a que é realizada por meio de um alfabeto fonético. Ainda para Urbano, o que se transcreve não é a língua falada, mas “os elementos segmentais e eventualmente os suprasegmentais, o que se pode enfim chamar de ‘oralidade’ da língua falada”. Assim, o estudioso em referência diz que “não se trata de uma transcrição fonética, mas sim eminentemente ortográfica, que atenta, no geral, às normas gráficas da escrita padrão. Entretanto, na tentativa de revelar alguma feição da produção oral, fazem-se algumas concessões às características da produção oral”.

Esse tipo de transcrição em apreço é susceptível da possibilidade de “apagamento do informante”, entendido como o não-registro de fatos susceptíveis de apreensão pelo transcritor. Esse apagamento pode manifestar-se tanto pela omissão de trechos quanto pela não explicitação de detalhes segmentais e/ou suprasegmentais presentes naquilo que foi dito pelo informante. Em relação a essas duas situações, considerem-se dois trechos de transcrições de dados pertencentes ao projeto “Variação e Mudança Lingüística: o Atlas Geo-sociolingüístico do Pará”. O primeiro ilustra o “apagamento” pelo corte de trecho importante, registrando o seguinte:

(...) “purquê eu nunca tinha passadu pruma festa tãu bunita comu eu passei essa agora, marcô muito na minha vida..., eu jamais vô isquecê (+)/ de lá eu (+) da... da... da... du clubi... eu fui tomá um banhu (+) pra voltá pru insaiu..., tevi um sigundo insaiu..., eu participei du sigundu insaiu i (+) eu insaiei (+)... di lá... eu peguei... fui pru salãu i eli... u... u cabelereru mi cumeçô a mi maquiá (+)... começô a mi maquiá (+)... fez u meu cabelo (+)/ FUI FIUMADA... duranti eu tá nu salãu... eu fui filmada... tenhu a fita (+)... foi mui/ foi um... uma... uma... uma aventura pra mim essi baili” (...)

em vez daquilo que de fato foi dito:

(...)“purquê eu nuncatinha passadu pruma fésta tãu bunita cúmu eu passeiessagóra... eu jamás vôisquecê(+)/ dilá i (+) da... da... da...duclúbi... eu fui tumáumbânhu (+) pravoutá-pruinsáiu... têviumsígunduinsáiu... eu participei dusígundu-insáiu i (+) euinsaiei(+)... djilá... eu peguei... fui... ~~insaiei... erumafaxadjisetchoras da nôitchi... ((?)) prusalãu... areufui-prusalãu iêli... ucabelelêru micumeçô a mimaquiá (+)... fêzumeucabêlu (+)/ FUIFIUMÁDA... durantchieutánusalãu... eu fuifiumáda... tênhuafita (+) foimui/ foium... uma...uma...umaventura pramim essibáili” (...)~~

O segundo, abaixo transcrito, revela um transcritor que se deixa seduzir pelo hábito de utilização da escrita padrão oficial, razão por que se lhe escapam detalhes importantíssimos da oralidade:

(...) “quando separei fiquei sem nada só com os filhos que hoje ainda tenho, não dei nenhum graças a Deus criei trabalhando com ajuda de meus irmãos e alguns da comunidades meus colegas também me ajudaram bastante aí com o passar do tempo meu pai veio a falecer aí ficou minha mãe, aí ficamos trabalhando ajudando sempre e nunca fomos desu-nidos e minha mãe um tempo veio a também adoecer bastante durou um ano e pouco doente mas andava né, depois ela prostou-se e passou muito tempo prostada depois veio a falecer já este ano e: isto doeu muito, mais vamos levando graças a Deus é: hoje estou satisfeita graças a Deus porque criei meu filhos passei tudo isso na vida e: hoje estou criando netos seis netos eu crio ai continuo trabalhando ganhado pouco mais vai dando pra viver” (...)

As situações de “apagamento” que, aqui, serviram de exemplo, é claro, foram detectadas pelo trabalho de revisão criterio-

sa que o projeto anteriormente referido, a cujo corpus os dados pertencem, leva a efeito.<sup>2</sup>

### O SONORO E O TRANSCRITO NO ALISPA<sup>3</sup>

A palavra transcrever, em si mesma, parece antecipar muito do que se pode discutir sobre a atividade científica a que ela se refere. Em primeiro lugar, ela é “trans”; isso sugere transporte, passagem, trajeto, viagem de um lugar para outro, e parece que é a idéia de ir de um lugar para outro o que ela é em primeiro lugar: o que está falado “viaja” para o grafado. Considere-se em seguida que “cre” lembra que é necessário crer-se no que se ouve, para transcrever: é rezar o credo através do ouvido para o que recolhido foi. Por outro lado, ela também é “ver”, ver o que foi transportado para o visível, para um meio visual. Não precisa dizer que “crever” implica acreditar no que se está vendo, no caso, o que foi transcrito. Mas sobretudo está denunciada no “trans”, acompanhado de “cre” e “ver”, a circunstância de que, para além do que se crê e vê, estão as “viagens”, os construtos do subjetivo, a atear o fogo – às vezes até ardente – das fantasias humanas.

Esse caráter impressionista do que se transpõe para o grafado, resultante da influência da ótica pessoal, preocupa os linguistas que pensam as transcrições. Maria Fernanda B. Nascimento e outros (*apud* URBANO(1993)), afirma a esse propósito que “apesar do rigor com que tentamos reproduzir o oral sem o desfigurar excessivamente, não ignoramos que passá-lo a escrito é fazê-lo sofrer, primeiro por efeito da gravação e, depois, da transcrição,

<sup>2</sup> HILGERT (1989) afirma que se deve “*submeter a transcrição a um revisor que, possivelmente também tenha experiência de transcritor. Ele conferirá a fidelidade desta com a gravação*”.

<sup>3</sup> Atlas lingüístico Sonoro do Pará.

uma série de ‘metamorfoses’, de que sai profundamente empobrecido.” Do mesmo modo, HILGERT(1989) reconhece:

“(...) o que mais compromete a fidelidade da transcrição é o fato de ela ser moldada pela subjetividade com que o transcritor compreende o texto falado.”

Não é somente em relação à árdua tarefa de transcrever que os estudiosos das línguas manifestam preocupação quanto ao caráter tênue dos resultados. Jaberg e Jud(1928) advertem as afoitas ilusões em ciência da linguagem com a tantas vezes repetida lição proferida nos anos vinte:

“Envoyer vingt dialectologues dans le même endroit et vous aurez vingt résultats différents.”

Na mesma medida, subjacente está o problema em Nelson Rossi, (*apud* BRANDÃO (1991)), na introdução do Atlas Prévio dos Falares Baianos:

“Se as cartas não apresentam a nitidez de limites internos que seria desejável, nem sempre nos cabe a nós a responsabilidade. Às razões de serem aquelas e não outras (quem saberia quais?) as perguntas formuladas; de sermos nós e não outros os inquiridores; de serem aqueles e não outros os informantes; de serem as que foram e não outras (quem poderia dizer quais?) as localidades escolhidas; de termos dedicado apenas o tempo que dedicamos a cada inquérito, se não estamos enganados, deve muitas vezes somar-se à própria natureza do tecido emaranhado que se procura deslindar.”

O deslindamento do “tecido emaranhado” a que se refere Rossi passa pela transcrição de dados.

É em virtude da problemática acima tratada que se tem indagado sobre a possibilidade de evitar a intermediação do transcritor representada pela transcrição, seja ela de que natureza for, e oferecer à análise do lingüista o material coletado tanto mais “ao vivo” quanto possível. Exemplifica isso o projeto “O Atlas Lingüístico Sonoro do Pará”(RAZKY e outros (coords.) 1999), que está sendo presentemente executado, cuja formulação esclarece:

“Após a aplicação do questionário fonético-fonológico e a coleta de dados, proceder-se-á à elaboração de um programa interativo através do software ToolBook 6.5. O programa apresentará um mapa das 10 (dez) cidades estudadas. Um sistema informatizado permitirá ao usuário consultar de uma forma instantânea as variações fonéticas sonoras no Estado do Pará, ou seja, ele vai ter a possibilidade de saber, de ouvir, como uma determinada variável é realizada em cada uma destas cidades pesquisadas. A vantagem deste programa é que se vai ouvir a voz do informante, isto quer dizer que seu desempenho lingüístico vai chegar ao usuário sem a intermediação da subjetividade do lingüista, tão comum quando das transcrições fonéticas (...)”

A conseqüência mais importante de se trabalhar nessa direção é sem dúvida a socialização dos *corpora* de modo efetivo, pois dessa maneira mesmo as pessoas leigas, que não têm possibilidade de interpretar alfabetos fonéticos, poderão acessar aquilo que foi registrado nas entrevistas com os 40 informantes de dez cidades do Pará – todos escolarizados até não mais que a 4ª série primária, em duas faixas etárias – tudo isso sem contar a possibilidade de trânsito virtual do *corpus*, a sua globalização instantânea. Além disso, os reparos em análises articulatórias ou acústicas ficam facilitadas. Do mesmo modo, oportuniza-se a comprovação de que os dados existem, e que existem da maneira que foram declarados.

Por outro lado, o armazenamento de dados sonoros facilita o tratamento de *corpus* sonoros e o mapeamento de dados, oportunizando uma “representação cartográfica mais próxima da realidade lingüística”, face à “apresentação de dados reais, que podem ser ouvidos mais de uma vez e retranscritos, se for necessário[...]. A maioria dos atlas linguísticos publicados hoje são atlas impressos em papel. Na última década, a criação de inclusive ferramentas computacionais contribuiu de forma significativa para a mudança de apresentação dos dados geolinguísticos. A classificação automática tornou-se acessível a todos os pesquisadores sem muita experiência na área de informática. A chegada de discos rígidos com capacidade cada vez maiores abriu um caminho interessante para o tratamento de dados sonoros. A ideia de atlas sonoro virou realidade a partir da evolução técnica em termos de espaços físicos cada vez menores com (CD-Rom, DVD) com capacidades maiores para armazenar arquivos de todos os tipos sonoros.” (sic) (Razky(2000)).

Mas a transcrição dos dados não deixa de ser aproveitada no contexto de um atlas sonoro. No ALISPA, por exemplo, será incorporada a transcrição fonética dos dados, de modo que essa mesma transcrição poderá ser acessada paralelamente às falas dos informantes, inclusive com a comodidade de verificar contextos previamente selecionados pelo programa. Esse aspecto “legendado”, no entanto, pertencerá à segunda fase do ALISPA. Presentemente, as tarefas são as de digitalização, limpeza, equalização, compressão, classificação e codificação dos dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

BRANDÃO, S. F. A Geografia Lingüística no Brasil. São Paulo: Ática, 1991.

- CAMARA JR, J.M.C. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes, 1986.
- HILGERT, J. G. *A Paráfrase: um procedimento de constituição do diálogo*. São Paulo, 1989. Tese (Doutorado), USP-SP, inédita.
- JABERG, K. e JUD, J. *Volume d'introduction sur la méthode d'enquête*. Halle: Niemeyer, 1928.
- LYONS, J. *As Idéias de Chomsky*. São Paulo: Cultrix, 1976.
- MARCUSCHI, L. A. *Análise da Conversação*. São Paulo: Ática, 1991.
- RAZKY, A. e outros (coords.). *Projeto "O Atlas Fonético-Fonológico do Pará"*. Belém: UFPA/ PROINT, 1999.
- RAZKY, A. *Procedimentos Metodológicos para a Elaboração do Atlas Lingüístico Sonoro do Pará, 2000* (esboço, inédito).
- SAUSSURE, F. de. *Curso de Lingüística Geral*. São Paulo: Cultrix, 1975.
- URBANO, H.: *Reflexões em Torno de um Tema Metodológico no Estudo da Língua Falada: a Transcrição*. In: COLINS, H. (org.). *Intercâmbio*. São Paulo: PUCSP, 1993. V. 3.
- VOTRE, S. J., NARO, A. J. *Mecanismos Funcionais do Uso da Língua*. *D.E.L.T.A.*, v.5.