

A elaboração de um *corpus* dialetal da língua espanhola falada na Argentina: contribuições dos gêneros discursivos e da análise textual automática

(The development of a dialectal *corpus* of Spanish language spoken in Argentina: contributions of discursive genres and automatic text analysis)

Leandro Silveira de Araujo¹

¹ Faculdade de Ciências e Letras – Universidade Estadual Paulista ‘Júlio de Mesquita Filho’ (UNESP)

leandrocbt@msn.com

Resumen: La composición de un *corpus* que permite la observación de la lengua española en las siete regiones dialectales de Argentina proviene de la ausencia de un material organizado con tal naturaleza. Se sumó a su proceso de formación la atención a las características del contexto enunciativo, proporcionándonos un conjunto de enunciados que permite el estudio más preciso de los fenómenos lingüísticos considerados. El análisis del género discursivo *entrevista radial* nos mostró que los enunciados recogidos nos proporcionan un material con un lenguaje más espontáneo, es decir, menos monitoreado. Por último, se consideró de gran relieve el apoyo de los *softwares* de grabación de audio *Audacity* y del procesador semántico *Tropes*, ya que nos permiten, respectivamente, la recogida y el análisis automático de textos.

Palabras clave: *Corpus*; Entrevista Radial; Lengua Española; Dialectología, Argentina.

Resumo: A composição de um *corpus* que viabiliza a observação da língua espanhola nas sete regiões dialetais da Argentina decorre da ausência de um material organizado com tal natureza. Somou-se a seu processo de constituição a atenção às características do entorno enunciativo, facultando-nos um conjunto de enunciados que possibilita o estudo mais apurado dos fenômenos linguísticos considerados. A análise do gênero discursivo *entrevista radiofônica* mostrou-nos que os enunciados coletados propiciam-nos um material com um uso linguístico mais espontâneo, isto é, menos monitorado. Finalmente, a apoio dos *softwares* de gravação de áudio *Audacity* e do processador semântico *Tropes* foi considerado de grande relevância por possibilitarem, respectivamente, a coleta e análise automática de textos.

Palavras-chave: *Corpus*; Entrevista Radiofônica; Língua Espanhola; Dialectologia; Argentina.

Corpora em espanhol

Se nos dedicarmos a uma busca de *corpora* em língua espanhola de acesso disponível na internet, encontraremos materiais que aparentemente poderiam servir de consulta e, conseqüentemente, resposta aos objetivos traçados em trabalhos sobre o espanhol argentino. No entanto, como observaremos, nenhum deles nos oferece a totalidade de informações extralinguísticas que julgamos imprescindíveis, isto é, não nos apresentam dados sobre idade, nível de escolaridade, gênero discursivo, sexo e, fundamentalmente, origem geográfica dos enunciadorees envolvidos. Há *corpora* que trazem algumas dessas informações, contudo não todas. O levantamento de projetos de criação de *corpus* em língua espanhola (PÉREZ HERNÁNDEZ, 2002) mostra-nos que a construção de *corpora* nesse idioma ainda está em desenvolvimento, constituindo um cenário longe do ideal. A seguir, o quadro mostra-nos um panorama de *corpora* existentes em espanhol:

Quadro 1. Do panorama de corpora em língua espanhola

Nome	Site	Instituição de Origem	Nº de Palavras	Características
1. Word Sketch Engine	http://www.sketchengine.co.uk/	University of Leeds (Reino Unido)	116.900.060	- Sem informação extralinguística; - Possui <i>corpora</i> de diferentes línguas: Inglês, Francês, Português, Espanhol, etc.
2. Corpus del Español	http://www.corpusdelespanol.org/x.asp	Brigham Young University (EUA)	101.311.682	- Textos desde o Séc. XII; - Separação simples em gêneros textuais; - Não há qualquer divisão dialetal ou apresentação de qualquer outra informação do enunciador.
3. Corpus de referencia del Español Actual – CREA	http://www.corpusdelespanol.org/x.asp	Real Academia Española – RAE (Espanha)	Mais de 160 milhões	- Apresenta-se o país de origem; - Há informação de gênero textual, temática e modalidade; - Não se informam idade, origem geográfica específica e escolarização;
4. Corpus diacrónico del español – CORDE	http://corpus.rae.es/cordenet.html	Real Academia Española – RAE (Espanha)	Mais de 250 milhões.	- Do início do idioma até o ano 1945; - Há informação de gênero textual, temática e modalidade; - Não se informam idade, origem geográfica e escolarização do enunciador.
5. Corpus oral de referencia del español contemporáneo - CORLEC.	http://www.llf.uam.es/ESP/Info%20Corlec.html	Universidad Autónoma de Madrid (Espanha).	1.100.000	- Destinado à variedade peninsular; - Modalidade oral somente; - Informam-se dados extralinguísticos; - Diversificam-se os gêneros textuais.
6. Corpus Oral y Sonoro del Español Rural – COSER	http://www.llf.uam.es:8888/coser/	Universidad Autónoma de Madrid (Espanha).	---	- 800 horas de gravação; - Destinado à variedade peninsular; - Informam-se dados extralinguísticos; - Grande diversidade de informantes;
7. Corpus lingüístico del español contemporáneo – CUMBRE	---	SGEI Editorial (Espanha)	Mais de 20 milhões de palavras.	- Conteúdo não disponível.
8. Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico – MC-NLCH	---	Universidad de Las Palmas de Gran Canaria (Espanha)	---	- Conteúdo não disponível; - 84 horas de gravação; - Amostra paralela de 12 cidades hispânicas: México, Caracas, Santiago de Chile, Santafé de Bogotá, Buenos Aires, Lima, San Juan de Puerto Rico, La Paz, San José de Costa Rica, Madrid, Sevilla y Las Palmas de Gran Canaria; - Amostras distribuídas por faixa etária e sexo.

9. Proyecto para el Estudio Sociolingüístico del Español de España y de América – PRESEEA	http://www.linguas.net/Default.aspx?alias=www.linguas.net/portal-preseea	---	---	- Conteúdo não disponível; - <i>Corpus</i> da modalidade oral, representativo do mundo hispânico em sua variedade geográfica e social.
10. Corpus de conversación coloquial - Val.Es.Co	http://www.valesco.es/	Universidad de Valencia (Espanha)	---	- Conteúdo não disponível; - 341 horas de gravação.

Apesar de mais da metade dos *corpora* ter seu conteúdo divulgado na rede mundial de computadores e ser sustentada por um eficiente sistema de processamento computacional, que viabiliza um manuseio rápido e preciso dos dados, as bases de dados a que tivemos acesso apresentaram algum tipo de deficiência na descrição de informações extralinguísticas, seja pela ausência desse conteúdo ou por sua sistematização incompleta. O interesse pela análise diatópica da língua espanhola falada na Argentina, força impulsionadora do estudo, fez com que todos esses *corpora* se tornassem inadequados. Isso porque ou não se oferece a informação dialetal, ou se oferece, com um particular interesse nas variedades peninsulares da língua espanhola. Daqueles a que tivemos acesso integralmente ao conteúdo (1 – 6), somente o *CREA* (3), o *CORLEC* (5) e o *COSER* (6) forneceram-nos essa informação. No entanto, o *Corpus oral de referencia del español contemporáneo* (5) e o *Corpus Oral y Sonoro del Español Rural* (6) restringem-se ao espanhol peninsular. Por seu turno, o *Corpus de referencia del Español Actual* (3) estende seu registro a dados de vinte e dois países hispanofalantes. Apesar de sua maior especificidade, o *CREA* não nos possibilita o estudo das diferentes regiões dialetais argentinas, pois seus dados são apresentados de modo generalizado para todo o país.

Frente à inadequação dos *corpora* existentes em língua espanhola para o conhecimento do espanhol falado nas regiões dialetais da Argentina, fez-se necessária a constituição de um *corpus* que nos conduzisse a esse estado e que, ademais, possibilitasse o conhecimento do informante (escolarização, idade, procedência geográfica, sexo) e de sua situação de enunciação (modalidade da língua e gênero textual). Somam-se a essas especificidades, as limitações espacial e econômica do pesquisador, obrigando-o a coletar os enunciados e a obter informações extralinguísticas a partir do Brasil. A opção pelo trabalho com *entrevistas radiofônicas* provém da certeza de encontrarmos uma situação que satisfaça as exigências acima relatadas, além de nos permitir observar o uso efetivo e prestigiado da língua espanhola; isso porque:

En Hispanoamérica, la radio se apoya en las normas de prestigio, nacionales y regionales. La radio permite el acceso instantáneo a las voces y a las ideas de todos los miembros de la sociedad, como oyentes o como participantes ocasionales. El examen de la lengua de los programas de radio es un buen índice de las actitudes lingüísticas, así como del uso real. (LIPSKI, 1994, p.162)

A seguir, o estudo do gênero entrevista radiofônica contribuirá para entendermos os textos que compõem o *corpus* e os benefícios de trabalharmos com eles.

A entrevista radiofônica: apresentação do gênero discursivo

A fim de refazer o percurso epistemológico da conceitualização do gênero *entrevista radiofônica*, Pérez Cotten e Tello (2004, p. 28) dizem haver “[...] um formato ou gênero que se denomina entrevista jornalística”,¹ e acrescentam que, “conforme sua difusão em suporte papel ou por meios eletrônicos a entrevista jornalística é escrita, televisiva ou radiofônica”.² Em outras palavras, verificamos nessas duas asserções a definição do *domínio discursivo* a que pertence o gênero em questão, bem como o *suporte* de sua circulação social. Isto é, podemos entender que dentro da *esfera jornalística* há um modo de ação comunicativa estável que se chama *entrevista*, o qual, como sabemos, define-se pelas especificidades, exigências e necessidades do jornalismo. A característica que aporta o suporte (*rádio*) ao gênero (*entrevista jornalística*) mostra-nos que apesar de o suporte não alterar o valor dos textos que divulga, ele auxiliará na definição do gênero. No caso da rádio, um suporte virtual convencional, a ausência da imagem (presente na televisão) ou da possibilidade de releituras (possível no jornal impresso), entre outros traços que veremos mais adiante, fazem com que a entrevista radiofônica se diferencie das entrevistas difundidas por outros suportes do mesmo domínio discursivo.

Como salienta Marcuschi (2008), a funcionalidade do gênero é um elemento que deve ser levado em consideração quando queremos descrevê-lo. Visando encontrá-la, verificamos novamente nas palavras de Pérez Cotten e Tello (2004) o possível propósito da entrevista radiofônica:

Sirve para la construcción discursiva de diferentes relatos periodísticos [...] y, también, para escuchar directamente la voz del entrevistado [...] en ambos casos, el objetivo inicial (y por cierto, final) es el de producir conocimiento [...]. (p. 29)

La entrevista, por lo tanto, es el principal recurso periodístico para acceder a la información, ampliar una noticia, obtener la voz de algún personaje. (p. 34)

A estruturação desse gênero relaciona-se, portanto, à reconstrução de um evento por meio do discurso, à apreensão e à avaliação de opiniões envolvidas com acontecimentos das mais diversas esferas da sociedade. Ou seja, a entrevista radiofônica se organiza em função do *informar*, fazendo da informação, *notícia*.³ O curioso, no entanto, é que, se nos detemos somente nessa característica, não avançamos muito no conhecimento das particularidades desse gênero, isso porque difundir informações pertence, supomos, à

1 “[...] un formato o género que se denomina entrevista periodística.” (PÉREZ COTTEN; TELLO, 2004, p. 28). Estabelecemos o padrão de traduzir para o português todas as citações em língua estrangeira menores de três linhas e, por isso, inseridas no corpo do texto. Os originais serão expostos em nota de rodapé. Assumimos a responsabilidade por todas as traduções feitas dentro deste padrão.

2 “[...] según se difunda en soporte papel o por medios electrónicos, la entrevista periodística es escrita, televisivo o radial.” (PÉREZ COTTEN; TELLO, 2004, p. 28).

3 Pérez Cotten e Tello (2004, p. 29) diferenciam *informação* de *notícia*. A primeira seria somente o relato de um evento ocorrido ou que está por acontecer, por sua vez, a *notícia* é a informação que se tornou difundida pelo interesse despertado em seu espectador. Logo, toda notícia implica necessariamente uma informação, o contrário não é verdadeiro.

essência do domínio discursivo em questão e, por tanto, é de se esperar que também verifiquemos essa finalidade em outros gêneros do discurso pertencentes à esfera jornalística. É por isso que a diferenciação da *entrevista radiofônica* vai se dando, de fato, pela identificação de novas características. Assim, como se lê nos dois fragmentos acima, o aporte de novas informações se dá, neste gênero, pela recepção da “voz” direta do entrevistado (“voz de algum personagem”). *Voz* que, por sua vez, responde outra, geralmente indagadora. A atitude responsiva verificada nesse embate resgata um caráter dialógico, que na entrevista é visto como:

[...] un encuentro de absoluta formalidad donde los roles están bien definidos. Hay un actor que propicia el desarrollo del conocimiento de un tema a través de otro actor. Pero este actor conoce algo sobre el tema, tiene algunas ideas. Éste, a través de preguntas facilita que se produzca conocimiento nuevo sobre determinado tema. (PÉREZ COTTEN; TELLO, 2004, p. 25)

Ou seja, o diálogo, além de possuir uma forma bem definida e conhecida por seus participantes, é instaurado por dois atores, cujos papéis estão bem estabelecidos: um que se acredita dominar o conteúdo que está na pauta da entrevista e outro que, a partir do conhecimento superficial que adquiriu sobre o tema, conduz o debate, tentando facilitar, por perguntas, o descobrimento de novas informações sobre o assunto. Na entrevista radiofônica, esse diálogo formalizado com fins informativos assume um caráter eminentemente *público*, pois se constrói fundamentalmente para difundir uma informação (transformando-a em notícia) e, logo, satisfazer o anseio que o ouvinte tem por determinado assunto. Não é por acaso que se caracteriza a entrevista nesta esfera discursiva como a mais pública das conversas privadas (PÉREZ COTTEN; TELLO, 2004).

A cena constituída por entrevistador e entrevistado ($Eu \Leftrightarrow Tu$) debatendo um tema (Ele) comportaria, aparentemente, uma situação suficiente para a concretização de enunciados. No entanto, isso não é assim no diálogo da entrevista jornalística. Graças ao traço *público*, há nesta entrevista, imprescindivelmente, a exigência de mais um personagem, o telespectador, quem motiva e, por fim, recebe toda a mensagem. A figura 1 auxilia-nos na compreensão do gênero sob o ponto de vista do diálogo (bakhtiniano).

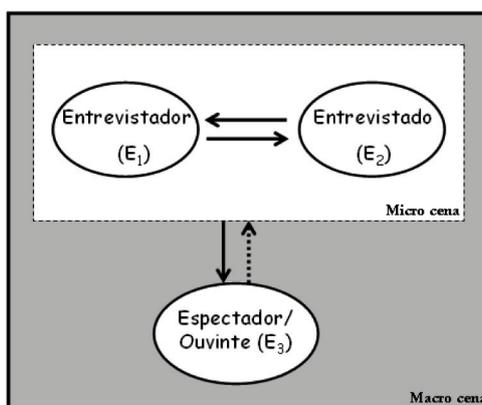


Figura 1. Do diálogo na entrevista radiofônica

Nela, percebemos uma micro cena, na qual instaura-se um diálogo entre entrevistador (E_1) e entrevistado (E_2) para satisfazer, unicamente, a necessidade informativa que tem o espectador/ouvinte (E_3), na macro cena. Por seu turno, a resposta do ouvinte (E_3) ao

enunciado composto por E_1 e E_2 não é necessariamente imediata, isso porque pode ser dada de diferentes formas: desde modos mais perceptíveis (como em comentários por chamadas telefônicas à rádio) até de maneiras menos conexas (por exemplo, com um diálogo traçado entre E_3 e um outro interlocutor (E_n) com quem possa interagir, até mesmo, assincronicamente). Por isso mesmo, escolhemos representar essa interação por meio de uma flecha pontilhada

Frente a tal cenário, parece propício considerarmos *secundário* o gênero *entrevista radiofônica*, pois por se valer da esfera *jornalística* e do *rádio* (“circunstâncias de uma comunicação cultural mais complexa e relativamente mais evoluída” (BAKHTIN, 1997, p. 281), este gênero apoia-se em uma rebuscada estruturação do diálogo, que, como se sabe, vai além da simplicidade verificada no diálogo cotidiano (*strictus sensus*). Assim, em um primeiro plano (micro cena), constrói-se um diálogo (entre E_1 e E_2) mais próximo do cotidiano, no qual as relações são sincrônicas, mais espontâneas e onde, aparentemente, se configura a entrevista. No entanto, como vimos, essa micro cena não tem valor por si mesma; sua finalidade será informar E_3 (macro cena) e, por tanto, estabelecer uma relação dialógica (*latus sensus*) com ele. Tanto é assim que os integrantes do primeiro plano têm consciência de que estão ali por causa do interlocutor (E_3), o ouvinte da rádio. As interações nesta macro cena ($[E_1 \leftrightarrow E_2] \leftrightarrow E_3$) são viabilizadas pelo entrevistador (E_1) e já não são necessariamente tão evidentes, isto é, sincrônicas.

Em outras palavras, o que aparentemente poderia configurar um gênero primário ($E_1 \leftrightarrow E_2$), torna-se componente do gênero secundário ($[E_1 \leftrightarrow E_2] \leftrightarrow E_3$), isto é, “[...] transforma-se dentro deste e adquire uma característica particular: perde sua relação imediata com a realidade existente [...]” (BAKHTIN, 1997, p. 281), à medida que se estrutura para satisfazer as necessidades presentes na realidade que envolve o ouvinte (E_3). Encontramos nas palavras de Farneda (2007, p. 2) a definição sintética do que tratamos nos parágrafos anteriores:

A entrevista radiofônica é um gênero jornalístico que diz respeito ao encontro de um entrevistador (jornalista) e um entrevistado (especialista em um assunto em particular), cujo interesse é fazer falar o *expert* a respeito dos diferentes aspectos de uma questão e, dessa forma, levar as informações fornecidas, por essa interação, a terceiros. Sendo contrária a uma conversa comum, a entrevista apresenta um caráter estruturado e formal, cujo objetivo é satisfazer as expectativas do destinatário.

Atentando-nos à figura do entrevistador, sabemos que ele deve buscar não só a simpatia de seu espectador, mas também a de seu entrevistado, quem lhe poderá conceder maior informação à medida que se estabeleça um vínculo de maior confiabilidade. Daí que se infere que “a entrevista é um ato de *sedução mútua*. O entrevistador seduz para obter mais e melhores respostas. O entrevistado procura convencer seu entrevistador, leva-lo a seu jogo, dizer o que quer dizer e evitar dizer o que não quer dizer.”⁴ Para que o entrevistador tenha controle da entrevista, sugere-se que ele não seja muito jovem, mas no caso de o ser, deve deixar claro que é um profissional competente. Outras estratégias sugeridas são: ter clara a finalidade da entrevista proposta, conhecer o perfil do programa

4 [...] la entrevista es un acto de *seducción mutua*. El entrevistador seduce para obtener más y mejores respuestas. El entrevistado busca convencer a su entrevistador, llevarlo a su juego decir lo que quiere decir y evitar decir lo que no quiere decir. (PÉREZ COTTEN; TELLO, 2004, p. 37).

que a receberá, escolher o entrevistado adequado e manter contato prévio, reunir informações sobre ele e verificar antecedentes sobre seu desempenho em outras entrevistas realizadas (PÉREZ COTTEN; TELLO, 2004, p. 88). Portugal e Yudchak (2008, p. 76) sugerem, ainda, que o entrevistador esteja armado de uma pauta de questões, mas alerta para o perigo de ficar limitado a ela, perdendo a possibilidade de seguir por caminhos abertos pelo entrevistado na prática discursiva, os quais até então eram inimagináveis⁵.

Outro aspecto da entrevista radiofônica que merece especial atenção é a, já citada, presença da voz do entrevistado. O uso do discurso direto (1) aproxima o ouvinte do entrevistado, já que aquele sabe que este fala para ele; (2) elimina a intermediação direta de um terceiro (jornalista), criando uma (3) aparente objetividade; (4) permite que o ouvinte tire suas próprias conclusões, protegendo, assim, o jornalista de eventuais asserções comprometedoras; (5) cria empatia ou antipatia entre ouvinte e entrevistado; além, é claro, de (6) permitir o conhecimento imediato da voz e os testemunhos envolvidos nos episódios noticiados (PÉREZ COTTEN; TELLO, 2004, p. 33).

Haja vista que a presença direta de pelo menos duas vozes (entrevistador e entrevistado) é imprescindível neste gênero, o estudo de sua *construção composicional* deverá considerar o modo como, formalmente, organiza-se o diálogo nessa interação verbal. Neste sentido, Farneda (2007, p. 2) mostra-nos que o “entrevistador e o entrevistado ocupam posições assimétricas”, mas que este “deve respeitar a agenda de perguntas prevista pelo jornalista. Os turnos de pergunta devem sempre terminar com uma interrogativa” e por sua vez, os turnos do entrevistado revestem-se unicamente de respostas. Ao entrevistador cabe a abertura e o fechamento da entrevista e não lhe compete “formular expressões de ratificação, opinião ou comentário, abstendo-se de formar opinião contra ou a favor do entrevistado”. A fim de transmitir uma neutralidade, o entrevistador controlará os turnos por meio de falas mais breves. O entrevistado, por sua vez, terá falas mais longas a fim de transmitir a informação que detém. A estas características composicionais, somam-se as instruções dadas por Portugal e Yudchak (2008, p. 76) para a condução da entrevista radiofônica. Para os autores, deve-se fazer uma pergunta de cada vez, de modo claro, curto e concreto.

Portugal e Yudchak (2008, p. 84) destacam ainda a preferência por uma linguagem simples e, a partir de relatos de importantes jornalistas, mostram-nos o *estilo* mais empregado neste gênero. Assim, Alfredo Leuco, jornalista cordobês, diz utilizar “uma linguagem mais simples, frases curtas, [apelar] à sabedoria popular, à linguagem da rua, à linguagem coloquial, [...] com o que se ganha maior familiaridade e proximidade”.⁶ Notemos como o uso de uma linguagem menos erudita possibilita a aproximação entre jornalista e sua audiência. Graciela Manscusó diz que “hoje a linguagem foi liberada, fala-se como na vida”⁷ e Chiche Gelblung afirma que “a rádio tem que falar como o ouvinte”.⁸ Esta jornalista

5 Comentando esse tipo de preparo prévio, Szymanski (2002) diz que essa ferramenta pode inibir o desenvolvimento da entrevista, ocasionando a não divulgação de importantes informações que poderiam ser dadas pelo informante em um diálogo fluido. A autora sugere que esse instrumental seja usado somente como apoio, em momentos em que o diálogo, por algum motivo, torna-se escasso.

6 “[...] un lenguaje sencillo, frases cortas, [] apelo a la sabiduría popular, al lenguaje de la calle, al lenguaje coloquial, [...] con lo que se gana familiaridad, cercanía.” (PORTUGAL; YUDCHAK, 2008, p. 84).

7 “Hoy se ha liberado el lenguaje, se habla como en la vida.” (PORTUGAL; YUDCHAK, 2008, p. 85).

8 “La radio tiene que hablar como el oyente.” (PORTUGAL; YUDCHAK, 2008, p. 85).

destaca ainda o caráter espontâneo que tem a linguagem nesta interação discursiva: “em realidade, a rádio tem que falar como você quer, nada substitui o pessoal”⁹.

É evidente que diante de uma *mídia* de ampla difusão e expressividade, não podemos acreditar que, de fato, seu uso linguístico é exatamente igual ao uso comum, rotineiro e casual¹⁰, no qual não se observa preocupação com a fala. Diante dessa situação de enunciação, o falante tende a monitorar, ainda que discretamente, sua fala a fim de alcançar um padrão linguístico que esteja de acordo com um meio de maior prestígio e desenvolvimento tecnológico, como é a rádio. Por outro lado, a existência de uma maior espontaneidade nesse gênero pode estar relacionada, em partes, à modalidade de concepção e transmissão dos enunciados, isto é, por pertencer ao domínio da oralidade. A impossibilidade de fazer releituras corretivas é outra característica que a modalidade atribui ao gênero, isto porque uma vez enunciada, a mensagem não pode ser apagada.

Marcuschi (2008), por meio da figura 2, mostra-nos que a *entrevista radiofônica* pertence ao domínio da fala, e que, por isso, figura abaixo da linha tracejada. O fato de estar mais ao lado esquerdo desta imagem, onde o monitoramento é mais baixo, mostra-nos mais uma vantagem da utilização, nos *corpus* de análise, de textos pertencentes a este gênero textual. Isso porque, o pouco monitoramento, reflexo da baixa formalidade e da intenção discursiva, aproxima o uso linguístico neste gênero ao uso das situações mais simples e cotidianas, isto é, do uso vernacular.

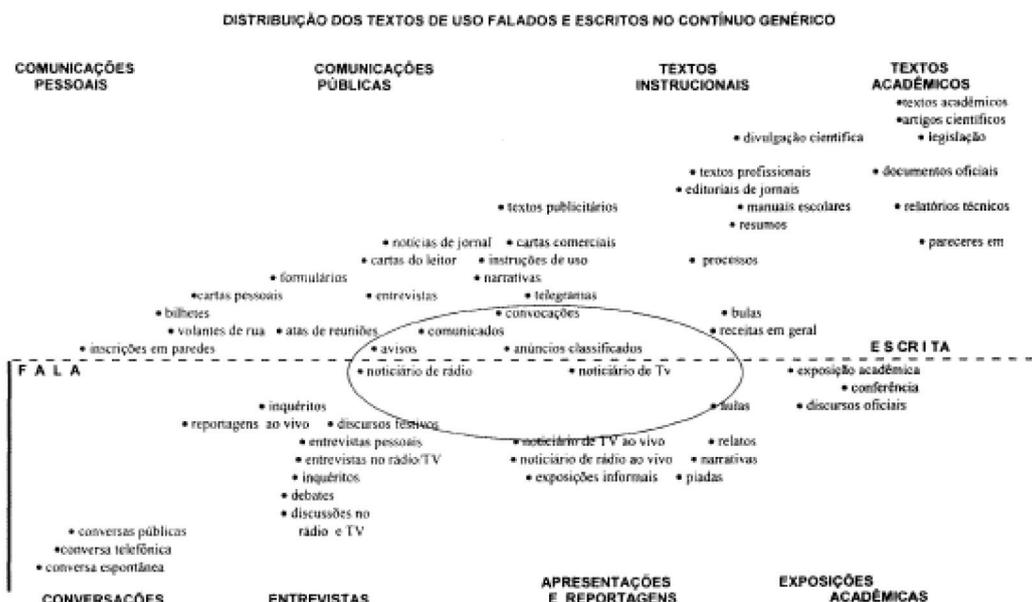


Figura 2. Da distribuição dos gêneros no *continuum* da relação fala-escrita (MARCUSCHI, 2008, p. 197)

Nosso próximo passo será explorar o *corpus* constituído, isto é, descrever os enunciados coletados e verificar como as características do gênero discursivo em foco realizam-se neles. Antes, porém, vamos nos ater a um rápido esclarecimento dos critérios de seleção das cidades representantes das regiões dialetais e à apresentação de dois *softwares*

9 “[...] en realidad la radio tiene que hablar como vos lo sientas, nada reemplaza a lo personal.” (PORTUGAL; YUDCHAK, 2008, p. 85).

10 Segundo Labov (2008, p. 111), “por fala casual, em sentido estrito, entendemos a fala cotidiana usada em situações informais, em que nenhuma atenção é dirigida à linguagem”.

que nos auxiliaram na montagem do *corpus* e que viabilizarão, de forma mais rápida e precisa, o processamento e compreensão de seus dados.

Corpus diatópico da Argentina

A variável diatópica

Um estudo da *dialetologia hispano-americana* (MORENO DE ALBA, 2000) deverá nos mostrar a inexistência de um acordo comum para a divisão dialetal do continente e tão pouco da Argentina; conjuntura que nos coloca em uma situação um pouco incerta sobre que proposta de fragmentação das áreas dialetais assumir a fim de realizarmos a coleta de enunciados conforme as diferentes zonas linguísticas do país. Devido à ausência deste acordo, decidimos partir da proposta de divisão dialetal feita por Fontanella de Weinberg (2004), quem divide o país em sete regiões dialetais: *bonaerense, de Cuyo, noroeste, central, nordeste, do litoral e patagônica*. Diante da impossibilidade de observarmos falantes de todas as cidades pertencentes a cada uma das sete regiões dialetais, procuramos nos orientar pelo princípio de *irradiação linguística* (COSERIU, 1977). O qual aponta os centros políticos, administrativos, culturais, comerciais e de comunicação como os principais centros propagadores de um padrão linguístico. É consciente disso que Moreno de Alba (2000) critica os atlas linguísticos que ignoraram a fala das grandes cidades e afirma que “são precisamente as grandes urbes, focos de irradiação linguística, que merecem especial atenção se o que se persegue é o conhecimento do estado que guarda a língua em um espaço dado”.¹¹

Tendo em vista esse princípio e nos orientando pelos dados fornecidos pelo censo argentino de 2010,¹² buscamos, em cada uma das sete regiões dialetais argentinas, a cidade de maior índice populacional – isso porque pressupomos que cidades detentoras dos principais serviços de ordem política, administrativa, comercial e cultural são as que tendem a atrair mais habitantes. Assumiremos, portanto, a *Capital Federal, a Região Metropolitana de Mendoza* e as cidades de *São Miguel de Tucumán, Córdoba, Posadas, Rosario, e Comodoro Rivadavia* como representantes, respectivamente, das sete regiões dialetais da Argentina. Feita a devida apresentação das cidades representantes de cada umas das regiões, passemos à apresentação das ferramentas tecnológicas que nos auxiliaram o estudo.

Ferramentas tecnológicas para estudos da linguagem

Audacity 1.3. Trata-se de um *freeware* de gravação e edição de áudio. Por sua facilidade de manuseio, serviu-nos para a gravação das entrevistas radiofônicas transmitidas por internet, edição das mesmas, além de nos auxiliar na transcrição dos enunciados. A informação do tempo de gravação é também um dado fornecido pelo programa por meio da linha do tempo. A seguir, expomos uma imagem da tela de entrada do Audacity, 1.3.

11 “Son precisamente las grandes urbes focos de irradación lingüística que merecen especial atención si lo que se persigue es el conocimiento del estado que guarda la lengua en un espacio dado.” (MORENO DE ALBA, 2000, p. 138).

12 Para mais informações, acessar <<http://www.censo2010.indec.gov.ar/>>. Acesso em 10 jan. de 2011.

A partir dos resultados (que apontam estilo geral do texto, frases relevantes, universos de referência, relações entre referências, categorias e lista de palavras preponderantes, entre outros), o *software* possibilita a (a) classificação automática das palavras do texto; (b) visualização direta dos fragmentos pertinentes; (c) filtragem dos temas em função de sua relevância; (d) gráficos que permitem visualizar cada referência detectada em seu contexto; (e) ferramentas que permitem localizar rapidamente as séries cronológicas no interior de um documento e compreender imediatamente a estratégia do narrador; (f) cenários semânticos que permitam produzir ferramentas de interpretação e explorar as classificações personalizadas.¹³ Como veremos mais adiante, as análises oferecidas pelo *Tropes* irão nos auxiliar na descrição das entrevistas radiofônicas, oferecendo-nos, mais especificamente, informações sobre o estilo e o conteúdo temático do gênero. Passemos, então, para a exploração, de fato, do *corpus* por nós elaborado.

Composição do *corpus*

Pérez Hernández (2002) explica-nos que a revolução tecnológica vivenciada pela humanidade nas últimas décadas viabilizou, entre tantas outras coisas, o estudo linguístico baseado na análise conjunta de enunciados pertencentes às múltiplas situações de interação discursiva, afastando-nos, deste modo, de análises baseadas unicamente em intuições linguísticas do pesquisador. Com auxílio desses avanços, constituímos o *corpus* de análise a partir de materiais linguísticos que registram “a linguagem natural realmente utilizada por falantes e escritores da língua em situações reais” (BERBER SARDINHA, 2000, p. 352). Mas quais as características que um *corpus* deve apresentar para que seja considerado como tal? Conforme a definição de Sanchez (1995), será *corpus* o

[...] conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistemizados según determinados criterios, suficientemente extensos en amplitud y profundidad de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con el fin de obtener resultados varios útiles para la descripción y el análisis. (SÁNCHEZ, 1995, p. 8-9)¹⁴

Em sua asseveração, percebemos que um *corpus*, além de constituir-se por enunciados, ou seja, por textos efetivamente realizados em língua natural, deverá seguir critérios que lhe assegurem representatividade de, pelo menos, algum âmbito do idioma. Estes critérios, por sua vez, se relacionam a condições de naturalidade, de autenticidade, de diversidade, de dimensões e a outros padrões que possam estar vinculados aos propósitos do investigador. A disponibilidade para o processamento digital é também uma das características que deve possuir o *corpus*, haja vista que com o auxílio da tecnologia podemos analisar, de forma muito mais precisa e ágil, uma grande quantidade de textos.

Tendo em vista nossas limitações espaciais e temporais, sem deixar de lado algumas destas características, acreditamos haver encontrado condição que satisfaça as exigências propostas por Sánchez (1995) em um *corpus* composto por *entrevistas radiofônicas*. Isto porque, além da possibilidade de obtermos estes enunciados por meio da internet – em

¹³ Dados fornecidos pelo site difusor do *software*: <http://www.semantic-knowledge.com/tropes.htm>.

¹⁴ Grifos nossos.

rádios das sete regiões dialetais que disponibilizam sua transmissão *on-line* –, este gênero, como já mencionamos, resgata uma variedade linguística não muito distante do vernáculo.

Conscientes de que enunciados pertencentes a um só gênero e de uma única modalidade da língua (falada), não podem constituir um *corpus* representativo da totalidade de usos linguísticos de uma comunidade de fala, as considerações provenientes da análise do *corpus* deverão estar limitadas a um importante âmbito do uso do castelhano na Argentina, no qual, se observa o domínio da oralidade, com pouco monitoramento e, logo, mais espontâneo. Além disso, o recurso a esse gênero nos garante a observação da fala de uma quantidade significativa de informantes das sete regiões dialetais do país, tornando viável um estudo diatópico.

Por assumirmos os pressupostos da Sociolinguística e da Dialectologia, torna-se imprescindível a obtenção de informações que nos descrevam os indivíduos e seu entorno de enunciação, principalmente no que diz respeito a seu local de origem. Mais uma vez, a opção por esse gênero e o apoio da internet possibilitaram o acesso a esses dados ora por inferência na própria entrevista, ora por contato direto com as rádios (*e-mail*) ou, até mesmo, por meio de rede de relacionamentos (*facebook*). Sobre a obtenção dos textos, o uso do *software Audacity 1.3*, como já mencionado, serviu-nos para gravação das entrevistas – as quais compartilham preponderantemente a característica de serem ao vivo. A seguir, dispomos um quadro que relaciona as regiões dialetais da Argentina com informações das entrevistas.

Quadro 2. Da descrição das entrevistas radiofônicas que compõem o *corpus*

Região	Rádio	Site	Programa	Nº Entrev.	Tempo de gravação total	Nº de palavras	Nº de Inform.	Idade (anos)	Sexo Femin.
1. Bonaerense (Buenos Aires)	Continental	www.continental.com.ar	La mañana	4	00:46'58"	7.770	6	37 - 70	2
	Palermo	www.radiopalermo.com.ar	Comunas en Plural						
			Entre nosotras						
2. Cuyana (Mendoza)	LV10	www.lvdiez.com.ar	La deportiva	6	00:47'14"	8.104	13	33 - 51	4
3. Noroeste (S.M. de Tucumán)	LV12 (Indep.)	www.lv12.com.ar	Tiempos Modernos	5	00:47'56"	8.467	8	37 - 59	2
	LV7	www.lv7.com.ar	Manyines en la radio						
4. Central (Córdoba)	Cadena 3	www.cadena3.com	La mañana de LV7	5	00:50'56"	9.088	11	21 - 67	0
			Tarde						
			Carrusel						
5. Nordeste (Posadas e outras)	Antena Uno	www.antenamisiones.com	Sucesos deportivos	5	00:48'40"	7.995	13	31 - 66	3
	República	www.republica99.com.ar	Meda de café						
	Misiones	www.notciasdel6.com	Bijo la radio						
			Estamos en vuelta						
6. Litoral (Rosario/M. Juana)	Meridiano	www.fmmeridiano.com.ar	La mañana	4	00:48'32"	8.441	7	24 - 60	1
	María Juana	www.radiomariajuana.com	Jorge Kurrie						
	La deportiva	www.lodeportiva.com.ar	XXX						
7. Patagonia (Com. Rivadavia)	Visión	www.fmradiovision.com	Zona de noticias	4	00:46'34"	7.204	9	37 - 55	0
	Plus 10	www.plus10.com.ar	Comunicándonos						
			Punta pie inicial						
TOTAL	14 rádios		22 programas	33	05:37'15"	57.069	67	21 - 70	12

Como podemos observar, as 5h37min15seg, referentes à gravação de 33 entrevistas radiofônicas (Nº Entrev.), forneceram-nos mais de 57 mil palavras,¹⁵ sendo, em média, mais de oito mil a quantidade de palavras provenientes de cada região. Vale a pena repararmos que dentro do grupo de 67 falantes (Nº de Inform.) encontramos indivíduos que possuem desde 21 anos até 70 anos. No entanto, os dois extremos não formam parte do grupo etário

15 As informações de tempo de gravação e quantidade de palavras foram obtidas, respectivamente, por meio da linha do tempo exposta no *Audacity 1.3* e pela ferramenta contar palavras, do editor de texto *Microsoft Office Word 2003*.

preponderante neste gênero. Se realizarmos a média aritmética da idade dos 67 indivíduos, encontraremos que a idade mediana dos falantes é de 44 anos. Em relação ao sexo dos informantes, nota-se claramente a preponderância de homens, número que chega aos 82%. Esses dois dados (idade e sexo) fazem-nos refletir, mais uma vez, sobre o gênero discursivo em questão. Seria a entrevista radiofônica marcada pela maior presença de uma população masculina e adulta?

Seguindo a tipologia proposta pela Linguística de *Corpus* (BERBER SARDINHA, 2000), nosso material de análise se identifica com o *modo falado*, pois tanto em sua concepção como em sua propagação faz uso da oralidade. Enquanto ao *tempo*, trata-se de um *corpus* sincrônico e contemporâneo, por abordar um único período: o corrente. É *dialetal e especializado*, por apresentar, respectivamente, um *conteúdo* que visa satisfazer um estudo dialetológico e por decorrer de um único domínio discursivo: o jornalístico. Por fim, é um *corpus* de *língua nativa*, já que seus *autores* também são nativos. Em relação ao tamanho do *corpus* elaborado e sua relativa representatividade, é pertinente observarmos que, como propõe Berber Sardinha (2000, p. 346), lidamos com um *corpus* pequeno, por possuir menos de 80 mil palavras. No entanto, o autor adverte que “a quantidade mínima de dados necessários para a formação de um *corpus* nunca foi estimada [...], sendo o critério de tamanho empregado subjetivamente na definição de *corpus*” (p. 345).

Para tratarmos algumas características relacionadas ao estilo, ao tema e à composição dos enunciados, recorreremos, como havíamos comentado, ao auxílio do *tropes*, *software* de processamento semântico. Segundo sua análise, o tipo textual¹⁶ preponderante na composição dos enunciados do *corpus* é o *argumentativo*, ainda que os demais tipos sejam encontrados com bastante frequência. Sobre o tema, foi-nos apontada a preponderância de debates sobre *questões sociais*. No entanto, também estiveram presentes temas ligados à *geografia e urbanismo, arte e cultura, política, ciências e tecnologia, economia e finanças, jogos e esportes, animais, alimentação, empresarial, saúde e doenças, direito e justiça, emprego e trabalho*, ou seja, temas que envolvem a vida cotidiana da sociedade e que, na esfera jornalística, são resgatados e transformados em notícias por serem de interesse do ouvinte. Por fim, sobre o estilo, o *software* aponta a preponderância de verbos *factivos e estativos*¹⁷, *que contribuem para a constituição de uma encenação dinâmica*. Por meio da categoria *modalização temporal*, a mais preponderante das modalizações no *corpus*, podemos verificar uma grande recorrência de advérbios e locuções adverbiais de tempo que envolvem, fundamentalmente, o âmbito da enunciação (*ahora, hoy, ya*), que expressam frequência (*siempre, una vez, de nuevo*) e fragmentação do dia (*tarde, mañana, anoche*), além de alguns advérbios de *posterioridade e anterioridade (mañana, ayer)*, entre outros.

Salientamos ainda a preponderância de formas pronominais de primeira e de segunda pessoas do singular – reflexo evidente da situação discursiva instaurada no gênero, na qual duas pessoas estão em diálogo direto –, e da tendência da variedade linguística usada neste gênero aproximar-se do uso menos monitorado e espontâneo da língua; isto é, do uso vernacular. Podemos citar, como exemplo, os seguintes fragmentos:

16 Diferentemente dos gêneros textuais, os *tipos* formam um grupo escasso com cinco, conhecidos como: *narração, argumentação, exposição, descrição e injunção*; sem tendência ao aumento (MARCUSCHI, 2008, p. 154). Eles são utilizados pelos gêneros no seu processo de composição, de modo que um único gênero pode se valer de mais de um deles.

17 Verbos que, segundo o *software*, expressam, respectivamente, ação e estado.

- (02) “Hola, Víctor Hugo, Tanto tiempo, ¡qué placer! me estaba riendo con los chistes de tu compañero.” <BsAs, Gr 02>.
- (03) “Para hablarnos de *Libertad*, como siempre, Rodrigo Carrizo, a quien saludamos. ¿Cómo Estás? Rodrigo, Buenas Tardes”. <MDZ, Gr 02>.
- (04) CAR: 2007, ahora sí, ahora sí, ahora sí. Acomodaditos
 MAR: Ya estamos todos acomodados.
 CAR: Impresionante este feriado.
 MAR: ¿Qué onda? <TUC, Gr 01>.

Nos enunciados apresentados, além de expressões que indicam uma situação de aproximação e espontaneidade (“*¡qué placer!*”, “*me estaba riendo*” e “*chistes*”), encontramos conjugação verbal (*estás*) e paradigma pronominal (tu compañero) referente à segunda pessoal do singular *vos*, cujo uso, na Argentina, também está relacionado a circunstâncias de menor formalidade e maior familiaridade. Destacam-se também *repetições com efeito cômico* (“*ahora sí, ahora sí, ahora sí*”), *uso de diminutivos* (“*Acomodaditos*”), *expressões de gosto pessoal* (“*¡qué placer!*”, “*Impresionante*”), *formas de questionamento e cumprimentos menos formais* (“*Hola*”, “*¿Cómo estás?*”, “*¿Qué onda?*”), entre tantos outros exemplos verificáveis na totalidade do *corpus*. Não obstante, eventualmente se pode encontrar marcas que apontam para um monitoramento linguístico. Assim, entre outras, são evidências de monitoramento:

- (05) **Repetições de palavras sem finalidade discursiva:** “[...] hemos progresado bastante en lo... en lo grupal y en en en lo técnico, en lo táctico. Así que estamos estamos bien (MAR, masc.; 33 anos, tesoureiro do clube 9 de julio, 3. COR)
- (06) **Pausas na fala:** [...] É... la cifra era monstruosa, ha bajado É... bueno é... [...]. (FAR; masc.; 51 anos; administrador, 4. ROS).
- (07) **“Vícios” de linguagem,** como uso excessivo de “*é*”, “*este*” e “*bueno*”: “É... bueno, una É... una sesión muy emotiva”; “[...] un romance que se mantiene y, bueno, é... [...]”; “Este... de estas personas que [avalaron] la desaparición de personas, la detención de hijos de desaparecido, la apertura, este... yo digo que este... nadie puede dormir con esta sensibilidad que nosotros nos la creamos. A mi parece importante la crítica seria este... É... A las gestiones que del gobierno. (TOR, masc.; 58 anos, senador, 4. POS)

Ou seja, não negamos a importante aproximação entre o uso vernacular e a variedade linguística usada no gênero *entrevista radiofônica*, no entanto, há de se levar em consideração que o falante é sensível à realidade linguística instaurada neste gênero e tende a se ater um pouco mais ao modo como fala. De maneira que não podemos considerar as conclusões provenientes da análise deste material como válidas, necessariamente, para um contexto casual de monitoramento quase nulo, como se espera, por exemplo, da conversa cotidiana entre amigos.

Em síntese, o estudo do gênero discursivo *entrevista radiofônica* permitiu-nos conhecer o material linguístico que servirá de base para conclusões sobre usos e valores atribuídos a diferentes fenômenos linguísticos nas sete regiões dialetais da Argentina. Consideramos essa abordagem teórica fundamental, pois nos fornece o entorno enunciativo destes “enunciados relativamente estáveis” (BAKHTIN, 1997). Por conseguinte, somos levados a conhecer fatores linguísticos e extralinguísticos que, de algum modo, possam estar relacionados ao uso dessas estruturas da língua. Finalmente, não podemos deixar de destacar o fato de o *corpus* ser constituído por enunciados da modalidade oral da língua – tanto em sua concepção, como em seu meio de propagação – e estar atento a variedades

hispano-americanas. Esse destaque se deve a que, além dos já conhecidos traços linguísticos típicos da oralidade, esse material apresenta-nos também a realidade linguística em uma significativa parte do cone sul americano, contribuindo, assim, para os estudos hispânicos no Brasil e no mundo à medida que registra e facilita o acesso à língua falada na respectiva área. Se resgatamos o cenário apresentado no início desse artigo, comprovaremos a relevância do trabalho, haja vista que são poucos os *corpus* orais preocupados com a América Hispânica, sendo a situação mais agravante quando nos atemos à qualquer localidade que não seja a capital dos tidos como principais países americanos.

REFERÊNCIA

BAKHTIN, M. *Estética da criação verbal*. São Paulo: Martins Fontes, 1997.

BERBER SARDINHA, A. P. Lingüística de *corpus*: histórico e problemática. *D.E.L.T.A.*, São Paulo, v. 16, p. 323-367, 2000. Disponível em: <http://www.scielo.br/pdf/delta/v16n2/a05v16n2.pdf>. Acesso em: 15 fev. 2011.

COSERIU, E. Geografia lingüística. In: _____. *El hombre y su lenguaje*. Madrid: Gredos, 1977. Cap. 4, p. 103-158.

FARNEDA, E. S. Perguntas e Respostas na Entrevista Radiofônica. *Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Lingüística e Literatura Letra Magna*, São Paulo, v. 6, p. 1-18, 2007. Disponível em: <<http://www.letramagna.com/entrevista-radio.pdf>>. Acesso em: 15 fev. 2011.

FONTANELLA DE WEINBERG, M. B. (Org.) *El Español de la Argentina y sus variedades regionales*. 2. ed. Bahía Blanca: Asociación Bernardino Rivadavia, 2004.

LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola, 2008.

LIPSKI, J.M. *El español de América*. Madrid: Cátedra, 1994.

MARCUSCHI, L. A. *Produção textual, análise de gêneros e compreensão*. São Paulo: Parábola, 2008.

MORENO DE ALBA, J. G. *El español en América*. Ciudad de México: FCE, 2000.

PÉREZ COTTEN, M.; TELLO, N. *La entrevista radial*. Buenos Aires: La Crujía, 2004.

PÉREZ HERNÁNDEZ, M. C. Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del Español*, Barcelona, v. 18, p. 1-75, 2002. Disponível em: <<http://elies.rediris.es/elies18/>>. Acesso em: 15 fev. 2011.

PORTUGAL, M.; YUDCHAK, H. *Hacer radio: guía integral*. Buenos Aires: Galerna, 2008.

SANCHEZ, A. et al (Org.). *CUMBRE* – Corpus Linguístico del Español Contemporáneo – Fundamentos, Metodología, y Aplicaciones. Madrid: SGEL, 1995.

SZYMANSKI, H. (Org.) *A entrevista na pesquisa em educação: a prática reflexiva*. Brasília: Plano, 2002.