

Building a *corpus* for comparative analysis of language attrition.

The aim of this presentation is to show the importance of *corpus* linguistic methodology in studying first language attrition. As first language attrition we mean "the non-pathological decrease in proficiency in a language that had previously been acquired by an individual, i.e. intragenerational loss" (Köpke and Schmid, 2004). This phenomenon can take the form of an increasing restructuring, convergence or loss of phonological, morpho-syntactical, lexical and pragmatic structures in the production skills of speakers of an L1 after a long contact with an L2, due to interference of L2 or to lack of input of L1 (Schmidt *et alii*, 2004). This study shows the results of a research focused on attrition Italian clitic pronominal system in contact with Brazilian Portuguese (BP). Very few works about attrition engage with data using *corpus* methodology: translation tests, judgement about grammaticality tests, sociolinguistic questionnaires, C-Tests with fill-in-the-blanks, Wug Test, guided narratives, Can-do scales and narrative of films are the preferred methods used in this field. The major advantage of using an oral *corpus* is the opportunity to obtain really spontaneous data. The most influential studies on language attrition are those by Seliger and Vago (1991) and Köpke and Schmid (2004), along with articles published in specialized journals. One of the most complete studies on this subject has been conducted by Keijzer (2007) on the L1 attrition of Dutch in anglophone Canada; nevertheless her research uses almost no data obtained by corpora or spontaneous data. On the other side, the investigation conducted by Hutz (2004), despite being a longitudinal corpus study, is of little interest to our intent, since it is neither oral nor spontaneous. In the specific field of Italian L1 in contact with BP, some studies by Raso and Vale (2007 and 2009) focus on spoken language corpora. Their researches looked into the linguistic attrition of high educated Italians in long-term contact with BP. Our study explores the results obtained by Raso and Vale and verifies them using a new corpus of 21298 words, assembled with more sophisticated criteria; this corpus was compared with similar texts extracted from C-ORAL-ROM (Cresti and Moneglia 2005). We examined the same clitics studied by Raso and Vale: *ci attualizzante*, *lessicalizzante* and *locativo*; *ne partitivo*, *argomentale* and *locativo* and the third-person accusative clitics *lo*, *la*, *li*, *le*, *l'*. The architecture of the new corpus follows the recommendations of Biber (1993), McEnery and Wilson (1996), Meyer (2004) among others.

Methodology of corpus compilation and of the research:

- selection of proper informants: to avoid the ambiguity between attrition and incomplete acquisition and to ensure their metalinguistic capacity, the informants should be Italian native speakers, born and raised in Italy and must have completed their studies there until obtaining at least a high school degree; they also must be graduated in or outside Italy. They have to be living in Brazil for at least eight years: this span of time is generally considered suitable for the identification of strong evidence of attrition;
- audio recording of the interactions with wireless equipments that

guarantees a high acoustic quality suitable for the prosodic analysis and allowing free movement of the speaker;

- choice of a wide array of interaction situations, reducing interviews, considered a not completely spontaneous situation;
- transcriptions of recordings in CHAT format and selection of the most appropriate ones for the study;
- manual examination of the corpus searching for the clitics studied (or their absence);
- analysis of the results obtained.

Our study mostly confirms the previous results, but it is exactly in the points of divergence between the two studies that lies the major interest of this research, leading to interesting methodological considerations. The corpus examined by Raso and Vale was mainly built with interviews about the same topic. The new corpus collates a much wider variety of genres, inducing the production of different structures in the two corpora. We exemplify here only two results. 1) The pronominal attrition has important consequences in the utterance structure. The anaphoric pronouns of left dislocated elements showed a 26,02% increase in the corpus studied by Raso and Vale, due to the fact the topicalization is a much more used strategy in BP, as shows the comparable texts extracted from the NURC corpus, but not in our research. We explained this behaviour with the pragmatic function of the left dislocation as a thematic function which identifies the semantic field of application of the illocutionary force. In monologues and situations with a weak link to the situational context, this kind of function is essential to communicate to the listener the semantic context the linguistic act refers to. On the other hand, in more interactive situations, where the situational context is essential, the semantic context of reference of the act is frequently given by the situation itself. 2) We also found discrepancies between the two Italian corpora used to compare the two Brazilian ones: the BADIP (De Mauro, 1993), used in the studies of Raso and Vale and characterized by much more planned interactions, and the Italian C-ORAL-ROM, much more spontaneous. The *ci locativo* presents an interesting divergence between the results of the two studies. The normalized frequency (over 10000 words) of this clitic in the BADIP corpus is 21,27, while in the C-ORAL-ROM is 8,95. The *ci attualizzante* also presents intriguing results: the normalized frequency in the C-ORAL-ROM is 129,09 when this clitic is joined with the verbs *essere* or *avere*, while in the BADIP corpus this frequency decreases drastically to 59,18. A possible explanation is that the use of the *ci attualizzante* indicates an informal and spontaneous kind of speech. In particular the *ci* in conjunction with the verb *avere* (*averci*) was barely used and avoided, even in the oral speech, until a few decades ago, for being considered substandard.

The results show how important corpus linguistic can be in studying attrition (and speech as a whole), but also the challenging accuracy required by the architecture of corpus compilation and by the methodology used in corpora

comparison.

Bibliography

- BIBER, D.. Representativeness in Corpus Design. In: *Literary and Linguistic Computing*, Vol. 8, n. 4, Oxford University Press, 1993.
- CRESTI, E.; MONEGLIA, M. (a cura di) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Language*. Amsterdam-Philadelphia: John Benjamin, 2005
- DE MAURO, T. *et alii* . *Lessico di frequenza dell'italiano parlato*. Milano: EtasLibri, 1993
- HUTZ, M. Is there a natural process of decay? A longitudinal study of language attrition. In: KOPKE, B.; SCHMID, M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin, 2004, pp. 189-206
- KEIJZER, M. *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. Vrije Universiteit, 2007
- KÖPKE, B.; SCHMID, M.S. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004
- MACWHINNEY, B. *The CHILDES project: tools for analysing talk*. Hillsdale: Lawrence Erlbaum, 1994
- _____ *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription format and programs. Volume 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000
- McENERY, T.; WILSON, A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996
- MEYER C.F. *English Corpus Linguistics. An introduction*. Cambridge: Cambridge University Press, 2004
- RASO, T. Erosione dei clitici e strutture tematizzanti in italiani colti in contatto prolungato col portoghese brasiliano. In: *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana, 2009. p. 384-399.
- RASO, T. ; VALE, H. P. A erosão lingüística em italianos cultos em contato prolongado com o português do Brasil: os clíticos e alguns efeitos na estrutura do enunciado. In: *Revista de Italianística*, v. 16, 2009, p. 1-22
- SCHMID M.S. "A new blueprint for language attrition research". In KÖPKE B. e SCHMID M.S., *First Language Attrition. Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2004
- SELIGER H.W. & VAGO R.M. (Eds.), *First Language Attrition* (pp. 175-188). Cambridge: CUP, 1991
- UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. Faculdade de Letras. *Projeto norma lingüística urbana culta - RJ: banco de dados*. Disponível em: <<http://www.lettras.ufrj.br/nurc-rj/>>. Acesso em: 21 mar. 2007.

VALE, H. P. . A erosão lingüística dos italianos cultos em contato prolongado com o português do Brasil: os clíticos. Belo Horizonte: Monografia apresentada no Curso de Graduação em Letras da Faculdade de Letras da Universidade Federal de Minas, 2007