Towards an experimental historical phonetics: Italian speech is changing

Massimo Pettorino

Università degli Studi di Napoli L'Orientale
Italy

1. INTRODUCTION

After many centuries of "Verba volant" age, at present we are living in the "Verba manent" age. The starting day of the new era was the 6th December 1877, when Thomas Alva Edison decided to give a demonstration of the prototype of a talking machine to his assistants and to the carpenter who had made the wooden model of phonograph. The carpenter was quite skeptical about the machine and decided to observe with great attention the inventor's manipulations. Edison let the brass cylinder turn around his axle and, speaking in the direction of an elastic membrane, he recited a nursery rhyme:

"*Mary had a little lamb*,
its fleece was white as snow,
and everywhere that Mary went,
the lamb was sure to go".

The vibrations of the membrane let a stylus oscillate on the tin-foil covering the cylinder. The result was a series of more or less deep grooves on the foil. At that point Edison put the cylinder back at the initial position and let the stylus run in the groove. As the roller started turning, the astonished audience could hear Edison's voice repeating "*Mary had a little lamb…*". The "Verba manent" age had just begun.

Until then, the only possibility to describe the phonetic characteristics of a language in former times was to read the descriptions of that language in literature. But can a spoken language be depicted in a written form in all its components? We know that spoken communication is much richer than the written text. The message that goes from the speaker to the listener is made of many ingredients and among them the words are only a little part: the same sentence, uttered in a different way, can assume a totally different meaning. It is question of rhythm, prosody and intonation. The variations of the speech rate, the fluency, the insertion of silent pauses of different duration, and of filled pauses in the spontaneous speech, all these elements occur to define the meaning of what the speaker wants to send to the listener. It is to these acoustic correlates that the transmission of emotions can be assured in the speech. And there must be a sort of shared system of rules that allows the listener to understand the message correctly. Now the question is: is it possible to know how the words, the sentences, were really uttered in the past? Is it possible to describe, by a written text, a tonal peak? A hesitation? A vowel lengthening? Now, the point is that all these ingredients of the speech are not something "optional", that can be added to the speech to give more expressivity to the voice. On the contrary, rhythm, prosody and

tonal contour are basic components of the linguistic message: without them, we could not communicate each other at all [1].
This is to say that we know very little of the spoken languages of the past. But, as we have already said, this situation has changed at the end of the XIX century: the '900 has been the first century in which it has been possible to capture the voice, to store it, to listen to it again every time we want. Starting from the mid-century, a vast database of recorded speech is available (radio and TV programs, movies, songs etc). Now, at the beginning of the third millennium, for the first time in the history we have the possibility of comparing voices produced in different ages. So, we think that it is time to give rise to a sort of Experimental Historical Phonetics. The aim of this branch of studies is to know how languages have changed in the last 50-60 years. The method to follow is quite simple: once we have chosen the two corpora, we can analyze both of them following the same criteria and procedures. The result will give us an idea of the changes occurred in the interval of time that separates the two corpora. Of course, things are never simple: it is not sufficient to take any recorded speech of '50s and to compare it with nowadays speech. There are some rules that have to be followed. Some of them are related to the corpus, some to the analysis procedure.

1.1. The corpora

We know that speech varies according to many factors. We speak in a different way according to the listener, to the contextual situation, to the relationship between speaker and listener. In a formal situation our speech will be more hyperarticulated: the intonative contour will be quite flat, the speech rate will be quite low. We will use frequent silences, that will give us the possibility to think to what we are going to say. In an informal situation, our speech will be more spontaneous, full of filled pauses, the speech rate will accelerate and decelerate according to what we are saying, the tonal contour will be more varied. Because of this, it is not convenient to compare the two corpora produced in a very different situation. The same happens if we compare a spontaneous speech to a read speech: also in this case the differences that we can find would be due to the different type of speech more than to the years passed between the two corpora. So, it is important to choose very carefully the speech to analyze: the more the variables are equal, the more the variable "time" will be isolated and detected.
Another factor that influences the way of speaking is the text. If a TV speaker must announce a tragedy, maybe an earthquake occurred somewhere

in the world, he will use a low articulation rate, a low tonal range and silent pauses. On the contrary, if he has to say that the national team has won the world cup, the tonal range becomes wider, with many tonal peaks, the speech rate will be more varied and some filled pauses will appear. So, also in a same type of corpus (for instance read speech, formal style as in the TV news), it is better to compare news of the same topic (politics, economy, sport, chronicle etc).

1.2. The analysis

To compare two corpora on the suprasegmental level, we propose to calculate some prosodic coefficients [5] [6] [11]. They are:

- Articulation Rate (AR). The AR is given by the ratio between the number of syllables and the time employed to produce them (syll/s). The AR does not consider the time of the pauses and its variations measure the accuracy of the articulatory gesture. When the AR increases, the average syllabic duration is reduced, with a consequent reduction of the articulatory movements [2].
- Speech Rate (SR). The SR is given by the ratio between the number of syllables and the total time of the utterance. The SR mostly varies according to the time of the pauses: the more frequent and the longer the silences are, the more the SR decreases.
- The fluency (F) corresponds to the number of syllables produced between two consecutive silences.
- The percentage of silence (SP). The SP represents the percentage of silence compared to the whole utterance.

2. AN EXPERIMENT

2.1. Some data

Some experiments of experimental historical phonetics have been made for Italian [4] [8] [9]. The authors have compared two corpora of read speech, taken from the radio news of ’50s and ’90s. In another work we compared two corpora taken from TV news of ‘60s and 2002 [7]. The results of all these analyses show that in the last decades the Italian language has

remarkably changed the rhythm and of the tonal contour. As we can see (Fig. 1), in the modern Italian the speech chains are longer, the silences are less frequent and shorter in duration, the tonal range is wider with a higher number of tonal peaks. The SR is accelerated of about one syllable per second (6 syll/s). On the contrary the AR has not changed, so that the articulatory quality of the speech is not altered. The acceleration is obtained only by the reduction of the silent pauses. The percentage of silence lowers from the 27% to the 7% for the radio news and from the 18% to the 10% for the TV news. Also the duration of silences has changed: once there were long and short pauses (for strong and weak sentence boundaries), today all silences have more or less the same duration, similar to the short pause in the past years. So we can say that there is a univocal tendency: Italian is going towards a more efficient way of speaking. The speaker exploits better the time at his disposal. This result is obtained through a new strategy in using the pauses [3]. Once they were mostly syntactical, nowadays, silences are very short and they occur anywhere in the text, being only functional to the respiratory mechanics.

2.2. A problem

As we have already said, the text modifies the rhythm and intonation. Now, when we compare radio or TV news of different periods, it is obvious that the text cannot be the same. So, perhaps it is possible that the trend found for Italian is only due to the fact that the text of the news was totally different. To eliminate this kind ob objection, the text of the old TV news should be the same of today news. Is this possible? The authors of this series of works have made it possible [10]. The procedure is the following: they have scripted the TV news of the '60s. Then they have asked the collaboration of the RAI (the Italian network) to realize an "artificial" TV News: they went to the RAI studios and two speakers, one male and one female, read the "old" news as they were the true news of the day. The result was a TV News in which the text of the news was the same of that uttered 40 years before. For the first time, it was possible to compare the same text read after 40 years. Moreover, the two texts were both read speech, uttered in a very similar contextual situation by professional speakers.

2.3. Further data

The analyses made instrumentally confirmed the tendency already found in the previous works. The percentage of silence decreases from 13% to 6%. This lowering is due to the co-occurrence of two different factors: reduction of the numbers of silent pauses and shortening of their duration. In the 60s, pauses generally marked the sentence boundaries: in other words we can say that silences corresponded to what, in the written language, are commas and full stops. In today's news, on the contrary, silences are often avoided, especially those corresponding to commas (70% less). When a syntactical pause disappears, the boundary is marked by a tonal peak on the following syllable.
The shortening of the duration is particularly evident in the silences corresponding to the full stops: if in the 60s there were two different durations (about 0.9 s and 0.4 s for the full stop and the comma respectively) today the full stop disappears, as all silent pauses have the same duration (about 0.3 s). So, it is a very productive kind of speech, in the sense that the time is better exploited. Also the Articulation Rate is quite high (6.5 syllables per second), even if it does not reach its highest values, as it usually occurs in the last sentence of advertisements of drugs (8 syllables per second): in this case, the intention of the speaker is completely different. In fact, as there are some possible collateral effects of the drugs that by law have to be said, the speaker voluntary accelerates its speech to the highest speed because the goal is "the less the listener understands, the better it is".

2.4. Conclusions

Thanks to a device it has been possible to compare a same text uttered in the same contextual situation at about 40 years of distance. The aim of the experiment was to understand what had changed in the spoken Italian of TV news on the prosodic and rhythmical levels. The data allow to say that the today's Italian is uttered very carefully but at a fast rate, often absolutely accelerated. The acceleration is obtained thanks to the reduction of silences inside the text. In fact silent pauses are quite rare and short. Pitch is generally high and the intonational contour is rather varied, with frequent tonal peaks. A kind of speech that only a few years ago would have appeared as frantic and convulsive, is today the speech that we hear every day on television. The question is: is this trend true only for the Italian language or these considerations can be applied also to other languages and other

countries? If it is so, are we going towards a sort a globalized speech that, on the rhythmic and prosodic level, has already found a shared model? Do languages that differ for the lexicon, the grammar, the syntax, tend to be pronounced at the same speed, with the same pauses and the same intonation? In order to try to answer to these questions, we think that it would be very interesting to do the same kind of research on other languages.

3. REFERENCES

[1] Federico A.., Mori L., Paoloni A., 2004. La velocità di articolazione come parametro identificativo: potenzialità e limiti per la caratterizzazione del parlante. *Atti AISV*, Padova.

[2] Giannini A. 2000. Range di variabilità della velocità di articolazione in italiano. *Atti del XXVIII Convegno Nazionale AIA*, Trani, 253-256.

[3] Giannini A. 2006. I silenzi del telegiornale. *Atti di Comunicazione Parlata*. Napoli: Liguori (in press).

[4] Giannini A., Pettorino M. 1998. I cambiamenti dell'italiano radiofonico negli ultimi 50 anni: aspetti ritmico-prosodici e segmentali. *Atti delle VII Giornate di Studio del GFS*, Venezia, 65-81.

[5] Grosjean F., Deschamps A. 1975. Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, 31, 144-184.

[6] Miller J. L., Grosjean F. & Lomanto C., 1984. Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica* 41, 215-225.

[7] Pettorino M. 2002. Il parlato radiofonico: linee di tendenza dell'italiano. In Amedeo De Dominicis (ed), *La voce come bene culturale*, Roma: Carocci, 141-157.

[8] Pettorino M., 2004. La velocità di articolazione. In: A. De Dominicis, A. Mori, L. Stefani, M. (eds), *Costituzione, gestione e restauro di corpora vocali*, Roma: Esagrafica, 227-232.

[9] Pettorino M., Giannini A. 1993. Aspetti prosodici del parlato radiofonico. In *Atti delle IV Giornate di Studio del GFS*, Torino, 19-28.

[10] Pettorino M., Giannini A., 2005. Italian TV Speech: a diachronic analysis. In: *Proceedings of IX International Symposium on Social Communication***,** Santiago de Cuba, 849-853.

[11] Zmarich C., Magno Caldognetto E., Ferrero F. 1996. Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza. In: *Atti delle 7e giornate del Gruppo di Fonetica Sperimentale*, 111-139.

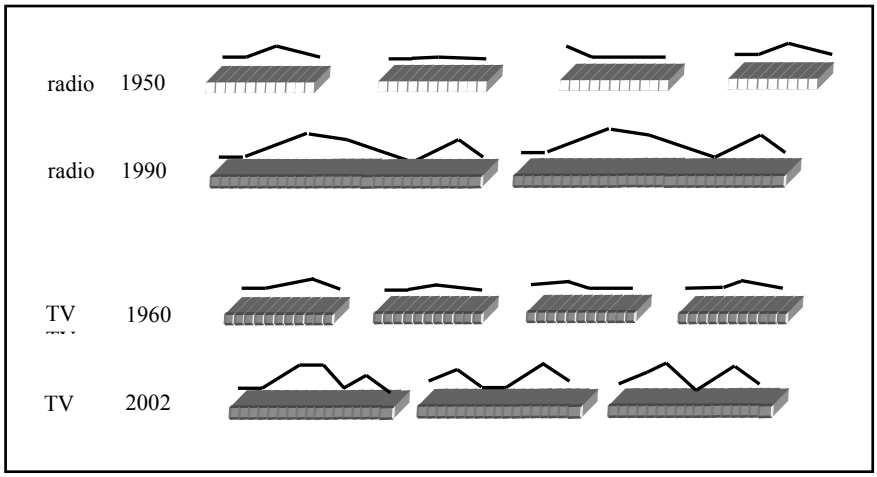

Figure 1: Schematic representation of radio and TV Italian speech