Where do accents come from? Factors affecting the degree of foreign-accented Italian.

Claudia Roberta Combei[1,2], Giovanna Marotta[2]
[1] University of Bologna, [2] University of Pisa

Foreign accents reflect our identities and origins, but at the same time they convey information about our sociolinguistic background. In addition, the degree of foreign accent may reflect how open we are to embrace the language and the culture of the foster country, or it could simply be a predictor of the quality and the quantity of the input we have had for the specific language we are learning (Flege, 2009). In recent years, non-native speech has been a topic of continuous research interest in theoretical literature, applied linguistics, and speech technology. Many studies in second-language acquisition (SLA) have covered production and perception of non-native speech, often targeting English (e.g. Piske, MacKay, & Flege, 2001; Moyer, 2013; etc.). Nonetheless, in the last decades, various studies have also addressed foreign-accented Italian (e.g. Marotta, 2008; Marotta & Boula de Mareuil, 2010; Bianchi & Calamai, 2012; Pellegrino, Caruso, & De Meo, 2015; etc.).

Differently from previous studies on spoken L2 Italian, we investigate whether speaker- and context-related factors affect the degree of the perceived foreign accent. Such an approach may be advantageous in explaining and understanding the production and perception of non-native Italian speech from a sociolinguistic perspective. At the same time, the results of this study may be beneficial in the field of second language acquisition, especially in teaching Italian pronunciation to foreigners.

In this respect, and in order for the stimuli to meet balancedness and representation criteria, when selecting the speakers from the *CorAIt* corpus (Combei, 2017), we controlled for the following variables: the type of speech sample (i.e. read and spontaneous), the speaker's mother-tongue (L1), the occupation (all speakers were enrolled as regular or exchange students in degree, Ph.D., or specialization programmes in Bologna), the age (ages range from 20 to 30 years old), the gender, the self-assessed proficiency level in Italian (ranging from B1 to C2 levels, based on *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment-CEFR*[1]), the length of stay in Italy (i.e. less than 12 months, less than 24 months, more than 24 months), the age of onset (i.e. the first exposure to the Italian language: infancy, adolescence, adulthood), the predominant language learning method (i.e. purely naturalistic and guided-naturalistic), and the presence or absence of specific pronunciation training. The accents considered here are French, Romanian, Spanish, English, German, and Russian, compared to native varieties of Italian. Ten-to-twelve-seconds-samples of read and spontaneous speech (*cf.* Combei, 2017, for a complete description of the speech samples), produced by a gender-balanced set of 42 speakers – 6 for each L1 –, were included in a perceptual experiment delivered to 288 native speakers of Italian; the stimuli were previously validated by five expert listeners who assessed the degree of foreign accent (inter-rater reliability: ICC (A, 1) = 0.76[2]; p-value 0; 95% – confidence interval), they were automatically randomized for each experimental session, and they could be played twice.

Keeping in mind that also a wide range of factors regarding listeners could affect the outcomes of this study, we built a computer-based experiment that was conducted over the Internet on a *WordPress*[3] platform. This allowed us to reach a large and varied pool of listeners. The experiment was password-protected, and the IP verification blocked multiple submissions. The first section of the experiment was compulsory and qualifiable for further rounds; thus, the questions were designed to collect background information on the listeners. We sampled for the region of origin, the gender, the L1 (only monolingual[4] Italian speakers were admitted), the age range (i.e. 18-30, 31-45, 46-65), the highest level of education achieved (i.e. high-school, B.A./B.Sc., M.A./M.Sc./M.B.A.), occupation (i.e. student, employee, unemployed), proficiency level in foreign languages (i.e. French, Romanian, Spanish, English, German, Russian), the degree of familiarity with the foreign accents investigated (i.e. French, Romanian, Spanish, English, German, Russian), and experience in the field of linguistics (i.e. having attended B.A. and/or M.A. courses in Linguistics). An adequate number of participants for each of the levels of the afore-mentioned variables was collected.

In the accent rating experiment, listeners were asked to assess the degree of foreign accent using a 6-gradients Likert scale (i.e. from 0 = "no accent/native speaker of Italian" to 5 = "very strong accent"). Before analysing the results, an inter-rater reliability test was performed on the listeners' judgements, yielding a positive outcome: ICC (A,1) = 0.68[5]; p-value: 0; 95% – confidence interval. After a treatment coding of the categorical variables, we verified the strength of the effect of listener-related factors on the score listeners assigned to speech samples, by attempting to fit a Linear regression model; its results may be summarised as follows: $F(26,261) = 1.07$, $p = 0.37$; $R^2 = 0.096$;

[1] A description of *CEFR* is available online at this webpage: https://europass.cedefop.europa.eu/it/resources/european-language-levels-cefr (accessed on the 12th of September 2018).
[2] According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.76 is considered 'excellent'.
[3] A description of this platform and its functions is available online at this webpage: https://it.wordpress.com (accessed on the 12th of September 2018).
[4] The term "monolingual" is used here to refer to speakers that are neither early nor late "perfect bilinguals"; however, these speakers may be proficient in other languages without being considered "bilingual" (Grosjean, 2010).
[5] According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.68 is considered 'good.

Adj. $R^2 = 0.006$, Residual standard error: 0.46 on 261 degrees of freedom. Given the model complexity, meaningful results were obtained only for the education variable, namely, the higher the listener's education level, the more severely they will judge the speakers' degree of accentedness (p-value: 0.03).

Following Piske, MacKay, & Flege (2001), we focus on assessing which factors contributed to the degree of foreign accent. For such purpose, we fitted Linear regression models, where we tested the effect of all the variables under control on the accentedness score (i.e. the dependent variable). The weakest factors in predicting the degree of foreign accent were gender and the age of onset and they were excluded from the final model. Generally, the samples of read speech were perceived as more accented than those of spontaneous speech (p-value: 0.04), presumably as a result of the reading errors that might have arisen due to conflicting orthographic norms between the speakers' L1 and Italian. At the same time, in spontaneous speech the speaker will tend to use familiar and less problematic constructions.

Next, all other grouping variables and levels being comparable, Romanian speakers outstripped the other participants, since, most of the times they were rated with "no accent" or "mild accent" scores. On the other hand, the samples produced by German speakers were perceived as the most accented. They were followed in this order by Spanish, French, English, and Russian. This is consistent with the SLA literature that has linked pronunciation deviations, and therefore, foreign accent, to blocking mechanisms ascribable to L1 (e.g. Flege, 1995). In fact, the speaker's L1 seems to predict the degree of foreign accent (p-value: 0.005). Furthermore, the speakers who received specific pronunciation training were perceived as less accented than those who did not (p-value: 0.001); this, could advocate for shifting the view towards the role of input in SLA (Flege, 2009). However, having learned Italian in a specific learning setting is not often associated to a better pronunciation score (p-value: 0.06). Additionally, the length of stay in Italy does not seem to have any role in predicting the accentedness (p-value: 0.18), probably because the difference between the levels is too small (i.e. 12 months) to display any significant effect. A similar rationale behind the rather feeble but still statistically significant effects of the proficiency level in Italian (p-value: 0.01), together with the fact that it may not be a fully reliable predictor, since speakers self-assessed their performance; nevertheless, the speakers that claimed they had a C2 level in Italian received indeed good ratings, suggesting that their pronunciation is close to that of a native speaker.

The outcomes of the final model – consisting in the following independent variables: the type of speech sample, the speaker's L1, the Italian learning method, the presence or absence of pronunciation training, the length of stay in Italy, and the proficiency level in Italian – may be summarised as follows: $F(13,22) = 11.34$, $p = 0.00$; $R^2 = 0.87$; Adj. $R^2 = 0.79$, Residual standard error: 0.59 on 22 degrees of freedom. According to our data, the degree of foreign-accented Italian seems to be determined especially by the following factors: the speaker's L1, the presence or absence of specific pronunciation training during the learning path, and the type of oral production. However, in order to confirm the afore-mentioned associations at a larger scale, in future research, other accents should be investigated as well.

References:

Bianchi, F., and Calamai, S. (2012). Voci italiane e straniere a confronto. Indagine sugli stereotipi associati agli accenti stranieri. In *Cambio*, Anno 2, No. 4, December 2012, pp. 149-164.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. In *Psychological Assessment*, 6 (4), pp. 284–290.

Combei, C.R. (2017). CorAIt – A non-native speech database for Italian. In *R. Basili, M. Nissim & G. Satta (Eds.), Proceedings of the Fourth Italian Conference on Computational Linguistics, 11-13 December 2017, Rome*, 113-118, Torino: Accademia University Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in Cross-Language Research*. Baltimore: York Press, pp. 233-277.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, Massachusetts; London, England: Harvard University Press.

Flege, J. (2009). Give Input a Chance! In Piske & Young- Scholten (editors), *Input Matters in SLA*. Bristol: Multilingual Matters, pp. 175-190.

Marotta, G. (2008). Sulla percezione dell'accento straniero. In Lazzeroni et al. (editors), *Diachronica et synchronica. Studi in onore di Anna Giacalone Ramat*, pp. 327-347. Pisa: ETS.

Marotta, G. and Boula de Mareuil, P. (2010). Persistenza dell'accento straniero. Uno studio percettivo dell'italiano L2. In Schmid et al. (editors), *La dimensione temporale del parlato, Atti del V Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Zurigo (Svizzera), 4-6 febbraio 2009, pp. 475-494. Torriana: EDK Editore.

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge, UK: Cambridge University Press.

Pellegrino, E., Caruso., and De Meo. (2015). Discriminating between pathological and foreign- accented speech: an acoustic-perceptual study on Italian. In Vayra et al. (editors), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*, pp. 271-287. Milano: Officinaventuno.

Piske, T., MacKay I.R.A., and Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29(2), pp.191-215.