Ensaio sobre a revisão da oralidade

Paula Taveira e Diana Santos
Linguateca e Universidade de Oslo

Abstract:
This paper discusses the revision of the transcription of oral interviews done by the Portuguese Museu da Pessoa to preserve the life histories of common people, often with little or no formal education, which had been transcribed by students without linguistic training. This task increased our awareness of how orality works and of how many choices a transcriber is required to make, being a sort of translator. In addition, it uncovered a set of «errors» or deviations in many elderly speakers from the North of Portugal which are often considered typically Brazilian, leading us to reconsider variety divergence.

Keywords: orality, transcription, norm deviations, Portuguese varieties, corpora
Palavras-chave: oralidade, transcrição, desvios da norma, variedades do português, corpos

Este trabalho resultou da opção de rever as entrevistas portuguesas do Museu da Pessoa incluídas no Acesso a Corpos/Disponibilização de Corpos (AC/DC) da Linguateca. Com o contributo voluntário das duas autoras, o seu fim é continuar a enriquecer os recursos servidos pela Linguateca, que em 2010 perdeu financiamento devido às regras europeias que entraram em vigor em Portugal. O AC/DC, veja-se Santos & Bick (2000), é o projeto mais antigo da Linguateca, rede internacional de recursos para a língua portuguesa: existe desde 1999, disponibiliza à comunidade interessada na linguística do português muitos corpos diferentes, todos eles com uma anotação sintática efetuada pelo PALAVRAS (Bick, 2000) e com anotação semântica, cf. Santos (2014a), e tem associados vários serviços de procura e comparação dos seus resultados (Simões & Santos, 2014).

Um dos corpos incluídos no AC/DC é o Corpo Museu da Pessoa, cujos usos vão dos estudos de linguística em geral ao ensino de português como língua estrangeira, passando pela Gramateca (novo projeto no âmbito da Linguateca para fazer gramática baseada em corpos, Santos (2014c), http://www.linguateca.pt/Gramateca/). É de qualquer maneira relevante salientar não fomos nós as responsáveis pela criação do corpo (escolha de falantes, temas) e que por essa razão não nos podemos pronunciar sobre a escolha dos textos.

1. O Museu da Pessoa e a decisão de rever as entrevistas

«*O Museu da Pessoa é um museu virtual e colaborativo fundado em São Paulo no ano de 1991. Desde sua origem, tem como objetivo registrar, preservar e transformar em informação histórias de vida de toda e qualquer pessoa da sociedade. Nosso acervo conta atualmente com mais de 16 mil depoimentos em áudio, vídeo e texto e cerca de 72 mil fotos e documentos digitalizados.*» Esta é a informação que consta na página http://www.museudapessoa.net/pt/entenda/o-museu-da-pessoa, consultada em outubro de 2015.

O Núcleo Português do Museu da Pessoa, sediado na Universidade do Minho, cedeu à Linguateca cerca de cem entrevistas digitalizadas por alunos estagiários (ver Almeida et al., 2000), mas essas páginas continham bastantes erros de ortografia (e provavelmente de transcrição também). Como era dos poucos casos em que o AC/DC tinha linguagem oral, pareceu-nos importante melhorar esse recurso.

Pensamos que obtivemos as entrevistas após terem sido reproduzidas fielmente do discurso registado, embora faltassem as indicações sobre risos, suspiros, assobios, possivelmente retiradas pelo processamento do XML para texto.

2. A tarefa a que nos propusemos, dúvidas e dificuldades

Dedicámo-nos a rever a língua, mantendo as marcas da oralidade, mas corrigindo os erros dos transcritores e dos falantes. Definir o objetivo de salvaguarda de «texto oral» ajudou a eliminar um tipo de dúvidas (apesar de estarmos a ler, o que líamos «era mais» para ser ouvido… e as aparentes quebras no discurso podem ter sido «visíveis» em gestos ou expressões), mas não todos. Ver o texto escrito escancara-nos o discurso: ao lermos, detetamos mais facilmente as

redundâncias e os erros de concordância facilmente cometidos a falar. Mas teria um dado lapso sido cometido por quem falou ou por quem transcreveu? Nesta primeira missão, ficámos sem saber. Tínhamos apenas o texto transcrito, sem acesso ao material sonoro.

Decidimos então manter os erros (desde que não fossem apenas erros de transcrição) mas com uma sugestão de correção (com <erro corr="o que deveria estar"> o que estava </erro>), e tentámos corrigir o menos possível para manter o interesse do ponto de vista linguístico e permitir investigar se as pessoas usavam a norma do País ou se havia fenómenos da oralidade. Nesse processo, acabámos por marcar os erros ou desvios relativos à norma padrão portuguesa conhecida e praticada pela primeira autora, revisora linguística de profissão, após discussão dos casos mais complexos pelas duas autoras, e sua documentação exaustiva, que pode ser consultada em http://www.linguateca.pt/acesso/revisao_mp.html.

Sabemos que «oral transcrito» é uma família de entidades (Santos, 2014b, Santos, 2016a) e que muitas outras formas de o preparar como corpo eletrónico seriam possíveis, veja-se Raso & Mello (2014) para uma proposta recente. O nosso objetivo era criar um texto para «publicação», no sentido de que repetições ou hesitações ou fragmentos ininteligíveis não prejudicassem a compreensão das histórias contadas, e que os termos mal grafados ou pronunciados não evitassem encontrar esses assuntos.[1]

A revisão não significou, contudo, que se perdesse a forma original de transcrição: mantivemo-la «escondida», mas acessível, no corpo (exceto as correções simplesmente ortográficas, ou seja, os erros introduzidos pelo transcritor).

3. Os resultados: primeira impressão

Os resultados da nossa revisão foram ao encontro da visão que temos sobre a língua portuguesa como um todo cheio de variedades, embora tenhamos ficado surpreendidas pela variedade de problemas e de erros encontrados.

De facto, muitas das propriedades elencadas pela variante brasileira para mostrar o seu afastamento da variante de Portugal foram encontradas em falantes idosos e pouco escolarizados

[1] Visto que praticamente apenas um traço dialetal, «num» por «não», tinha sido codificado pelos transcritores, resolvemos desfazê-lo. Vejam-se outros exemplos na secção 7.2.

do Norte do País... tradicionalmente mais conservador em termos linguísticos. Escolhemos apresentar aqui a posição dos clíticos, as relativas sem preposição, o uso de diferentes preposições ou falta destas, e várias questões relativas a concordância, acabando por apresentar as orações finitas iniciadas por *depois*.

3.1. A posição dos clíticos

(1) Mas ajudou-me muito, era um homem de Linguística e eu nunca gostei de Linguística, de maneira que ele *ajudou-me* (em vez de me ajudou) imenso.

(2) Não havia pessoal e as coisas *tinham-se que fazer* (em vez de tinham que se fazer).

(3) Era pela rádio que a gente se informava e muitos portugueses foram para lá combater porque estiveram em Sevilha, essa parte, e o que a gente queria é que a coisa se fosse afastando, e felizmente *se afastou* (em vez de afastou-se).

(4) Era muito engraçado quando aqueles mais reguilas não acertavam na rainha e esta *virava-lhe* (em vez de lhes virava) as costas.

(5) Na altura em que *iam-no* (em vez de o iam) levar para a prisão, pelo monte fora, apareceu uma bruta serpente que aterrorizou os mouros.

3.2. A falta de preposição nas relativas

(6) Depois houve uns tempos (em) *que* eu pertencia aos escuteiros.

(7) No dia (em) *que* o meu filho fazia um ano fui-me embora outra vez.

(8) Uma atividade engraçada, (a) *que* pouca gente ligava, era a criação de abelhas.

(9) Agora, em vida, as coisas (a) *que* eu vou assistindo.

3.3. A falta de concordância ou o uso de uma concordância não padrão

(10) Foi então... estava a começar, os terrenos de Guimarães foi no meu tempo que se compraram, porque tínhamos lá uma casa emprestada, mas a universidade... *foram* (em vez de foi) necessário encontrar o local e comprar os terrenos.

(11) 3 Quer dizer que a maioria dos seus clientes *é pescador*? (em vez de são pescadores).

(12) Sobre a cozinha portuguesa, eu gosto muito, embora coisas do tipo papas de sarrabulho ainda *são* (em vez de sejam) um pouco raras para mim!

(13) *Olha* (em vez de Olhai), ide prò caraças.

(14) Agora, por acaso até estão em terra, que não há pesca, nem há tempo para eles *ir* (em vez de irem) pescar.

(15) E acontece que a partir do momento (em) que eles começassem a orar, se eles *curasse* (em vez de curassem) a filha, o rei mouro fazia uma festa no dia 24 de Junho, que é o dia que *calhe* (em vez de calha) o S. João, calhe o dia que calhar.

3.4. Diferentes preposições e falta delas

(16) Portanto, isto nasceu assim de pequenino e, como disse, foi um prémio e daí para cá tenho participado todos os anos, salvo quando tive um acidente e não pude estar, não pude ir de bugio, independentemente de eu ter pedido ao médico, onde estava hospitalizado, para que me deixasse vir a casa, que gostava de participar *da* (em vez de na) festa.

(17) O ouro está *da* (em vez de na) minha mão, está guardado, é para a minha Maria Helena…

(18) Ao fim e ao cabo, é cumprir uma meta, digamos assim, uma satisfação plena que a gente sente, aquela satisfação de meta, de estar em cima, *ao* (em vez de no) auge da carreira.

(19) Que diga-se que é a realidade, porque há muita gente que sabe que é verdade, eu já fui criticado *a* (em vez de por) mim próprio, portanto, não posso levar a mal se os outros me criticam.

(20) Mas também nunca tive propensão *por* (em vez de para) isso, não sei se por cobardia se por formação cristã, nunca percebi porque é que eu nunca dei para ser um marginal.

(21) Vim aqui *nas* (em vez de às) ruas de São Paulo e vi uma no colo de uma senhora ainda muito jovem, que levaria uma miúda com seus 2, 3 anos atrás pelo seu pé e levava ao colo um recém-nascido ainda de olhos fechados, ainda com as pálpebras muito inchadas.

(22) Esses têm medo da água e não aparecem aqui [para] tomar banho.

3.5. *Depois* com oração finita

(23) Ora bem, enquanto eu fui pequeno, vivi sempre lá, *depois que casei* (por depois de casar) é que vim morar para o Porto.

(24) *Depois que as coisas acalmaram* (por Depois de as coisas acalmarem) lá levámos o volfrâmio.

(25) *Depois que abri* (por *Depois de abrir*) a taberna, vinha cá abrir de manhã, ficava aqui um bocado e depois ia trabalhar e a minha mulher ficava aqui.

(26) Agora, *depois que eu fui* (por depois de ter sido) operada é que me proibiram de lavar.

(27) *Depois que a minha mãe entravou* (por depois de a minha mãe ter ficado entrevada) é que eu vim para casa para tomar conta dela e aí é que reconheci que tinha mãe.

Convém indicar aqui que estamos conscientes de que *depois* apenas com infinitivo é muito provavelmente norma de Lisboa, e que no resto do País a forma finita é aceite e praticada, pelo menos a partir de Leiria.[2]

4. Revisão pressupõe interpretação

Nesta secção apresentamos vários casos em que a nossa intuição de falantes nos serviu para propor alternativas a textos incompreensíveis ou pelo menos suspeitos.

4.1. Em nomes próprios

Detetámos questões de transcrição que não têm que ver com a variante nem com a língua, mas com a dificuldade de interpretar nomes próprios (estrangeiros ou nacionais), quer por os falantes os terem travestido, quer por os transcritores os desconhecerem, ou ambas as razões:

(28) *Papa Piedoso* para *Papa Pio XII*.

(29) *Linhais da Serra* para *Unhais da Serra.*

(30) *Passo Sousa Alves* para *professor Sousa Alves*.

(31) *São Pedro Alfa* para *São Pedro de Alva*

2 Agradecemos a Anabela Barreiro, natural de Leiria, essa informação.

(32) *Escola Comercial de Mil Navarra, em Almada* para *Escola Comercial Emídio Navarro, em Almada*

Repare-se que em alguns casos não podemos afirmar que a nossa correção não seja hipercorreção... O primeiro falante poderia de facto estar convencido de que o verdadeiro nome do Papa em questão era Papa Piedoso, e sempre se referir a ele assim.

Por outro lado, falando de uma instituição de ensino no Minho, *Passo* poderia ter sido corrigido para *Padre*, em vez de *professor*.

4.2. Em frases

Em algumas partes do texto, tivemos dúvidas sobre o sentido do que fora dito. Seguem-se dois exemplos com duas interpretações que levam a duas sugestões de correção diferentes:

(33) Os operários devem saber o que podem exigir na empresa sem que a empresa vá abaixo, porque os patrões estão a explorar os operários quando paga a menos é porque não pode pagar.

(33.1) Os operários devem saber o que podem exigir na empresa sem que a empresa vá abaixo, porque os patrões estão a explorar os operários quando pagam a menos e é porque não podem pagar.

(33.2) Os operários devem saber o que podem exigir na empresa sem que a empresa vá abaixo, porque os patrões estão a explorar os operários... Quando (o patrão) paga a menos, é porque não pode pagar.

(34) Estes restaurantes já são um bocadinho antigos, embora estejam mais embelezados, não é, talvez consoante vai subindo vão ficando mais embelezados, mas digo-lhe uma coisa: já havia estes restaurantes, havia na mesma.

(34.1) Estes restaurantes já são um bocadinho antigos, embora estejam mais embelezados, não é, talvez consoante (os restaurantes) vão subindo (de categoria) vão ficando mais embelezados, mas digo-lhe uma coisa: já havia estes restaurantes, havia na mesma.

(34.2) Estes restaurantes já são um bocadinho antigos, embora estejam mais embelezados, não é, talvez consoante (o senhor) vai subindo vão ficando mais embelezados, mas digo-lhe uma coisa: já havia estes restaurantes, havia na mesma.

Esclareça-se que as expressões entre parênteses são apenas para dar ao leitor deste artigo uma melhor compreensão das interpretações em questão, não tendo sido adicionadas ao texto.

5. Ler o texto como se o ouvíssemos

Também sentimos necessidade de analisar partes deste «texto oral» como se o estivéssemos a ouvir para chegarmos ao sentido do que fora dito, mas mal transcrito:

(35) (…) *e se* arte portuguesa. / (…) isso é arte portuguesa.

(36) Fazia capas, aventais de riscado, burel *por priosar*, saiotes com rendas de lã feita à agulheta. / Fazia capas, aventais de riscado, burel para o prior usar, saiotes com rendas de lã feita à agulheta.

(37) A minha mãe dizia-me: «Ó Amadeu, vai à ti Ana do Barroso que te dê *uma rátele* de arroz.» / A minha mãe dizia-me: «Ó Amadeu, vai à ti Ana do Barroso que te dê *um arrátel* de arroz.»

6. Marcas da oralidade?

6.1. Redundâncias

(38) E então, eu com as Ciências e a dona Maria José, que atualmente é secretária do atual administrador da universidade, começamos a organizar as nossas pastinhas e passado algum tempo estávamos com total autonomia.

(39) (Para) além desse *hobby* de lazer, o que gosta de fazer mais?

(40) Aqui há uns anos atrás!

Este último caso é muito interessante porque é um dos pleonasmos mais criticados no Brasil, mas cujo uso parece ter-se tornado praticamente universal nesse país. Diz Evanildo Bechara a este propósito: «(…) Podem-se suprimir as palavras *atrás* ou *passado(s)* que aparecem com o

verbo haver, uma vez que este já indica tempo decorrido: "Há três dias atrás ou Há três dias. / Há três dias passados ou Há três dias."» (Bechara, 2009:614)

6.2. Maior liberdade (no uso de regras)

Como o «texto oral» nos permite ler aquilo que poderíamos ouvir como foi dito, notámos a maior liberdade da oralidade em relação a um maior condicionamento da escrita. Diversos exemplos:

(41) As pessoas é que não podiam ser presas, largavam o carro e fugiam. Quando [eu/alguém/uma pessoa?] sentia que não podia fugir, abandonava o carro e fugia [o entrevistado não estava a usar a primeira pessoa, mas, a falar, se calhar até misturou as pessoas e falou dele próprio].

(42) Foi uma operação que fiz, [em] que me tiraram uma pedra da vesícula que podia *gerar em* (por gerar ou 33] cancro porque já estava em ferida.

(43) (…) porque há gente amargurada que é muito difícil tratar com elas [(…) porque há gente amargurada com quem é muito difícil tratar / porque há gente amargurada e é muito difícil tratar com elas=essas pessoas=essa gente].

(44) E a senhora lembra-se *o* (por do) que pensou nessa noite, quando o seu marido estava no mar?

(45) Como é que *descrevia* (por descreveria) a educação que recebeu?

(46) Torres Novas fica a uns cem quilómetros de Lisboa e Torres Vedras é (por a) bem menos, é (por a) uns quarenta quilómetros de Lisboa.

(47) Ó Teresa, manda-me o programa do nosso encontro agora em Novembro porque eu guardei tão bem guardadinho o programa, que eu não sei aonde é que eu meti. / Ó Teresa, manda-me o programa do nosso encontro agora em Novembro porque o guardei tão bem guardadinho, que não sei onde é que o meti.

(48) Sempre me meti e eu, na minha empresa, fui tudo [o] que era possível ser...

(49) Comecei a *trazer* (por trazê-la) porque a aflição para ela era a água diretamente.

O último caso, (49), é um famigerado exemplo de objeto nulo, considerado mais uma vez um apanágio do português brasileiro, mas que aparece também no Norte de Portugal. Veja-se Jansen (2016) para uma confirmação de que essa ausência não é só brasileira.

Em (48) temos o problema de escolher entre duas categorizações: a do verbo *meter (em)* e a do verbo *saber* (transitivo), a primeira sugerindo *aonde*, a segunda exigindo *onde*, caso já discutido em Santos (2004).

6.3. Quanto aos erros lexicais ou morfológicos

Exemplos de «erros» lexicais ou morfológicos:

azilhargas, argelianos, imediático, asterose, escandinávios, trupedeados, pequeninha, auguinha, perpetuzinhos, pouquexinhos, prume, safes, parenta, furecidade, desconfrar, antão, Trás-dos-Montes, interviu, bebeides, comedes, deziam, camurflada, mandavam-los, manifestaria-se, nos as, deveria-me acompanhar, trazia-nas, disse-lo, consta-se, tinham quem lhes socorresse, abono-lhe, a gente semos, deviam haver, mais grandes, caneta permanente

Parece-nos especialmente interessante a mudança de desinência de *parente* para *parenta*, marcando o feminino; o excesso de clíticos iniciados por *-l* ou a sua falta; e a conjugação errónea de *vós*, assim como a falta dos mesoclíticos nos condicionais. São todos eles, diga-se de passagem, erros típicos de aprendizes de português como língua estrangeira.

7. E quem fala assim…

Há termos que denunciam quem fala.

Pudemos constatar em alguns casos arcaísmos ou regionalismos que, embora tenhamos corrigido para a norma padrão, convém indicar aqui.

7.1. … talvez seja «antigo»

(50) Carrava cestos de erva e cântaros de água às lavradeiras, ajudava nos trabalhos do campo durante todo o ano.

O verbo *carrar* não está registado em dois dicionários atuais disponíveis na Internet (*Priberam* e *Dicionário da Língua Portuguesa* da Porto Editora), mas está, por exemplo, na edição de 1913 do *Novo Diccionário da Língua Portuguesa*, de Cândido de Figueiredo. Deixou, contudo, marca na expressão frequente *carradas de...*

7.2. … talvez seja do Norte

(51) Saímos do lameiro, [disse] diz assim um passador: «Num podeides (por Não podeis) ir por tal sítio, por acolá, tendes de ir por acolá, tendes que marchar que estais denunciados à frente.»

(52) Toca a andar, andar, fomos andando, lá num caminho disseram-nos assim: «Tendes que subir aqui.» Começámos a subir, e para onde subimos?, para um chouto (por souto) de castanheiros…

(53) Nós já sabemos há uns poucos de dias (por dias). Já eram bascos, era a Guarda de San Sebastian. «Agora vindes connosco que depois nós vamos buscar os vossos colegas e vós esperais aqui por nós.»

(54) Ai a rica freeesca!! Sardiinha bibiinha (por viviinha)!

(55) Mas era muito repentino e estava sempre a oferecer-nos nas bentas (por ventas), mas não nos batia porque o meu pai não dava asas aos irmãos andarem (por não dava azo a que os irmãos andassem) a bater uns nos outros.

(56) Isto foi em janeiro, no inverno, e quando me ponho a pé de manhã fiquei barado (por varado) da vida, porque vejo aqueles montes todos brancos: aquilo era só neve!

Traços comuns do falar do Norte são, além do *num* já mencionado, a troca do «v» pelo «b» e o uso da segunda pessoa do plural, já extinto no resto do País... Além disso, não pudemos deixar

de observar o uso frequente de palavras regionais em Portugal que são nacionais no Brasil, como *botar* e *ruim*.

8. Descrição quantitativa

8.1. Tamanho do material

Ao todo, fizemos 1528 alterações em que mantivemos o original. Relembramos que as simples correções ortográficas (como de *heide* para *hei-de* ou de *à* para *há*, ou adição de pontuação ou de maiúsculas) não foram contabilizadas.[3]

O texto que foi alvo de correção, relativo a 105 entrevistas em português de Portugal[4], continha 346 mil unidades (palavras e sinais de pontuação), correspondendo a 22.853 frases.[5]

Os entrevistados, 42 mulheres e 49 homens, podem ser sumariamente descritos pelas figuras seguintes: a Figura 1 mostra a distribuição das idades dos entrevistados pelo sexo e pelo país (entre Portugal e Brasil), enquanto o Quadro 1 apresenta as origens geográficas mais frequentes dos entrevistados em Portugal.

[3] A este respeito convém documentar que usámos, nesta primeira fase da revisão, a grafia portuguesa antes do Acordo Ortográfico, visto que as entrevistas foram originalmente transcritas antes da sua entrada em vigor mas estamos a considerar alterar essa decisão na segunda fase da revisão.

[4] Embora as 109 entrevistas que revimos tenham sido gravadas em Portugal (continental), excluímos do presente cômputo aquelas feitas a brasileiros (E072 e E091) e a estrangeiros residentes em Portugal (E008, E051 e E052), não por não serem igualmente interessantes do ponto de vista linguístico e pessoal (e foram, aliás, também revistas), mas porque nos interessa aqui discutir a diversidade no âmbito do português de Portugal. Convém, contudo, também realçar que muitos dos entrevistados, sobretudo masculinos, tinham passado parte da sua vida no estrangeiro, como se pode apreciar lendo as próprias entrevistas.

[5] Estamos naturalmente conscientes de que «frase» não é a melhor categoria para analisar o oral, embora o seja tanto melhor quanto mais próximo do escrito o discurso estiver. A nossa contagem de frases é feita automaticamente pelo segmentador da Linguateca (ver http://www.linguateca.pt/acesso/atomizacao.html) e foi apenas pontualmente revista ao fazer a revisão do texto. Conforme será referido adiante, a questão mais complexa do discurso direto ainda não foi sistematicamente tratada, mas calculamos que a sua identificação aumente consideravelmente o número de frases distintas no material.

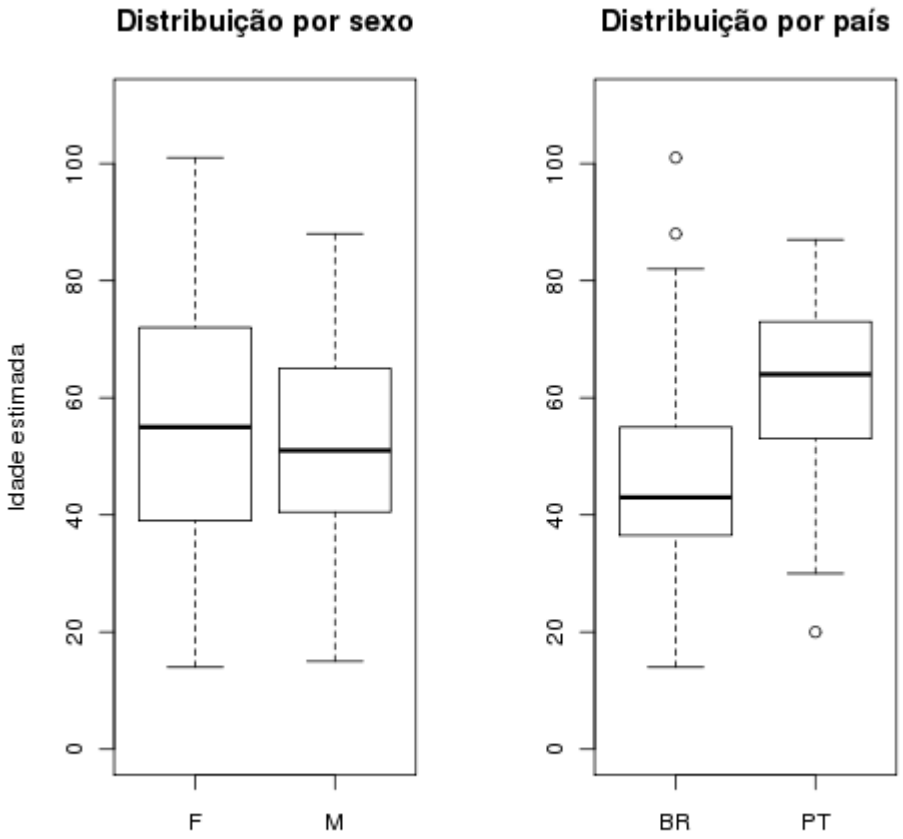

Figura 1: A idade dos falantes, por sexo e por país (de todo o corpo Museu da Pessoa)

Repare-se na escolha consistente de pessoas mais velhas no caso de Portugal, com um valor central (a mediana) superior a 60 anos. Mencione-se também o facto de em algumas entrevistas falarem várias pessoas, razão pela qual o número dos falantes é maior do que o das entrevistas. Há quatro entrevistas com duas pessoas, uma de cada sexo (E008, E035, E050 e E073), e uma entrevista com quatro pessoas, um homem e três senhoras, E085. Além disso, numa outra entrevista, E060, fala um homem e dois clientes, enquanto na E078 a conversa é com quatro

pescadores da Afurada. Nem sempre conseguimos identificar a idade, por isso a Figura 1 apenas apresenta as idades identificadas.[6]

Local	Entrevistados
Afurada	11
Sobrado	7
Porto	6
Vilar de Perdizes	3
Vila Nova de Gaia	3
Bustelo	3
Friões	3
Lisboa	2
Telhado	2
Montalegre	2

Quadro 1: Locais de nascimento com mais de um entrevistado (apenas dos portugueses)

A anotação no próprio corpo da idade e género dos falantes, que se encontra de qualquer forma acessível em http://www.linguateca.pt/acesso/metadadosMP.html, permite cruzar procuras linguísticas com o género e idade, embora certamente a escolha dos entrevistados não se possa dizer aleatória, e por isso têm de ser cuidadosas as generalizações, sobretudo em relação à idade.

Para obter acesso às correções, a expressão de procura é <corr> []+ </corr>, como ilustrado na Figura 2. Se se quiser também ver a forma original, antes de ser corrigida, basta selecionar a opção «Mostrar texto original nas correções» (em baixo).

[6] Ainda outra complicação sobreveio nesta tentativa de cartografar as idades dos entrevistados portugueses: é que, se em algumas entrevistas é especificamente perguntada a idade do entrevistado, na maior parte dos casos o que sabemos é a data de nascimento, e não temos (por agora?) acesso à data exata da entrevista. Estimámos por isso arbitrariamente que todas as entrevistas foram feitas no ano 2000.

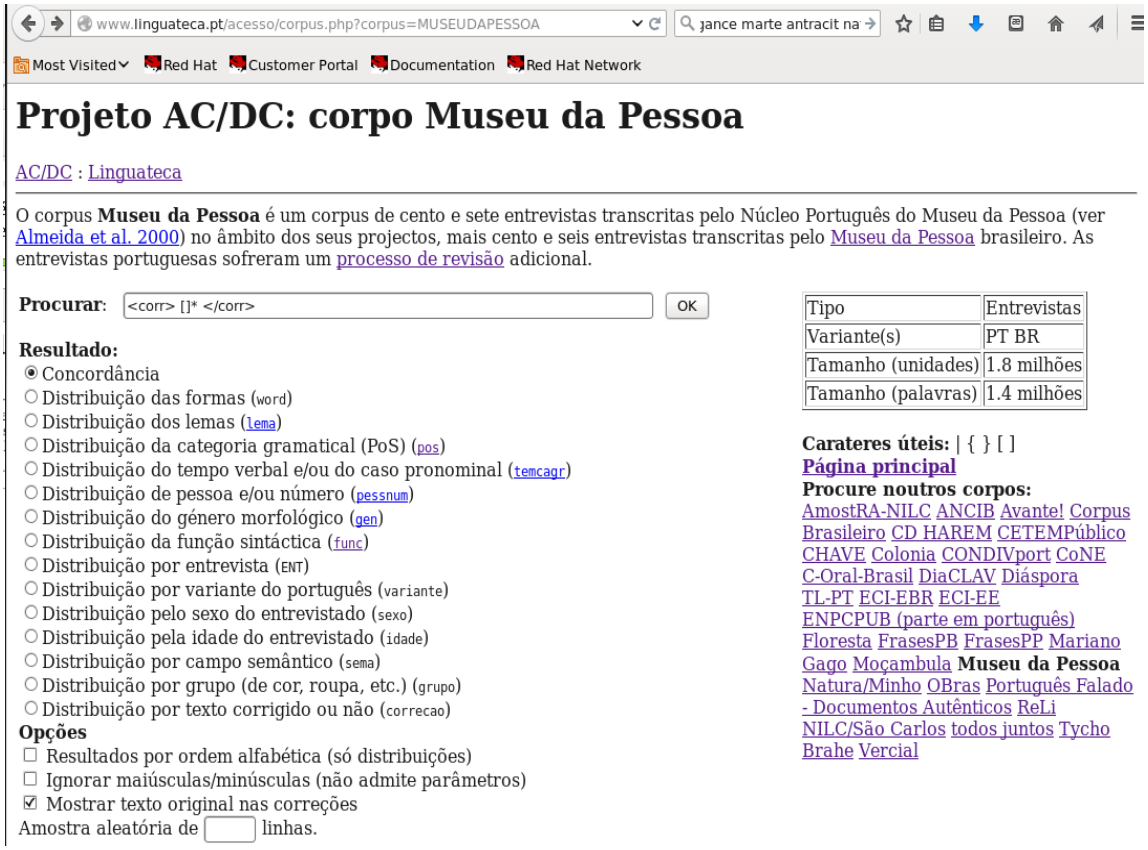

Figura 2: Interface de procura relativa ao Museu da Pessoa no AC/DC

8.2. Distribuição

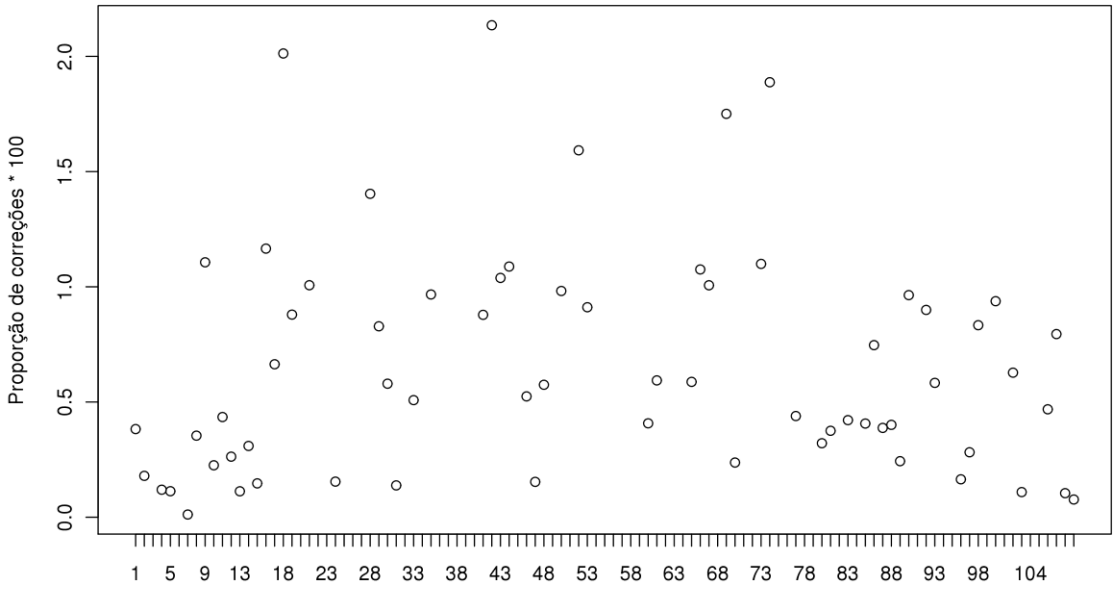

Figura 3: Proporção de correções por entrevista

Distribuição das correções por entrevista: houve 89 entrevistas que sofreram pelo menos uma correção, sendo a E074 (com 249), a E020 (com 120) e a E041 (com 118) as que tiveram mais correções. Embora não seja trivial encontrar a unidade ideal para quantificar o número de correções, visto que as entrevistas diferem significativamente em tamanho, usámos a proporção do número de unidades modificadas[7] por tamanho da entrevista em unidades na Figura 3. Note-se que esta figura não inclui a entrevista mais modificada de todas, a E034, que teve um índice de 12% de correções.

[7] Modificadas pode querer dizer simplesmente transferidas para outra posição na frase. No caso de simples adições, ou seja, casos em que a correção consistiu simplesmente em adicionar palavras que faltavam, isso foi contado como uma unidade modificada. No caso de palavras com clíticos ou com contrações, contámos como uma palavra só, seguindo a atomização defendida em Santos & Bick (2000) e documentada em http://www.linguateca.pt/acesso/atomizacao.html.

9. Comentários finais

Foi muito interessante observar a língua em uso sem passar pela monitorização da escrita. Claramente a questão da falta de concordância e da variabilidade da posição dos clíticos – muito bem documentados para o português brasileiro e para o português de Moçambique, veja-se, por exemplo, Kato & Roberts (1993), Brandão (2011) e Gonçalves & Stroud (1997-2000) ou Jon-And (2011) – são casos comuns a todas as variantes do português, ou pelo menos estão presentes no português de Portugal também. Da mesma forma, a variação entre *nós* e *a gente* (Silva, 2010) ou a variabilidade na subcategorização (Ferreira, 1996) foram identificadas sem estarmos necessariamente à procura destes casos. Isso levou-nos a ter mais confiança na viabilidade, e na pertinência, de desenvolver um português internacional, veja-se Santos (2016b).

Em relação à própria tarefa a que nos propusemos, apercebemo-nos claramente de que não é fácil delimitar até onde se revê e que marcas do oral se devem manter, para manter a naturalidade e a fluidez da entrevista, mas parece óbvio que há diferentes níveis de autenticidade que se podem defender.

Podemos, de facto, perguntar: o que é um texto (final)?

A nossa conclusão é essencialmente pragmática: para tipos de estudos linguísticos diferentes, diferentes versões seriam preferíveis. Mantendo ambas, permitimos pelo menos dois tipos de estudos: o dos desvios à norma (escrita), e o do conteúdo e forma normalizados.

Estamos, contudo, dolorosamente conscientes de que o papel do revisor e do transcritor, sendo de mediador entre o falante e o leitor, não pode ser 100% objetivo. De facto, gostaríamos aqui de insistir na analogia com a tradução: tal como um tradutor traduz o que compreende e o que pensa ser a mensagem do autor, mas que outro tradutor poderia exprimir diferentemente, também um revisor ou editor, ao interpretar o texto, não pode abdicar da sua compreensão e por isso mesmo interpretação.

Concluímos com uma declaração de intenções em relação a este projeto de revisão, que não é mais do que a descrição de trabalho futuro sobre este material.

Nesta primeira revisão apresentada aqui, pela necessidade de levarmos a bom termo um projeto que nos foi revelando tantas questões interessantes, decidimos não nos debruçar sobre a

questão do discurso direto, indireto ou livre, que cedo observámos ter sido tratado (ou grafado) diferentemente pelos diferentes transcritores, e que compreendemos que seria, ou poderia ser, difícil de identificar sem acesso ao som. Ficámos, contudo, interessadas em efetuar uma nova revisão focando precisamente as questões que se prendem com o discurso direto e indireto, e que esperamos que possa ser iluminado depois de um trabalho (Freitas *et al.*, 2016), recentemente concluído também no âmbito da Gramateca, nos ter aguçado o interesse pela complexidade do discurso relatado em português.

Concomitantemente, pretendemos classificar as várias correções em vários grupos, produzindo assim uma meta-anotação do tipo de correção, de forma a permitir uma quantificação mais fina e também procuras parcelares: quantas correções relativas a clíticos? Quantas de concordância? Quantas puramente lexicais?

Finalmente, chamamos a atenção para a semelhança desta forma de tratamento do texto com as práticas filológicas que têm sido identificadas, e postas em prática, pela comunidade que se dedica à compilação de corpos diacrónicos do português, por exemplo em Sousa (2007). E sugerimos a pertinência de estudos como os de Biber (1988) e Biber & Gray (2010) para caracterizar estatisticamente as dimensões de variação entre o português oral e escrito, na senda dos trabalhos pioneiros do Português Fundamental (1987) para Portugal e do projeto NURC para o Brasil (por exemplo Ilari, 2014). Esperamos, assim, que este pequeno corpo, público, com esta revisão, possa contribuir para uma maior compreensão da oralidade em português e da sua relação com a variabilidade no seio da língua.

Agradecimentos

Estamos muito agradecidas ao nosso colega José João Dias de Almeida por ter apresentado este trabalho no Encontro da APL em Braga, e o ter enriquecido com os seus comentários e opiniões. Estamos obviamente gratas ao Núcleo Português do Museu da Pessoa por nos ter cedido o material e o ter tratado, assim como ao Museu da Pessoa brasileiro por nos ter dado autorização para usar o seu material no AC/DC. Estamos gratas a Paulo Rocha por ter compilado mais algumas entrevistas e a Lise Bianchini por ter criado os metadados relativos aos entrevistados brasileiros. Finalmente, continuamos gratas à Fundação para a Computação

Científica Nacional (FCCN) por continuar a manter os recursos do AC/DC e da Linguateca públicos através da sua infraestrutura computacional, e ao Research Infrastructure Services Group da Universidade de Oslo por nos facilitar o processamento e desenvolvimento desses mesmos recursos.

Referências

Almeida, J. João, J. Gustavo Rocha, P. Rangel Henriques, Sónia Moreira & Alberto Simões (2000) Museu da Pessoa – arquitectura. In *Atas do Encontro da ABAD - Associação Bibliotecários e Arquivistas*.

Biber, Douglas (1988) *Variation across speech and writing*. Cambridge University Press.

Biber, Douglas & Bethany Gray (2010) Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9 (1), pp. 1-82.

Bick, Eckhard (2000) *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus: Aarhus University Press.

Brandão, Silvia Figueiredo (2011) Concordância nominal em duas variedades do português: convergências e divergências. *Veredas* (UFJF. Online) 15, pp. 164-178.

Ferreira, Maria Isabel Aldinhas. (1996) Fenómenos de Alternância na Estrutura Argumental de Predicadores Verbais: um Problema na Descrição Lexicográfica. In *Actas do XI Encontro da Associação Portuguesa de Linguística* (Lisboa, 2-4 de Outubro de 1995), APL, pp. 237-245.

Freitas, Cláudia, Bianca Freitas & Diana Santos (2016) QUEMDISSE?: Reported speech in Portuguese. In *Proceedings of LREC 2016.* Portoroz: ELDA. http://www.linguateca.pt/Diana/download/FreitasetalLREC2016.pdf

Gonçalves, Perpétua & Christopher Stroud (orgs.) (1997-2000) *Panorama do Português Oral de Maputo*. INDE, Moçambique, 4 volumes, 1997-2000.

Ilari, Rodolfo (org.) (2014) *Palavras de classe aberta*. Vol 3. da *Gramática do Português Culto Falado no Brasil*, Editora Contexto.

Jansen, Heidi (2016) *Objeto nulo no português: observações sobre a sua problemática*. Tese de mestrado, Universidade de Oslo.

Jon-And, Anna (2011) *Variação, contato e mudança linguística em Moçambique e Cabo Verde*. Tese de doutoramento, Universidade de Estocolmo.

Kato, Mary & Ian Roberts (eds.) (1993) *Português Brasileiro: uma viagem diacrônica*. Campinas: Editora da UNICAMP.

Português Fundamental, Volume II, Métodos e Documentos, tomo 1, Inquérito de Frequência, Lisboa: Centro de Linguística de Universidade de Lisboa, 1987.

Raso, Tommaso & Heliana Mello (2014) C-ORAL-BRASIL: Description, methodology and theoretical framework. In Tony Berber Sardinha & Telma São Bento Ferreira (eds.), *Working with Portuguese corpora*. Bloomsbury, pp. 257-276.

Santos, Diana & Eckhard Bick (2000) Providing Internet access to Portuguese corpora: the AC/DC project. In Maria Gavrilidou et al. (eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000,* pp. 205-210.

Santos, Diana (2004) Aonde vamos em relação a *aonde*. *The ESPecialist* 25 (1), pp. 85-103.

Santos, Diana (2014a) Corpora at Linguateca: Vision and roads taken. In Tony Berber Sardinha & Telma de Lurdes São Bento Ferreira (eds.) *Working with Portuguese Corpora*, Bloomsbury, pp. 219-236.

Santos, Diana (2014b) Como estudar variantes do português e, ao mesmo tempo, construir um português internacional? Apresentação no *Contact, Variation and Change: corpora development and analysis of Iberoromance language varieties workhop*, Estocolmo, 7-8 de abril de 2014. http://www.linguateca.pt/Diana/download/VariantesPIGSCP.pdf

Santos, Diana (2014c) Gramateca: corpus-based grammar of Portuguese. In Jorge Baptista, Nuno Mamede, Sara Candeias, Ivandré Paraboni, Thiago A.S. Pardo & Maria das Graças Volpe Nunes (eds.), *Computational Processing of the Portuguese Language, 11th International Conference, PROPOR 2014, São Carlos/SP, Brazil, October 6-8, 2014, Proceedings*, LNAI 8775. Springer, Heidelberg, pp. 214-219.

Santos, Diana (2016) Comparando corpos orais (transcritos) e escritos na Gramateca. In Bardel, Camilla & Anna De Meo (eds.), *Parler les langues romanes / Parlare le lingue romanze / Hablar las lenguas romances / Falando línguas românicas. Atti del Convegno Internazionale GSCP 2014*. Napoli: Università di Napoli L'Orientale, Il Torcoliere. No prelo.

Santos, Diana (2016) Português internacional: alguns argumentos. In José Teixeira (org.), *O Português como Língua num Mundo Global: problemas e potencialidades*, Centro de Estudos Lusíadas da Universidade do Minho, pp. 51-68.

Silva, Caio Cesar Castro da (2010) A variação *nós* e *a gente* no português culto carioca. *Revista do GELNE* 12 (1), Piauí, pp. 67-74.

Simões, Alberto & Diana Santos (2014) Nos bastidores da Gramateca: uma série de serviços. In Atas do *Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish, at PROPOR 2014, São Carlos, Brazil*, 9 de outubro de 2014, pp. 97-104.

Sousa, Maria Clara P. (2007) Sistema de edições eletrônicas do corpus histórico do português Tycho Brahe: fundamentos, diretrizes e procedimentos. Setembro, 2007. http://www.tycho.iel.unicamp.br/~tycho/corpus/manual/prep/pdf/manual_2007_print.pdf

Dicionários, gramáticas e outros recursos usados na revisão

Bechara, Evanildo (2009) *Moderna Gramática Portuguesa*. 37ª Edição, revista, ampliada e atualizada. Rio de Janeiro: Nova Fronteira.

Cunha, Celso & Lindley Cintra (1985) *Breve Gramática do Português Contemporâneo.* 16ª Edição. Lisboa: Edições João Sá da Costa.

Ciberdúvidas da Língua Portuguesa — https://ciberduvidas.iscte-iul.pt

Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa (2001). Lisboa: Academia das Ciências de Lisboa. Editorial Verbo.

Dicionário Priberam da Língua Portuguesa http://www.priberam.pt/DLPO/

D'Silvas Filho (1995) *Prontuário Universal. Erros corrigidos do Português*. 2.a Edição. Lisboa: Texto Editora.

Figueiredo, Cândido de (1913) *Novo Diccionário da Língua Portuguesa*. http://www.dicionario-aberto.net/dict.pdf

Gonçalves, Francisco Rebelo (1966) *Vocabulário da Língua Portuguesa.* Coimbra: Coimbra Editora

Infopédia, Dicionários Porto Editora http://www.infopedia.pt

Mateus, Maria Helena Mira, Ana Maria Brito, Inês Duarte & Isabel Hub Faria. *Gramática da Língua Portuguesa*. 2.ª Edição Revista e Aumentada. Lisboa: Editorial Caminho.

Peres, João Andrade e Telmo Móia (1995) *Áreas Críticas da Língua Portuguesa*. 2.ª Edição. Lisboa: Editorial Caminho.

Portal da Língua Portuguesa http://www.portaldalinguaportuguesa.org