

ATA DA REUNIÃO DA EQUIPE DE CORPUS DO PHPB

Data: 01.06.2010

Horário: 18h 30min a 21h

Presentes: v. Lista de presença anexa

Pauta:

- I. Definição dos conjuntos do corpus comum mínimo no eixo diacrônico (séc. XVIII, séc. XIX e séc. XX)
 - a. Corpus comum mínimo – manuscritos
 - b. Corpus comum mínimo – impressos
 - c. Corpus comum diferencial
- II. Definição sobre número de palavras a serem coletadas
- III. Criação de fichas sociolinguísticas para cada documento a segundo os critérios sugeridos pelas equipes da Bahia, Rio de Janeiro e São Paulo
- IV. Siglário com orientações para a rotulação dos textos
- V. Circulação dos materiais
- VI. Tratamento computacional dos textos
- VII. Prazos para a entrega dos corpus comum mínimo da equipes regionais

- I. Definição dos conjuntos do **corpus comum mínimo**

Foram contrastados os levantamentos feitos pelo Prof. Afrânio Barbosa no vol. *História do Português Brasileiro* (CASTILHO & HORA org., 2010, versão preliminar, p. 35-49) do Corpus Geral do Projeto Nacional PHPB (versão junho 2009) com a proposta de SIMÕES & KEWITZ (comunicação oral no VIII Seminário do PHPB, João Pessoa). Procurou-se buscar uma convergência entre as propostas, levando também em consideração as contribuições dos GT de História Social feitas durante o referido seminário.

A partir dessas considerações, os participantes da reunião discutiram a viabilidade de distribuir a coleta de materiais respeitando a disponibilidade e representatividade das tipologias textuais em cada uma das 11 regiões que compõem o Projeto PHPB. Nesse sentido, a opção foi de compor três conjuntos de textos divididos da seguinte maneira:

- a. *Corpus* comum mínimo – manuscritos
- b. *Corpus* comum mínimo – impressos
- c. *Corpus* comum diferencial

O grupo decidiu que haverá prioridade para a coleta dos dois primeiros conjuntos (manuscritos e impressos) para atender à agenda de análises para a publicação dos volumes do PHPB que consolidarão os trabalhos das várias equipes. Parece ser consenso que esses dois conjuntos são de fácil acesso nos arquivos para coleta e edição ou já estão editados por algumas equipes, facilitando assim o trabalho de coleta e edição desses materiais pelos novos integrantes do Projeto (AL, PE, RN, SC).

A tabela 1 descreve as tipologias textuais que compõem cada um dos conjuntos (a, b, c):

- ▶ **CORPUS COMUM MÍNIMO - MANUSCRITO**
- ▶ 1. Testamentos
- ▶ 2. Processos-crime
- ▶ 3. Atas de Câmara
- ▶ Cartas
 - 4. Particulares
 - 5. Administração privada
 - 6. Oficiais
- ▶ **CORPUS COMUM MÍNIMO - IMPRESSO**
- ▶ Textos jornalísticos a partir do séc. XIX
 - 7. Cartas de redatores/editoriais
 - 8. Cartas de leitores
 - 9. Anúncios
- ▶ **CORPUS COMUM DIFERENCIAL**
- ▶ 10. Inventários
- ▶ 11. Memórias históricas e diários
- ▶ 12. Entremeses e outros textos teatrais
- ▶ 13. Inquéritos orais

Foram abordados ainda os seguintes tópicos:

- a. Recolha de materiais de três séculos: séc. XVIII, séc. XIX e séc. XX
- b. Redistribuição de determinados tipos de textos no *corpus comum diferencial*

Em vista da dificuldade de ampliar o corpus comum mínimo para os séc. XVI e séc. XVII, optou-se por restringir a coleta dos materiais para os séculos XVIII, XIX e XX.

Questões:

1. A Prof.a Tânia Lobo (UFBA) questionou a necessidade neste momento de considerar a 2.a metade do séc. XX para o corpus comum mínimo (manuscritos e impressos) em função do cronograma de entrega dos materiais pelas equipes regionais.
2. O Prof. Leonardo Marcotu
3. lio (UFRJ) observou que para o RJ é difícil identificar a nacionalidade brasileira para os autores de cartas da 1.a metade do séc. XVIII e faz a **sugestão** de coletar para esse período também as cartas de portugueses no Brasil.

Além disso, dada a dificuldade de encontrar i) inventários, e ii) memórias/relatos históricos e diários históricos de viagem em todas as 11 regiões, optou-se por transferi-los para o *corpus comum diferencial*, deixando assim as equipes dispensadas da obrigação de integrá-los tanto ao corpus comum mínimo quanto ao corpus comum diferencial. Os iii) entremeses e outros textos teatrais e iv) os inquéritos orais (NURC) também serão considerados no *corpus comum diferencial*, haja vista a restrição da existência das memórias em arquivos públicos e de inquéritos orais NURC em algumas capitais. Observou-se que, para o futuro, seria interessante prever a coleta desses materiais para incrementar o conjunto de textos para as análises.

- II. Definição sobre número de palavras a serem coletadas

A partir da proposta acerca do número de palavras a serem coletadas por tipologia textual, decidiram-se questões:

- a. Segmentação dos séculos em dois períodos de 50 anos: 1.a e 2.a metades
- b. Número de palavras por tipos de textos em cada uma das metades dos séculos

Dada a dificuldade de levantar grandes quantidades de textos da variada tipologia para os séculos pretéritos, decidiu-se que seriam levantadas 5.000 palavras para cada tipo de texto por cada período de 50 anos.

Questão:

1. A Profa. Sílvia Cavalcante ponderou sobre a necessidade de aumentar a quantidade de palavras recolhidas para 10.000 palavras. O Prof. Ataliba Castilho manifestou-se contra esta sugestão, argumentando que o conjunto final de dados teria uma representatividade suficiente para as pesquisas.

III. Criação de fichas sociolinguísticas para cada documento a segundo os critérios sugeridos pelas equipes de História Social da Bahia, Rio de Janeiro e São Paulo

Ficou acordado que serão organizadas fichas sociolinguísticas que atendem às necessidades de cada tipo de documento, conforme as reflexões pontuadas ao longo das apresentações das equipes de História Social e de Linguística de *Corpus*. Sugere-se que seja adaptada a ficha catalográfica sugerida pelo Prof. Afrânio Barbosa no vol. *História do Português Brasileiro* (CASTILHO & HORA org., 2010, versão preliminar, p. 53) da qual farão parte fatores como informações sobre o letramento dos escribas, critérios objetivos de erudição como o registro ou estilo do escriba (latinismos, BARBOSA 2005), ampliação das categorias sociais (grau de parentesco ou hierarquia social entre os escribas ou interlocutores).

IV. Siglário com orientações para a rotulação dos textos

[C 18 1 AI ____]
[(a) (b) (c) (d) (e)]

- (a) as primeiras letras referem-se ao gênero textual: carta (C), peça teatral (TEA), memória histórica (M), anúncio (A), inquérito do Projeto NURC (D2SP-330);
- (b) o primeiro número refere-se ao século;
- (c) o segundo número indica a 1ª. ou 2ª. metade do século em questão;
- (d) a notação em letras indica o documento de origem e
- (e) o último campo poderá indicar uma categoria específica de maior precisão e controle do pesquisador.

V. Circulação dos materiais

A Profa. Maria Clara Paixão de Sousa (USP) sugeriu que a equipe de *Corpus* pense em estratégias de circulação tanto dos *corpus comum mínimo* como da totalidade dos *corpora* regionais através de sua disponibilização em um portal na internet. Dessa forma todos os pesquisadores do PHPB terão acesso a todos os textos disponíveis e não somente os materiais do *corpus comum mínimo*.

Questão: Qual o tipo de acesso ao material: regulado com senha ou livre?

VI. Tratamento computacional dos textos

Proposta da Profa. Maria Clara Paixão de Sousa (USP): dividir a equipe de Corpus entre um grupo responsável pela coleta e edição filológica de manuscritos e a) a equipe de processamento computacional dos textos que preparará os materiais para as edições anotadas morfossintaticamente.

VII. Prazos para a entrega do corpus comum mínimo da equipes regionais

A partir da proposta de número de palavras para os três conjuntos de *corpora*, a Profa Maria Clara informou que a partir da entrega da edição filológica dos materiais, a edição anotada poderia ser executada em um prazo de 6 meses.

Questão: Que prazo vamos definir para que as equipes regionais entreguem os materiais do *corpus comum mínimo de manuscritos e de impressos* editados filologicamente.

Proposta (da Profa. Tânia Lobo) de que o volume sobre Tradições Discursivas trabalhe efetivamente sobre o que foi estabelecido efetivamente sobre o que foi estabelecido pela Equipe de Corpus. Ou seja, precisamos conhecer as tradições discursivas das cartas, dos testamentos, dos processos-crime etc. Os volumes não podem ser uma coletânea de textos. É preciso preservarmos a organicidade da obra coletiva de 15 anos da história de um projeto de pesquisa.

Questões em aberto:

1. Que tipos de cartas de leitores e anúncios vamos coletar para o séc. XX?
2. Não considerar os anexos dos testamentos, inventários e processos-crime com o mesmo rótulo e sim como texto individual, p. ex., carta particular.
3. Atas de Câmara e certos tipos de cartas: a partir do séc. XX esses materiais passam a ser impressos. Como considerar o registro gráfico desses materiais?