

## **A constituição de um corpus de italiano falado para o estudo de pedidos e pedidos de desculpas: considerações sobre a validade dos dados**

Elisabetta SANTORO (USP)

O presente trabalho pretende discutir questões ligadas à constituição de um corpus de italiano falado, coletado a partir de gravações em áudio e vídeo, para um estudo que se insere no âmbito da pragmática linguística e procura investigar dois atos de fala específicos, a saber, pedidos e pedidos de desculpas.

Poderia ser útil começar analisando as próprias definições de corpus, inclusive colocando-as em relação com os objetivos de pesquisa e com o tipo de análise a ser realizada, mas aqui iremos nos concentrar em considerações relativas à validade interna e externa dos dados coletados.

De fato, em muitas pesquisas que se propõem a constituir corpora para a investigação da pragmática linguística, especialmente quando ligadas à mais ortodoxa análise da conversação, atenta-se principalmente para a "validade externa" dos dados a serem estudados, isto é, procura-se coletá-los de modo que haja a maior correspondência possível entre os fenômenos observados ao longo da investigação e os que acontecem, ou se presume aconteçam, na vida real. Em outras palavras, os dados considerados de maior relevância para o estudo da pragmática das línguas, principalmente em contextos cotidianos, são os ditos "dados naturalísticos", coletados de preferência sem que o informante tenha ciência, no momento em que os fornece, de participar de uma pesquisa e, possivelmente, no caso de gravações só em áudio, com os aparelhos escondidos, de modo que o informante nem mesmo saiba que sua fala está sendo gravada.

Não citaremos aqui as questões éticas e legais que procedimentos como esses envolvem (ver, entre outros, Bazzanella, 1994), mas é preciso lembrar que a própria definição de dado naturalístico não é isenta de problemas. É suficiente pensar nas observações sobre o "paradoxo do observador" de Labov (1970) ou nos questionamentos de Ochs (1979) sobre a impossível neutralidade do processo de transcrição, para concluir que a realidade linguística não poderá nunca ser colhida em toda a sua complexidade e que o pesquisador sempre irá intervir para recortar do material coletado as partes mais significativas para o seu projeto de pesquisa.

Mesmo assim, é necessário diferenciar entre as possíveis maneiras de eliciar os dados, sempre considerando também os objetivos de cada pesquisa. Para esse fim, há estudiosos que prepararam listas para propor hierarquizações das metodologias de coleta dos dados, colocando-as em uma ordem que vai do menor ao maior grau de controle sobre a produção dos dados, isto é, da maior "validade externa" à maior "validade interna" (Pallotti, 2001).

Diferente do que acontece nas pesquisas citadas, quando a perspectiva é a da pragmática intercultural ou interlinguística, é prática comum coletar os dados utilizando metodologias com um elevado grau de controle sobre as variáveis e, portanto, são frequentes os casos nos quais, para a coleta dos dados, se escolhem até DCT (Discourse Completion Tests) escritos nos quais os informantes escrevem o que diriam em determinadas situações (ver, por exemplo, Hudson/Detmer/Brown, 1995). Com essas metodologias as produções dos informantes são muito controladas e, além disso, os dados assim coletados requerem um baixo dispêndio de tempo e energias, pois não precisam de equipamentos de gravação em áudio e vídeo e podem ser gerados em grande número até em uma única sessão.

Para a constituição de um corpus de italiano falado que se propõe a analisar pedidos e pedidos de desculpas, optamos por uma metodologia de coleta dos dados que se coloca em uma

posição intermediária. Trata-se do roleplay que, em relação a outras opções, possui a vantagem de criar uma verdadeira interação oral entre dois interlocutores, mantendo, portanto, as características da língua falada, embora a interação seja induzida pelo pesquisador.

Coletamos os dados com a metodologia do role play aberto e com a recriação do contexto (setting) (Nickel, 2006), para que os informantes pudessem mais facilmente evocar as rotinas linguísticas utilizadas em situações do mesmo tipo. Com um fim parecido, procuramos também definir contextos e situações nos quais todos os informantes poderiam se encontrar na vida real, de modo a não levá-los a representar um "papel".

Essa metodologia permitiu que fossem controladas algumas variáveis independentes. De fato, se é verdade que as escolhas quanto a respeitar a identidade e a relação real entre os informantes limitou ou impossibilitou a seleção de situações com claras diferenças de poder relativo e distância social (para imaginá-las teria sido necessário pensar em contextos como o ambiente de trabalho, nos quais isso é mais evidente), é também verdade que o grau de imposição, variável que pode produzir notáveis diferenças nos atos de fala, pôde ser incluído.

Procuramos, portanto, organizar as situações dos roleplays em pares nos quais sempre havia uma situação com um baixo grau de imposição(-I) e outra, no mesmo contexto, com o alto grau de imposição +I).

As situações foram colocadas em três contextos e havia, para cada um deles, dois pedidos e dois pedidos de desculpas, chegando-se assim a 12 situações gravadas por cada uma das 30 duplas de informantes que participaram da pesquisa e realizaram interações orais a partir do mesmo input (dado a apenas um dos dois informantes de cada dupla). Cabe acrescentar que foi considerado na elaboração dos role plays que haveria diferentes graus de familiaridade entre os participantes e foi assim decidido dividi-los em duas categorias.

Para os pedidos realizamos também gravações em estabelecimentos públicos e comerciais, de três diferentes cidades italianas, nos quais pudemos contar com a participação das pessoas que habitualmente atendem o público. Para essas gravações foi dada aos informantes uma instrução verbal reduzida ao essencial.

Além de permitir o controle das variáveis e, portanto, uma validade interna elevada que possibilita um estudo sistemático das ocorrências, o corpus coletado é caracterizado pela replicabilidade. Pretendemos, de fato, constituir corpora com as mesmas características para o português brasileiro e com aprendizes brasileiros de italiano que poderão representar a base para estudos de pragmática intercultural e interlinguística.

## **Bibliografia**

Bazzanella, C. *Le facce del parlare*. Firenze: La Nuova Italia, 1994.

Hudson, T.; Detmer, E. & Brown, J. D. *Developing Prototypic Measures of Cross-Cultural Pragmatics*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i, 1995.

Labov, W. *The study of language in its social context*. *Studium Generale*, 23, p. 30-87, 1970.

Nickel, E. L. "Interlanguage Pragmatics and the Effects of Setting". In: Bardovi-Harlig, K.; Félix-Brasdefer & Omar, A. S. (orgs.). *Pragmatics & Language Learning*, vol 11, Honolulu: University of Hawai'i, p. 253-280, 2006.

Ochs, E. "Transcription as theory". In: Ochs, E./Schieffelin, B. (orgs.). *Developmental Pragmatics*. New York: Academic Press, 1979.

Pallotti, G. "L'ecologia del linguaggio: contestualizzazione dei dati e costruzione di teorie". In: Albano Leoni, F./Stenta Krosbakken, E./Sornicola, R./Stromboli, C. (orgs.). *Dati empirici e teorie linguistiche. Atti del XXXIII Congresso Internazionale di Studi della Società di Linguistica Italiana*, Roma: Bulzoni, p. 37-57, 2001.