

Mapping Paulistano Portuguese: SP2010 project

Ronald BELINE MENDES (USP)

Livia OUSHIRO (USP)

The Portuguese spoken in the city of São Paulo has not yet been extensively studied from a sociolinguistic perspective (Mendes, 2009; Rodrigues, 2009). This may be due to the difficulties of building a representative speech corpus of a heterogeneous and multicultural city with more than 11 million people. This fact raises questions as to which social parameters are most relevant for linguistic differentiation and stratification and how to reach speakers of varied social networks. In addition, it raises interest in developing methodologies for handling a large amount of spoken linguistic data.

Whereas Paulistano Portuguese has already been documented and analyzed through broad and significant research projects such as Projeto NURC-SP (Castilho & Preti 1986, 1987; Preti & Urbano 1988, 1990) and Projeto Para a História do Português Paulista (Castilho 2007), most works within these projects aim at analyzing "Brazilian Portuguese", either in contrast with European Portuguese (e.g., parametric variation), or in relation to its internal processes of change (e.g., grammaticalization). Among the very few works about Paulistano Portuguese in its social context, Rodrigues (1987) analyzed variable subject-verb agreement (e.g., "nós vamos" vs. "nós vai" 'we go') in the speech of 40 (semi-)illiterate speakers in two favelas, and Coelho (2006) analyzed variable use of 1PP pronouns ("nós" vs. "a gente" 'we') in the speech of 24 speakers living in a lower class community. Yet little is known about the linguistic production and perception of many other (supposedly) typical Paulistano variants (e.g., the realization of coda /r/ as a tap in words such as "porta" 'door', the diphthongization of nasal /e/ in words such as "fazenda" 'farm') and other variants in the city, as well as their social distribution and evaluation in the speech community at large.

This paper reports the objectives, methods, and results from the project SP-2010 (Mendes, 2011), currently under execution by the Grupo de Estudos e Pesquisa em Sociolinguística da USP, which has thus far collected and transcribed 89 sociolinguistic interviews. The main objectives are (i) to build a contemporary and representative speech sample of Paulistano Portuguese; (ii) to develop studies of sociolinguistic variation in the city; and (iii) to make the corpus of recordings and transcripts available online for a wider group of researchers. The social parameters for constituting this sample are sex/gender (men and women), three age groups (20-34 y.o.; 35-59 y.o.; 60+ y.o.), and two levels of education (up to high school; college). In a second stage, we will focus on social class, an understudied social factor in Brazilian sociolinguistic studies due to lack of criteria for categorizing speakers in different socioeconomic groups (Rodrigues, 2009; Mendes, 2011). Finally, the corpus will also be stratified according to speakers' generation in the city, in order to examine the contribution of different groups of migrants and immigrants in the community.

Speakers to be recorded have been contacted through the "friend of a friend" method (Milroy, 2004). The experience so far has revealed that speakers in the city are very resistant to talking to a "stranger" (the researcher), but when introduced by a common acquaintance, speakers tend to be much more receptive and solicitous, a fact that also has consequences for naturalness of speech. After a speaker has been recorded, the researcher asks them to suggest another speaker. Although this method does not yield a truly random sample of the community, we have been able to reach speakers of 59 different neighborhoods, living in the five zones of the city (Central, North, South, East, West).

Criteria for transcribing the recordings follow a simplified semiorthographic approach in order to make the material more easily available in a written medium. Transcripts do not contain any

special formatting such as boldface, italics, tab stops, columns, and are saved in plain text (.txt) with UTF-8 encoding. These criteria aim at facilitating the manipulation of text files in softwares such as R (Gries, 2009; Hornik, 2011) to automatically identify and extract tokens of a variable into a spreadsheet program.

Based on this corpus of 89 sociolinguistic interviews, the research group has been developing studies of sociolinguistic variation in Paulistano Portuguese (Mendes, 2009, 2010; Rocha, 2011; Oushiro, 2011; Mendes & Oushiro, 2011), which contribute to minimizing a gap in Brazilian sociolinguistic studies.

References

- Castilho, A. T. & D. Preti (eds.) (1986) A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. I – Elocuções Formais. São Paulo: T.A. Queiroz.
- Castilho, A. T. & D. Preti (eds.) (1987) A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. II – Diálogos entre dois informantes. São Paulo: T.A. Queiroz/FAPESP.
- Castilho, A. T. de (2007). Projeto para a História do Português Paulista. Projeto Temático de Equipe (Proc. FAPESP n. 06/55944-0).
- Coelho, R. F. (2006) É nós na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana. O pronome de primeira pessoa do plural e a marcação de plural no verbo. Dissertação de Mestrado. FFLCH-USP.
- Gries, S. Th. (2009) Quantitative Corpus Linguistics with R. New York: Routledge.
- Hornik, K. (2011) R FAQ. Available at <<http://cran.r-project.org/doc/FAQ/R-FAQ.html>>.
- Mendes, R. B. (2009) Who sounds /r/-ful? The pronunciation of coda /-r/ in the city of São Paulo. Paper presented at New Ways of Analyzing Variation - NWAV38. University of Ottawa.
- Mendes, R. B. (2010) Sounding Paulistano: variation and correlation in São Paulo. Paper presented at New Ways of Analyzing Variation - NWAV39. University of Texas at San Antonio.
- Mendes, R. B. (2011) SP-2010 – Construção de uma amostra da fala paulistana. Projeto de Pesquisa.
- Mendes, R. B. & L. Oushiro (2011) Production and perception of retroflex /r/ in São Paulo Portuguese. Paper presented at Variation and Language Processing – VaLP 2011. University of Chester.
- Milroy, Leslie (2004) Social networks. In: Chambers, J.K., P. Trudgill & N. Schilling-Estes (eds.) The Handbook of Language Variation and Change. Oxford: Blackwell.
- Oushiro, Livia (2011) Identidade na pluralidade: produção e percepção linguística na cidade de São Paulo. Projeto de doutorado.
- Preti, D. & H. Urbano (eds.) (1988) A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. III – Entrevistas. São Paulo: T.A. Queiroz/FAPESP.

Preti, D. & H. Urbano (eds.) (1990) A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. IV – Estudos. São Paulo: T.A. Queiroz/FAPESP.

Rocha, R.S. (2011) Negação pós-verbal no português paulistano: definição do envelope de variação. In: Seminário do GEL, 59, Programação, Bauru: GEL, 2011. Available at <<http://gel.org.br/detalheResumo.php?trabalho=7564>>.

Rodrigues, Angela C. S. (1987) A concordância verbal no português popular em São Paulo. Tese de doutorado. FFLCH-USP.

Rodrigues, Angela C. S. (2009) Fotografia sociolinguística do português do Brasil: o português popular em São Paulo. In: Castilho, Ataliba T. (ed.), História do Português Paulista. Campinas: Instituto de Estudos da Linguagem/UNICAMP.