

CHAPTER III: METHODOLOGY

3.1 General Overview

The previous chapter focused on relevant literature for the present study. In this chapter the emphasis will be given to the participants, procedures and materials that I used in the adaptation of Derwing, Munro and Wiebe's (1998) study.

The first section of this chapter (3.2) focuses on the methodology carried out by Derwing, et al. (1998). In section 3.3 I will describe the characteristics of the participants of the current study such as speakers, listeners and the instructor. I will also describe the procedure of data collection (3.4) in terms of recordings (3.4.1.1), selection of speech samples (3.4.1.2) and the ratings of such audio stimulus (3.4.1.3). I will also present the materials used for the pronunciation instruction delivered to the students of the experimental group (3.5). Finally the procedure under which the data were analyzed will be briefly explained (3.6).

3.2 Derwing, Munro and Wiebe's (1998) Methodology

In terms of the speakers, Derwing, et al. (1998) collected speech samples of 48 adult students, at an intermediate proficiency level in a full-time ESL program. Their ages ranged from 18 to 44 years with a mean age of 31.7 years. The researchers collected speech samples from the ESL participants near the beginning of their course (time 1) and again 11 weeks later (time 2).

Three instructors were in charge of the pronunciation instruction. The teachers self-selected the approach they preferred to teach: global or

suprasegmental and segmental. The control teacher, the one who did not have to include the pronunciation component in the class had not studied linguistics and did not feel comfortable teaching either had.

Forty-eight Canadian NESs played the role of listeners and they were in charge of rating the speakers' speech in terms of intelligibility, comprehensibility and foreign accent.

The conditions under which the instruction of pronunciation took place will now be presented. Three conditions were settled for each group where one group had to receive a segmental approach, another had to receive a suprasegmental approach and the third one did not receive any kind of pronunciation instruction, which was called the control group. Since the participants were enrolled in a full-time ESL program, they attended ESL classes for 20 hours per week. The only difference between these three groups was the pronunciation component, the segmental and global group received approximately 20 minutes per day of explicit pronunciation instruction.

One of the limitations regarding this issue is the fact that this study was not comprised of randomly selected students, they were intact groups that were taken as they were and given different types of pronunciation instruction (segmental, suprasegmental). In order to lessen this limitation Derwing et al. (1998) carried out a pre-test for the participants of each group to ensure that all three groups were of a similar proficiency level regarding pronunciation before the study began.

The to-be-rated speech samples consisted of two tasks: a recording of simple statements and an extemporaneous narrative description of a standard

picture story. The sentences used for the controlled reading consisted of a single clause, the speakers were given time to read the sentences silently and then they were recorded onto tape. Immediately afterward, the speakers recorded the extemporaneous speech. Additional recordings of four Canadian NESs were also made, which served as the control recordings.

The ratings took place in several group listening sessions held over a period of two weeks. The listeners heard the stimulus, which consisted of single-clause sentences and rated for comprehensibility and accentedness on a 9-point scale. After hearing each stimulus once, there was a pause that gave time to the listener to decide how difficult the utterance was to understand rating from '1' (very easy to understand) to '9' (impossible to understand). In the second part of the same study and during a different listening session, the same listeners evaluated how accented an utterance from the extemporaneous speech was by rating it from '1' (no accent) to '9' (very strong accent).

For the extemporaneous speech, the listeners assigned comprehensibility and accent ratings in the same manner as in the section above.

Derwing et al. (1998) found that comprehensibility, and accentedness could be improved depending on the type of pronunciation instruction (segmental or suprasegmental) given to the learner. Despite the fact that each group of learners showed improvement in comprehensibility and accentedness, only the suprasegmental group improved in both aspects.

The results showed that although the three groups improved, the segmental group had improved after pronunciation training in the controlled

reading task, whereas the suprasegmental group was significantly better than the other two in the production of extemporaneous speech. This can be due to the fact that all the participants were exposed to a great amount of input just by being in an English speaking country, which makes us wonder whether the improvement shown by participants was due to the specific pronunciation instruction received, or if this improvement was the result of the amount of linguistic input to which they were exposed. One of the reasons I decided to replicate this study was to observe how the amount of input would affect in the improvement of intelligibility and comprehensibility in a situation where students have a limited exposure to the target language.

The following section presents the setup of the present study, its participants, data collection, and data analysis.

3.3 Participants

In this section, the most relevant characteristics concerning the speakers, the listeners and the instructor in charge of the pronunciation instruction will be presented.

3.3.1 Speakers

34 Native Spanish Speakers (NSSs) participated in this study. They were students at a private university in Central Mexico enrolled in an EFL program. As part of their plan of studies, students have to take three mandatory English courses such as English ID101, English ID102 (both high-intermediate level) and ID201 (advanced level). Two groups participated in this study from the course ID102, which is an upper-intermediate level class, comprising a sample

of 34 students. Since these groups were already formed, I randomly chose one to be the group who was about to receive the pronunciation instruction (experimental group) and another one who would not (control group).

	Control Group	Experimental Group
Students participating in the study	18 students	16 students
Gender	7 male 11 female	8 male 8 female
Range of Ages	17 to 23 years-old	18 to 29 years old
Mean age and mode	$x = 20.2$ Mo = 19	$x = 21$ Mo = 19

Table 3 – Characteristics of the Speakers

Table 3 describes the characteristics of each group; it presents the number of students per group, their ages, mean age and mode. This information was gathered through a questionnaire that was applied at the beginning of the semester (Appendix C)

As this table shows, there was a total of 34 students; 18 in the control group and 16 in the experimental group. Their ages ranged from 17 to 29 years-old, and the mode was 19 years-old.

3.3.2 Listeners

The number of listeners who participated in my study was 8; all of them were Native English Speakers studying Spanish at the same private university as the speakers, during the spring 2009 semester. The students who participated in this research as listener-raters were the ones with the lowest levels of Spanish.

The listeners were contacted through their Spanish professors during the 1st week of classes of the Spring 2009 semester; the project was briefly explained to them and were asked to fill in a questionnaire (Appendix D). The purpose of this questionnaire was to gather information regarding the participants' backgrounds.

The characteristics of the listeners in terms of gender and age are shown in the following table.

Gender	4 male	4 female
Ages	20 to 21 years-old	19 to 22 years-old
	Mean age: 20.75	Mean age: 20.25
	Mode: 20	Mode: 20

Table 4- Characteristics of the Listeners-Raters

The 8 students who participated as listeners-raters were from the United States. Their ages ranged from 19 to 22 years old. The mean age was 20.25 years-old and the mode was 20 years old. According to their responses from the questionnaire, none of them had Spanish heritage, but the majority of them

had taken at least 3 courses of Spanish. Only two of these students said that this was the first time they were studying Spanish. According to the questionnaire, it was also found that they had very little contact with Spanish speakers in their country and this was their first time in a Spanish speaking country.

For the purpose of this study, the ideal listener-raters should be Spanish students who are at the beginner level of proficiency in Spanish. By having little contact with the language, there is a probability they would have had less contact with the language features of Spanish and therefore, be less acquainted with Spanish accents –especially with Spanish accents in English. This means they should have taken no more than two courses of Spanish (information elicited through the questionnaire). It was difficult to find such a population, especially when there were no students registered in the classes belonging to the beginners' levels.

In terms of speakers and listener-raters, it has to be acknowledged that one of the limitations of this study was to put together a sample of students with the ideal characteristics, equal to those participants from Derwing et al. (1998)'s study. People who have participated in this project are the ones who were available (speakers and their classes' schedules) and who wanted to be part of it (listener-raters). This is potentially limiting my study since the small number of listener-raters (for example) resulted in a smaller amount of data to assess the intelligibility and comprehensibility of the non-native speakers which, in turn, limits the conclusions that can be drawn from this study. However, I could not have in my study US American students who had lived in Mexico for more than

2 months; this would have gone against the characteristics described above necessary to carry out this research.

The next section deals with the characteristics of the experimental and control groups' instructor, who was also the researcher of this study.

3.3.3 Instructor

In the current study, I was the instructor and researcher. I am a Mexican female whose first language is Spanish. I have 4-years experience as a language teacher and I have studied English since I was 6 years old. I have also lived abroad, in the United States for a total time of 9 months. In my interaction with other NESs, I have been acknowledged in several occasions to have a native-like accent. My command of the language fits the profile required to teach a high-intermediate level course.

The availability of groups within the language department and the control that I needed to have in this study were the reasons why I was also in charge of the experimental and control groups. Although this can represent a limitation for my study, I considered it necessary. By being in charge of the two groups I could be confident about the fact that one of these groups was receiving explicit pronunciation training, and the other was not. This concern was raised when the pilot study of the present study took place in which another professor was in charge of the control group and I realized that I could not be 100% sure that no pronunciation instruction took place, perhaps because the professor could have taught it unconsciously.

3.4 Data Collection

The collection of data took place during the fall semester of 2008 for the speech samples uttered from the speakers, and also during the spring semester of 2009 when the ratings were carried out by the listener-raters. Each of the instruments used in the study were first piloted during the summer of 2008 by applying them to random students at the same university in order to validate them.

The questionnaires applied to listeners-raters were piloted with the NESs during the 2008 spring semester at the Language Department within the same university. The same case applies to the questionnaire applied to the Native Spanish speakers.

The material used for the elicitation of speeches from the participants and the Likert scales were adapted from different sources. These procedures and the creation of material will be explained in-depth in the following sections.

3.4.1 Procedures

In order to collect information regarding the speakers of this study, two questionnaires were applied; one for the speakers and a different one for the listeners-raters. The application of the questionnaires to the speakers took place at the beginning of the semester in their classroom. Students were told that they were invited to participate in a study carried out by the language department, which was about the acquisition of a foreign language, as extra-credit. The questionnaires applied to the listener-raters also took place in their classroom. An appointment was made with the Spanish instructor beforehand.

Speech samples from the participants were collected near the beginning of their English course (time 1) and again 12 weeks later (time 2). The participants recorded the reading of a paragraph and an extemporaneous speech elicited from the researcher, in which they had to talk about their family or themselves for at least one minute.

Once the recordings of both groups were completed, NESs evaluated the speech samples in terms of intelligibility, comprehensibility and accentedness.

3.4.1.1 Recordings

In order to collect samples of speech from the participants, two recordings were carried out. The first one was recorded at the beginning of the semester (time 1). The second recording was conducted at the end of the semester, 12 weeks later (time 2). The recordings were carried out in the *Grupo de Investigación en Lingüística Aplicada* (GILA), a place located in the first floor of the Language Department in the university. GILA was chosen due to its location within the department, which is an isolated and quiet space, which purpose is to carry out investigations from the language department. These characteristics provided recordings with minimal interference.

The recording sessions were performed individually and with the help of a tape recording machine and a microphone. They were then transferred to a CD, in order to be edited. Since the tape recording needed to be operated manually, someone had to be physically present when the recordings took place. Therefore, the researcher's assistant carried out the recordings of the experimental and control groups. She was a female colleague from the

language department. The purpose of the study was not mentioned to the speakers. They were only told that this was a language department research project to study the factors intervening in the acquisition of a second language. No questions were asked on behalf of the students after this explanation.

The tasks performed by the speakers were first, the reading of a paragraph (Appendix E, taken from Grant, L. (2007)), and immediately after, a mini monologue they talked about their family or themselves for at least one minute. However, the extemporaneous speech, which consisted of a mini monologue, was the only data taken into consideration for the ratings of intelligibility, comprehensibility and accentedness. The reason why I decided to disregard the data obtained from the controlled reading derives from the fact that I was interested in observing the speakers' intelligibility in running speech and not in a controlled production of the language.

Additionally, according to the research assistant, the students were a little nervous because the paragraph reading was their first task to complete. This behavior was noticeable in the participants' sitting positions, the way they played with the paper containing the paragraph and how their perceivable hands were shaking. The fact has been considered that, during the second task, the students would have been less nervous and this would not have affected their performance.

In preparation for the recording, the student first had to read the paragraph silently, along with some prompts (Appendix F) in the form of questions. The speaker could use the prompts in order to talk about his/her family or him/herself for the one-minute mini monologue, the second task. If the

student did not have any questions about unfamiliar words, the student was then asked to go inside the recording room to read the paragraph aloud onto the voice recorder. Later, s/he talked about his/her family, which was the extemporaneous speech section. This was a monologue of approximately one minute. The person recording (a colleague) usually did not speak during any of the tasks. However, if the participant could not talk about his/her family or him/herself for one complete minute, the researcher assistant was then able to ask the participant one or two questions to finish the task.

Additional recordings were made of 3 US English speakers and 1 Australian English speaker (2 male, 2 female). They served as a guideline for the individual listeners' use of rating scales in the listening tasks. It was expected that all raters would consistently assign very good scores to NESs. Failure to do so might indicate a misunderstanding of the instructions.

Once the recordings from the pre-test and post-test were completed, the recordings were then recorded onto computer. This process occurred in the Audio booth, located in the university's Humanities building, where they were also edited. The editing procedure involved choosing the audio samples to be presented to the listeners, which will be explained in-depth in the following section.

3.4.1.2. Selection of Speech samples

The selection of the speech samples to be presented to the listener-raters was a laborious and meticulous procedure. First, the recordings from each student were recorded on the computer. In order to facilitate the identification of each

file, they were given the students' real names, for example 'Veronica1' for the reading of the paragraph and 'Veronica2' for the extemporaneous speech, and so on. All files, which ended in '1,' were not used because they were the paragraph readings. The files ending in '2' were edited, as it is described below.

Speech samples from the mini-monologue had a mean duration of 75 seconds. Each file was divided into three equal parts, according to their length. In other words, if the file lasted 60 seconds, each part had a length of 20 seconds. The second section of each file was the one which was taken into account in order to select the final stimulus. One of the main reasons behind this decision was the fact that during the first 15 to 20 seconds the students had used all the prompts given to them in order to complete the minute. For example, it was noticed that if they had to speak about their family, the majority of them were talking about their siblings during the 10th to 15th second. As a consequence, during the second part of the speech, they were forced to talk about other things that were not part of the prompt questions. Therefore, the topics and situations they were talking about were a bit different. Having different issues being discussed by the speakers made the process of selecting the data stimuli a little easier, since not all of them were saying that they had x number of brothers and sisters along the time that was taken into consideration to select the final audio stimuli. This procedure was carried out for each speech sample, for the pre-test and the post-test.

After having determined the part of the speech sample that was to be heard in order to get the final stimulus, shorter excerpts were selected that were of sufficient utterance length duration, which will be explained next. These

selections were to be transcribed by listeners after a single listening. In order to fulfill this requirement, Munro and Derwing's (1999) selection of stimulus example was followed. They mention in their study that they selected utterances from 4 to 17 words with a mean length of 10.7 words (Munro and Derwing, 1999). In the end, the final stimulus for this study was a selection of 73 utterances, with a mean length of 8 words and a range from 4 to 11 words. There were 68 speech samples from both groups, including the pre and post-test and 5 speech samples from the NESs. Table 5 describes the number of audio files per group and per time.

Once the final stimulus was selected and the listener-raters were contacted, the rating sessions took place. The way in which they were carried out is described in the following section.

Group	Pre-test	Post-test	Total
Experimental	16	16	32
Control	18	18	36
NESs		5	5
Total number of Speech Samples			73

Table 5 - Total Number of speech Samples.

3.4.1.3 Ratings

Two group listening sessions were completed in order to rate the speech samples produced by the participants. They took place in a room at the *Centro de Aprendizaje de Lenguas*, aka CAL. It is located next to the language department building, and it has many computers along with other materials for language learners. The room in which the sessions took place was equipped with individual computers, headphones, an overhead projector and a screen.

The first session consisted of a short training session and the completion of the intelligibility and comprehensibility tasks. The training session consisted of a presentation on behalf of the researcher that included the defining of intelligibility and comprehensibility. For the intelligibility task, it included a couple of examples in which the task the listener-raters were about to complete was practiced. Listeners had to transcribe in standard orthography exactly what they heard from the extemporaneous speech and had to write each utterance word for word. They were presented with stimuli collected during the spring semester.

For the comprehensibility task, the raters had to define how comprehensible the speakers were by using a 4-scale rating system. Listeners were given a table which described the levels of comprehensibility they could give to each speech sample. The scale was adapted from several sources, such as: the Massachusetts speaking assessment criteria (Chicago Board of Education, 2000), the International English Language Testing System speaking band descriptor (IELTS), the Complete speaking test rubric (New York State

English as a Second Language Achievement Test, 2005), and the Explanation of level 2 Speaking rubric (Fairfax County Public Schools, 2000).

The scale used for this study consisted of 4 bands in order to rate the participants' speech: 1-easy to understand, 2-a bit difficult to understand, 3-very difficult to understand, and 4-impossible to understand (Appendix G). There were descriptors and benchmarks for each level. Since the audio files were located in each computer, NESs were able to listen to each benchmark through their headphones as many times as they needed in order to identify why the file sample was chosen to exemplify a certain descriptor. During the presentation, some of them had questions which were addressed, after that they started with the tasks. First they had to orthographically transcribe what they heard and immediately after rate the degree of comprehensibility of each file (Appendix H). They first completed 34 speech samples, had a break of 10 minutes and continued with the second half of the audio files.

On a second day, held 2 days later, the listener-raters were asked to rate the degree of foreign (non English) accent with a 4-point Likert scale (See Appendix I) that corresponded to the same audio files they listened to when they performed the intelligibility and comprehensibility ratings. The scale to rate foreign accent was also adapted from different sources (same as the comprehensibility scale mentioned above). This scale also consisted of 4 bands: 1-no foreign accent, 2-mild foreign accent, 3-strong foreign accent, and 4-very strong foreign accent. The listeners were also given benchmarks for each band and had the opportunity to do some practice before they started rating the data. Just as in the first session, listener-raters rated the first 34 speech samples, had a 10-minute break and continued with the last part.

Students who participated in this part of the study were not paid, but instead were offered some snacks and drinks from the researcher, which were gratefully accepted by the NESs.

3.5 Instructional Materials

The groups Native-Spanish-Speaking students participating in this study took the same upper-intermediate level course, English ID102. Both groups followed the same book which they bought in the previous semester when they covered units 1 to 6 in English ID101. During the English ID102 course, the students review units 7 to 12. The content of this course includes aspects of grammar, vocabulary, reading and listening comprehension based on the book. Speaking ability is implicitly included through an oral evaluation in the middle of the semester and through some exercises that the book presents throughout the content units.

For the explicit pronunciation instruction, the participants from the experimental group were presented with different kinds of materials: slides, copies of exercises, links from the internet, and books. As a reminder to the reader, the control group did not receive any kind of pronunciation training. The materials focused on the instruction of segmental as well as suprasegmental features. Sessions of 8 to 10 minutes were given at the end of class over a period of 12 weeks. Due to time constraints, the sessions were given once or twice a week. The agenda of the sessions on pronunciation, given to the experimental group, is presented in table 6. It includes the phonological components, which were taught during the semester:

Segmental Features	Suprasegmentals
Sounds of 'th' at the beginning, middle and end of words (voiced and voiceless sound of theta)	Counting syllables and syllable stress.
Consonant clusters starting with the phoneme /s/	Reduction and linking words.
Past tense of irregular verbs ending with 'ed' and its sounds /t/, /d/, /ld/	Thought groups and sentence stress
Difference between /s/ and /z/	Rhythm - stress words
Difference between /v/ and /b/	Old and given information
Nasals /m/, /n/, /ng/ in the middle and the end of words	Intonation
Tense vs. lax vowels	
Front, middle and back vowels	

Table 6- Agenda for Pronunciation Instruction

After three sessions, recycle sessions that included the aspects of phonology already studied in class were planned. However, this was not possible due to the syllabus of the course. In spite of this, the experimental group's instructor put emphasis on the target sounds, as needed, during the

course of the class. For example: in the case of suprasegmentals, if we were reviewing 'reported speech' I would draw their attention to the intonation of the two different types of questions (yes/no or wh- questions). The segmental and suprasegmental features of English were presented at random, that is, not all segmental features at once and then all the suprasegmental ones.

The methodology carried out followed the one presented by Derwing, et al.'s (1998) study with some changes. The changes resulting from the adaptation were derived from the availability of materials and human resources of an EFL setting, such as: having two groups (control and experimental group) instead of three (segmental group, global group and no instruction group), availability of NESs, and the access to the original materials used to elicit the control reading and the extemporaneous speech. Another mismatch between Derwing, et al., (1998)'s study and mine was the amount of pronunciation instruction the speakers received. While the participants from Derwing, et al. (1998) received 100-minutes of instruction per week, my students only received around 20 minutes. Once again, these differences were consequence from the setting in which each study took place: ESL and EFL.

3.6 Data Analysis

The treatment under which this data was analyzed will be described in the current section. Due to the fact that my first two research questions addressed the idea of improvement after explicit pronunciation training, the intelligibility and comprehensibility scores used to show such improvement were submitted to the same quantitative analysis.

3.6.1 Quantitative Analysis carried out for Intelligibility and Comprehensibility Scores

Since the data presented to the rater-listeners was at random, the first step in order to carry out the data analysis was the grouping of speakers according to the group they belonged to. That is, all the speakers from the control group were put together, as well as those who belonged to the experimental group. Within each group, the scores obtained were divided into two sub-categories: pre-test and post-test.

Mean comprehensibility scores were computed for each speaker during the pre-test and the post-test. Based on the type of data collected and the design of this study a statistical test was chosen as the best option. One of the objectives of a statistical test is “to test a hypothesis concerning the values of one or more population parameters” (Wakerly, Mendenhall, Scheaffer, 2002, p. 461). In other words the purpose of this test is to show evidence to support a hypothesis or to reject it. A two-sample statistical test (t-test) was carried out in order to state that the experimental and the control groups were on equal conditions in terms of intelligibility and comprehensibility before the study began.

As mentioned in the introduction of this chapter, two hypotheses concerning the improvement of intelligibility and comprehensibility were formulated in which I hypothesize that there will be an improvement after explicit pronunciation instruction. For this reason, each set of scores was submitted to a paired t-test. The purpose of carrying out this test was to find out if there was an

improvement in terms of intelligibility and comprehensibility. The data submitted for this test was the mean scores obtained during the pre-test (1st set of data) and the mean scores of obtained during the post-test (2nd set of data) for each group. As a result, four paired t-tests were carried out according to the dependent variables: intelligibility (control and experimental group) and comprehensibility (control and experimental).

Finally, a simple t-test was carried out in order to see which of the two groups had improved more. The data submitted for this test was the mean difference calculated from the subtraction of the mean scores obtained from the post-test minus the mean scores obtained from the pre-test of each group.

3.6.2 Quantitative Analysis for Foreign Accent

Although it was not the objective of my study to see a reduction in perceived foreign accent through pronunciation training, I decided to submit the scores obtained from the listener-raters to the same procedure described above. The reasons of performing such analysis were twofold. First, I wanted to compare my results to those obtained from Derwing et al. (1998), where they affirm that their participants showed an improvement on foreign accent after explicit pronunciation instruction. Second, I wanted the information to be displayed so that the reader could have the data that would support the section related to the correlation existing (or not) between accentedness and comprehensibility.

3.6.3 Correlation Between Foreign Accent and Comprehensibility

In order to find out if there was an existing correlation between the variables of comprehensibility and foreign accent, the Pearson r coefficient was computed. Since I was not interested in seeing if the control or experimental group were better than the other, I submitted all the scores obtained from each group during the pre and posttest. The results obtained would determine if there is a correlation between the aforementioned variables or not.

By carrying out such analysis I am addressing the third hypothesis of this study, which states that having a very strong foreign accent does not affect the comprehensibility of the message uttered.

In addition, not all the research design was based on a quantitative analysis. The qualitative analysis that took place addressed the study of intelligibility errors and their categorization.

3.6.4 Qualitative Analysis for Orthographic Transcriptions: Intelligibility

Each speech sample was orthographically transcribed by the researcher by listening to each one of the audio files as many times as needed. After the transcriptions were done, the number of words per utterance was counted. An intelligibility score was calculated for each of the 73 speech samples on the basis of the number of words that exactly matched the corresponding transcription. An intelligibility score was also computed for each of the utterances by taking the mean of the 8 listener-raters' scores for the utterance.

It was necessary to carry out a qualitative type of analysis was necessary to carry out in order to find the intelligibility problems encountered by the listener-raters. Even though this data analysis does not address any of the research questions of this study, I consider it important since it could yield interesting results regarding the errors that may cause intelligibility problems to the speakers. This analysis consisted of transcribing each utterance, as heard by the listener-raters, and pointing out the errors in transcriptions. A categorization of the type of errors made in the transcriptions was also carried out.

The results obtained from this data analysis are presented in Chapter 4, the interpretation of the latter will be discussed in Chapter 5.