Second Language Learner Speech and Intelligibility: Instruction and Environment in a University Setting

Sara Kennedy

Department of Integrated Studies in Education
McGill University, Montreal

September, 2008

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Education

Your file Votre référence
ISBN: 978-0-494-66708-8
Our file Notre référence
ISBN: 978-0-494-66708-8

ABSTRACT

The purpose of this study was to investigate changes in the pronunciation and intelligibility of instructed and uninstructed second language (L2) learners over time, and to identify instructional, environmental, and methodological factors playing a role in pronunciation and intelligibility.

Seventeen L2 graduate students at an English-medium university recorded three personal anecdotes over five months. The students also regularly logged their exposure to and use of English. Nine of the students (instructed group) were concurrently taking an oral communication course focussing on suprasegmental pronunciation. Classroom instruction was regularly observed and recorded. All 17 students were interviewed at the end of the study.

L1 listeners heard anecdotes from three instructed and three uninstructed students, matched for length of residence and first language (L1). Listeners also heard anecdotes from four L1 English speakers. One group of listeners retold each anecdote after hearing it (discourse-level task). The other group paused the recording of each anecdote whenever a word was unclear (word-level task). Each group of listeners also rated excerpts for accentedness, comprehensibility, and fluency.

Results of quantitative and qualitative analyses showed that: (a) no unambiguous changes in the pronunciation or intelligibility of either L2 learner group occurred over time; (b) word-level intelligibility measures more consistently differentiated L1 and L2 groups, and the instructed and uninstructed L2 groups; (c) compared to the instructed group, the uninstructed group logged relatively more English exposure/use for academic activities and relatively less for interactive social activities; (d) many instructed L2 learners did not believe that their pronunciation had noticeably improved, but almost all expressed satisfaction with their ability to communicate in English; (e) at the end of the study, many uninstructed learners reported persistent difficulties in communicating in English.

The results suggest that instruction in suprasegmental aspects of pronunciation sometimes may not lead to improved intelligibility or pronunciation. In addition, some L2 learners can be as intelligible as L1 speakers, depending on the listening task. Finally, results suggest that L2 learners' perceptions of their communicative ability and their patterns of L2 exposure/use are related. Implications for university preparation and support programs for L2 graduate students are discussed.

RÉSUMÉ

Cette étude enquête sur le développement de la prononciation et de l'intelligibilité des apprenants d'une langue seconde (L2) instruits et non-instruits, et sur l'identification des facteurs d'instruction, du milieu, et de méthodologie qui jouent un rôle dans la prononciation et l'intelligibilité.

Dix-sept étudiants du deuxième et troisième cycle qui parlaient anglais comme langue seconde ont enregistré trois anecdotes personnelles sur une période de cinq mois. Les étudiants ont aussi tenu un journal de leur emploi de l'anglais. Neuf étudiants suivaient aussi un cours axé sur les aspects prosodiques de la prononciation. Tous les 17 étudiants ont participé dans une entrevue à la fin de l'étude.

Des participants anglophones ont écouté les anecdotes de trois étudiants de chaque groupe (instruits et non). Les participants ont aussi écouté les anecdotes de quatre anglophones. Un groupe a écouté les anecdotes et les a racontées de nouveau (tâche au niveau de discours). L'autre groupe a arrêté le fichier sonore chaque fois qu'un mot n'a pas été compris (tâche au niveau des mots).

Les résultats ont démontré que : (a) il n'y avait aucun changement dans la prononciation ou l'intelligibilité des deux groupes d'étudiants ; (b) les mesures au niveau des mots étaient plus fiables pour différencier les anglophones des groupes d'étudiants, et les deux groupes d'étudiants ; (c) par rapport au groupe instruit, le groupe non-instruit a employé l'anglais plutôt en activités universitaires, et moins en activités sociales; (d) beaucoup des étudiants instruits n'ont pas cru que leur prononciation s'est améliorée, mais la plupart étaient satisfaits de leurs habiletés en communication ; (e) beaucoup des étudiants non-instruits ont souligné des difficultés tenaces en communication.

Les résultats suggèrent que l'instruction d'aspects prosodiques de prononciation n'entraîne pas nécessairement l'amélioration de l'intelligibilité ou de la prononciation. De plus, l'intelligibilité d'étudiants de L2 peut atteindre le niveau de celle d'anglophones, selon le type de tâche d'écoute. Finalement, les perceptions des étudiants de L2 de leurs d'habiletés de communication sont reliées aux types d'emploi de la L2. Les implications pour les programmes de préparation et soutien pour les étudiants de L2 des deuxième et troisième cycles sont examinés.

ACKNOWLEDGEMENTS

Doctoral dissertation research can be somewhat like a relay race/obstacle course, where a group of people guide, push, and sometimes outright carry the doctoral student over pitfalls, through swamps, and past dead-ends. This dissertation would not have been completed without those people, first among whom is Lise Winer, my supervisor. A doctoral student's vision of her research can sometimes lead to inflexible or impractical approaches. Lise was always patient and resourceful in identifying possible obstacles or wrong turns and in suggesting alternative means of accomplishing the same ends. She also provided a wealth of expertise in aspects of sociolinguistics and discourse, especially oral storytelling and its analysis. Finally, her sense of good writing and organization pushed me to make this dissertation clearer and more coherent.

I also thank my committee members: Roy Lyster, Mela Sarkar, and Glyne Piggott. Their questions and comments at all stages of the research have made this dissertation immeasurably better.

Laura Collins, Tracey Derwing, and Norman Segalowitz graciously provided advice and comments on various methodological issues. Media Services in the Faculty of Education were professional and personable in providing recording equipment, and Elizabeth Gatbonton very kindly lent some of her own research equipment in order to make data collection easier. My colleagues, classmates, friends and family have all encouraged, supported, and sustained me throughout my doctoral studies.

Randall Halter and Pavel Trofimovich have helped in almost every way imaginable, from discussing theory and methodology to providing equipment to advising

on statistics to troubleshooting formatting. They must be at least as relieved as I am that this dissertation is finished.

I particularly thank the course instructor, the native speaker talkers and listeners, and the graduate student participants for their openness and for their diligence and thoughtfulness in completing their research tasks. The graduate students, who had bravely leaped into the challenge of living and studying in another language and culture, were especially humbling to talk to, and I am glad to have met them. I have no doubt that their courage and their drive will ultimately lead them to success, and that their contributions will make our country better.

This dissertation is dedicated to Pavel, who was in on it from the beginning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Non-Native Graduate Students

Canadian universities have opened their doors to the world. From 1996 to 2006, the enrolment of full-time international graduate students, who had come to Canada expressly to study, doubled from 11,000 to 22,000; this latter figure makes up close to 20 percent of full-time graduate students in Canada (Association of Universities and Colleges of Canada, 2007b, p. 16). Many graduate students born outside of Canada not only study, but have settled here as well. Over 35,000 recent immigrants in the Canadian workforce in 2006 had earned a Master's or PhD degree in Canada (p. 34). More and more foreign-born (both visa and immigrant) graduate students are studying at universities in Canada, and their numbers are expected to grow in the future (p. 28).

This increase has not been driven just by the interest of potential graduate students. Government, universities, and industry are also pushing to attract and retain foreign-born students. In its 2007 report to Parliament on international education, the Association of Universities and Colleges in Canada advocated that the federal government intensify its efforts to recruit and support international graduate students.

> This growing pool of global talent has the potential to bring significant economic and societal benefits to host countries as a future source for highly qualified people that can supply the labour market/workforce and feed the research and innovation pipeline (Association of Universities and Colleges of Canada, 2007a, p. 2).

Foreign-born graduate students are valued for the knowledge and expertise that they can bring to Canadian universities and the Canadian workforce. However, they face many obstacles in completing their studies and possibly finding a job in Canada. One of the most frequent and important obstacles is difficulty in understanding and using the language of instruction and of the greater society.

Foreign-born graduate students in Canada who do not speak English or French as a first language (L1) have reported that their struggles with language make it more difficult for them to be successful academically (Cheng, Myles, & Curtis, 2004; Schutz & Richards, 2003). Indeed, for some non-native (NN) students, their self-reported level of language difficulties was significantly negatively correlated to their GPAs (Berman & Cheng, 2001).

Being Understood

NN graduate students must know about and use their second language (L2) in many different ways in order to successfully complete academic tasks. However, all NN graduate students who are successful in using an L2 meet at least one basic requirement: other people can understand what they are saying. People who interact with NN graduate students, whether faculty, staff, or undergraduate students, sometimes say that it is difficult to understand these graduate students' speech (Hinofitis & Bailey, 1981; Damron, 2001; Gallego, 1990). NN graduate students themselves also note this phenomenon, as researchers have found. "[Hua] told me that sometimes what she said, how she said it, and how listeners understood her had no relationship to each other" (Wang, 2003, pp. 57-58).

When people have trouble understanding the speech of NN graduate students, those students may not be able to fully inform faculty or staff about questions they have or problems they are experiencing in their academic work or field of study (Schutz & Richards, 2003). Possibly, the students might not attempt to get help at all, for fear of the difficulties involved in getting their message across. Some NN graduate students may consequently experience setbacks in their work or have incomplete or inaccurate knowledge of their field. Students who struggle to be understood may not be very successful or confident in their interactions with people who do not speak their L1, whether within the academic setting or outside it (Schutz & Richards, 2003). This may lead to negative evaluations of their professional competence or their social abilities. As a result, these students may avoid opportunities to use or be exposed to the L2. They may restrict their recreational and social contact to friends and acquaintances who speak their L1, thus limiting their chances to learn how to communicate successfully in the L2. Under these circumstances, NN graduate students may have difficulty meeting the demands of their graduate programs, finding jobs after they graduate, or participating in Canadian society.

General Questions

This raises important questions: Does the L2 speech of NN graduate students become more intelligible (better understood) and more favourably evaluated over the course of their graduate studies? Does this change happen when NN graduate students are receiving instruction which targets L2 communication? Does it happen when they are

simply following their program of study within the university setting, and not receiving instruction which targets L2 communication? In the study described in this thesis, these general questions are investigated with respect to English, the language of instruction at the university where the study was conducted.

In Chapter 2, some basic terms will be defined, previous research related to the questions raised above will be reviewed, and specific research questions will be presented. In Chapter 3, the instruments and methods for collecting and analyzing the data will be described. In Chapter 4, the results will be presented, and in Chapter 5, the results will be discussed and interpreted. Chapter 6 contains concluding remarks.

Definition of Terms

Because I am concerned with how listeners understand and evaluate the L2 English of NN graduate students, this section focuses on the ways in which *intelligibility* (understanding) and *evaluation* of L2 English were defined in this thesis research. The first three definitions are related to listeners' subjective *evaluations*, involving their own internal criteria, and the last definition is related to *intelligibility*, involving objective criteria for measuring understanding.

Accentedness is a listener's subjective evaluation of how closely a speaker's accent (the phonological characteristics of their speech) approaches that of a native speaker. The term "native speaker" itself has become a controversial one (Canagarajah, 2007; Davies, 2003); in this study, the nature of native speaker characteristics was left to the judgment of each listener.

Comprehensibility is a listener's subjective evaluation of how easily the listener understands a given speaker.

Fluency is a listener's subjective evaluation of how "fluid" and "smooth" a speaker's speech is. It includes considerations of pauses, hesitations, incomplete words, and stammering.

In this study, intelligibility and understanding are taken to be equivalent terms. This measure involves a listener's understanding of a speaker; the criteria for assessing this understanding are external to listeners, unlike comprehensibility, accentedness, and fluency ratings. The listener completes a task which is meant to demonstrate the listener's

understanding of the speaker, and the listener's performance is scored according to pre-set, objective criteria. For example, a listener could identify or repeat words or sentences, answer comprehension questions, or re-tell a story. The listener's production would then be scored according to criteria external to the listener's own *perceptions* of understanding. In contrast, for comprehensibility, accentedness, and fluency ratings, each listener gives ratings according to her own individual criteria.

Overview of Research

In this chapter, the research reviewed covers four broad areas. The first two areas address the first two research questions which will be posed, while the last two areas address the last two research questions which will be posed:

1. the elements of speech which are linked to the intelligibility and listener evaluation of L2 English;
2. the effects of particular types of instruction on the intelligibility and listener evaluation of adult speakers' L2 English;
3. the relationship between adult speakers' use of and exposure to L2 English and listener evaluation of those speakers' L2 English;
4. the views of NN graduate students on their challenges and experiences using L2 English in an academic setting.

Understanding and Evaluating Key Elements of Speech in L2 English

Before describing research on how certain types of instruction have affected the intelligibility and listener evaluation of L2 English, it is important to first review studies

which identify elements of speech which might contribute to the intelligibility and listener evaluation of an L2 English speaker. I do not concern myself here with factors such as the individual differences of speakers or listeners (e.g., differences in age, gender, short-term memory) or the situational context of the speech (e.g., an academic lecture vs. a conversation). I consider here only elements internal to L2 speech itself because they are assumed to be under the control of the speaker and so possibly amenable to change through instruction.

Research in which multiple linguistic elements were analyzed has shown mixed findings. Munro and Derwing (1995, 1997), looking at statistical relationships between multiple linguistic elements (e.g., grammar and phonemic errors, intonation scores), intelligibility, and listener evaluations, found no clear correlational link between any of the linguistic elements and intelligibility; nevertheless, for listener evaluation, there seemed to be a link at some level between better accentedness and comprehensibility ratings on the one hand, and more accurate production of grammar, segmental, and suprasegmental elements on the other. This relationship, though, may be stronger when speakers are at lower proficiency levels.

Anderson-Hsieh, Johnson, and Koehler (1992) also looked at multiple elements related to pronunciation. They found significant negative correlations between favourable listener evaluations and error rates for segments and syllable structure, as well as positive correlations between favourable listener evaluations and goodness of prosody ratings. The prosody ratings were the strongest predictors of the listener evaluations.

When suprasegmental elements were investigated in isolation, some studies seemed to show a consistent relationship to accentedness ratings, comprehensibility

ratings, intelligibility, or a combination of these. Trofimovich and Baker (2006) found that L1 Korean speakers' rhythm, peak alignment, speech rate, pause frequency, and pause duration all significantly correlated to listeners' accentedness ratings, and that speech rate and pause frequency were significant predictors for ratings of accentedness. Munro and Derwing (1998, 2001), too, found significant correlations between speech rate and both accentedness and comprehensibility ratings. They also observed that when non-native speech was speeded up to a native speaker rate or slowed down below the average non-native rate, it received relatively worse ratings. Unlike the two previous studies, Anderson-Hsieh and Koehler (1988) did not find a difference in accentedness ratings for different rates of non-native speech, but did observe that faster speech was objectively harder for listeners to understand (i.e., less intelligible).

When suprasegmentals (e.g., tonic stress) have been manipulated to reflect native speaker norms, speech which is more native-like has proved to be significantly more intelligible to listeners than speech with non-native-like properties (Hahn, 2004).

J. Jenkins (2000) also found that tonic stress and vowel length in particular were important for L2 speakers' intelligibility in English, but that the production of certain segmental sounds also had an impact.

When considering only segmental sounds and their relation to listener evaluation, the picture is much less clear than it is for suprasegmentals. Studies by Riney and others (Riney & Flege, 1998; Riney & Takagi, 1999) demonstrated that although accuracy of segmental production (measured acoustically) may sometimes increase as accentedness ratings improve, some learners with accurate (native-like) segmental production may still be rated as having a strong non-native accent.

Overall, the linguistic elements which are linked most consistently to intelligibility and listener evaluation in the studies mentioned are suprasegmentals. However, it is important to note that of the research done on non-native *segmental* production in English, little research explicitly explores the connection between segmental production, intelligibility, and listener evaluation. The focus has usually been on the impact of other factors (typically, individual differences), such as language learning experience or age of L2 learning onset, on segmental production.

Effects of Instruction

In the section above, it was shown that suprasegmental aspects of English pronunciation such as speech rate and tonic stress were the elements with the most consistent links to intelligibility and listener evaluation of L2 English. Are these findings reflected in the types and effects of instruction received by ESL learners? This section will cover research on different types of instruction received by adult ESL learners, and the effects of that instruction on the intelligibility and listener evaluation of their English.

This section begins with research on the effects of instruction with no specific focus on L2 speech, such as a general English as a second language (ESL) course or a content course in which English is the language of instruction. The section then continues with research on the effects of different types of adult ESL instruction which targets L2 speech. Each type of instruction targets the intelligibility and pronunciation of adult ESL learners, but each involves different pedagogical approaches and/or linguistic elements. A summary of the findings and general conclusions comprise the final part of the section.

To my knowledge, there is no research on the relationship between intelligibility for adult L2 English speakers and L2 instruction with no specific focus on L2 speech.

Most of the few studies on this type of instruction have not shown any benefits for the way listeners evaluate L2 speech. Neither Purcell and Suter (1980), Thompson (1991), Elliott (1995), nor Flege et al. (1999) found that amount of instruction on or in the L2 was a significant predictor of the accentedness ratings received by adult L2 English speakers. However, Flege and Fletcher (1992) did find that the number of years of instruction in English received by L1 Spanish learners had significant (albeit low) predictive value in English accentedness ratings.

Effects of Instruction Targeting L2 Speech

A considerable number of studies describe the effects of instruction targeting L2 speech on the intelligibility and listener evaluation of L2 English speakers. This research, reviewed below, is sorted into four broad categories of pedagogical approaches used in L2 English speech instruction, and the effects of those approaches. The first category is instruction which targets articulation. The second category covers training using drama and more general listening and speaking practices. The third category covers instruction in suprasegmentals. The fourth category covers research in which several different instructional approaches were used for different groups of learners. In these studies, the intelligibility and listener evaluation of these different learner groups were assessed in light of their exposure to a particular instructional approach.

In a growing trend in articulatory training, instruction has focused not simply on the articulation of particular phonemes, but on the ways in which speakers of a specific language typically position their speech organs (e.g., tongue, jaw, lips). This positioning is termed a speaker's "articulatory setting". For example, Kerr (2003) examined how attempts to modify an ESL learner's articulatory setting affected the learner's intelligibility. She speculated that a Cantonese learner's posterior and nasal focus of resonance, with more tension in the back of the tongue, interfered with the production of alveolar sounds in English and a less nasal (and more nativelike) focus of resonance.

The learner undergoing training was a 56-year-old Cantonese male who had been living in Australia for nine years. The training program comprised 12 sessions over five months and was based on four principles: anterior focus of resonance (for more a more nativelike voice quality); vowels rather than consonants; attention to the difference between spoken and written English; and deliberate use of certain techniques. For example, the learner was encouraged to relax his speech organs and to attend to kinaesthetic feedback of the position and movements of bones in his head and face. Kerr focused the learner's self-monitoring on obvious movements of the lips and mouth, such as whether the lips were spread or round. During training, the techniques were practised first with field-specific vocabulary, then eventually in mock job interviews and conversations.

Before and after the training, the learner recorded a Weiss Comprehensive Articulation test.[1] No significant differences were found in the number of pronunciation

[1] This test is norm-referenced and is designed to reveal articulation disorders, misarticulation patterns, and talker-specific features of articulation for L1 English speakers.

errors in the pre- and post-tests. Four native listeners of Australian English transcribed sentences excerpted from both tests. These listeners were able to transcribe more of the post-test sentences than the pre-test sentences, but no tests of significance were conducted. Kerr suggested that the learner's improved level of intelligibility was due to factors at the level of the phrase or sentence, and not due to phoneme-level improvements.

Drama and Speaking and Listening Practices

Some instructional approaches focus on training learners through production of extended pieces of discourse (often plays and speeches), and training learners to use and reflect on speaking and listening strategies and techniques to improve overall communicative success.

A wide-ranging approach of this kind was taken by Acton (1984), who described a course he taught in both holistic and specific listening and speaking practices designed to modify pronunciation in English which was fossilized. The L2 learners were professionals and had resided in an English-speaking country for an average of five years. The course consisted of a four-hour class once a week for 12 weeks. The four main assumptions behind the instructional approach were that: learning outside the classroom was the most significant learning for pronunciation; learners should be taught to use their own resources to adapt their pronunciation; learners were responsible for their own improvement; and intelligibility is transmitted paralinguistically as well as linguistically.

Learners began with a focus on controlling their states of affective and physical readiness by discussing their inner thoughts and attention while speaking English and by practising breathing and postural exercises to control their emotions and physical speech

settings. Training in monitoring strategies included recording read-aloud speech and analyzing it for problems, monitoring pronunciation through kinaesthetic feedback, and using a native speaker informant in their workplace for information and feedback about pronunciation. The learners also were instructed in the contribution of body language to meaning, and practiced mirroring the body language of other people. To encourage transfer from the classroom to the outside world, learners had to report every week on what steps they had taken in their workplace to modify their pronunciation.

Acton reported that only one third of the learners both finished the course and put adequate effort into finishing assignments and using the strategies. However, for these learners, independent listeners found a noticeable improvement in the pronunciation of each from pre-test to post-test recordings (the content and nature of the recordings and assessment were not explicitly mentioned). Supervisors and work colleagues also reported by the middle of the course that they noticed a difference in the speech of the learners. In this research, the evaluation of learners' improvement was only impressionistic, and the study did not include a control group for comparison.

A more intensive course was described by Derwing, Munro, and Wiebe (1997). A "general speaking improvement" course was offered for high-proficency adult ESL learners of mixed L1s. Most of the thirteen learners had chosen to take the course, while four graduate students had been directed to the course by their supervisor. The course lasted 12 weeks, with the class meeting twice a week for three hours each class.

The instructors did not focus on specific sounds, but targeted general improvement in speaking. This included training in body language, voice quality, volume and rate of speech, discourse markers, stress, intonation, and rhythm. Learners worked in

groups, giving and receiving peer corrective feedback. It appears that the researchers' knowledge of the course was based on the instructors' self-reports and materials, and that the researchers did not actually themselves observe or record any part of the course.

Learners recorded true and false sentences at the beginning and end of the course. Thirty-seven native listeners transcribed each sentence and rated it for comprehensibility and accent. Scores for intelligibility (accuracy of transcription) improved significantly from pre- to post-test. True sentences had significantly better comprehensibility and accentedness ratings over time, but not false sentences. Three of the learners improved significantly in scores on intelligibility, accent, and comprehensibility, while eight learners significantly improved in at least one measure and four learners had significantly lower scores over time in one measure. The instructors reported that the learners who demonstrated an openness to working to adapt their pronunciation improved the most. The researchers suggested that some learners who did not significantly improve in their intelligibility scores may have already been highly intelligible. Because of the absence of a non-native control group, it is again not clear whether gains in pronunciation were due to instruction. However, the average length of residence of the learners in an L2 environment (10 years) points to instructional effects and not simply exposure as being an important factor for the change.

Stevens (1989) described a course for international teaching assistants (ITAs) which focused on speaking, listening, and teaching practices, but with an orientation towards drama. The 15 future teaching assistants had scored below 250 on the standardized oral proficiency SPEAK test[2] (Educational Testing Service, 1996), and were

[2] In this test for NN speakers in academic settings, testtakers must respond orally to printed and recorded stimuli. Their performance is scored by local raters.

required to take a 40-hour intensive course to improve their intelligibility. The training targeted teaching skills such as maintaining concentration and focus in adapting to changing situations in the classroom, using spontaneity and creativity in the preparation and explanation of lessons, and monitoring and adapting the manner of certain actions such as entering or moving around the classroom. The ITAs also practised producing fluent, energetic, and confident speech and more nativelike use of suprasegmentals. Activities included choral chants, mirroring others' speech, and "controlled shouting matches" (p. 189). The learners also observed professors' body language in teaching sessions and practiced interpreting the expressions and postures of undergraduate students.

Concurrent with the 40-hour course, each learner was also tutored individually for 30 hours by an ESL specialist. Segmental sounds which individual learners found difficult were modelled by the tutors. Learners listened to recordings of broadcasts or lectures and then recorded their own versions. Outside of the class, the learners interviewed native speakers and transcribed and analyzed portions of the interviews, looking at specific aspects of language or pronunciation.

The effects of the training were evaluated with three tests and impressionistic self-ratings. The SPEAK oral proficiency test, as well as the TEACH test, was administered to the ITAs before and after training. The TEACH test measured "cultural appropriateness, pedagogical effectiveness, linguistic clarity and accuracy within an instructional setting" (p. 190). The last test consisted of standardized university course evaluations by undergraduate students on the learning and teaching that had occurred in their ITA-led courses. The scores for intelligibility-trained ITAs were significantly higher

from the first SPEAK test to the second by an average of 36 points. The TEACH scores were given on a scale of zero to three. The intelligibility-trained ITAs were assessed on the TEACH test together with a group of ITAs who hadn't received intelligibility training. From pre- to post-test, TEACH scores rose significantly for the *combined* group by an average of 0.83. However, the change in TEACH scores for only the intelligibility-trained ITAs is not mentioned. All but one of the eighteen undergraduate classes evaluated their newly-trained ITA's language proficiency as adequate. The one ITA whose language was not rated as adequate at the end of the training had been placed as an ITA against the recommendations of the course instructors. The ITAs themselves at the end of the course reported that their production of suprasegmentals in pronunciation had improved noticeably.

Although intelligibility was not evaluated in isolation, the communication skills of the ITAs seem to have clearly reached a more acceptable standard after the course. The lack of a control group means that the effect of incidental exposure to and practice in English on the ITAs' communication skills is not known. However, the case of the unsatisfactory ITA who had not been recommended for teaching suggests that the performance requirements for the intelligibility course reflected the communicative demands for the undergraduate classroom.

Instruction in Suprasegmentals

In the research described below, a focus on suprasegmentals was the primary principle of the instruction and practice. Wajnryb, Coan, and McCabe (1997) report on a short-term pronunciation clinic developed for L2 learners at a language centre in Australia. The clinic lasted for one hour a week for six weeks. Ten learners of

intermediate proficiency or above, all L1 speakers of Asian languages, chose to participate.

The overall focus of the clinic was on rhythm and stress, although some individual work was done with specific sounds which were problematic. The instructor also targeted smooth speech through elision, reduction, and intonation. The learners listened to recorded texts, marking the stress, linking, and intonation. They then read along with the recording and recorded their own version of the text, subsequently receiving feedback from the instructor. The same nursery rhyme was practised weekly for practice in stress placement. A substantial amount of out-of-class practice was reported, and attendance was high.

The learners' pronunciation was evaluated by recording the same short paragraph from a commercial textbook at the beginning and end of the short clinic. Each of the five sentences from the text was evaluated by an experienced teacher with previous experience with L2 phonology, who marked the sentences for sentence stress, word stress, reduction, elision, intonation, and articulation. The average improvement for all six aspects of pronunciation from pre- to post-test ranged from 7 to 27%, with an overall average of 20.8%. The increase in scores was higher for the learners who began with the lowest scores. No tests of significance were conducted. Similarly, the aspects of pronunciation which had the lowest scores in the pre-test (reduction and elision) showed the highest increase in scores at the post-test. Sentence and word stress had both improved to a similar extent by the time of post-testing. Eight of the ten learners believed they had improved in their pronunciation by the end of the course; most of the learners were able to identify areas in which they had improved. Some learners reported more

confidence when they spoke English. However, the absence of a control group makes it more difficult to rule out simple exposure over time as a primary factor in learners' improvement.

Comparison of Approaches

In all the studies previously cited, the focus was on one particular course or training program in which learners received instruction and practice in pronunciation. In the final two studies, different types of training were provided to separate groups of learners. The results were then compared for differences in instructional effects. Both the length of training and the type of speech elicitation task were factors in whether training effects were seen.

The effect of short-term training was investigated by Macdonald, Yule and Powers (1994). Twenty-three L1 Chinese graduate students who had been identified as having problems with pronunciation participated in the training. The students recorded two different mini-lectures on the metric system; instructions and the words and phrases which had to be included in all the lectures were provided. The students recorded the first lecture by themselves immediately before the training (Time 1), then the same lecture immediately after the training, with a native speaker audience member (Time 2), then a different lecture, as a continuation of the first lecture, two or three days after the training (Time 3).

The students underwent one of four training conditions: teacher-directed drill, modified interaction, self-study drill, and silent review. The teacher-directed drill had an instructor modelling the words, phrases, and sentences provided to the students for the lectures. The students repeated the models and received corrective feedback from the

instructor. This condition lasted for ten minutes. The interaction condition had the students presenting their mini-lectures individually to the instructor, who made clarification requests when the students produced the selected words and phrases. This condition also lasted ten minutes. The self-study drill condition had students listening to a tape with the selected words, phrases, and sentences and repeating them. No external corrective feedback was available. This condition lasted 30 minutes. In the last condition, silent review, the students reviewed their notes for ten minutes.

Ten of the same words or phrases from each of the learners' three lectures were selected. These 30 items were paired within the same speaker, with same words or phrases from Times 1 and 2, or from Times 1 and 3. Each of the 120 L1 English raters heard six different speakers saying each of the ten word or phrase pairs. After hearing each pairing, raters indicated whether the first or second utterance was more nativelike.

The judgements were analyzed by training condition and by speaker. When analyzed by training condition, the only significant difference was found at Time 2, when significantly more listeners gave ratings of "nativelike" to the items from the self-study condition than to the items from the silent review condition. No significant differences were found between conditions for items at Time 3. The researchers report that for only one condition, self-study, was pronunciation rated as nativelike more often at Time 3 than at Time 2. However, this statement is based only on the amount of difference from Time 1 utterances, as no direct comparison was ever made of utterances at Times 2 and 3.

When analyzed by speaker, no unambiguous pattern of instructional effects emerged. The researchers concluded that no training condition was equally helpful for every learner. They suggested that learners' individual differences may have a greater

influence than instructional approach on the learning of pronunciation. The researchers pointed to the greater amount of time allowed for the self-study condition, which may have tipped the scales in favour of this condition. The relatively small number of participants for each training condition, the isolated nature of the assessment of pronunciation, and the short time allowed for each condition may also be important factors in the lack of clear findings.

Three instructional conditions were examined in Derwing, Munro, and Wiebe's (1998) research, but the instruction lasted much longer and targeted more participants. Forty-eight intermediate-level ESL learners of mixed L1s were divided into three classes of 16, roughly matched for L1, gender, age of arrival, and length of residence. Each class was in an 11-week intensive ESL course for 20 hours a week which covered speaking, listening, reading, and writing. The three training conditions lasted 20 minutes per day, and were described as a segmental approach, a global approach, and no instruction specific to pronunciation. The segmental approach had a focus on the pronunciation of words and smaller units of sound, while the global approach targeted pronunciation in discourse, particularly the suprasegmental aspects of stress, intonation, and rhythm. The content and procedures for the training were negotiated between the researchers and instructors, who met at the beginning of the course and halfway through for progress reports.

Besides suprasegmentals, the instructor for the global approach also focused on speaking rate and projection. Jazz chants and group presentations were activities consistently used, both for perception and production. No instruction or feedback was

given for individual sounds. The instructor for the segmental approach used perception and repetition tasks with language lab materials and teacher-centred exercises.

The learners' pronunciation was recorded near the beginning of the course and at the end. They read simple one-clause statements and told a story from a set of pictures (the same statements and pictures were used for both recording times). The sentences evaluated included one from each speaker from the beginning and end of the course. Forty-five-second excerpts of each speaker's picture story from the beginning and end of the course were also selected for evaluation.

Forty-eight L1 English speakers rated the statements only, evaluating the comprehensibility and accentedness of each utterance. The learner groups were not rated differently in comprehensibility or accentedness at the beginning of the course. However, the groups receiving the global and segmental training both improved significantly in comprehensibility ratings at the end of the course. All three groups improved significantly in accentedness ratings at the end of the course, but the group receiving segmental training improved significantly more than the other two groups.

Six experienced ESL instructors rated the picture story excerpts on accent and comprehensibility, as well as fluency. The three learner groups were not rated differently on any aspect at the beginning of the course, but at the end of the course, only the group receiving global training had significantly higher comprehensibility and fluency ratings. No significant differences were found for accentedness ratings.

The researchers concluded that both the listening task and the instructional approach significantly affected whether learners' speech showed improvement. They highlighted the significant improvement in comprehensibility and fluency by the global

group for the picture story task. The transfer of training to *spontaneous production*, say the researchers, was evident only for the group receiving global training, whose picture story excerpts were rated significantly more comprehensible at the end than at the beginning of the course. As in their 1997 study, it must be remembered that it appears the researchers did not themselves observe what occurred in each classroom, but relied on each instructor's self-report. The different groups of raters for each task, with different experience and possibly different expectations of L2 speech, might also be a factor in the lack of significant differences for the segmental group for the story task.

Summary: Instruction Targeting L2 Speech

A very general summary of the results of the instructional practices described above is that some noticeable improvement in learners' pronunciation, awareness of pronunciation, or confidence in speaking almost always follows instruction and practice targeting L2 pronunciation, whether through general speaking practices, suprasegmental training, or other approaches. Because of the widespread lack of significance testing and control groups in most research, it is not clear whether that effect is generally significant or not, or even whether it is due to the instruction received. Even when focusing on specific approaches and instructional practices, general trends in instructional effects are sometimes difficult to discern.

General Conclusions: Instruction

From the pedagogical research reviewed above, it was found that when ESL adult learners received instruction which was *on* or *in* English, but which did not target speaking specifically, the amount of instruction generally did not predict the accentedness ratings the learners received (e.g., Flege et al., 1999; Purcell & Suter, 1980). However,

for almost all types of instruction targeting L2 speech, ESL learners made some improvements, although very brief training seems to have had little effect (e.g., McDonald et al., 1994). This leads to the unsurprising conclusion that some type of instruction and structured practice targeting L2 speech is probably better than none, with the caveat that very short training periods do not seem effective.

As discussed in the previous section on linguistic elements, when various linguistic elements were investigated for their relationship to intelligibility scores and listener evaluation, the elements with the most consistent links were suprasegmental aspects of pronunciation. Significant improvement in *spontaneous* speech was only seen with extended training in suprasegmentals and in an approach combining drama, teaching, segmentals, and suprasegmentals. Based on the limited research available, it appears that instruction using these approaches may be effective in helping adult ESL learners to become more intelligible and more favourably evaluated by listeners.

However, all of the studies cited above have certain methodological shortcomings. These studies did not meet at least one of the three conditions which are necessary to explore possible relationships between instruction, intelligibility, and listener evaluation. These conditions are (a) thorough observation of the instruction over time, (b) detailed measurement of L2 learners' production over time, and (c) at least one comparison group of L2 speakers in order to identify changes in L2 speech due to non-instructional factors. The present study was designed to meet these three conditions, leading to the first two research questions.

Research Question 1

What is the nature of classroom instruction for NN graduate students in an ESL course targeting oral communication?

Research Question 2

How does the speech of NN graduate students develop, in terms of its intelligibility and evaluation by listeners:

(a) from the beginning to the end of an ESL course targeting oral communication?

(b) when not taking an ESL oral communication course, but only studying in an English environment?

Language Exposure and Use

NN speakers often seek out ESL instruction in spoken English partly because it provides them with regular opportunities to both hear and use spoken English. However, there are other means by which NN speakers can use or be exposed to English. It is therefore important to explore how adult ESL learners' spoken English and their *use of* or *exposure to* English are related to each other.

The only study known to me which relates intelligibility to L2 exposure and use is Liu (2001), who found that for mainland Chinese graduate students at a U.S. university, the students' reported frequency of contact with Americans and American culture predicted 50% of the variance in their score in an oral proficiency (SPEAK) test. However, there are many studies which link *accentedness* ratings to exposure/use. Flege et al. (1999) found that when L1 Korean English learners were matched for age of arrival in the U.S., length of residence, and amount of U.S. education, the learners who used

English relatively often in their daily lives were rated with significantly more nativelike accents than learners who used English relatively infrequently. Similar results were found by Flege, Frieda and Nozawa (1997), Guion, Flege and Loftin (1999, 2000), Piske, Mackay, and Flege (2001) and Tahta, Wood, and Lowenthal (1981). In contrast, Thompson (1991) found a significant simple correlation between accentedness ratings in English and L1 Russian speakers' use of English at home, but use of English was not a significant predictor of accentedness ratings in a multiple regression. Flege and Fletcher (1992) also did not find any significant correlation between the percentage of English used daily by L1 Spanish speakers and accentedness ratings.

Summary: Language Exposure and Use

The majority of research on language use and exposure for adult NN speakers points to a consistent advantage in listener evaluation, and perhaps the intelligibility, of learners who use English relatively more than those who use it relatively little. However, in all of these studies, English exposure and use was measured by NN speakers' one-time reports, which cannot capture fine-grained patterns of language exposure/use and are also vulnerable to speakers under- or over-estimating how much they are exposed to or use English. If we are to investigate how NN graduate students' spoken English develops over time, then we should also investigate how they are exposed to and use their L2 *over time*. This brings me to the third research question:

Research Question 3

In what ways do NN graduate students at an English-medium university come into contact with and use English on a weekly basis? The longitudinal tracking of NN speakers' exposure and use is an innovative aspect of this research.

Graduate Student Reports of Their Language Skills and Experiences

In order to learn more about how NN graduate students' L2 English speech develops over time, it is important to measure how they are understood and evaluated by listeners, to observe the targeted instruction they might receive, and to track their English exposure and use. However, these measurements and observations do not convey the whole story. Language learning is not simply a stimulus-response mechanism; learning is clearly influenced by the learners themselves, their experiences and perspectives (Towell & Hawkins, 1994). Non-native graduate students have inside knowledge about how their English skills affect and are affected by their experiences within the university and in other settings.

In order to fully explore NN graduate students' learning and use of English, it is crucial to record their own views on the matter. Therefore, numerous researchers have used surveys and interviews to investigate NN graduate students' experiences at university and in society. The following section reviews research on how NN graduate students' English skills, particularly those related to their speaking ability, are related to their academic experiences.

Most research on L2 English graduate students describes graduate students in the U.S. From surveys and interviews, a common finding is that NN graduate students who report having problems with the language of instruction often also report having difficulties academically (Sun & Chen, 1997). In a survey of upstate New York colleges and universities, Xu (1991) found that international graduate students' self-rating of their

proficiency in English was the major predictor of their self-rated level of academic difficulty in doing speaking, listening, reading, and writing tasks.

In Canadian contexts, Fletcher and Stren (1989) surveyed international graduate students from a large university in Ontario. Similarly to Xu's (1991) findings, they found that students' reports of their language difficulty in their coursework were negatively correlated with their satisfaction with their academic program, and with their GPA. Sun's (1987) language needs analysis survey for Chinese graduate students at a Canadian university revealed that overall, students believed that aural/oral skills, such as understanding a lecture, giving a seminar, or talking on the phone, were the most important skills, both academically and socially. In another needs analysis study, Cheng, Myles, and Curtis (2004), surveyed and interviewed international graduate students at an Alberta university. In the survey, students rated academic skills for their importance and their difficulty. The majority of skills rated as quite important had to do with reading. However, of the five skills receiving the overall highest ratings for importance, four had to do with understanding or producing oral language, such as leading class discussions. The authors noted that skills received higher ratings for importance than for difficulty, with only one skill out of 31 rated "somewhat difficult" and the rest rated lower. However, of the five skills receiving the overall highest ratings for difficulty, three involved oral production (e.g., participating in class discussions) and two involved written production (e.g., writing long reports).

In Cheng et al.'s interviews with students, the students reported becoming noticeably better in their listening comprehension over time, especially for lectures. They still had difficulty understanding Canadian classmates, TV programs, or movies. Giving

presentations in front of class was described as being the most difficult skill because of students' difficulties with language and their corresponding lack of confidence. According to many students, this skill could be improved with practice; however, some students were not satisfied with their overall progress in English.

> As to my English language, actually, although I don't find any difficulties in communicating, like participating and listening to all the lectures, but I do find I do not improve my spoken English at all and I also find my English, although I do not have any difficulties, I mean my English level still stay there, like generally speaking the language proficiency do not improve at all. It just stay there (Cheng et al., p. 64).

In the following three studies, researchers conducted numerous interviews to explore Asian graduate students' experiences in learning and using their L2 in a university setting in North America. In each study, a common theme was how powerfully students' language difficulties could affect their views of themselves and their own competence.

> I was an eloquent speaker and capable writer at home...I have been proud of my ability to speak with fluency and to write with confidence. However, since I came to Canada, I lost not only the eloquence in my speech and writing but also the source of my self-pride. Whenever I speak something with a broken English, my sense of dignity subsides. Whenever I make a mistake because I misunderstand someone's speech, my self-esteem recedes. It is really difficult for me to accept the current "I". It is definitely remorseful for me to say it to myself, over and over, like chanting a spell, that "I am more than that" (Lee, 2006, p. 75).

The demands of conversation, responding to an interlocutor by forming a message into appropriate words and grammar, could greatly reduce students' spontaneity and sense of easy participation.

> I don't think I've had many opportunities to converse with others in English. What I mean is a real conversation, not chatting or any other sort of small talk...Because English is not my language, I have to form entire sentences before uttering them. Otherwise I easily get lost in the middle. But in this laborious moment, I already miss the streaming of conversation and quickness of response...My Canadian colleagues often call me "thoughtful". But I know that's not always a compliment. It's also a humble expression of how difficult [it is] to converse with me on their parts (p. 86).

Some students found it so difficult to converse with others that they would abandon communication altogether when they had difficulty. For example, Wang (2003) described a student who was asked what happened if someone said something to her and she didn't know how to answer. "[S]he told me she would tell the listener not to worry about it" (p. 82).

Those students with friends who were both proficient in the language of instruction and familiar with the university setting reported that these relationships were very important. "I have a really good friend...he can help me a lot. So, whenever I talk with him, we talk about things, [he] just says it slowly, slowly, and [gives an] explanation, and so he can help me" (Schutz & Richards, 2003, p. 60).

In the majority of survey studies described above, a consistent finding was that when NN graduate students reported difficulties with the language of instruction, they also found their academic program or academic tasks difficult. Of the tasks rated the most difficult by students, listening and speaking tasks were at the forefront. In the interview studies, some students reported that they became better at understanding oral speech over time, but seemed to plateau in their speaking abilities. Other students described how they lost pride in themselves when they could not communicate successfully.

Although Liu (2001) found that graduate students' oral language proficiency was linked to the amount of contact they had with Americans or American culture, some students in the interview studies found extended conversations with native speakers were such a challenge that they were rarely attempted.

It is clear that the development of NN graduate students' English skills shapes and is shaped by students' studies and their lives. It is therefore not sufficient to simply measure the development of their spoken English over time. This would provide a picture of the graduate students' language development; however, it would leave out graduate students' insider knowledge, their own assessments of struggles and successes in learning and using English. To truly understand how their abilities in English affect graduate students in their daily lives, it is necessary to ask the students themselves. I therefore ask the fourth and final research question.

Research Question 4

What do NN graduate students think about their development in English over a five month span and during their graduate studies in general?

Summary

In this chapter, I defined the terms used in this study to describe listeners' understanding and evaluation of L2 speech. I described research on elements of speech which are linked to intelligibility and listener evaluation. As well, I reported on studies on the effects of instruction and exposure on the intelligibility and listener evaluation of L2 English. I also described research on the views of NN graduate students on using L2 English. Drawing on these areas of research, four research questions were developed. In the following chapter, the methodology for investigating these four questions will be described.

CHAPTER 3: METHODOLOGY

In this chapter, I report on the methodology of conducting the study. I describe the recruitment and characteristics of the study participants, the materials which were selected and developed, and the procedure for collecting and analyzing data.

Recruitment

Talkers

The recruitment of two groups of talker participants was done primarily through class visits and notices posted in buildings at an English-medium university. To recruit control group participants, beginning in November 2005 notices were posted around campus. When potential participants contacted me by email, they were sent a draft consent form telling them of the conditions of the study; I then discussed with them by phone or email the content of the consent form. A final version of the consent form was emailed in December, and a first meeting with each participant was arranged in the third week of January 2006. Please see Appendix A for the research ethics certificate and Appendix B for the control group consent form.

To recruit participants being trained in oral communication (treatment group), in the second week of January 2006 I visited two class sections of an oral communication course designed for NN graduate students, and told the classes about the chance to participate in a long-term study on the development of speaking in English; participants would be compensated financially and with a tutorial at the end of the study, and the instructor would not be told about a given student's participation or performance in the study. All students received a consent form to look over (Appendix B). Interested students provided their email addresses.

I also informed the students of one section that, with the permission of their instructor, their class would be videotaped every second class in order to accurately record what and how the instructor taught; it might then be learned how the instructor could make useful pedagogical changes which would help students develop their speaking. The students of that section were also advised that it was the instructor, not the students, who was the main focus of the videotaping, and that if a student did not want to appear on camera, that student should not sign a consent form giving permission to be videotaped in class (Appendix B). The student would then not appear on camera at any time over the study. Only one student chose not to give consent to be videotaped, and that student never appeared on-camera during any class.

Potential participants for the treatment group were contacted by email reminding them of their expressed interest. If the potential participant emailed back to agree to participate, a first meeting was arranged for the third or fourth week of January 2006.

Listeners

Listeners were recruited by posting notices around campus and electronically on a university student employment website, and by word of mouth. Potential listeners qualified for the study if they confirmed by email that they were native speakers of English and had at least one parent who was a native speaker of English.

Participants

Talkers

Twenty-one NN graduate students were originally recruited to participate as talkers. In addition to their graduate work, eleven students were in one of two sections of

the oral communication course (treatment group). Ten students from the same university were also doing graduate work but were not currently taking the oral communication course (control group). However, five of the control group students had taken the course two or more semesters before. Due to attrition, data from two students each from the treatment and control group had to be discarded, leaving nine students in the treatment group and eight in the control group. When the study began, the average ages of students in the treatment and control groups were 27 and 31, respectively. Treatment group students had L1s of Mandarin, Spanish, French, Russian, and Tamil, while control group students spoke Mandarin, Farsi, and Spanish as their L1s. The treatment group students, including one Quebecois student, had been studying and/or working in an English environment for an average of 17 months, and the control group students for an average of 27 months. The NN students' personal biographical information can be seen in detail in Tables 1 and 2. All names listed are pseudonyms.

All NN talker participants completed a questionnaire about their level of proficiency in English, self-evaluating their strengths and weaknesses in English and reporting on their goals for English learning and their level of exposure to English every day. This information, in the participants' own words, is shown in Tables 3 and 4. The complete questionnaire can be seen in Appendix C.

In addition to the NN graduate students, four native speakers of North American English participated as talkers in order to provide a baseline for measuring intelligibility. These participants, two men and two women, were university graduates or were currently studying at the undergraduate level, with a mean age of 33. Three of the participants had experience teaching English or French as a second language.

Table 1

Biographical Information of Participants in the Treatment Group

Name	Sex	L1	L2	Program	Age of arrival[a]	Age[a]	Length of residence[b]
Tai Ning	M	Mandarin	English	M.Sc., Science	23	23	4
Javier	M	Spanish	English	Ph.D., Engineering	26	26	5
Sigman	M	Tamil	English	M.A., Social Science	25	27	16
Marie-Pier	F	French	Spanish	M.Eng., Engineering	n/a	24	16
Hui	M	Mandarin	English	M.Eng., Engineering	22	23	16
Piotr	M	Russian	Ukrainian	Ph.D., Science	22	23	17
Bao	F	Mandarin	English	Ph.D., Medicine	31	33	19
Xiao	F	Mandarin	English	L.L.M., Law	29	32	24
Christine	F	French	English	Ph.D., Social Science	22	28	40

Note. [a]in years. [b]in months.

Table 2

Biographical Information of Participants in the Control Group

Name	Sex	L1	L2	Program	Age of arrival[a]	Age[a]	Length of residence[b]
Ahmed	M	Farsi	English	Ph.D., Engineering	25	26	4
Lupe	M	Spanish	English	Ph.D., Engineering	30	30	5
Esteban[c]	M	Spanish	English	M.Eng., Engineering	23	24	12
Ma[c]	M	Mandarin	English	Ph.D., Engineering	39	40	12
Ping	F	Mandarin	English	M.Eng., Engineering	29	31	18
Feng[c]	M	Mandarin	English	M.Eng., Engineering	30	33	42
Jiao[c]	M	Mandarin	English	Ph.D., Computer Science	27	31	60
Xing[c]	M	Mandarin	English	Ph.D., Engineering	29	35	66
Ahmed	M	Farsi	English	Ph.D., Engineering	25	26	4

Note. [a]in years. [b]in months. [c]These participants had also completed the oral communication course two or more semesters before.

Table 3

Language Learning Information for Participants in the Treatment Group

Name	Strengths	Weaknesses	Goals	Exposure		
				School	Home	Social
Tai Ning	preparing speech	grammar, pronunciation	pronunciation, grammar, native English words	4 hrs/day	0	1 hr/day
Javier	pronunciation	rhythm	rhythm	2-3 hrs/day	30 min/day (TV)	30 min/day
Sigman	grammar	speak very fast, primary stress a problem	bring down pace of communication	4-5 hrs/day	10-15 min/day	5 hrs/week
Marie-Pier	talking about normal stuff, confidence	accent, sound, grammar	to be perfectly understandable	10-18 hrs/day	0	about 3 hrs/day
Hui	topics that interest me or academic study	topics I'm not interested in	speak English as natural as my mother tongue	3-5 hrs/day	‹ 1 hr/day	‹ 1 hr/day
Piotr	speaking fluently	vocabulary, pronunciation, grammar	make it 100% understandable for others	4-5 hrs/day	2 hrs/day	1 hr/day

Table 3 (*continued*).

Name	Strengths	Weaknesses	Goals	Exposure		
				School	Home	Social
Bao	grammar	pronunciation	pronounce clearly	5 hrs/wk	1 hr/wk	10 hrs/wk
Xiao	mimicking	express	expression and communication	4-5 hrs/day	0.5 hrs/day	seldom
Christine	communicate with people	accent, intonation, rhythm	same as weaknesses	3 hrs/day	1 hr/day	1 hr/day

Table 4

Language Learning Information for Participants in the Control Group

Name	Strengths	Weaknesses	Goals	Exposure		
				School	Home	Social
Ahmet	reading, technical speaking	listening	everyday conversation	2 hrs/day	0	1 hr/day
Lupe	some confidence, social English	academic communication, attending classes	accent	5 hrs/day	2 hrs/day	1 hr/day
Esteban	explaining my ideas	public speaking	reduce accent and public speaking	1 hr/day	1 hr/day	1 hr/day
Ma	n/a (left blank)	worrying about making mistakes	express my thinking freely	~2 hrs/day	30 min/day	a few min/day
Ping	confidence	vocabulary	think in English, more words in mind	2 hrs listening, 30 min speaking/day	30 min listening/day	10 min/day

Table 4 (*continued*).

Name	Strengths	Weaknesses	Goals	Exposure		
				School	Home	Social
Feng	I can be understood	not confident, grammar mistakes	speak fluently	1 hr/day	30 min/day	2 hrs/day (work)
Jiao	confidence	tense, pronunciation, vocab	presentation ability	2 hrs/day	15 min/day	0 min/day
Xing	organizing a sentence	oral English and informal expressions	understanding others better	30 min/day	30 min/day	0

Oral Communication Course

The oral communication course was offered by a university department whose mandate is to offer credit courses in English to any registered university student. The department has several courses offered specifically to graduate students, one of which is the oral communication course. The course was a one-semester, three-credit course offered only to graduate students who did not speak English as an L1. It was offered for no extra charge to registered graduate students. The stated goal of the course was "developing pronunciation and communication skills, including aspects of pronunciation that most affect intelligibility, and with verbal and non-verbal techniques for effective presentations" (Graduate Courses, para. 2).

Instructor

The instructor, Michelle, was a native speaker of English with a graduate degree in Second Language Education and over ten years' experience teaching English as a second and foreign language. She had been in her full-time teaching position for four years and was the sole instructor for the two sections of the oral communication course at the time of the study. The sections met on the same days, and Michelle covered the same content on the same days for each section. Michelle was known in her department for her expertise in teaching pronunciation.

Listeners

Altogether, 87 participants did one of two listening tasks. Data from 11 of these participants were discarded because it was later discovered that they did not qualify for the study despite the advertised requirements for participants; these ineligible participants were either NN speakers of English, had two NN speaker parents, or were native speakers

of a variety of English not from North America. Therefore, the final pool of listeners comprised 76 participants. Almost all listeners were undergraduate students; only six were not undergraduates (one staff member, one computer programmer in a lab, one research associate, one postdoctoral fellow, and two graduate students). These six listeners were distributed throughout the listener groups. Listeners were assigned to each group according to their self-reported frequency of exposure to non-native, accented English. That is, each group contained equal numbers of participants with the exposure ratings *never, rarely,* and *occasionally* and participants with the ratings *fairly frequently, frequently,* and *very often.* For the first task, the average age of the 46 listeners was 21, with 9 males and 37 females, and for the second task, the average age of the 30 listeners was 22, with 8 males and 22 females. Information on listeners for each of the two tasks is shown in Appendix D.

Materials

The materials used in the study had five main purposes:

1. to collect background information on participants (questionnaires),
2. to record the classroom training received by the treatment group (observation scheme),
3. to elicit speech from talker participants (short films and prompts),
4. to document talker participants' exposure to English (English exposure/use log),
5. to measure L2 speech intelligibility and evaluation by listener participants (story analysis schemes and ratings scales).

Purpose 1: Collecting Background Information on Participants

Background information on both talker and listener participants was collected through written questionnaires. The questionnaire for talker participants was a modified version of a questionnaire from Li, Sepanski, and Zhao (2006), while the questionnaire for listener participants was one piloted in a previous study (Kennedy, 2006). Please see Appendices C and D for samples of these questionnaires. The information from these questionnaires allowed for subsequent matching of participants in both talker groups and listener groups, and provided an initial baseline to potentially track changes in talkers' self-reported weaknesses (e.g., confidence), strengths (e.g. aural comprehension ability), proficiency, and exposure to English.

Purpose 2: Recording the Classroom Training Received by the Treatment Group

Class observations were documented and categorized with the Communicative Orientation to Language Teaching scheme (COLT), Part A (Spada & Fröhlich, 1995). This scheme classifies classroom activities according to various categories, such as the focus of the activity, the type of interaction, and the nature of the text used. By using this scheme, the observer can code in a systematic way patterns of classroom interaction, instructional focus, materials used, etc. Part A of the COLT coding scheme is shown in Appendix E.

Purpose 3: Eliciting Speech from Talker Participants

The prompts for eliciting speech from talkers were of three types: true-false sentences, personal anecdote prompts, and short animated films and film retell prompts. The true-false sentences, 45 in total, were taken from a bank of true-false sentences in Munro and Derwing (1995). These sentences were five to eight words long, with three to

six content words per sentence. A word range analysis (Cobb, n. d.) showed that 79% of the words were among the first 1000 most frequent words in English (West, 1953). An equal number of true and false sentences were used for each recording session (see Appendix F). The sentences were included in the speech prompts so that changes in intelligibility of isolated, read-aloud sentences could potentially be investigated.

Most studies on L2 intelligibility focus solely on isolated, read-aloud words and/or sentences. However, generalizing the findings of these studies to authentic communication with L2 speakers is problematic because much of L2 speech, especially outside the classroom, is extended, speaker-generated speech. In order to explore the intelligibility of this type of speech, the two other types of prompts focused on narratives: namely, personal anecdotes and short film retells.

Six personal anecdote prompts had been piloted in an earlier study with NN university students (Kennedy, 2006). Two of the prompts had been found to elicit sufficient speech, with coherent narrative structure. The four other prompts were not used for the main study because some prompts elicited anecdotes well over five minutes long, and some prompts were not clear and had to be explained to pilot talker participants. For the main study, five additional prompts were generated using three criteria: (a) the prompt was different from any other prompt used, (b) the prompt focused on relatively concrete experiences which would be familiar to any given talker or listener participant, and (c) the prompt provided a specific scope for the narrative while giving talkers freedom to choose the content and structure of their anecdotes. The complete list of anecdote prompts used in the main study appears in Appendix G.

To elicit film retells, seven short animated films had been piloted in an earlier study (Kennedy, 2006). Two of the films were misunderstood or considered childish by the pilot talker participants, and so were not used for the current study. The other five animated films were used in the main study. One film (2 minutes and 30 seconds in length) was used for a warm-up activity; the other four films (four to eight minutes long) were used for the main tasks. All films had a strong narrative structure, often with some twist at the end of the story. Two films had no or little dialogue, while the remaining three had periodic or running narration, with one of the three having sung narration. The complete listing of film prompts used in the main study appears in Appendix G.

Purpose 4: Documenting Talker Participants' Exposure to English

The instruction in oral communication received by the treatment group was taperecorded with a video camera and documented using COLT, Part A. However, formal classroom instruction was not the only exposure to English that was available to talkers. Both treatment and control groups were exposed to and used English in various domains outside the classroom. In order to measure this exposure and use, an English log for exposure and use was modified from Ranta and Meckleborg's (2002) log for non-native graduate students' long-term exposure to English. Ranta and Meckleborg's original log was delivered through custom-made software on a personal digital assistant. Because the log for the present study was originally planned to be paper-based, the log categories were reduced to decrease the amount of time needed to complete a daily log. The five categories comprised (a) activities, (b) sub-activities, (c) the proportion of English use/exposure in relation to other languages during an activity, (d) the time spent

using or exposed to English, and (e) participant comments. The modified log, created in Microsoft Word, appears in Appendix H.

During the first recording session, when the paper-based log was introduced to talker participants, a participant offered to transfer the log categories to Access, a database application available on many PCs with Windows operating systems. The creation of the log database was completed before the first week of scheduled log completion, and was emailed to other participants. Some participants chose to record their log on computer using the original Word file, and some used the paper-based version. The participants thus had the choice of using either the paper-based version or a computer-based version, either in Access or Word.

Purpose 5: Measuring L2 Speech Intelligibility and Evaluation by Listener Participants

In order to explore the development of NN graduate students' L2 speech, samples of their speech (sentences, personal anecdotes, and film retells) were recorded so that measures of intelligibility and evaluation could later be collected from listeners. The procedure and timeline for recording, as well as for the subsequent processing and selection of the recorded speech samples, are described below. Because the intelligibility measures were refined during analysis of the speech samples, the intelligibility measures are described below in the Data Analysis section.

The evaluation measures were based on three rating schemes used by Munro and Derwing (1995) for measuring comprehensibility, accentedness, and fluency. All ratings were subjective judgements by the listener. Comprehensibility is here defined as how easy it is for the listener to understand a given talker. Accentedness is defined as how closely a given talker's pronunciation approximates a native speaker norm. Fluency is

defined as the smoothness and fluidity of a given talker's oral production (e.g., speech free of pauses, repetitions, false starts, and incomplete words). The ratings were on 9-point Likert-type scales, seen in Appendix I.

Procedure

Data Collection

The data collection unfolded in three phases. In Phase 1, I observed the oral communication classes and periodically recorded talker participants, and talker participants periodically completed English exposure logs. In Phase 2, the talker participants and I together verified unclear utterances, and I conducted one-time interviews with talker participants and the oral communication course instructor. In Phase 3, I selected speech samples, and had listener participants record story retells, guess unclear words, and rate (evaluate) speech samples.

Phase 1

Class observation. This research is unique in including longitudinal data collection not only of talkers' oral performance, but also of the training received by the talkers in the treatment group. Thus, the focus of instruction and the types of practice activities available to the treatment group talkers week by week were recorded. The classes for both sections of the course met two times a week; one section of the course was observed once a week except when mid-term presentations, individual teacher-student meetings, Reading Week, and final presentations occurred. The instructor taught the two sections on the same days at different times, covering the same material on the

same days. A total of 8 classes (out of the 27 scheduled classes over the term) were observed, comprising 611 minutes.

By prior agreement with the instructor, observations began in the third week of the winter term of 2006. I recorded classes with a video camera and the documented classroom instruction using the COLT, Part A observation scheme, which allows for coding of teaching and learning activities in real time. More detailed information on the nature of instruction and practice activities was recorded in written field notes and upon review of the videotape. The camera was focused mainly on the instructor, although during pair, group, and individual activities, students were videotaped as well. The aim of the videotaping was to create an overall record of the pedagogical activities, so micro-interactions between students or student and instructor were not targeted.

Recording of talker participants. Recording of talkers also began in the third or fourth week of January 2006, depending on the talker's availability. The first session lasted between 50 to 90 minutes (see Appendix J for the timeline and tasks for each recording session). The individual talkers and I met in a quiet room at the university. The first recording session began with the perusal of the consent form, which all talkers had already been emailed. Once the consent form was signed, the talkers completed a questionnaire about their biographical information and language learning history. Apart from these first two tasks and the final log task in the first recording session, all recording sessions had the same general structure as outlined below.

The recording part of each session consisted of three main tasks: reading sentences aloud, retelling a short film, and telling a personal anecdote. The first task was reading sentences aloud. Talkers were given cards, each with one true-false sentence

typed on it. Talkers were instructed first in writing, then orally, to read aloud all the sentences and to ask me about words which were unfamiliar in meaning or pronunciation; I would then explain or model words, which talkers usually repeated. After reading aloud all the sentences, talkers were instructed in writing and orally that they would read each sentence again while being recorded.

The digital recording was done directly onto a laptop using a Plantronics (DSP-300) head-mounted microphone and CoolEdit, a speech editing software (Johnston, 2000) at a sampling rate of 44.10 kHz and a resolution of 16 bits. Talkers were asked to read sentences at a normal speed. In order to ensure that each sentence was read as an individual item and not as an item in a list, I handed each sentence card to the talkers after a pause. If talkers mis-read, stammered, or stumbled over a word in a sentence, they were asked to repeat the entire sentence. An initial set of five practice sentences familiarized talkers with the task and allowed me to check recording levels. After recording the practice sentences, the entire set of ten trial sentences was recorded. The talkers were then asked to read the ten sentences aloud again. This provided them with another opportunity to produce their best version of the sentences.

The second task was either the recording of a short-film retell or a personal anecdote. Prompts for all speaking tasks are presented in Appendix G. For each session, a warm-up task always preceded the other two extended speech tasks. At every recording session, the order of the anecdote and film tasks was counterbalanced across the talkers in each group.

The warm-up task was similar to the anecdote and film tasks in that talker participants were required to tell a story. The warm-up task was done to familiarize

participants with story-telling tasks so that the participants would understand the form and requirements of the two main tasks. Talkers received a written prompt for the warm-up. They were allowed to think and to make notes for one minute, but were not subsequently allowed to look at the notes while speaking, though they were allowed to keep the prompt. I did not make audible comments during the recording of any of the extended speech tasks.

When talkers did the film retell task, they were given written, then oral instructions that they would watch a short film twice in order to retell the story later. They were not allowed to take notes during the film, but only after the two viewings. After viewing the film twice on a laptop, they were asked whether there was any part of the story they did not understand. If a participant indicated that s/he did not understand or misunderstood the storyline, I discussed it with the participant until it was understood (this happened infrequently for all film retells in all sessions). The talkers were then instructed orally and in writing that they had five minutes to make notes and/or plan the story retell, but that they would not be able to use the notes during the retell. If talkers asked for a vocabulary item, they were instructed to use their own resources to tell the story, and if they asked about the length or amount of detail in the retell, they were instructed to tell the story until they were finished. These last two instructions also held for the personal anecdote task.

When the talkers did the personal anecdote task, they were given written instructions with a prompt to think about a particular type of experience in their lives, with some questions to help them generate a story. They had five minutes to plan or make

notes, but were not allowed to use the notes while recording the anecdote, though they could look at the prompt.

The final part of the first session involved explanation of the English log and the binders provided to talkers. Each binder contained a page with important dates, such as weeks for recording sessions and for log completion. These dates were brought to the attention of participants. The binders also contained a sample log for use of and exposure to English over one day, a listing of the categories and sub-categories for possible activities, and blank logs for one week (Appendix H). I explained the purpose of the English exposure log, and the characteristics of the categories and sub-categories were discussed. The talkers and I looked over the sample language log and discussed some of the entries, going over the instructions provided below the sample. If talkers were willing, they practised entering their previous day's activities into the log. The first session then ended.

As mentioned above, the overall procedure for recording was basically the same at all recording sessions. For the sentence recording, the five practice sentences remained the same across all sessions, but a new set of ten trial sentences were recorded every session. For the recording of the personal anecdotes and film retells, new films and anecdote prompts were used every session, and the order of the two tasks was counterbalanced in each group. However, whatever the order, the first of the two tasks was always preceded by a (new) warm-up task.

One change that took place in the second recording session was the time taken for planning stories. Some talkers were ready to record their stories almost immediately, and some became quite annoyed at having to wait for five minutes when they were ready to

record. Because the extreme irritation of some talkers might have negatively affected their telling of stories, it was decided to allow talkers to record when they said they were ready to record, up to a maximum of five minutes planning time. This protocol was followed for the subsequent recording sessions.

An additional task was added in the third and fourth recording sessions. After the film retell and personal anecdote tasks had been completed, another film retell task was done. The short film seen two sessions before the current session was shown once more, and talkers retold the film after a planning session of between one and five minutes. This repeated retell was done in order to be able to investigate the effects of repetition of the same content on oral performance.

At the end of every recording session, talkers were asked whether they were having any trouble filling in the English log, and were reminded of the schedule for the next log completion week and the next recording session. If I had questions about a log a talker had already handed in, (e.g., missing or unclear information) that log was discussed at the end of the session.

English exposure/use logs. The English logs were scheduled to be completed approximately one week out of every four weeks. The schedule was modified somewhat around the time of Reading Week and final exams to allow talkers time off needed for vacation or study. The schedule for log completion is seen in Table 5. Before the first scheduled day of log completion that week, talkers were sent an email reminding them that they should complete the English log every day for the next seven days. For the next seven days, talkers were sent a daily email encouraging them to complete their language

log for that day. At the end of the seven days, talkers could email their language log for that week to me, or could bring it to the next recording session.

Table 5

Schedule for Log Completion

Dates	Log Entries
Jan. 22–28	every day
Feb. 11--19	every day
Mar. 12–18	every day
Apr. 9–15	every day
Apr. 30–May 6	every day

Phase 2

Transcription and verification of unclear utterances. The stories told by the talker participants would later be heard by listener participants, who would retell the talkers' stories. Therefore, the talkers' stories first needed to be accurately transcribed to allow for comparison with the listeners' retells. All talkers' stories, with the exception of the warm-up stories, were transcribed by me. Some utterances were unclear and could not be accurately transcribed (348 out of 84,110 total words, or 0.4 % of the total words from the 17 participants). Another native speaker of English with 5 years of ESL teaching experience checked 65% of the transcripts against the recordings. The percentage agreement for the transcripts ranged from 96-100%, with most differences centring on

function words which had been repeated ("And, and…). The basis for the subsequent analyses were my transcriptions, along with the talkers' verifications, described below.

Talker participants were scheduled for a verification session, as well as a one-on-one interview and tutorial. Out of the 17 talker participants, 14 took part in the verification session, interview, and tutorial. The three other participants did not meet me; one had health problems, one repeatedly cancelled appointments, and one was no longer in email or phone contact and seemed to have left the country. Only one of these three participants (Hui) had his stories included in the final story set heard by listeners, which is described in the next section. In Hui's selected stories, only two (function) words had been flagged as unclear; this level of accuracy was considered acceptable.

The verification sessions were done only with talkers whose stories contained some words which were not clearly identifiable; these talkers saw transcripts of their stories with the unclear words highlighted, then listened to those sections using CoolEdit 2000 audio editing/playback software and Plantronics (DSP-300) headphones. They were asked to determine what they had said, and the transcript was changed when they proposed a different word than the one in the original transcript. Sometimes repeated listening was necessary, and sometimes talkers were not able to say with confidence what they had said. In that case, the unclear word was left as a question mark and was not included in data analysis. For all samples verified, out of the 348 unclear words, 120 were confirmed, 130 were changed, and 98 remained unclear. The verification sessions lasted between 5 and 50 minutes, depending on the number of unclear utterances for a participant.

One-on-one interviews. The one-on-one interview for the talker participants was usually done the same day as the verification session unless the verification session lasted a long time. The semi-structured interview was intended to provide a richer and more detailed picture of the participants' experience of language learning over the course of the study, both in the communication course and in the wider environment. The semi-structured interview with Michelle, the instructor, was done after the end of the course and the submission of final marks. This interview allowed Michelle to discuss her views on the progression of the course that term and her overall approach to teaching oral communication. Interviews lasted between 10 and 40 minutes. Interview questions were prepared, but other areas were also pursued if they came up in conversation. The interview questions are shown in Appendix K. After the interview with Michelle was transcribed, she reviewed the transcript and added or clarified comments from the original. This revised transcript was then analyzed.

Tutorials. Tutorials for talker participants were also conducted in this phase. They served no data collection purpose, but were part of the compensation offered for participation. The tutorial was usually done the same day as the verification session and the interview. Measures of vocabulary range and fluency (words/min) were calculated for each participant's speech at each recording session, and patterns of non-native production of vocabulary, grammar, and pronunciation were identified. These measures and patterns were shown to each participant, and instances of systematic non-native pronunciation from their recordings were played. The participant and I did exercises and practice activities focussing on several of these patterns, and the participant was given advice on activities and strategies to work on improving these areas. The participant was given a

record on paper or CD of the non-nativelike patterns of production that had been discussed, and often received recordings of their stories and sentences.

Phase 3

Selection of speech samples. Talker participants had recorded speech at four different times, and any changes in their intelligibility and pronunciation needed to be measured. First, speech samples had to be selected for the listening tasks. Three challenges had to be addressed: (a) selecting samples that would be of value for intelligibility and L2 speech research, (b) keeping listening sessions to a reasonable length, (c) grouping samples in ways that allowed parametric statistical analysis. The responses to these three challenges are described below.

The first challenge was that the selected samples needed to be appropriate for measuring intelligibility. The film retells posed a problem in this regard, since listeners could not hear more than one retell of the same film without having previous knowledge of the content, thus possibly leading listeners to understand more and more with each successive retell heard. Each listener group could have heard one retell of each of the four films (total: four retells), but in order for retells from multiple talkers to be heard, a large number of listener groups would be needed. Therefore, the film retell samples were not used for intelligibility measurement in the current study.

The remaining speech samples were the true-false sentences and the personal anecdotes. Many previous studies have made use of isolated sentences for measuring intelligibility of L2 speech (e.g., Bent & Bradlow, 2003; Munro & Derwing, 1995; Wijngaarden, Steeneken, & Houtgast, 2002). However, the intelligibility and pronunciation rating of extended L2 speech has been explored very little to date. In order

to further the state of knowledge in this area of intelligibility and pronunciation research, only the personal anecdotes (and not the isolated sentences) were selected for the listening tasks in the current study. Both the film retells and the isolated sentences will be analyzed in future research.

The second challenge concerned the length of the speech samples. For each of the 17 talker participants, there was a total of 4 unique personal anecdotes. The participants varied in the length of their anecdotes, with some telling anecdotes of over ten minutes and some telling anecdotes of just over a minute. For the treatment group, the total length of personal anecdotes was 96 minutes, and for the control group, 150 minutes. The length of anecdotes for each participant is shown in Appendix L. When the anecdotes from the native speaker group (27 minutes) were added to the total, it was clear that simply listening to all anecdotes from all talkers would have required a substantial amount of time from the listeners. Therefore, it was decided to reduce the number of talkers heard by listeners, at the same time matching the talkers from the treatment and control groups for L1 and length of residence in an English-speaking environment. Flege and Fletcher (1992) suggest that length of residence in an L2 environment has an effect on the pronunciation of adult learners who are within their first year of arrival.

Three talkers each from the treatment and control groups were selected. Their L1s, ages, and length of residence are shown in Table 6. Sometimes, more than one talker from a group could have been matched with a talker from the other group. In that case, a talker with the smaller average length of anecdote was chosen.

Table 6

Selected Non-Native Talkers for Speech Samples

Treatment group				Control group			
Name	L1	Age[a]	Length of residence[b]	Name	L1	Age[a]	Length of residence[b]
Javier (M)	Spanish	26	5	Lupe (M)	Spanish	30	5
Hui (M)	Mandarin	22	16	Ping (F)	Mandarin	29	18
Xiao (F)	Mandarin	29	24	Feng[c] (M)	Mandarin	30	42

Note. [a]in years. [b]in months. [c]This participant had also completed the oral communication course two or more semesters before.

The amount of time required to listen to all the anecdotes of all the selected talkers would still have been considerable (66 minutes from the non-native talkers + 27 minutes from the native speakers). Therefore, only the anecdotes from the first, third, and fourth recording sessions were selected (Time 1, Time 3, and Time 4). This corresponded to near the beginning and end of the oral communication course and four weeks later. The lengths of the selected anecdotes for each selected talker are shown in Table 7. The mean durations of the stories were 2.0, 3.6, and 2.3 minutes for the treatment, control, and native speaker groups, respectively.

Table 7

Length of Personal Anecdotes for Selected Talkers

Group and talker	Length (min)			
	Time 1	Time 3	Time 4	Total
Treatment Group				
Javier	1.48	3.40	2.77	7.65
Hui	0.90	2.83	2.28	6.01
Jiao	1.00	1.67	1.67	4.34
Total	3.38	7.90	6.72	18.00
Control Group				
Lupe	2.20	3.62	2.42	8.24
Ping	1.43	4.62	6.02	12.07
Feng	5.20	1.93	4.48	11.61
Total	8.83	10.17	12.92	31.92
Native Speaker Group				
Aileen	1.61	1.62	1.39	4.62
Brenda	1.95	5.22	3.44	10.61
Daniel	2.62	3.05	3.72	9.39
William	0.82	1.20	1.00	3.02
Total	7.00	11.09	9.55	27.64
Grand Total	19.21	29.16	29.19	77.56

The final challenge related to finding an appropriate grouping of the speech samples for carrying out parametric statistical analyses. Even with the removal of the samples from the second session, the total time of all recordings was over 77 minutes, which was a significant length of time. In addition, each talker was represented three times in three different speech samples. Significant talker familiarity effects on intelligibility have been reported by Bradlow and Bent (2003). This means that later samples from a given talker could be better understood by a given listener than earlier samples, simply through the listener's greater familiarity with that talker's speech. It was therefore important that a listener hear a talker only once.

One possible solution would have been to create semi-random sets of samples, in which one sample from each talker appeared in each set, in random order, but that sample (out of a possible three) was chosen randomly. However, this would have led to a difficulty in parametric statistical analysis. Analysis could not have been done by listener, since each listener would have heard a different sample set, with the same talkers, but different samples in different orders. The listeners thus would not have been comparable. Moreover, analysis could not have been done by talker, since only three talkers from each NN group were represented.

For that reason, it was decided to group the listeners within each listening task by the time of recording of the speech samples, with one listener group hearing all samples from the first recording session, another listener group, from the third session, and the third listener group, from the fourth session. In this way, statistical analysis could be done by listener, since all listeners from one group had heard identical samples. The performance of the listeners could then be compared across listener groups to investigate

intelligibility changes in the treatment and control groups over time, or within listener groups to investigate differences in the treatment and control groups at a given time. In order to control for order effects, listeners in a given group heard the same samples, but in individually randomized order for each listener. Every listener group also heard the same warm-up sample, which was an anecdote recorded by an additional NN talker in the second session, using an additional prompt.

Speech sample processing. The selected anecdotes were edited to remove any extraneous noise at the beginning or end of the recordings, such as any of my instructions. Using CoolEdit 2000, the recordings were normalized for peak intensity to reduce differences in perceived loudness. Two versions of each anecdote were prepared. In the version used for intelligibility measurement, the anecdote was presented in its entirety. In the version used for listeners' subjective evaluations (ratings), the first 20 seconds of each anecdote was excised, because in a pilot study (Kennedy, 2006), it had been shown that listeners rated talkers within the first 20 seconds of hearing their speech. Each anecdote and 20-second excerpt was bookended by a brief recording of the order of the story (e.g., "Story One") at the beginning, and a brief recording signalling the end ("Stop") at the end.

Listening tasks. After signing a consent form (Appendix M) , listeners completed a questionnaire for contact information, language learning history, and exposure to non-native accent (Appendix D). The listener groups were composed according to listeners' self-reported frequency of exposure to non-native, accented English. Participants with the ratings *never, rarely,* and *occasionally* were equally distributed in groups with participants with the ratings *fairly frequently, frequently,* and *very often.*

In listening to the full anecdotes, listeners did one of two tasks measuring L2 intelligibility. As mentioned above, intelligibility of L2 speech has often been measured using word or sentence identification or transcription. Because the speech samples chosen were anecdotes, identification or transcription were not practical tasks because the speech was extended.

Answering comprehension questions was a possible task, but because all the anecdotes were unique, it was not clear whether it was possible to design questions for each anecdote which would carry the same weight across stories. For example, anecdotes had different numbers and types of characters and key events. An anecdote with a hypothetical element, when the talker discussed possible but unreal events, would require different sorts of questions than an anecdote with events taking place only in reality. It was not clear how to develop criteria which would allow valid and reliable scoring of answers to different questions from different anecdotes.

Therefore, story retells were chosen as the first task to measure intelligibility at the levels of discourse, sentences, and words. Retells could be analyzed for the extent to which they faithfully reflected a given anecdote, whether that story had, for example, hypothetical events or not. In order to make sense of an anecdote so as to retell it, a listener could use several different levels of analysis. If a listener found a word or words to be unclear, the listener could use the surrounding context of the sentence to try to figure out the words, and similarly, if a listener found a sentence or a particular event to be unclear, the listener could make use of the greater story context, and possibly the following events, to understand or speculate on the unclear sentence or event.

With these sources available to understand an anecdote, it could be that a talker would pronounce some words in English in a very non-nativelike way, but the listener could still understand the larger anecdote from the larger sentence or story context. Therefore, in order to also investigate the intelligibility of talkers' *words*, the second listening task measured on-line (immediate) word intelligibility; listeners heard the same anecdote, stopping the recording when a word was unclear. These two tasks provided a means of investigating how word-level unintelligibility is related to higher-level unintelligibility.

Story retells. Listeners were given written, then oral instructions that they would hear one practice story and ten other stories, and retell each story in turn. They were instructed to take detailed notes while they were listening, to pause the recording only when a story was finished, to look over their notes, then to record the story using their notes. They received explicit instructions to state when they did not understand something in the story. The instructions also told listeners of the overall topic for the stories. See Appendix N for story retell instructions.

The stories were recorded onto CDs as audio files, with each CD containing stories from one of the three recording sessions. In a quiet room, each listener played their own individually randomized CD on a JVC stereo CD player using Nexxtech stereo headphones. The practice story was used to familiarize listeners with the task and to adjust the volume. Listeners took notes on blank sheets provided to them, and recorded their retells with an Olympus DS-2 digital voice recorder with a Sony ECM-T2 lapel microphone. After completing the warm-up task, I answered any questions about the task requirements and adjusted the volume, if necessary.

The story retell part of the listening session lasted between 40 and 75 minutes, depending on the set of stories heard and the speed of the listener in retelling. After completion, the listener's written notes and recorded story retells were collected.

On-line word intelligibility task. Listeners were given written, then oral instructions that they would hear one practice story and ten other stories, and should pause the recordings when they did not understand or were not sure of a word. They should then talk about what the word might be. They were also instructed that if they unpaused the recording, then later understood a word they had not understood before, they should again pause the recording, identify the word, and talk about how they had come to understand it. It was emphasized that the listeners were not judging whether the talker told a good story or whether they liked the story; the listeners were only to stop the recording when they didn't understand a word. See Appendix N for instructions for the on-line word intelligibility task.

The participants did this task together with me. The recordings were on a laptop computer as audio files, and each listener heard an individually randomized order of stories from one of three recording sessions. The stories were played on SoundScriber (Breck, 1998), an audio transcriber application with time display. The files were played at normal speed, but rewound two seconds after pausing and unpausing. This was to ensure that listeners with slow reaction times who paused a recording would be able to resume listening at the point at which they did not understand. Listeners had control of the mouse and listened to the stories with Nexxtech stereo headphones. I also listened in with other headphones to determine the context of the unintelligible word. The listener recorded

comments with an Olympus DS-2 digital voice recorder, with a Sony ECM-T2 lapel microphone.

When the listener paused a recording, I would note the time of the pause and the listener's comments on a sheet in order to have a written record, in addition to the audio recording, of the listener's comments. The on-line word intelligibility part of the listening session lasted between 20 and 45 minutes, depending on the recording session heard and the frequency of the listener's pauses. After completion, I collected the listener's recorded comments and my notes.

Ratings (listener evaluation). After listening to the full stories and doing one of the two tasks, all listeners then did a rating task to determine how they evaluated the talkers. Listeners were given written and oral instructions that they would hear one practice excerpt and ten other excerpts from the same ten anecdotes heard previously; they would then rate the talkers on 9-point scales for comprehensibility, accentedness, and fluency. Instructions and example scales for the ratings task are seen in Appendix I. Descriptors for each endpoint were given, and fluency was further described as the smoothness and fluidity of a talker's speech, with descriptions of some characteristics of dysfluency, such as pauses, incomplete words, and word repetitions. It was emphasized that ratings on all three scales had to be made for each excerpt.

Listeners from the story retell task heard the 20-second excerpts as audio files played from CD on a JVC stereo CD player and Nexxtech stereo headphones. Listeners from the on-line word intelligibility task heard the excerpts as audio files on a laptop with Nexxtech stereo headphones, played from CD on Windows Media Player. At the end of every excerpt, which was immediately followed by "Stop", both sets of listeners paused

the player to rate. The rating task lasted between 5 and 10 minutes, depending on the time the listener took to rate. After completion of the rating, the rating sheets were collected.

Data Analysis

Classroom Observation

Not all categories in the COLT, Part A observation scheme (Spada & Fröhlich, 1995) were useful for the purposes of this study. The categories of interest were those that allowed a descriptive summary of the activities seen, their focus, and their length, as well as a record of macro-patterns of classroom interaction. These categories were: Activities and Episodes, Participant Organization, and Content. A category which was not used, for example, was Materials Source. Field notes and review of the videotape were also used to complete the summary. The summary provided the date, time, and focus, and content of the classroom activities, along with a brief description of how the activities unfolded and the patterns of interaction observed between students and teacher. Each class lasted 90 minutes. The full descriptive summary of all eight classes can be seen in Appendix O.

English Exposure/Use Log

Because only two of the eight participants from the treatment group completed logs for the final log period, logs from the final (fifth) log period were not analyzed. Participants who completed only one or two logs overall (one from the treatment group and two from the control group) were removed from the analyses in order that those participants would not have an undue influence on the overall numbers. Each of the remaining 14 participants' English logs was first examined by log period (each period

was seven days). For each log period, the amount of time (in minutes) reported for each category was tabulated. Then, for each participant, the total minutes of exposure overall and the total minutes of exposure in each category of activity overall were tallied for all the logs submitted by that participant. The logs of all the participants within each NN group were combined, and the activities which involved only a little or some English (1 or 2 in the English Use/Exposure column) were removed. The language use/exposure reported for the remaining activities was mostly or wholly in English.

For each of the two NN groups, the number of minutes overall and for each category was tabulated, first for each log period, then for the total number of logs. For each log period, the mean minutes per talker were also calculated. As well, for every log period, the percentage of minutes for each category was calculated relative to the total minutes for that log period. Finally, selected sub-categories were classified into four areas: interactive social, interactive academic, non-interactive social, and non-interactive academic (see Appendix P). For each log period, the percentage of time for each of these areas was calculated relative to the total time for that log period.

Recordings

Two listening tasks needed analysis: the story retells and the on-line word intelligibility task. The analysis of both tasks will be described in turn.

Story Retells: Analysis Schemes

A story can be analyzed in different ways to explore how intelligible it was to the listener, such as with multiple-choice questions, (Smith, 1992), comprehension questions (Matsuura, Chiba, & Fujieda, 1999), or listeners' ratings (Derwing, Munro, & Wiebe, 1998). The story retells in the current study were analyzed in three basic areas: (a)

accurate content, (b) inaccurate content, and (c) listeners' statements of lack of understanding. The first step was to divide the original stories and retells into manageable units for coding. The narratives were first divided syntactically into clauses, using Berman and Slobin's (1994, p. 660) protocol for clausal analysis. They were then categorized semantically, using a combination of story elements from three different narrative analysis schemes modified by me. The three schemes, described below, used a story grammar approach to identify units for narrative analysis.

A story grammar is "a grammar which attempts to represent a processor's internal organization of story material" (Stein & Glenn, 1979, p. 56). It is made up of different functional categories, here referred to as semantic story elements, classifying the types of information which occur in stories, such as setting, outcome, and evaluation. These story elements are further described in Table 8. Since the early 1970s, researchers have modified and extended story grammars from previous research to address perceived shortcomings and to focus on the researchers' own interests. Therefore, the three narrative schemes which were modified also owe much of their content and organization to previous research. The three schemes are from Labov (1972), van Dijk (1976), and Trabasso, van den Broek and Suh (1989). Labov's (1972) narrative analysis scheme was generated during a study targeting black English vernacular in New York City. Spontaneous stories told by speakers of African-American Vernacular English were analyzed for their overall structure. Van Dijk (1976) posited that there was a "logic of action" reflected in narrative discourse, especially in simple narratives like fairy tales, which could be analyzed structurally using a rule-based system. The story elements from

Trabasso et al.(1989) were part of a theoretical framework for goal-directed narratives (also usually fairy tales) which contained causally-related story elements

Because each analysis scheme was generated for a different purpose, none of the schemes was suitable on its own to analyze the talker participants' stories. The talker participants' stories were often much more variable in their narrative structure than the stories from Van Dijk's and Trabasso et al.'s schemes especially, so the selection of particular semantic story elements from the three schemes was based on how well the elements "fit" the stories of the talker participants. Table 8 shows the semantic story elements used from the three schemes, with duplicate elements also identified.

Table 8

Elements in Story Narration in Three Analysis Schemes

	Analysis scheme		
Story element	Labov (1972)	Trabasso et al. (1989)	van Dijk (1976)
Abstract	Abstract:[a] a short summary of the story		Abstract: a summary of the narrative
Setting	Orientation: identification of the time, place, persons, and their activity or situation.	Setting (S):[a] introduces characters, gives relevant characteristics of persons and circumstances; establishes "causal field" of story.	Initial state: description of the initial circumstances of a reported course of events.
Action	Complicating action: main body of narrative clauses.	Initiating event (IE):[a] character undergoes change in state; this event determines the main causal chain of states and actions in a story. Internal response (IR):[a] character's feelings, cognitions, beliefs in response to IE, A, or O.	Basic action: event which is intentional, under control of actor, and purposeful.

Table 8 (*continued*).

Story element	Labov (1972)	Trabasso et al. (1989)	van Dijk (1976)
		Goal (G):[a] character wants something that will resolve the problem; either state, activity, or object. Attempt (A):[a] tries at attaining the goal.	
Other events			Complicating event (CE):[a] unexpected, unnormal event that leads to a state contradictory to the purposes of a main character.
Outcome	Result: the end of that series of events.	Outcome (O):[a] result, positive or negative, in attaining goal.	Resolution: success or failure of character in resolving problem.
Evaluation	Evaluation (E):[a] narrator's judgments, beliefs, emotions about events, characters, or circumstances in story.		Evaluation: narrator expresses attitudes with respect to reported course of events.
Moral			Moral: consequences of reported course of events on future actions of agent/narrator or hearer.

Note. [a]Elements chosen for the modified analysis scheme.

Table 9 shows a sample division of an original narrative into clauses and story elements.

Table 9

Coding of Semantic Story Elements

Clause	Story element
Well, when I, when I first arrived in Montreal last year in 2006, in 2004.	setting
My friend, my friend told me...	initiating event
there was a holiday named Boxing Day.	
And on that day everything, everything is going to be cheaper than, than they were before.	setting
So I had a plan to buy a laptop two months before Boxing Day.	goal
And on that day, I went to the Future Shop located in Angrignon...	attempt
and I waited outside to buy the laptop.	
I, I remembered I spend one thousand two hundred dollars on it...	outcome
which is three hundred dollars less than its original price.	
I think it, it was really a bargain.	evaluation

As is seen in Table 9, more than one clause could be attached to one semantic story element. The opposite was also true: for some clauses or sentences, some content in

the clause or sentence could be classified as one semantic story element, while other content in the same clause or sentence could be classified as another story element.

Analysis procedure. There were three steps to analyzing the original stories and retells: (a) analysis of the original stories, (b) analysis of the listener retells, and (c) analysis of the similarities and differences between them.

Analysis of original stories. The first set of original stories to be classified syntactically and semantically was the set of stories at Time 4 (chosen randomly from the three sets of recordings), from approximately four weeks after the end of the oral communication course. Each story was first divided into clauses, then into semantic story elements. In the course of analysis, some clauses in the stories were found not to fit into any of the pre-existing semantic story element categories. Therefore, I created another semantic story element category, which was added to the overall scheme. This element was called "Discourse" and referred to elements of the story that served to signal its narrative structure. For example, Javier, describing a decision to change his academic concentration, asked the rhetorical question, "What were the reasons?" He continued, "The reasons were..." This rhetorical question identified the following content as the motivation for his action, and was thus classified as a discourse element.

Analysis of listener retells. The story retells from Time 4 were also divided into clauses, then semantic story elements. The semantic analysis of the original stories and retells from Time 4 was recursive; the criteria for identifying various semantic story elements were fine-tuned over time. Therefore, after all the stories and retells from Time 4 had been analyzed, they were re-checked to see that the analyses of all the stories and retells were consistent. Some clauses in the retells could not be classified under the

existing scheme of semantic story elements. These unclassified elements were addressed in the next step, that of comparing the original stories and retells.

Both original stories and retells from Time 4 had been divided into clauses and classified for semantic story elements. Beginning from number 1, each semantic element in each original story was given a number. In some cases, one or two semantic story elements were made up of multiple clauses or sentences which were related to each other, but the clauses or sentences had quite separate pieces of information. These clauses or sentences were given sublettering, for example, 1a, 1b, 1c. The semantic story element "outcome" in Hui's story in Table 9 is an example of an element which would receive sublettering for its two clauses (e.g., I remembered I spend one thousand two hundred dollars on it [a] which is three hundred dollars less than its original price [b]).

Analysis of similarities and differences. As described above, the retell analysis targeted three aspects: the accurate content (AC), the inaccurate content (IC), and listeners' statements of lack of understanding (LU). To determine the first two aspects, each retell was compared to its original story to see which semantic content from the original story appeared in the retell. Retell elements with content derived from the original were given a number matching that of the original element with the same content, and further examined for the completeness and accuracy of the content.

The analytical categories for the completeness and accuracy of semantic elements of a retell came out of the listener retell data. Table 10 shows the analytical categories for the completeness and accuracy of semantic elements of a retell. Examples of this coding are shown in Table 11.

Table 10

Analytical Categories for Completeness and Accuracy of Retell Elements

Category	Criteria
Y	(Yes) Content of retell element retold fully and accurately from the original
RE	(Repeated) Second or third appearance in a retell of the same semantic content
PM	(Part Missing) Retell element missing a major part of the original content
DM	(Detail Missing) Retell element missing a minor part of the original content
RM	(Relation Missing) Retell element complete in itself, but missing connection to other retell element
W	(Wrong) Retell element with quite inaccurate semantic content
PW	(Partly Wrong) Retell element with some inaccurate semantic content
DW	(Detail Wrong) Retell element with slightly inaccurate semantic content
RW	(Relation Wrong) Retell element accurate in itself, but connection to other element is inaccurate

Table 11 shows the completeness and accuracy coding of one listener's retell of the story illustrated above in Table 9. Some retell elements had both missing and inaccurate semantic content. This caused a problem because the retell intelligibility measures were calculated as ratios, with story elements as the unit of analysis (this is described in the Equalizing Raw Numbers section below). Coding retell elements for

both missing and inaccurate content would, in effect, give those elements twice the weight of retell elements which had *only* missing *or* inaccurate content. That is, a retell element with both missing and inaccurate content would be counted twice. In order to avoid this, those retell elements were coded for inaccuracy alone, not missing content. Semantic story elements from the original which were completely missing in a retell did not receive any code, but were listed by their original story number under the heading "Missing" at the end of each retell analysis. Besides the retold semantic elements, listeners made other comments about the original stories. Coding for these elements is shown in Table 12.

Table 11

Sample Coding of Completeness and Accuracy of Retells

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
1	Well, when I, when I first arrived in Montreal last year in 2006, in 2004.	setting	1	This person first arrived in Montreal in 2004.	Y	setting
2	My friend, my friend told me… there was a holiday named Boxing Day.	initiating event	2	A friend told him about a holiday named Boxing Day.	Y	initiating event
3	And on that day everything, everything is going to be cheaper than, than they were before.	setting	3	Where everything is cheaper.	Y	setting
4	So I had a plan to buy a laptop two months before Boxing Day.	goal	4	And since he wanted a, a laptop two months before that day.	Y	goal
5a	And on that day, I went to the Future Shop located in Angrignon…	attempt	5a	On that day he, he went to a store…	DM	attempt
5b	and I waited outside to buy the laptop			I couldn't understand where it was.	DU	
			5b	He said he waited…	DM	attempt

Table 11 (*continued*).

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
6a	I, I remembered I spend one thousand two hundred dollars on it	outcome	6a	and he bought a laptop for thirteen hundred dollars…	DW	outcome
6b	which is three hundred dollars less than its original price		6b	and it was two hundred dollars less than the original price.	DW	
7	I think it, it was really a bargain	evaluation	7	And he thinks that it was a really good bargain.	Y	evaluation

Table 12

Coding of Additional Retell Elements Based on Listener Comments

Code	Criteria
DU	Didn't understand some part of the story.
HU	Hard to understand some part of the story.
DUS	Didn't understand significance of some part of the story.
TS	"I think s/he said..." or something similar.
SL	"It sounds like..." or something similar.
AS	Assuming the occurrence of state/event unmentioned in original.
IN	Interpreting the meaning of state/event in story.
AD	Giving additional information not included in original.
DS	Mentioning areas where talker did not provide information.
DIS	Giving discourse information.
EV	Approving of, disapproving of, or otherwise judging state/event in story.

These codes could co-occur with codes in Table 10. For example, a retell element could have semantic content that corresponded completely and accurately to an original element, thus being coded with a Y. However, if the listener said that s/he had trouble understanding or s/he thought the talker said this, the retell element could also be coded HU or TS at the same time. Appendix Q provides examples of coded story elements for

both talkers and listeners. The same steps for coding originals and retells from Time 4 were followed for coding originals and retells from Times 1 and 3.

Normalizing raw numbers. For each retell of each listener, the number of the complete and accurate retell elements (Y) was tabulated, as well as the number of other types of retell elements (e.g., the number of DW elements). However, although all talkers were given the same prompts, the stories that the talkers told were different, particularly with regard to length.

Talker A's original story might have a much greater number of original story elements than Talker B's story (Talker A's story would thus be longer). Therefore, a listener retelling the two talkers' stories would likely retell a greater number of elements for Talker A's story than for Talker B's story. If some elements in the listener's retell of each story had inaccurate or missing content (e.g., DW), there might be more of those types of retell elements for Talker A's story, simply because Talker A's story had a greater number of original elements than Talker B's story.

Therefore, the number of each type of retell element for a given retell was divided by the total number of elements in the original story, yielding ratios of retell elements to the total elements in the original story (presented as percentages). This calculation was used to normalize the raw numbers of elements for stories of different lengths. An example is shown in Table 13.

Table 13

Sample Normalization Calculation for Retell Elements

	Number of story elements		
Talker	Original story	Retold story[a]	Normalization formula
Lupe	23	2	(2/23)*100% = 8.70%
Javier	26	1	(1/26)*100% = 3.85%

Note. [a]Number of "W" retell elements in each retell.

Reliability of retell analyses. After the original stories and retells were coded and analyzed, a second coder also analyzed a sample of 10% of the original stories and retells. One talker from each of the three talker groups was randomly selected, and transcripts from each of the three talkers' original stories from Time 4, along with all retells of that story, were provided to the second coder. Time 4 was chosen because the second coder was trained with some other stories and retells from Times 1 and 3. The one-on-one training was done over three one-hour sessions, with the second coder doing additional practice coding on her own.

After becoming familiar with the categories for semantic story elements and for the accuracy and completeness of the retell elements, the second coder re-analyzed the original stories and retells, then checked her analyses against the original analyses. Reliability of the analyses was calculated through percentage agreement of the two sets of coding. Results are shown in Table 14. The greatest disagreement was over the coding of semantic story elements, which ranged between 81-100% for the original stories and 86-100% for the retells. However, these categories were not used in the calculation of the intelligibility ratios. The intelligibility ratios were calculated from the coding of the

accuracy and completeness of the retell elements, and the agreement between coders in the coding of these elements was very high, between 99-100%. The basis for the calculation of intelligibility ratios was thus my coding of the original stories and retells.

Table 14

Percent Agreement of Coding for Original Story and Retell Samples

	Coding type	
	Semantic story elements	Accuracy and completeness
Original stories		
Treatment	81.2	
Control	100.0	
Native speaker	91.7	
Retells		
Treatment	85.7	100.0
Control	94.8	98.8
Native speaker	92.5	99.5

Note. Because accuracy and completeness coding was defined in relation to the original stories, this coding applied only to the retells.

On-Line Word Intelligibility Task

In this task, inaccurate content (IC) and listeners' statements of lack of understanding (LU) were also analyzed through the intelligibility of words. Listeners paused the recording whenever there was a word they did not understand or were not sure

of, or when they understood a word they had not understood before. The analytical categories for the comments made by the listeners emerged from the data. The categories for on-line processing are shown in Table 15.

Table 15

Listener Comment Categories from the On-line Word Intelligibility Task

Category code	Criteria
AR	Listener proposes correct word
AW	Listener proposes incorrect word
SL	Listener says, "It sounds like..." or something similar
EO	Listener gives two possible alternatives for the word
DK	Listener doesn't know what word could be
GI2	Listener didn't understand word at moment of pausing recording, but understood word an instant later (before unpausing)

As with the retells, some categories could co-occur with each other for the same word or in the same pause. For example, if a listener said, "It sounds like..." and the proposed word was correct, the comment would be coded as both SL and AR. If a listener paused and proposed a stretch of words, a correct suggestion for a word in that stretch would be coded AR, while an incorrect suggestion for another word in that stretch would be coded AW.

For each playing of a story which was paused by a listener, the following data were collected: (a) the points at which the recording was paused, (b) content of comments, (c) suggested words, and (d) target words. In most cases, the word or words targeted by the listener could be identified based on the recorded times of pauses and the words the listener remembered hearing before the unknown word/s. However, sometimes it was not possible to identify which word/s the listener did not understand, especially if the listener could not venture a guess and could not remember the words they had heard just before even when prompted by me. Sometimes listeners heard some stories and did not pause the recording at all, so there were no comments to analyze for those stories heard by those listeners.

The comment categories which were analyzed were those that related to words or stretches of speech which were not simply difficult to understand at the time of listening, but actually still unintelligible to the listener by the end of the story. This means that the number of instances when a listener almost immediately understood a word after pausing (GI2), gave two possibilities, one of which was correct (EO), or did not understand a word but later identified it correctly (AR or SL) were not tabulated in the analysis. The categories tabulated were those when listeners made no guesses at all (DK), made incorrect guesses (AW or SL) which were not later corrected, or proposed two possible words, both of which were incorrect (EO) and were not later corrected.

Normalizing raw numbers. The issue of story length again appeared with the on-line intelligibility task. The more words in a story, the more possibilities a listener had to not understand a word. The calculation performed was similar to that for the retell task. Since the unit of analysis in this task was misunderstood words, the number of words in

each original story was counted. For each listener hearing a given story, the number of instances of each tabulated comment category was divided by the number of total words in the original story, yielding ratios of unintelligible words to the total number of words (presented as percentages). An example is shown in Table 16.

Table 16

Sample Normalization Calculation for On-Line Comment Categories

	Number of words		
Talker	Original story	Retold story[a]	Normalization formula
Lupe	323	1	1/323*100%= .31%
Javier	337	3	3/337*100% = .89%

Note. [a]Number of "DK" words in each retell.

Interviews

The interviews were first transcribed verbatim. Each transcript was then explored for various themes that emerged in the interview. During a first pass-through, notes were made in the margins of the transcript broadly categorizing the topic(s) related to language learning, teaching, and communication which were covered in that part of the transcript. The transcript was then read through again to refine or modify the categories for the topics, if necessary. The refined topic categories were placed as theme headings in a Microsoft Word file, with various subheadings added. The subheadings covered specific information or quotations which were both part of the overall theme and closely related to each other in content. For example, in the instructor's interview, the theme of Teaching

Methodology emerged. One subheading of this theme was Desired Future Modifications, while another subheading was Feedback. Line numbers showing the location of the information or quotation in the transcript were also included.

Sometimes, specific information was included in more than one theme. For example, the instructor's wish to improve students' access to technology was put under the theme of Technology and the theme of Course Design. After each transcript had been analyzed and relevant information placed in files under headings and subheadings, all the files from the talkers' interviews were read through again, and headings and subheadings were harmonized so that different headings and sub-headings were not used for similar themes across the different files.

Summary

In this chapter, I detailed the recruitment and characteristics of the talker and listener participants. I described the selection and development of materials, which had five main purposes: to collect background information on participants, to record the classroom training received by the treatment group, to elicit speech from talker participants, to document talker participants' exposure to English, and to measure L2 speech intelligibility and evaluation by listener participants. Then, I outlined the three main phases of data collection: observing classes and recording talker's speech and English exposure and use, verifying transcripts and conducting interviews, and collecting intelligibility and evaluation data from listeners. I described the methods by which the data were analyzed, as well as the narrative analysis scheme used to segment and analyze original stories and retells. In the following chapter, the results of the data analyses are presented.

CHAPTER 4: RESULTS

The results to be discussed pertain to four areas, relating to the four research questions:

1. the instruction received by the treatment group, as well as the instructor's views on the course (Research Question 1),
2. NN graduate students' development of L2 English speech (Research Question 2),
3. the students' use of and exposure to English (Research Question 3),
4. the students' views on their language learning and language use (Research Question 4).

Nature of Classroom Instruction: Overview

The first research question was the following: What is the nature of classroom instruction for NN graduate students in an ESL oral communication course emphasizing intelligibility?

The stated focus of the oral communication course was on "developing pronunciation and communication skills, including aspects of pronunciation that most affect intelligibility, and with verbal and non-verbal techniques for effective presentations" (Graduate Courses, para. 2). The course was roughly divided into two parts: in the first part, for six observed classes, the focus of instruction was on suprasegmental aspects of English pronunciation. In the second part, for two observed classes, the format and content of presentations, as well as strategies for giving

presentations, was the focus. To describe the nature of classroom instruction, first the topics of instruction, general patterns of interaction, and instructional focuses will be depicted. Then, typical learning activities and materials will be described, followed by a report of an extensive interview with the instructor. A summary of the predominant patterns of interaction, instructional focuses, and learning activities over the eight observed classes is seen in Appendix O.

Topics of Instruction

Because the class was observed once a week, but met twice a week, not all of the classes were observed. Therefore, a description of the topics of instruction in the course may leave out some topics which were covered during the other classes. Nevertheless, it is possible to establish aspects of oral communication which were the focus of instruction, and how much time was spent on one aspect compared to another. The terminology used for these aspects is the same terminology used by the course instructor. Table 17 shows the aspects covered and the percentage of instructional time spent on each aspect over the entire eight classes observed.

It is clear that students were instructed in many different facets of oral communication in these classes. Moreover, in these eight classes at least, the majority of instructional time was spent on both prosody- and fluency-based aspects of suprasegmental pronunciation (Trofimovich & Baker, 2007), such as lexical stress, linking, and pausing. This focus on suprasegmentals is in line with many current pronunciation and speaking courses in English as a second language, where suprasegmental aspects are often highlighted. Students were also taught some strategies for the genre of oral presentations, such as periodically looking up while reading a text,

checking the audience's comprehension, and using effective body language. Finally, the discourse structure of oral presentations was explored, for example, the rhetorical organization of a presentation's conclusion.

Table 17

Aspects of Oral Communication Covered in Class

Aspect	Percent of time spent over 8 classes
Lexical stress	36.0
Linking	17.0
Focus words (phrasal/sentence stress)	13.3
Pausing/thought groups	7.2
-ed endings	6.2
Reading while looking up	6.0
Effective body language in presentations	5.2
-s/es endings	4.4
Presentation conclusions	4.1
Comprehension checks	2.8
Problem-type presentation structure	2.6
Total	104.8

Note. Total is 104.8% due to activities with a combined focus.

Patterns of Interaction

Part A of the COLT observation scheme includes a category called "participant organization". It allows the observer to code classroom interaction as teacher-centred or student-centred. The sub-categories used in this analysis were T-C, S-C, Group, and Indiv. The first two of these sub-categories describe whole-class interactions between a teacher and the class, or between a student and the class. The last two sub-categories describe interactions between students in groups or individual work by students. Over the eight classes, the ratio of different patterns of interaction was fairly consistent. Forty-five to 65 % of class time (a mean of 58 %) was spent with the instructor interacting with the whole class (T-C), with the rest of the time spent almost entirely in group interaction. The ratio of teacher-centred instruction decreased in later classes from a high of 65% to a low of 49%, but swung up again as the time for final presentations grew near. Please see Appendix O for the ratios of interaction patterns for each class and in total.

Focus of Classroom Talk and Activities

Part A of the COLT scheme also includes a category in which the classroom talk or activities are classified as focussing either on classroom management or on language. The two management sub-categories are Procedure (procedural directives) and Discipline (disciplinary statements), and the four language sub-categories are Form (formal aspects of language, such as grammar, vocabulary, or pronunciation), Function (communicative acts such as requesting or apologizing), Discourse (the cohesive and coherent structure of language above sentence level), and Sociolinguistics (forms or styles appropriate to particular contexts).

In the first six classes observed, the majority of time spent in instruction and activities (85% or more) was categorized under the Form sub-category, with time spent on Procedure coming second. Instruction and activities in the first six classes centred almost entirely on suprasegmental aspects of pronunciation. However, in the seventh class, when the final presentation was the focus, instruction and activities classified under the Function and Discourse subcategories made up the majority of class time (80%). In the eighth and final class, the majority of time (76%) was again spent on instruction and activities classified as Form, with some time also spent focussing on Discourse and Procedure. Please see Appendix O for the ratios of content focus for each class and in total.

Classroom Instruction and Activities

Apart from the general patterns of interaction and content of the course, what did a typical class look like? What sorts of activities and materials were used? The main course text was *Accurate English* (Dauer, 1993), a pronunciation textbook aimed at intermediate- to advanced-level English learners. This book contains numerous explanations and exercises (usually at sentence-level, but with some paragraphs and dialogues) for both segmental and suprasegmental aspects of English pronunciation. There are also a number of graphic illustrations of tongue and mouth positions; some chapters include instructions for preparing a short oral presentation while practising an aspect of pronunciation. This text was used for both in-class and out-of-class work.

When focussing on a certain aspect of pronunciation in class, Michelle, the instructor, usually started with awareness-raising and hypothesis-testing, followed by practice. The students would be expected to have read the relevant pages in the textbook

describing that aspect of pronunciation. In class, Michelle would usually begin work on that aspect with written words or sentences on the board, in the text, or in a handout. The written models would be read aloud by Michelle or by students. Often, Michelle would use leading questions to elicit from the students characteristics or patterns of the aspect of pronunciation; she would sometimes introduce topics by having students do basic tasks (for example, counting syllables in words with and without –ed endings) which would uncover patterns of pronunciation.

After the particular pattern had been elicited, Michelle discussed it with the class. Often she modelled some example words or sentences again, sometimes with visual or aural aids. For example, to demonstrate the length of a stressed syllable, she stretched a rubber band, and she used a metronome to show how various utterances of different lengths but the same number of focus words could be said in approximately the same number of beats.

Following the awareness-raising, the students completed written exercises, which were almost always done in pairs or small groups. The exercises usually consisted of words or sentences which targeted a particular aspect of pronunciation. Students took turns reading aloud the words or sentences, sometimes first marking them to show specific suprasegmental patterns. Students were encouraged to give each other feedback on their oral production; the instructor also circulated around the class, answering questions and providing feedback. The exercises were then partially taken up as a class, with the instructor again answering questions and giving feedback on students' production. Feedback was targeted towards the suprasegmental patterns which had been

covered, although in the last two classes observed, a few instances of feedback on segmental production were seen.

When focussing on communication strategies (e.g., comprehension checks) for presentations, Michelle would often begin by eliciting and/or describing and modelling the strategy, and would then discuss the reason for using the strategy in oral presentations. When there was more than one way of engaging in a strategy, such as using different kinds of comprehension checks or body language, she would discuss with the class when particular techniques might be appropriate or inappropriate. The students would then usually complete a worksheet requiring them to analyze the suitability of certain techniques, and the responses would be discussed by the whole class. Sometimes, the students would then practise using the techniques in groups. Instruction on the discourse structure of presentations was much the same, with students completing a worksheet before or after discussing the nature of the particular element, such as conclusions.

Evaluation

The students were evaluated on the suprasegmental pronunciation patterns and the presentation strategies and structures they had been taught. Michelle provided students with explicit guidelines and evaluation rubrics long before the evaluations, which usually took the form of recorded readings or recorded presentations in front of the class. Students were graded not only on how they orally produced the suprasegmental patterns and used the presentation strategies, but also on how they annotated the written texts (readings and presentation scripts) to predict, according to the guidelines they had learned, how suprasegmental patterns would occur in the texts. Students also had to

complete self-evaluation sheets after watching their video-recorded presentations, grading themselves on particular aspects of pronunciation and presentation skills, and identifying strengths and weaknesses. This self-evaluation targeted their skills at monitoring their own speech and body language.

Computer-Aided Learning

Although the focus of the observation was instruction in the classroom, it is important to mention that the oral communication course included extensive optional exercises available in a language laboratory. The language laboratory was located in another building less than five minutes' walk from the classroom, and was accessible only to students taking for-credit language courses. The optional exercises were delivered through the CAN-8 software system, an application which allowed students to read and listen to recorded sentences and texts, and to record and listen to themselves. Exercises were directed towards the aspects of oral communication covered in the course. Course assignments were also recorded by students on CAN-8 for evaluation by Michelle.

Summary: Nature of Classroom Instruction

In general, the topics covered during the observed class sessions reflected current practice in pronunciation teaching, focusing largely on suprasegmental aspects of pronunciation. The work on presentations provided students the opportunity to practise using the patterns of suprasegmental pronunciation they had learned, through a speech genre which was authentic and important in academic communication. Most of the time, instruction and instructional activities centred on form-related aspects of language, suprasegmental aspects in particular. Michelle usually tried to have students analyze modelled words or sentences in order to come up with patterns of suprasegmental

pronunciation, but communication strategies and structures of presentations were often explained before any analysis. Analysis of language and rhetorical patterns was prominent in classroom activities. Between a third to a half of class time was given over to group work. During that time, students almost always worked on word- or sentence-level exercises from the textbook, receiving feedback from the teacher and other students.

Instructor Interview

Following the end of the course and after the grades had been submitted, Michelle participated in a semi-structured interview with me to discuss the course and Michelle's thoughts about teaching and learning oral communication. The interview was tape-recorded in a quiet room and lasted about one hour. The discussion covered the goals, design, and content of the course, some of the teaching and learning activities done, and more general thoughts on conditions for teaching and learning. Six important themes surfaced in the analysis of the interview: goals for students, course design, course content, methodology, technology, and learning conditions. These themes will be discussed in turn.

Goals for Students

Michelle mentioned several objectives which she hoped the students would attain by taking the course. The overall goal was for graduate students to become more intelligible when doing their academic work. However, Michelle was emphatic that "there's no way that in three months, students are going to markedly improve their communication." Therefore, the intermediate goals were for students to become aware of nativelike patterns of suprasegmentals in English, and to learn about and use independent

learning strategies, such as self-monitoring, so that they could later make use of the knowledge they had acquired about suprasegmentals. The instructor emphasized that students needed to start paying close attention to how native speakers use suprasegmentals, especially rhythm.

Course Design

Michelle was the sole instructor for the course, and she had put much time and effort into re-designing the course to improve its effectiveness. She had not designed the course from the ground up, but had begun by teaching one section of an already-existing course, while another instructor who was experienced with the course taught the other section. After one semester of teaching, Michelle suggested various changes to the course; one change was replacing course content targeted towards international teaching assistants (e.g., communicating in tutorials) with course content targeted towards all graduate students (e.g., giving presentations). This was because few of the students in the course were actually teaching undergraduate students. Another change was a greater use of technology and computer-based applications to deliver and supplement the course.

Both suggestions were followed. Giving presentations in English was thus included as a speaking task in itself and also, Michelle stated, as a "vehicle for exercising the pronunciation skills that were taught and learned in the course." In addition, the use of WebCT, an Internet-based course delivery system, was much more developed. Michelle cited the example of posting videos of past student presentations on WebCT to serve as realistic models and guides for current students. A site map of the class web-site can be seen in Appendix R.

Since Michelle was the sole instructor responsible for both course sections, continual experimentation was possible and fairly frequent. "I seldom teach the same thing the same way again and again and again." For example, the skill of reading aloud while frequently looking up during presentations was one new aspect added to the course which was observed during this study. Modifications judged to be successes were integrated into the course. When Michelle tried something that did not seem to work, it was re-worked and tried once more.

Because the course was partly designed around presentation skills, one enduring challenge Michelle cited was time, specifically the amount of class time taken up by mid-term and final presentations. This difficulty of insufficient class time had still not been resolved; however, she had tackled the problem of time in another way. Since Michelle believed that the fourteen weeks of the course were insufficient to bring about substantial change in students' oral skills, she wanted students who finished the course to continue to have access to purpose-made practice materials. However, the CAN-8 software system containing the course exercises was in a language lab which was restricted to students registered in for-credit language courses. Therefore, Michelle designed and implemented a new for-credit course to follow the original oral communication course. This credit course consisted of self-study exercises using discipline-specific material. Many of these exercises were on CAN-8, but some were available via the Internet on WebCT. This means that "they can access [the exercises] better."

Since students should have left the oral communication course with a knowledge of suprasegmental patterns and some independent learning strategies such as self-monitoring, Michelle believed they should be capable of completing these exercises, and

they could spend substantially more time practising the suprasegmental patterns covered in the first course. In addition, she hoped to one day transfer all the original exercises for the oral communication course from the CAN-8 laboratory to an Internet-based software, so that all students could practice wherever they had an Internet connection.

The ideas for these many modifications stem partly from Michelle's experience and intuition. However, she also sought out opportunities for professional development, reading research on teaching and applied linguistics, going to conferences, and discussing pedagogical principles, methodologies, and techniques with other teachers and researchers.

Course Content

When Michelle was asked which suprasegmentals she focused on the most, she said that rhythm received the most focus; work on rhythm began at the start of the course and continued to be emphasized throughout. Michelle defined rhythm as comprising pausing, linking sounds and focus words. She cited lexical stress as an aspect which also received a lot of focus, but not as much as rhythm. The actual overall time spent on topics of instruction is shown in Table 17 above. When the three areas of pausing, linking, and focus words are combined, their total instructional time is indeed slightly higher than the time spent on lexical stress. Michelle stated that rhythm received the most focus and was introduced at the beginning of the course because students need more time to work on it, because "there's more variation and fewer fixed rules." In contrast, she believed students did not need as much time to be successful with lexical stress because there was less variation in the placement of stress, and students could themselves use a dictionary to determine a word's stressed syllables.

Michelle was asked about the reasons for choosing to focus on the particular content of the course. In other words, why was this particular content chosen? She first talked about the requirement for graduate students to regularly make presentations. "Graduate students present at conferences, they speak at seminars." The decision to include suprasegmentals was influenced by several factors. One was that it was already part of the syllabus when Michelle began teaching the course, but "I can tell you that I believe in the choice that was made." She stated that recent research on teaching pronunciation suggested that there was a greater probability that adult learners would become more intelligible if they improved suprasegmental, as opposed to segmental, aspects of speech.

Michelle noted that almost every student she had taught had no explicit awareness of elements of suprasegmental pronunciation in English prior to taking her course. She also reported that there was not enough time to focus on both segmental and suprasegmental aspects during the course. Nevertheless, "there are people in the course who desperately need work with phonetics and the ability to produce certain sounds in English, there's no question about that." However, those students could receive individualized help from Michelle and were also able to use software in the language laboratory to work on specific sounds.

When asked about the course textbook, Michelle stated that "the book is O.K.," and was useful because it provided the students with opportunities to practice, "something to work with." She praised the quality of information on suprasegmental patterns: "I wish students would read it more." However, she brought up several shortcomings, including having only cassette tape recordings as opposed to a CD ROM,

and the use of one IPA symbol /ə/ to represent both schwa and the mid-central unrounded vowel /ʌ/, which commonly occurs in stressed, as oppose to unstressed, syllables.

Methodology

Michelle was asked about her technique of presenting students with model words or sentences, then encouraging them to induce a pattern of use for suprasegmental aspects. For her, it was an approach that was second nature. She believed that it helped students be successful in their learning because "students actually retain better and they can apply things better if they actually go through the process of figuring them out." Michelle cited her own extensive experience in learning other languages, and how she preferred to "take things apart and see how they work." She recognized that her own favoured learning style influenced her teaching style, but she also stated that allowing students to first analyze a model by themselves gave them more time to produce language, rather than Michelle doing all the talking. She cited one other benefit of this approach, which was that more work in analyzing language made students better able to give feedback to their peers, and indeed, she had sometimes noticed spontaneous peer corrective feedback during group and pairwork.

In relation to work in groups, Michelle was asked whether students were worried about getting input or feedback which was non-nativelike while working in groups or pairs with other students. She confirmed that "definitely they worry about the mistakes." She suggested that if students improved their skills at monitoring oral production, they would be able to monitor their own speech if a partner was not providing helpful input or feedback.

The next question was about Michelle's own priorities for corrective feedback. She said that, "I try to direct feedback to what is actually being practised in class, or what has been." When a particular aspect had been previously covered, feedback on that aspect could be provided from that point to the end of the course. I had noticed when observing the classes that Michelle had provided some corrective feedback on segmental production in the last two classes observed. When asked whether she had done this consciously, Michelle was surprised and said that she was not aware that she had done so and did not know why she might have given feedback on specific sounds, as it was not a priority for her.

Michelle was asked whether she would ever depart from the lesson plan for the day, which was always carefully structured. She answered that in theory it was possible, but because of the limited amount of time available to cover the material, modifications of the lesson plan did not often happen. If, however, she judged that students really needed more time on a particular aspect, that section would be extended, and other material would be covered more quickly. The evaluation would then be adjusted to reflect the instructional time. Conversely, if students were very quick in understanding or mastering a certain aspect, that section of the lesson would be shortened. Students were free to come to either the morning or afternoon section of the course, taught on the same day, so Michelle tried to ensure that each day the same material was covered in both sections.

Michelle also discussed her own goals for modifying the way she teaches or thinks about teaching the class in the future. She had a strong desire to ensure that students spent more time speaking to each other, and that she spent relatively less time

speaking to the class. However, she believed that, "I have to accept that it's O.K. for me to do a bit more speaking in class [...] it's sometimes necessary." She also wished to use more choral repetition in the class, which she herself did not like doing, but which, she believed, helped students with their confidence.

Technology

Michelle made no explicit statements about the value of technology in language learning, but her regular mention of how technology was used as part of the course implicitly revealed her support for using technology as an aid to language learning. During Michelle's graduate studies, "I decided very deliberately to go into two areas that would lead to employment." One of these areas was pronunciation, but the other was technology in language learning. It is clear that Michelle viewed technology as an important aid to independent study and the use of learning strategies for language learners; she had created many exercises for her students on software accessible to students in the language laboratory or via the Internet. Students' presentations were video-recorded and students were expected to watch their presentations and formally evaluate themselves. As mentioned above, she had also created an entire course, delivered via WebCT, in order to provide students with more accessible and appropriate practice opportunities. The use of technology for out-of-class practice and evaluation was an integral part of the course.

Learning Conditions

Michelle was asked what, in her view, were the ideal conditions for an adult to learn or improve pronunciation in another language. She stated that students had to be highly motivated, such as needing to become more intelligible to get a job to feed their

family, for example. Such students would also need time to develop, and guidance at the beginning of the process. "They have to develop their ear, they have to know what to listen for, which is where the guidance comes in, and perhaps some feedback...about how well they're producing."

Michelle was then asked how she tried to bring about something approaching those conditions in her course. She was quick to say that "I'm not sure I do have the time factor in there," because the schedule for the course was so full, but she believed this was a challenge common to many teachers. She attempted to get around this limitation by encouraging students to remember what they were learning so that they could "apply it later or at a time when they have more time." In terms of guidance, Michelle was confident that she did provide students with a lot of guidance, and that the students were not afraid to "ask questions, [...] make mistakes," or receive feedback in front of the class.

For the final question, Michelle was asked about her ideal pronunciation course, and what it would look like. She did not think that she would add much more content to the syllabus, but stated that the course "would be more than one semester" so that students would have more time for practice and she could give them more feedback. It is surely no coincidence that Michelle had developed a new course following on from the oral communication course that did provide students with many more opportunities to practise what they had learned.

Summary: Instructor Interview

A point which was frequently mentioned throughout this interview was time, and the lack of it. Michelle's goals for students were based on her belief that a one-semester

course was not long enough to see noticeable development in their pronunciation. She hoped to give students knowledge and strategies that they could use after the course was finished, when they continued to have input from native speakers, and could recognize and try to reproduce suprasegmental patterns. Michelle did not seem to be as concerned about time for the learning of discourse-level presentation skills and structure, as she did not explicitly mention time as a problem for this area; this was perhaps because presentations are not open-ended, spontaneous communication, but are almost always prepared beforehand, giving students the chance to rehearse and modify their presentations so that they could reproduce the discourse-level structures and skills they had been taught in class.

The design and content of the course had been carefully planned and were based on Michelle's awareness of current research and pedagogical approaches, of the needs of her students, and of the relative success or failure of certain aspects of the course. There seemed to be two imperatives driving the design of the course and the way it was implemented: (a) raise students' awareness and knowledge; and (b) provide students with opportunities to continue practising what they have learned. Technology played an important role in out-of-class assignments and practice opportunities. The content of each class was tightly structured, but Michelle could modify the rate at which the material was covered. She believed that adults could best learn pronunciation when they had high motivation, sufficient time, guidance, and feedback. She thus worked to create many of these conditions for her students through her design of the oral communication course and through the creation of an entirely new Internet-based course.

Development of NN Graduate Students' L2 Speech: Overview

It was shown above that the classroom instruction received by the treatment group NN graduate students was carefully designed and implemented to reflect current practice in pronunciation teaching and Michelle's perceptions of students' communicative needs and learning processes. Michelle herself did not believe that students' skills in communication would noticeably improve over the short period of instruction. In this section, I investigate the L2 speech development of students who received the instruction, as well as students who did not.

Research Question 2

How does the speech of NN graduate students develop, in terms of its intelligibility and evaluation by listeners:

(a) from the beginning to the end of an ESL course targeting oral communication?

(b) when not taking an ESL oral communication course, but only studying in an English environment?

Intelligibility

There were two listening tasks used to measure intelligibility: story retells and on-line word intelligibility. The story retells were analyzed for: the accurate content (AC), the inaccurate content (IC), and listeners' statements of lack of understanding (LU). The on-line word intelligibility task was analyzed for inaccurate content (IC), measured by incorrect guesses of words, and listeners' statements of lack of understanding (LU), measured by unintelligible words.

Accurate Content (AC)

As described in Chapter 3, accuracy ratios were calculated for the story retells, with the accurate retell elements divided by the total number of original story elements. These mean ratios for each time and each speaker group are shown in Table 18. Somewhat surprisingly, the ratio of accurately retold elements to total story elements was at all times lower in retells for the native speaker (NS) group than for the two NNGS groups. That is, relatively fewer accurate elements appeared in the story retells for the NS group than for the other two groups.

Table 18

Accurate Content Ratio in Story Retells

	Testing time			
Group	Time 1	Time 3	Time 4	Mean
Treatment	0.72	0.65	0.68	0.68
Control	0.72	0.58	0.69	0.66
Native Speakers	0.62	0.55	0.63	0.60
Mean	0.68	0.60	0.67	

In order to be able to statistically compare the development of the NN treatment and control group retells in their ratio of accurate content, the retells for both groups first had to be at similar levels of accurate content at Time 1. Therefore, a t-test was run comparing the means of the retells for the two groups at Time 1, with a Bonferroni

correction for subsequent additional t-tests on the retell data. There was no difference between retells for the treatment and control groups at Time 1. Therefore, the accuracy ratios of the treatment, control, and native speaker group retells were submitted to a two-way analysis of variance (ANOVA) with time (Time 1, Time 3, Time 4) as a between-subjects factor, and talker group (NS, NN treatment, NN control) as a within-subjects factor. Talker group was here analyzed as a within-subjects factor and time as a between-subjects factor because the analysis was by the listener groups (their retells), not by the talker groups. Each listener group heard the same groups of talkers, but heard different stories recorded at different times.

Significant main effects of time, $F(2, 43) = 4.70$, $p = .014$, and of speaker group, $F(2, 86) = 21.44$, $p = .0001$ were found, with no significant interaction. Tests exploring the effect of time, with a Bonferroni correction for multiple comparisons, showed that story retells at Time 3 had lower accuracy ratios (fewer elements accurately retold) than at Time 1 ($p = .018$). Tests exploring the effect of speaker group, also with a Bonferroni correction, revealed that story retells for the NS group had lower accuracy ratios overall than retells for both the treatment and control groups of NNGSs (all $ps < .0001$).

Inaccurate Content (IC)

Inaccurate content was measured in the story retell task by ratios comparing the number of retell elements which were in some way inaccurate (W, PW, or DW)[3] to the total number of original elements; inaccurate content was measured in the on-line word intelligibility task by ratios comparing the incorrect word guesses which were not later

[3] As described in Chapter 3, W .refers to a retell element with quite inaccurate semantic content. PW refers to a retell element with some inaccurate semantic content. DW refers to a retell element with slightly inaccurate semantic content.

corrected (AW, SL, or EO)[4] to the total number of original words. The results for inaccurate content from each task will be presented in turn.

Story retells. In looking at inaccurate content in retells, it was possible to investigate differences by analyzing ratios containing every retell element with any inaccurate content at all (W, PW, and DW), or by analyzing ratios containing certain types of those elements. The first analysis which was done used ratios containing all retell elements with any inaccurate content at all (W, PW, and DW). In order to determine whether retells for the treatment group and the control groups were similar at Time 1 in their inaccurate content, a t-test was run comparing the means of the two groups' retells at Time 1, with a Bonferroni correction for additional t-tests on the retell task data. There was no difference between retells for the treatment and control groups at Time 1 ($p = 1.2$, n.s.).

The data for all three groups were then submitted to a two-way ANOVA with time (Time 1, Time 3, and Time 4) and talker group (native, NN treatment, NN control) as between- and within-subjects factors, respectively. No significant main effects were revealed for time or for talker group. The next analysis removed the inaccurate detail elements (DW), using only W and PW elements. A t-test was again run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the retell task data. Again, there was no difference between retells for the treatment and control groups at Time 1 ($p = 2.65$, n.s.). The data for all three groups were then submitted to a two-way ANOVA. Figure 1 shows the mean frequency of W and PW elements in the retells for each talker group; the higher the bar, the more inaccuracies in

[4] As described in Chapter 3, AW refers to a listener's inaccurate guess in the on-line task. SL refers to a listener saying, "It sounds like..." in the on-line task. EO refers to a listener providing two alternatives for the unknown word in the on-line task.

the retells. No effect for time was found, but a significant main effect was found for talker group, $F(2, 86) = 6.35$, $p = .003$, with a significant interaction, $F(4, 86) = 2.89$, $p = .027$.

The main effect of talker group was explored in follow-up tests, with a Bonferroni correction, revealing that the retells of control group stories had more inaccurate content (W and PW elements) than the retells of the native speaker group stories ($t(90) = 3.57$, $p = .002$).

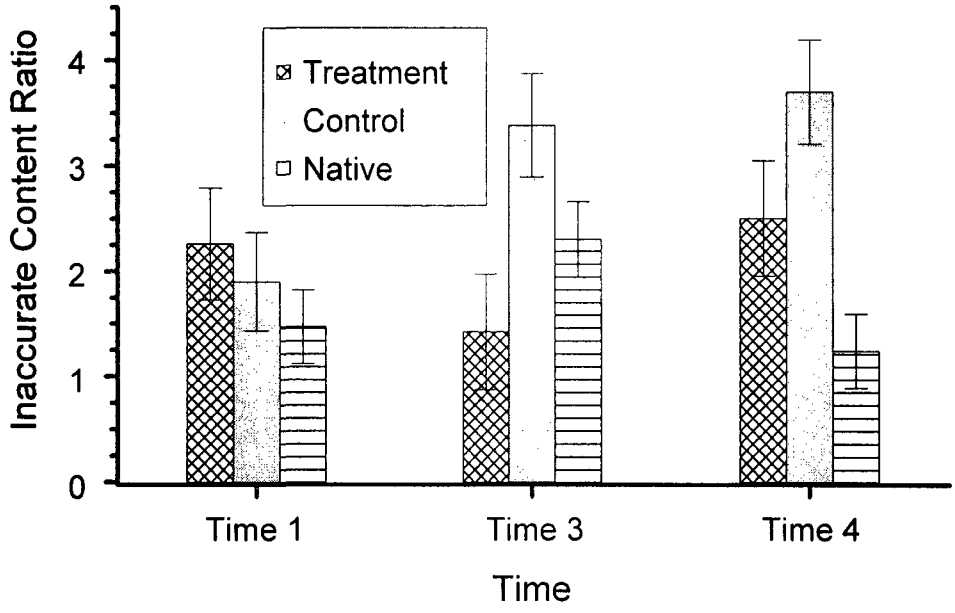

Figure 1. W and PW (wholly and partly inaccurate) elements in story retells. Because some of the original numbers for the ratios were small, the ratios and the standard error have been multiplied by 100 in order to better demonstrate differences between groups. The proportions across time and between groups remain the same.

Tests of simple main effects conducted to explore the significant interaction, with a Bonferroni correction, showed that at Time 3 the control group retells had more inaccurate content than the treatment group retells ($p = .02$), but this difference was not

maintained at Time 4. At Time 4, the control group retells had more inaccurate content than the native speaker group retells ($p = .001$).

The other measure of inaccurate content was incorrect guesses about words in the on-line word intelligibility task. This measure is presented next.

On-line word intelligibility task. The intelligibility ratios for incorrect guesses about words which were not later corrected (mostly coded AW [inaccurate], with some coded SL [sounds like...]or EO [either ...or...]) were calculated for each of the three times and talker groups. These ratios are shown in Figure 2; the higher the bar, the more inaccurate guesses by the listener group. In order to determine whether listeners at Time 1 had similar ratios of incorrect guesses for the treatment group and the control group, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for an additional t-test on the on-line task data. With the correction, there was no difference in ratios of incorrect guesses for the treatment and control groups at Time 1 ($p = .066$, n.s.).

The data for all three groups were submitted to a two-way ANOVA, with time (Time 1, Time 3, and Time 4) and talker group (native, NN treatment, NN control) as between- and within-subjects factors, respectively. No main effect for time was found, but there was a significant main effect for talker group, $F(2, 54) = 19.72$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that all talker groups were significantly different from each other: when listeners heard anecdotes from the native speaker group, there were fewer inaccurate guesses than for the treatment group ($p = .003$) or the control group ($p = .0001$). Listeners also made fewer inaccurate guesses for the treatment group than for the control group ($p = .001$).

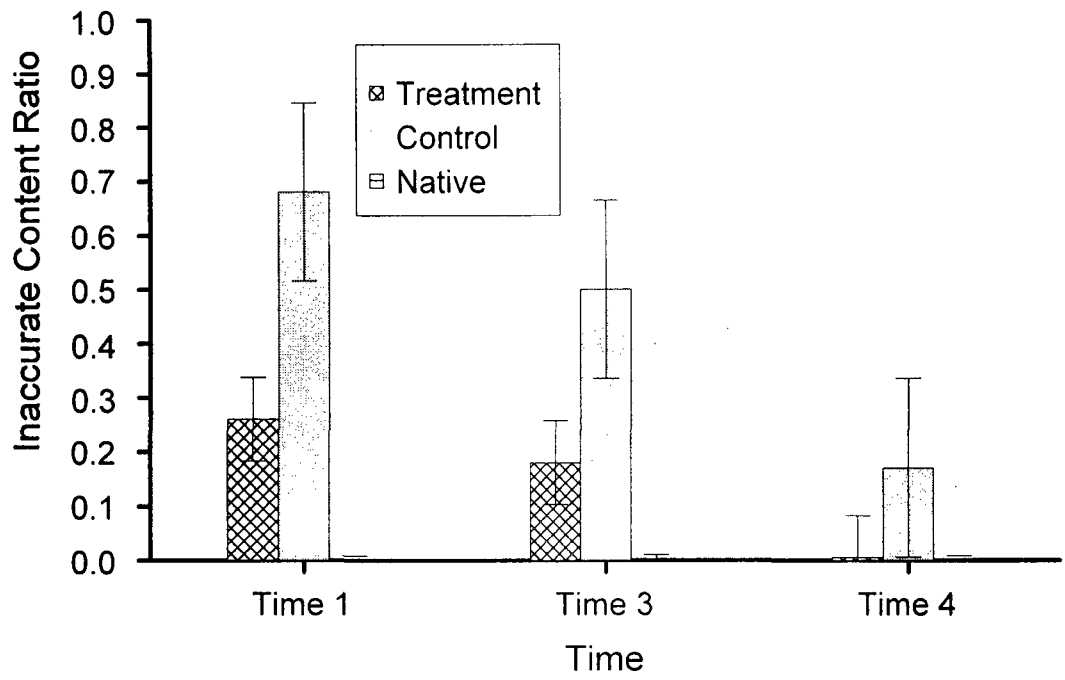

Figure 2. Inaccurate guesses from the on-line word intelligibility task. Because some of the original figures for the ratios were extremely small, the ratios and the standard error have been multiplied by 100 in order to better demonstrate differences between groups. The proportions across time and between groups remain the same.

Summary: Accurate and Inaccurate Content

In measuring the ratio of accurate and inaccurate content in story retells, no consistent difference in scores appears between the native speaker and the NN groups; this suggests that these measurements may not show differences in intelligibility based simply on nativeness or non-nativeness. Retells for the native speaker group stories had relatively less accurate content than retells for stories of the two NN groups. In addition, for inaccurate content, removing elements with inaccurate details (DW) from the statistical analyses revealed more differences between the retells for native speaker and

NN groups, since the control group retells had significantly more inaccurate content than the native speaker group.

Unlike the inaccurate elements in the story retells, the incorrect word guess measure in the on-line word intelligibility task revealed many more definite differences in the intelligibility of speaker groups. However, in this task, no significant changes over time were revealed, whereas in the retell task, the treatment group had fewer inaccurate elements than the control group, but only at Time 3.

Listeners' Statements of Lack of Understanding (LU)

The third way of measuring intelligibility was through listeners' statements of lack of understanding(LU).

Story retells. In the story retell task, these listener statements were coded DU (listener comments that they didn't understand a part of the story), HU (that they had difficulty understanding a part of the story), or DUS (that they didn't understand the significance of a part of the story). It was possible to explore LU statements by analyzing all three types of statements (DU, HU, and DUS), or by separating out one or two types. In the first analysis, all three types of statements (DU, HU, and DUS) were summed and divided by the total number of original story elements.

In order to determine whether the treatment group and the control group retells were similar at Time 1in their ratios of LU statements, a t-test was run comparing the means of LU statements in retells for the two groups at Time 1, with a Bonferroni correction for additional t-tests on the retell data. There was no difference in LU statements between the treatment and control group retells at Time 1 ($p = 1.50$, n.s.). The data for all three groups were submitted to a two-way ANOVA, with time as a between-

subjects factor and talker group as a within-subjects factor. No main effect for time was found, but there was a significant main effect for talker group, $F(2, 86) = 23.38$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, revealed that when listeners heard anecdotes from the native speaker group, they made significantly fewer LU statements than they did for both the treatment ($p = .0001$).and control groups ($p = .0001$). This analysis included listener statements that they had difficulty understanding a part of the story (HU), that they didn't understand a part of the story (DU), or they didn't understand the significance of a part of the story (DUS).

In the second analysis for story retells, the focus was only on listeners' statements that they didn't understand a part of the story (DU). The number of DU statements was divided by the total number of story elements. In order to determine whether the mean DU statements for the treatment group and the control group retells were similar at Time 1, a t-test was run comparing the means of the retells for the two groups at Time 1, with a Bonferroni correction for additional t-tests on the retell data. There was no difference in DU statements between the treatment and control group retells at Time 1 ($p = 1.00$, n.s.). The data for all three groups were submitted to a similar two-way ANOVA. The results were similar to the previous analysis, in that no main effect for time was found, but there was a significant main effect for talker group, $F(2, 86) = 13.63$, $p = .0001$, with no interaction.

Tests exploring the effect of talker group, with a Bonferroni correction, showed that, as in the first LU analysis, retells for the native speaker group had significantly fewer DU statements than retells for both the treatment ($p= .0001$)and control groups ($p = .0001$), but the treatment and control group retells were not significantly different from

each other. Figure 3 shows the DU statements in each talker group retell over time; the higher the bar, the more DU statements in listeners' retells.

On-line word intelligibility task. In the on-line word task, LU statements were those in which listeners could not venture any guess at all about a word or words they hadn't understood (coded DK). The number of LU statements was divided by the total number of words in the story. In order to determine whether the treatment group and the control group were similar at Time 1 in their ratios of words not understood by listeners, a t-test was run comparing listeners' mean LU statements for the two groups at Time 1, with a Bonferroni correction for an additional t-test on the on-line data. With the correction, there was a significant difference in listeners' LU statements for the treatment and control groups at Time 1 ($p = .036$), with the treatment group receiving significantly fewer LU statements. Since the treatment and control groups were different at Time 1, they could not be compared together in a two-way ANOVA.

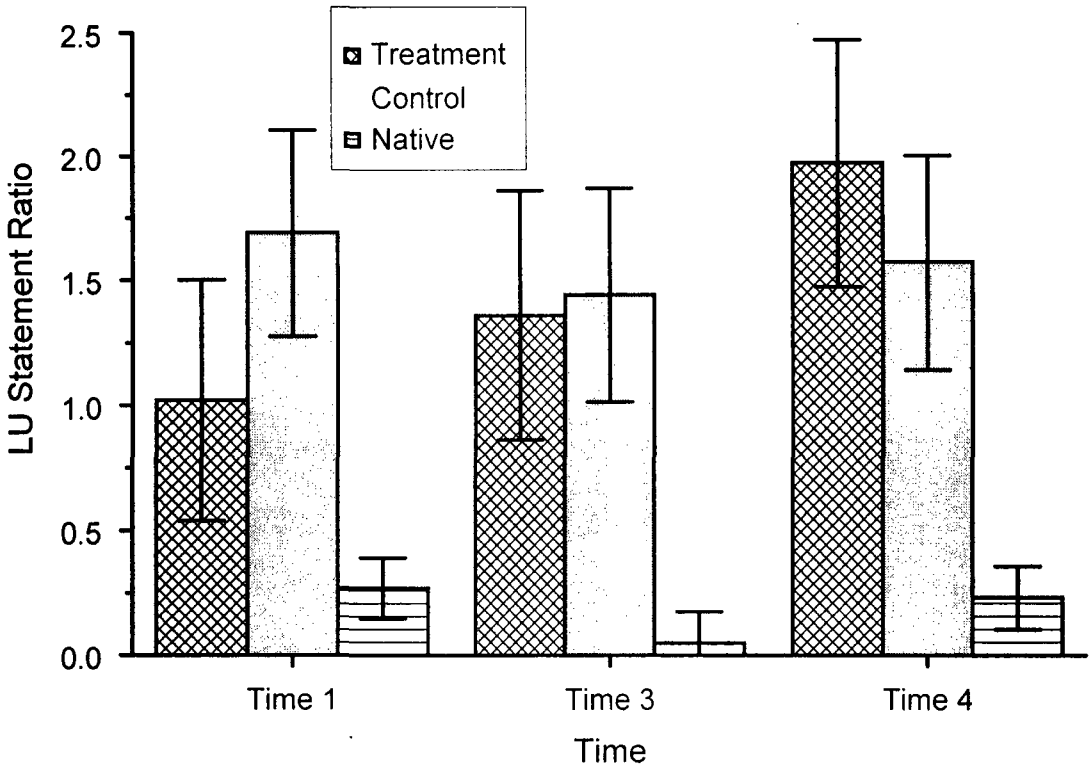

Figure 3. Statements of lack of understanding (DU elements only) in story retells. Some LU statement ratios appear to be greater than 1.0. Because some of the original numbers for the ratios were extremely small, the ratios and the standard error have been multiplied by 100 in order to better demonstrate differences between groups. The proportions across time and between groups remain the same.

Therefore, two separate two-way ANOVAs were done: one comparing the LU statements of listeners for the treatment and native speaker groups, and one comparing the LU statements of listeners for the control and native speaker groups. For both two-way ANOVAs, time and talker groups were the between- and within-subjects factors, respectively. Figure 4 shows listeners' LU statements for each talker group over time; the higher the bar, the more LU statements by the listener group.

For the two-way ANOVA comparing the treatment and native speaker groups, there were significant main effects for both time, $F(1, 27) = 3.72, p = .037$, and talker group, $F(1, 27) = 13.23, p = .001$, with a significant interaction $F(2, 27) = 3.72, p = .037$. Tests exploring the effects of talker group and of time, with Bonferroni corrections, showed that the native speaker group received significantly fewer LU statements than the treatment group ($p = .001$) and that there were significantly more LU statements overall at Time 4 than at Time 1 ($p = .034$). Tests of simple main effects conducted to explore the significant interaction, with a Bonferroni correction, revealed that there were more LU statements about words not understood in the treatment group stories at Time 4 than at Time1 ($p = .034$).

For the two-way ANOVA comparing the control and native speaker groups, there was a main effect for time, $F(2, 27) = 3.42, p = 0.047$, and for talker group, $F(1, 27) = 40.91, p = .0001$, with a significant interaction $F(2, 27) = 3.42, p = 0.047$. Tests exploring the effects of talker group and of time, with Bonferroni corrections, showed that, in the case of talker group, the native speaker group received significantly fewer LU statements than the control group ($p = .0001$). Regarding the significant effect of time and the significant interaction, no times and no combinations of groups and times were found to be different on the post hoc tests. The significant differences found in the omnibus analysis (i.e., the significant effect of time and the significant interaction) may be due to differences between some combination of times or some larger combination of times and groups.

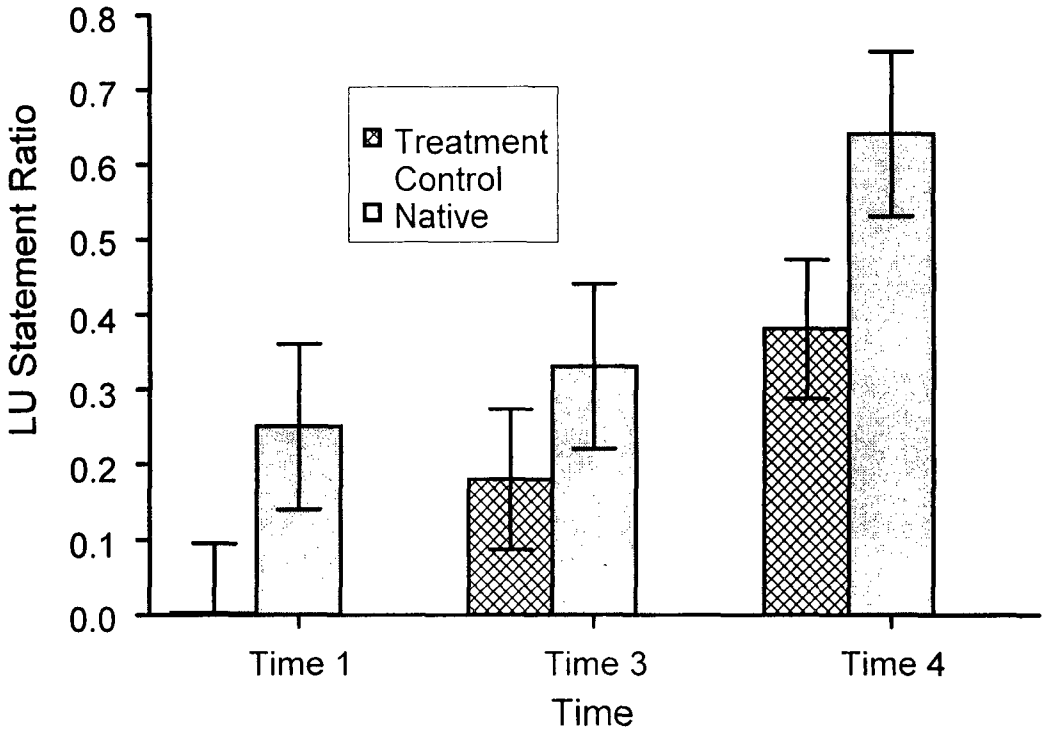

Figure 4. Statements of lack of understanding in the on-line word intelligibility task. Bars for the native speaker stories cannot be seen because the ratio at each time was virtually zero. Because some of the original numbers for the ratios were extremely small, the ratios and the standard error have been multiplied by 100 in order to better demonstrate differences between groups. The proportions across time and between groups remain the same.

Summary: LU Statements

In both the retell task and the on-line word intelligibility task, there were significantly fewer LU statements when listeners heard native speaker stories than when listeners heard stories from the two NN groups. In the retell task, no differences appeared between stories for the NN treatment and control groups, while in the on-line task,

differences appeared at Time 1, so the two groups were analyzed separately. In the on-line task, the treatment group stories received more LU statements overall than the native speaker group, and there were more LU statements for the treatment group stories at Time 4 than at Time 1. Like the treatment group, the control group stories received more LU statements than the native speaker group overall, but unlike the treatment group, the ratio of statements did not change over time.

Summary: Group Intelligibility

Overall, in the analysis of the retell elements, the (non)nativeness of talker groups did not *consistently* lead to differences in ratios of accurate and of inaccurate content. Listeners' scores for the native speaker group were often better than one, but not always both, NN talker groups. In terms of development over time, some differences between groups at specific times appeared, though these differences did not remain stable from one recording time to another.

In the analysis of the categories in the on-line task, differences between groups were much more stable. In terms of listeners' scores, the native speaker group always came out better than both NN groups, and when both NN groups were similar at Time 1 and could be compared, the treatment group was better overall than the control group. However, for the treatment group, retells at Time 4 did contain more LU statements than retells at Time 1.

Individual Talker Analysis

When groups of talkers are compared at particular points in time, we acquire information about the speaking abilities of the talkers relative to the other groups and to the passage of time. However, a group may include talkers of different levels of ability,

and in looking at the performance of the group, the development of individual talkers may be lost. To explore the development of the individual NN talkers, post-hoc analyses were conducted. Each talker's data on inaccurate content (IC) and listener comments of lack of understanding (LU) were submitted to one-way ANOVAs, with time as the between-subjects factor. Results from these tests will be presented in order of talkers' length of residence. The differences are summarized in Table 19.

Table 19

Intelligibility of Individual Talkers

	Length of Residence					
	Less than 2 years		1.5 years		0.5 years	
Task	Feng (C)	Xiao (T)	Ping (C)	Hui (T)	Lupe (C)	Javier (T)
	Inaccurate content (IC)					
Retell	n.s.	n.s.	T1 < T3 T1 < T4	T1 < T3	n.s.	n.s.
On-line	T4 < T1	n.s.	n.s.	n.s.	n.s.	n.s.
	Lack of understanding (LU)					
Retell	n.s.	n.s.	n.s	n.s.	n.s.	T1 < T4
On-line	T3 < T4	n.s.	n.s.	n.s.	n.s.	n.s.

Note. C = control group; T = treatment group; n.s. = no significant difference; significantly fewer inaccurate guesses or LU statements occurred at the time to the left of the < symbol than at the time to the right.

Length of Residence: Over 2 Years

Feng (control). No difference over time appeared for inaccurate retell elements (WPW). However, for inaccurate word guesses from the on-line task, there was a significant *F* ratio, $F(2, 27) = 4.36$, $p = .023$. Tests exploring the significant F ratio, with a Bonferroni correction, showed that significantly fewer inaccurate word guesses appeared at Time 4 than at Time1 ($p = .039$). There were no differences over time for LU statements in the retells. For similar statements in the on-line task, the one-way ANOVA yielded a significant *F* ratio, $F(2, 27) = 4.68$, $p = .018$. Tests exploring the significant F ratio, with a Bonferroni correction, revealed that there were significantly fewer statements when no word could be guessed at Time 3 than at Time 4 ($p = .016$).

Xiao (treatment). No difference over time appeared for inaccurate retell elements or for inaccurate word guesses. The ratio of LU statements both in the retells and in the on-line task did not change over time.

Length of Residence: One and a Half Years

Ping (control). The one-way ANOVA for inaccurate retell elements yielded a significant F ratio, $F(2, 43) = 11.42$, $p = 0001$. Tests exploring the significant *F* ratio, with a Bonferroni correction, showed that there were significantly fewer inaccurate retell elements at Time 1 than at Time 3 ($p = .0001$), and than at Time 4, ($p = .002$). The number of inaccurate word guesses did not change over time, nor did the LU statements in retells and in the on-line task.

Hui (treatment). Like Ping, for inaccurate retell elements, the ANOVA yielded a significant *F* ratio, $F(2, 43) = 3.93$, $p = .027$. Tests exploring the significant *F* ratio, with a Bonferroni correction, revealed that significantly fewer inaccurate elements appeared in

retells at Time 1 than at Time 3 (p = .028). However, there was no difference in inaccurate word guesses over time. No differences appeared over time in LU statements in the retells or in the on-line task.

Length of Residence: Half a Year

Lupe (control). The number of inaccurate retell elements and inaccurate word guesses did not change over time. No differences appeared over time for LU statements in retells or in the on-line task.

Javier (treatment). There were no differences in the number of inaccurate retell elements or inaccurate word guesses over time. A significant F ratio for LU statements in retells was seen, $F(2, 42)= 3.98$, p = .026. Tests exploring the significant F ratio, with a Bonferroni correction, revealed that there were significantly fewer such statements at Time1than at Time 4 (p = .045). No differences appeared for LU statements in the on-line task.

Summary: Individual Talker Analysis

Generally, there were not many differences over time for any talker, and the differences which did appear were not consistent. Listeners hearing Feng's anecdotes made fewer inaccurate word guesses at Time 4 than at Time 1, an improvement, but there were more words which were completely unknown to listeners hearing Feng at Time 4 than at Time 3. The most frequent differences surfaced between Time 1 and later times, with relatively better scores at Time 1. This pattern will be explored in the discussion in Chapter 5.

Listener Evaluation

Non-native graduate students' intelligibility for listeners was measured using retell story elements and word guesses by listeners. These objective measures of listeners' understanding are very helpful in exploring NN speakers' ability to communicate a message. However, measures of listeners' understanding of speakers are not necessarily the sole criterion on which those speakers can be evaluated. Listeners use their own individual criteria to judge speakers. Therefore, listeners' evaluations were recorded by having them rate excerpts of the speech samples using nine-point Likert-type scales on three dimensions: comprehensibility (how easy it was to understand the talker), accentedness (how closely a given talker's pronunciation approximated a native speaker norm), and fluency (the smoothness and fluidity of a given talker's oral production).

Listeners for both intelligibility tasks rated the same excerpts. However, the two intelligibility tasks made different demands on the listeners. For the retell task, listeners had to understand and reproduce the story elements, whereas for the on-line word task, listeners had to identify individual words which were unclear and guess at possible candidates for the words. Because the demands of the retell task were more global in focus than those of the on-line word task, listeners for the two tasks may have approached the subsequent rating task differently. Therefore, the ratings of listeners from the two tasks were analyzed separately and will be presented in turn. For each rating scale, data points which were more than two standard deviations above or below the mean for that group of listeners were removed as outliers. Outlier data accounted for 13% of the data points for listeners' ratings. Removing the outliers increased the internal consistency of

the data set which was analyzed, thus improving the reliability of the results of statistical tests.

Story Retells: Comprehensibility

Rater reliability. The interrater reliability between listeners in the retell task who rated comprehensibility was calculated for each of the three different listener groups. The intraclass correlation coefficient was calculated for ratings of the two NN talker groups, but not for ratings of the native speaker group, whose ratings were so similar that the intraclass correlation coefficient test could not be run. Generally, there was high interrater reliability for retell listeners rating comprehensibility. Table 20 shows the listeners' interrater reliability coefficients in rating the two NN talker groups.

Table 20

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Comprehensibility Ratings in the Retell Task

	Group	
Testing time	Treatment	Control
Time 1	.87 (.0002)	.90 (.0001)
Time 3	.74 (.0024)	.93 (.0001)
Time 4	.88 (.0001)	.93 (.0001)

Comprehensibility. In order to determine whether the treatment group talkers and the control group talkers were similar at Time 1 in the ratings they received for comprehensibility, a t-test was run comparing the means of the two groups at Time 1,

with a Bonferroni correction for additional t-tests on the rating data. With the correction, there was a significant difference between the treatment and control groups at Time 1 (p = .009), with the treatment group being rated more comprehensible. Since the treatment and control groups were different at Time 1, they could not be compared together in a two-way ANOVA. Therefore, two separate two-way ANOVAs were done, one comparing the treatment and native speaker groups, and one comparing the control and native speaker groups. For both two-way ANOVAs, time and talker groups were the between- and within-subjects factors, respectively. Figure 5 shows the comprehensibility ratings for each talker group over time; the higher the bar, the more difficult listeners thought it was for them to understand.

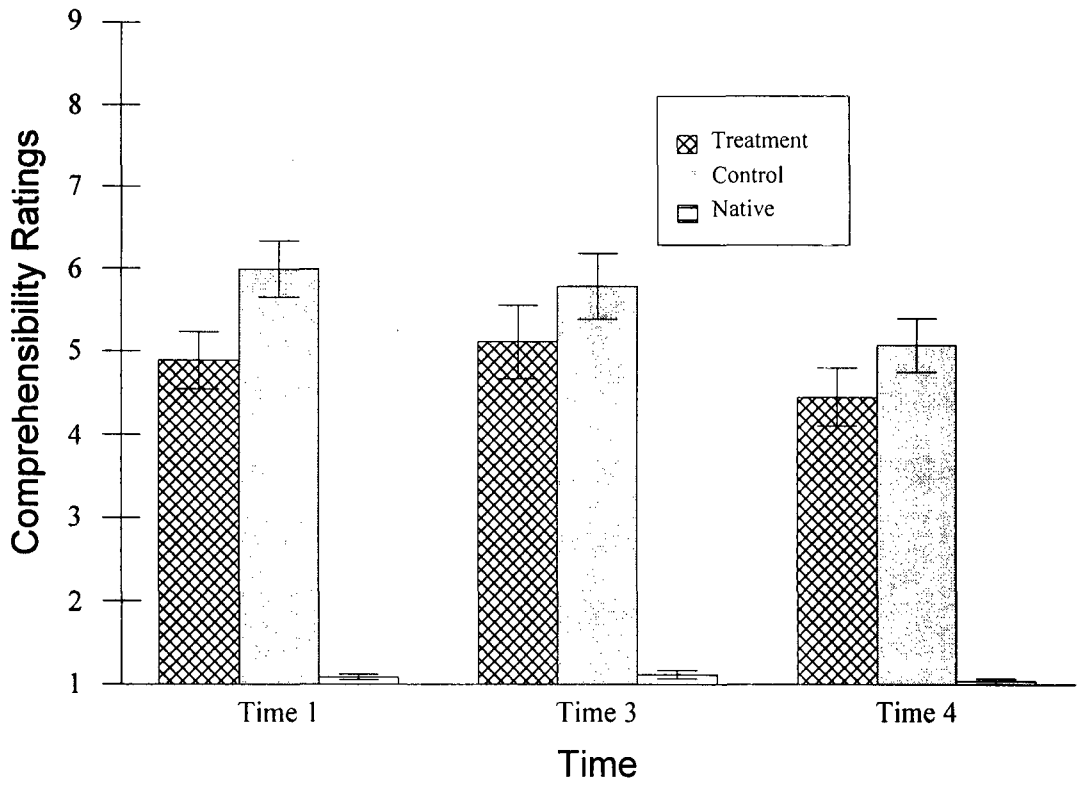

Figure 5. Comprehensibility ratings from story retells. Scale endpoints: 1 = very easy to understand, 9 = very difficult to understand.

For the two-way ANOVA comparing the treatment and native speaker groups, there was a significant main effect for talker group, $F(1, 34) = 244.26$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that the native speaker group was rated easier to understand than the treatment group ($p = .0001$). For the two-way ANOVA comparing the control and native speaker groups, there was again a significant main effect for speaker group, $F(1, 33) = 459.50$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that the native speaker group was rated easier to understand than the control group ($p = .0001$).

Story Retells: Accentedness

Rater reliability. The interrater reliability between listeners in the retell task who rated accentedness of NN talker stories was calculated for each of the three different listener groups. Again, the ratings of native speaker stories were so similar that the intraclass correlation coefficient test could not be run. Interrater reliability was generally high, with some lower coefficients of reliability, especially for the control group. Table 21 shows the listeners' interrater reliability coefficients in rating the two NN talker groups.

Table 21

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Accentedness Ratings in the Retell Task

	Group	
Testing time	Treatment	Control
Time 1	.87 (.0003)	.73 (.022)
Time 3	.83 (.0004)	.93 (.0001)
Time 4	.79 (.0001)	.62 (.05)

Accentedness. In order to determine whether the treatment group and the control group were similar at Time 1in their accentedness ratings, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the rating data. With the correction, there was no difference between the groups (p = 0.99, n.s.). The accentedness ratings for all three groups were submitted to a two-way ANOVA, with time and talker groups as the between- and within-subjects factors, respectively. A main effect was found for talker group $F(2, 58) = 1147.14$, $p = .0001$, but not for time, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, revealed that the native speaker group was rated less accented than both the treatment group ($p = .0001$) and control group ($p = .0001$). Accentedness ratings are seen in Figure 6; the higher the bar, the stronger the foreign accent.

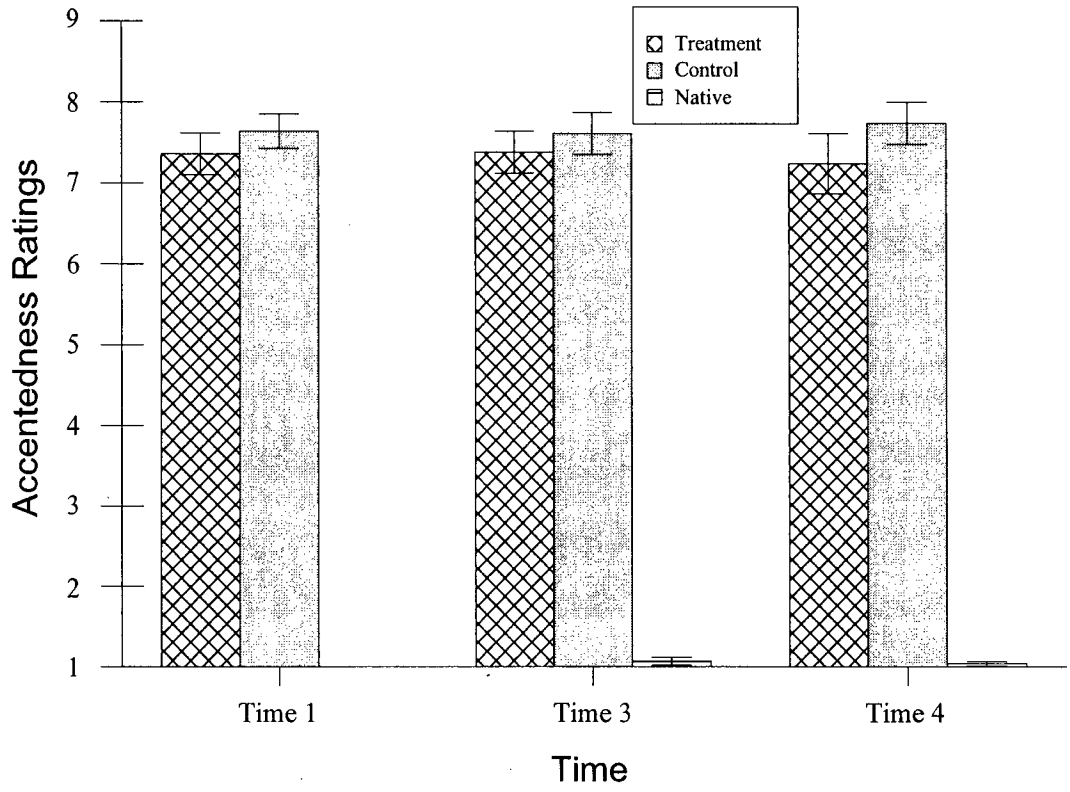

Figure 6. Accentedness ratings from story retells. Scale endpoints: 1 = no foreign accent, 9 = very strong foreign accent.

Story Retells: Fluency

Rater reliability. The interrater reliability between listeners in the retell task who rated fluency of NN talker stories was calculated for each of the three different listener groups. Again, the ratings of native speaker stories were so similar that the intraclass correlation coefficient test could not be run. Interrater reliability was much more variable across time and speaker groups, with low reliability coefficients especially for the treatment group. Table 22 shows the listeners' interrater reliability coefficients in rating the two NN talker groups.

Table 22

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Fluency Ratings in the Retell Task

	Group	
Testing time	Treatment	Control
Time 1	.69 (.011)	.98 (.0001)
Time 3	.02 (.369)	.95 (.0001)
Time 4	.90 (.0003)	.72 (.05)

Fluency. The fluency ratings for all three talker groups are seen in Figure 7; the higher the bar, the more dysfluent the talker. In order to determine whether the treatment group and the control group were similar at Time 1in their fluency ratings, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the rating data. With the correction, there was a significant difference between the treatment and control groups at Time 1 ($p = .015$), with the treatment group being rated more fluent. Since the treatment and control groups were different at Time 1, they could not be compared together in a two-way ANOVA.

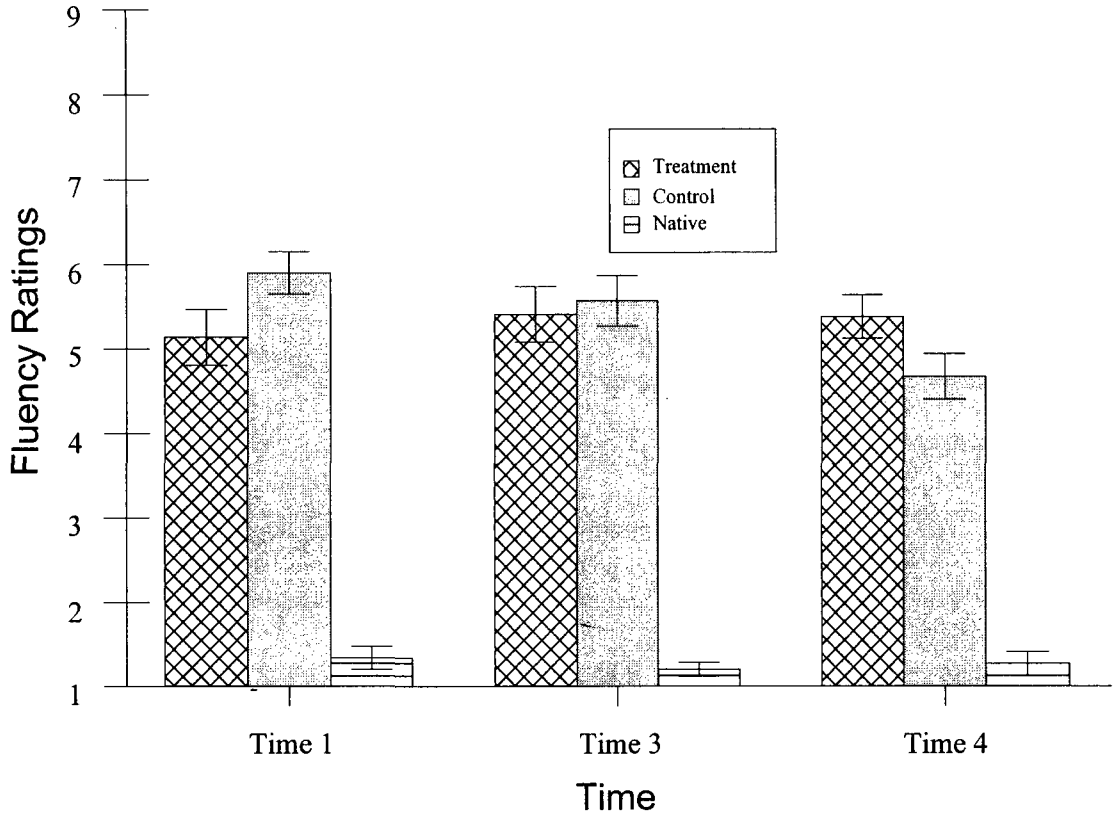

Figure 7. Fluency ratings from story retells. Scale endpoints: 1 = very fluent, 9 = very dysfluent.

Therefore, two separate two-way ANOVAs were done, one comparing the fluency ratings of the treatment and native speaker groups, and one comparing the control and native speaker groups. For both two-way ANOVAs, time and talker groups were the between- and within-subjects factors, respectively. For the two-way ANOVA comparing the treatment and native speaker groups, there was a significant main effect for talker group, $F(1, 34) = 321.53$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that the native speaker group was rated more fluent than the treatment group ($p = .0001$). For the ANOVA comparing the control and native speaker groups, there was a significant main effect for talker group,

$F(1,34) = 522.26$, $p = .0001$, a significant main effect for time $F(2, 34) = 3.96$, $p = .028$, and a significant interaction $F(2, 34) = 4.51$, $p = .018$. Tests exploring the effect of talker group and of time, with Bonferroni corrections, showed that the native speaker group was rated more fluent than the control group ($p = .0001$), and that stories at Time 4 received higher fluency ratings than stories at Time 1 ($p = .024$). Tests of simple main effects conducted to explore the significant interaction, with a Bonferroni correction, revealed that the control group was rated more fluent at Time 4 than at Time 1 ($p = .012$).

On-line Word Intelligibility Task: Comprehensibility

Rater reliability. The interrater reliability between listeners in the on-line task who rated comprehensibility of NN talker stories was calculated for each of the three different listener groups. The ratings of native speaker stories were so similar that the intraclass correlation coefficient test could not be run. The interrater reliability for the treatment group was low at Time 1, and was in fact at the level of chance at Time 3. Table 23 shows the listeners' interrater reliability coefficients in rating the two NN talker groups. The listeners are divided into the three groups who each heard one of three different sets of samples.

Table 23

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Comprehensibility Ratings in the On-line Word Task

	Group	
Testing time	Treatment	Control
Time 1	.71 (.03)	.82 (.005)
Time 3	.20 (.26)	.79 (.004)
Time 4	.94 (.0004)	.93 (.0001)

Comprehensibility. The results for the three ratings scales for this task are shown in Figures 8, 9, and 10. Comprehensibility ratings are shown in Figure 8. In order to determine whether the treatment group and the control group were similar at Time 1 in their comprehensibility ratings, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the rating data. With the correction, there was a significant difference between the treatment and control groups at Time 1 ($p = .045$), with the treatment group being rated more comprehensible.

Since the treatment and control groups were different at Time 1, they could not be compared together in a two-way ANOVA. Therefore, two separate two-way ANOVAs were done, one comparing the comprehensibility ratings of the treatment and native speaker groups, and one comparing the control and native speaker groups. For both two-way ANOVAs, time and talker groups were the between- and within-subjects factors, respectively. For the two-way ANOVA comparing the treatment and native speaker

groups, a main effect for talker group was found, $F(1, 22) = 229.01$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, revealed that the native speaker group was rated easier to understand than the treatment group, $p = .0001$. For the two-way ANOVA comparing the control and native speaker groups, a main effect for talker group was found, $F(1, 22) = 403.04$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, revealed that the native speaker group was rated easier to understand than the control group, $p = .0001$.

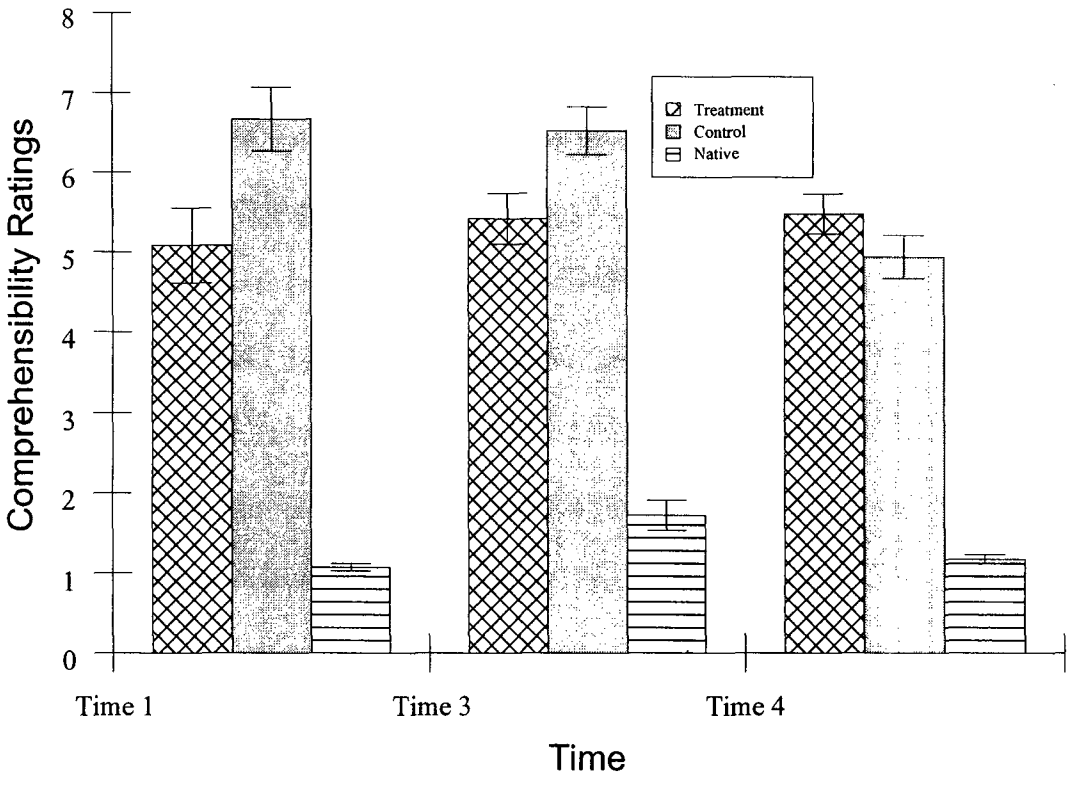

Figure 8. Comprehensibility ratings from the on-line word task. Scale endpoints: 1 = very easy to understand, 9 = very difficult to understand.

On-line Word Intelligibility Task: Accentedness

Rater reliability. The interrater reliability between listeners in the on-line task who rated accentedness of NN talker stories was calculated for each of the three different listener groups. The ratings of native speaker stories were so similar that the intraclass correlation coefficient test could not be run. The interrater reliability was high, with the exception of the rating of the treatment group talkers at Time 3, which was at the level of chance. Table 24 shows the listeners' interrater reliability coefficients in rating the two NN talker groups.

Table 24

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Accentedness Ratings in the On-line Word Task

	Group	
Testing time	Treatment	Control
Time 1	.92 (.0001)	.84 (.0005)
Time 3	.44 (.20)	.85 (.007)
Time 4	.91 (.0004)	.94 (.0002)

Accentedness. Figure 9 shows the accentedness ratings for the on-line word task. In order to determine whether the treatment group and the control group were similar at Time 1 in their comprehensibility ratings, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the rating

data. With the correction, there was no significant difference between the groups (p = .39, n.s.). Accentedness ratings from all three groups were thus entered into a two-way ANOVA, with time and talker groups as the between- and within-subjects factors, respectively. No main effect appeared for time, but a main effect for talker group was revealed, $F(2, 36) = 2100.17$, $p = .0001$, with no interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that the native speaker group was rated less accented than the treatment (p =.0001) and control groups (both ps = .0001). The treatment group was also rated less accented than the control group (p =.0001).

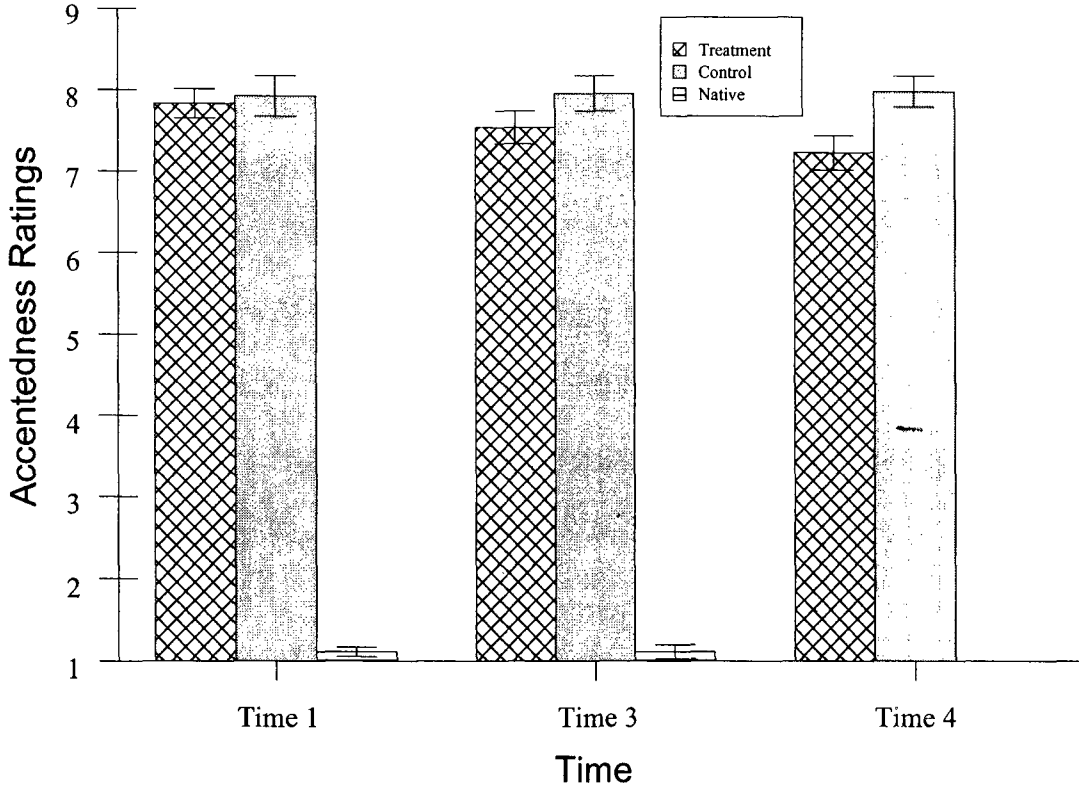

Figure 9. Accentedness ratings from the on-line word task. Scale endpoints: 1 = no foreign accent, 9 = very strong foreign accent.

On-line Word Intelligibility Task: Fluency

Rater reliability. The interrater reliability between listeners in the on-line task who rated fluency of NN talker stories was calculated for each of the three different listener groups. The ratings of native speaker stories were so similar that the intraclass correlation coefficient test could not be run. Similar to the other two rating scales, ratings of treatment group talkers at Time 1 and Time 3 varied so greatly that reliability was not higher than chance level. Table 25 shows the listeners' interrater reliability coefficients in rating the two NN talker groups.

Table 25

Intraclass Correlation Coefficients and Probability Values (in Parentheses) for Fluency Ratings in the On-line Word Task

	Group	
Testing time	Treatment	Control
Time 1	.48 (.64)	.93 (.0001)
Time 3	-2.17 (.70)	.94 (.0001)
Time 4	.82 (.018)	.92 (.0003)

Fluency. Fluency ratings are shown in Figure 10. In order to determine whether the treatment group and the control group were similar at Time 1 in their fluency ratings, a t-test was run comparing the means of the two groups at Time 1, with a Bonferroni correction for additional t-tests on the rating data. With the correction, there was a

significant difference between the groups, (p = .018), with the treatment group being rated more fluent at Time 1. Since the treatment and control groups were different at Time 1, they could not be compared together in a two-way ANOVA. Therefore, two separate two-way ANOVAs were done, one comparing the fluency ratings of the treatment and native speaker groups, and one comparing the control and native speaker groups. For both two-way ANOVAs, time and talker groups were the between- and within-subjects factors, respectively. For the two-way ANOVA comparing the treatment and native speaker groups, there was a main effect for talker group, $F(1, 21) = 329.64$, p = .0001, with no significant interaction. Tests exploring the effect of talker group, with a Bonferroni correction, showed that the native speaker group was rated more fluent than the treatment group (p = .0001).

For the two-way ANOVA comparing the control and native speaker groups, main effects were found for both time, $F(2, 20) = 13.62$, p = .00001, and talker group, $F(1, 20)$ = 463.49, p = .0001, with a significant interaction, $F(2, 20) = 7.91$, p = .003. Tests exploring the effect of time and talker groups, with Bonferroni corrections, showed that the listeners of stories at Time 4 rated talkers significantly more fluent than did listeners of stories at Time 3 (p = .0001) or Time 1 (p = .002). In addition, the native speaker group was rated significantly more fluent overall than the control group (p = .0001). Tests of simple main effects conducted to explore the significant interaction, with a Bonferroni correction, revealed that for the control group, the talkers were rated significantly more fluent at Time 4 than at Time 3 (p = .004) or at Time 1 (p = .001). The native speaker group was also rated more fluent at Time 1 (p = .023) and Time 4 (p = 0.33) compared to Time 3.

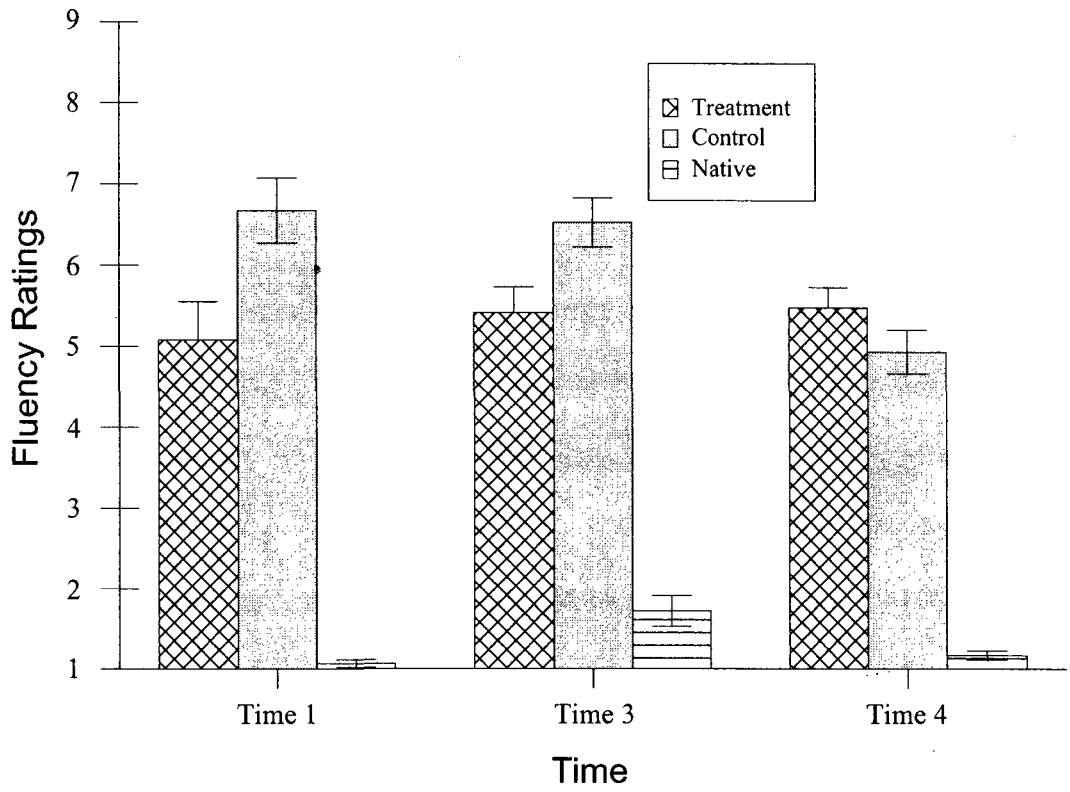

Figure 10. Fluency ratings from the on-line word task. Scale endpoints: 1 = very fluent, 9 = very dysfluent.

Summary: Listener Evaluation

Rater reliability. It is clear that listeners for both tasks did not act as monolithic groups in rating talkers. Listeners' inter-rater reliability for evaluating talkers' comprehensibility tended to be high for both listener groups. When listeners' ratings were variable, though, the variability usually occurred when they were rating the treatment group talkers. The reliability of fluency ratings for treatment group talkers was often low

for listeners for both tasks. In addition, listeners for the on-line word task who heard treatment group talkers at Times 1 and 3 had generally low inter-rater reliability.

Rating patterns. In evaluating talkers on their comprehensibility and accentedness, listeners for the retell and on-line word tasks rated in similar ways: at Time 1, the treatment group was rated easier to understand than the control group. Listeners rated the NS group easier to understand overall than both the treatment and the control groups. Listeners also rated the NS group less accented overall than the treatment and the control groups. For fluency, listeners for the retell and on-line word tasks were somewhat different in their rating patterns. For listeners for both tasks, the treatment group was rated more fluent than the control group at Time 1, and each NN group was rated less fluent overall than the NS group. For listeners for the retell task, the control group was rated more fluent at Time 4 than at Time 1, while for listeners for the on-line word task, the control group was rated more fluent at Time 4 than at either Time 3 or Time 1. The NS group was also rated more fluent at Time 1 than at Time 3.

English Exposure and Use

In comparing the intelligibility and listener evaluation of the NN talker groups, the treatment group often received higher intelligibility scores and better ratings than the control group. The treatment group had the benefit of being exposed to regular instruction in oral communication. However, over the duration of this study, there were various ways in which both NN groups were exposed to and used English. Apart from instruction, the nature of NN talkers' English exposure and use could also have influenced their L2

speech development. The third research question concerned the NN groups' English exposure and use.

Research Question 3

In what ways do NN graduate students at an English-medium university come into contact with and use English on a weekly basis?

Completion of Logs

Because the completion of the logs was done on a voluntary basis and not under direct supervision, not all talkers managed to complete all the language logs in every log period. Table 26 shows the number of talkers who completed the language logs for each log period.

Data from the first four logs were the subject of analysis; data from the fifth log were not analyzed because only two talkers in the treatment group completed the log. To ensure that talkers who completed only one or two logs did not have undue influence, talkers who did not complete at least three of the four logs (one talker from the treatment group and two from the control group) were removed from the analyses.

Table 26

Number of Talkers Who Completed Language Use Logs

	Group	
Log Period	Treatment (n = 8)	Control (n = 9)
Log 1 (Jan. 22-28)	7	7
Log 2 (Feb. 11-19)	7	7
Log 3 (March 12-18)	7	7
Log 4 (April 9-15)	5	6
Log 5 (April 30-May 6)	2	8

Note. Data from Log 5 were not analyzed because only two talkers from the treatment group completed their logs.

Analysis

The relatively small number of logs from each group meant that the amount of English exposure/use between groups could not be compared parametrically. However, even without parametric tests, the patterns of English exposure and use for each group could be described. Therefore, the data were analyzed in three ways. First, the total minutes of English exposure/use per weekly log and mean number of weekly minutes per talker were calculated for each group. Second, for each weekly log the percentage of time taken up by each of the six categories of use was calculated. Third, aural/oral sub-activities were selected and categorized under two domains - academic and social, and two modes - interactive and non-interactive. The percentage of time spent in each of

these sub-activities out of the total time of English exposure/use per week was calculated. The total and mean number of minutes of English exposure/use per log is shown for each group in Figure 11.

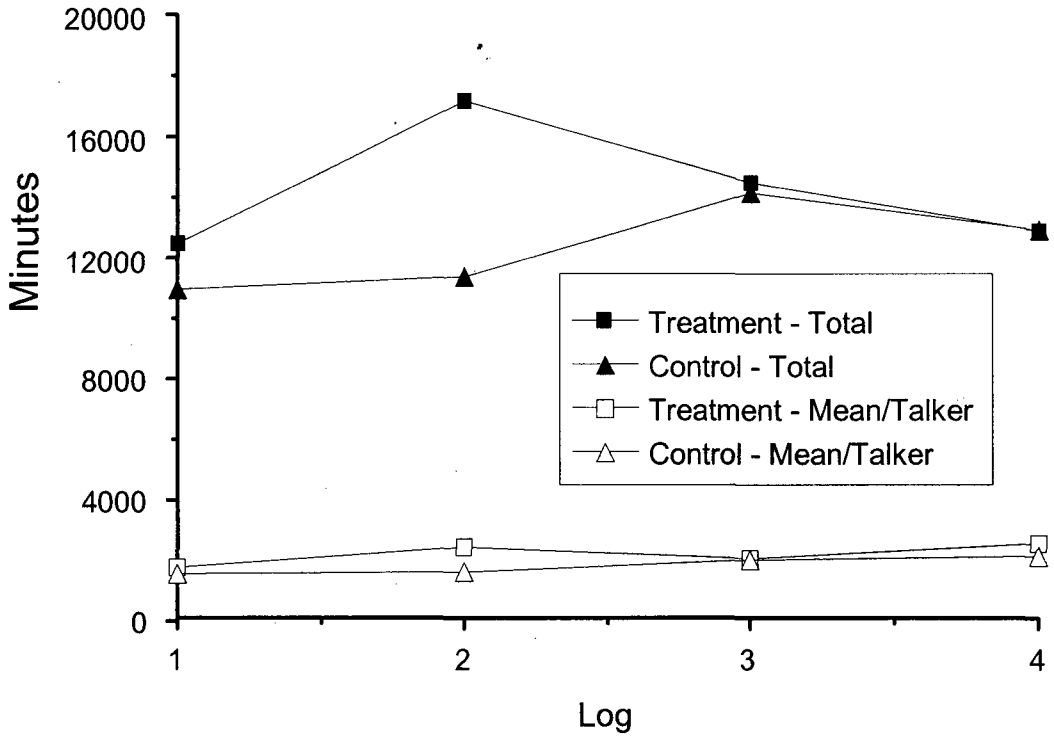

Figure 11. Total and mean English use/exposure for NN groups.

Total and Mean Minutes

The total number of minutes at each log period was higher for the treatment group than the total for the control group, ranging from 341 to 5788 minutes' difference. The mean number of minutes per talker in each log period was also higher for the treatment group (263 minutes on average, ranging from 48-827 minutes).

Percentage of Time in Categories of Use

In each log, there were six general categories of use: daily living, social interaction, academic work, attending class, RA/TA (research and/or teaching assistantship), and recreation. For each of the four log periods, Figures 12 through 19 show the percentage of time for each category relative to the total minutes for the treatment group and the control group's logs, respectively.

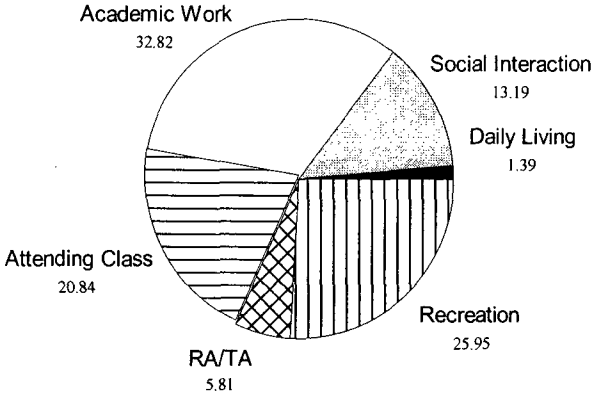

Figure 12. Treatment group log 1.

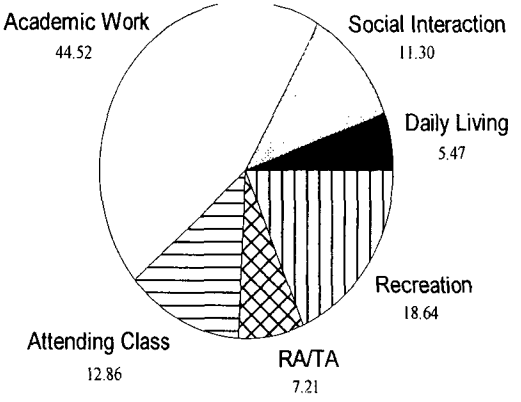

Figure 13. Control group log 1.

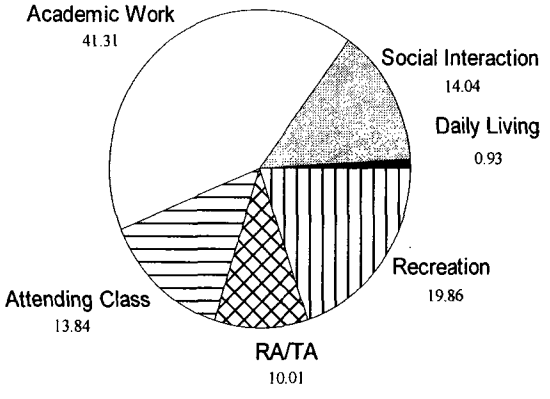

Figure 14. Treatment group log 2.

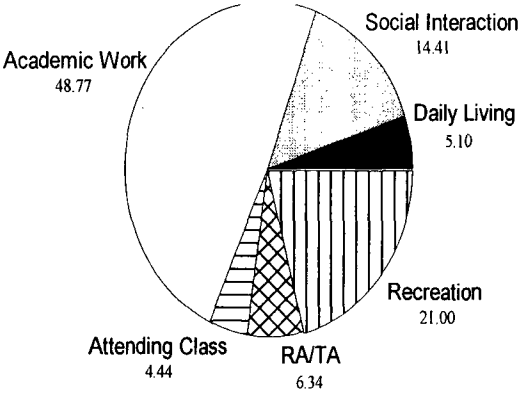

Figure 15. Control group log 2.

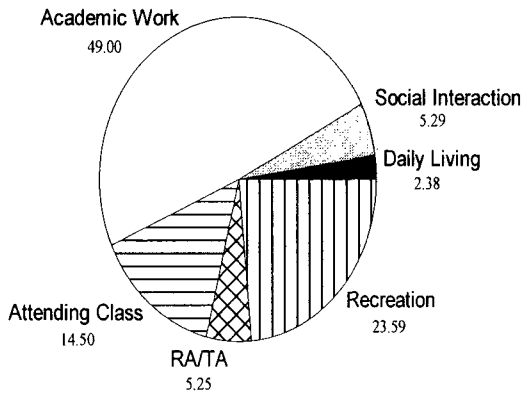

Figure 16. Treatment group log 3.

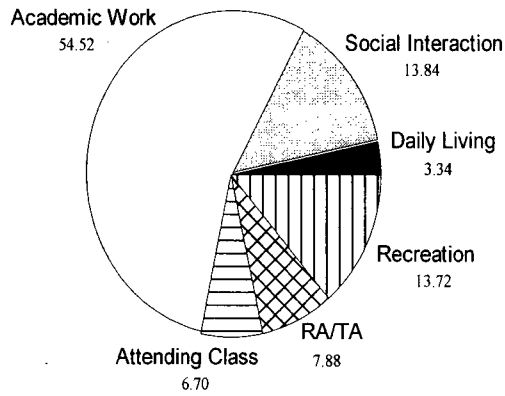

Figure 17. Control group log 3.

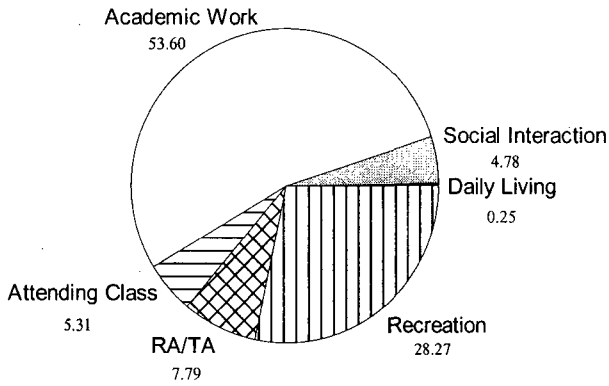

Figure 18. Treatment group log 4.

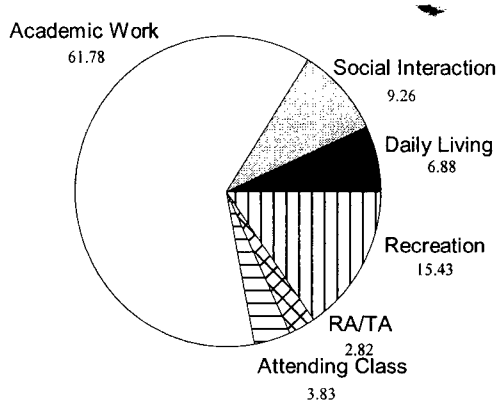

Figure 19. Control group log 4.

The most obvious pattern is that the highest percentage of English is used for the category of academic work, with the category *daily living* often having the lowest percentage. For both groups, as time goes on, more and more of their recorded use of or exposure to English is in *academic work*, starting from percentages in the low thirties and forties and ending up in the low fifties and sixties; this may reflect the more intensive work put in during the end of the semester to study for exams or to complete final papers.

Although both groups were using an increasing percentage of their total English for *academic work* as time went on, the control group was consistently using more of

their total English (5-10 % more) in the *academic work* category than was the treatment group. On the other hand, over most of the logs, the treatment group recorded a higher percentage, often 10% higher or more, of use/exposure in the *attending class* category than the treatment group; this probably reflected the higher number of MA students than PhD students in the treatment group than in the control. The treatment group generally recorded a greater percentage of their total English use/exposure in the *recreation* category than did the control group, usually ranging from 7-10% more.

Domains and Modes of Use

How much of talkers' exposure/use was interactive, when two-way communication was integral to the activity, within the academic and social domains? Selected sub-categories of activities were classified into four areas: interactive social, interactive academic, non-interactive social, and non-interactive academic. Interactive activities were those in which the communication was two-way, such as a discussion or a telephone conversation. Non-interactive activities were those in which much or most of the communication was one-way, such as listening to a lecture or giving a presentation. "Social" here means any activity which was not primarily for an academic purpose. All activities involved a primary focus on aural/oral English. The selected sub-categories and their descriptions are shown in Appendix P.

The patterns of exposure and use for the treatment group and control groups are shown in Figures 20 and 21, respectively. First, the interactive activities will be discussed, then the non-interactive activities.

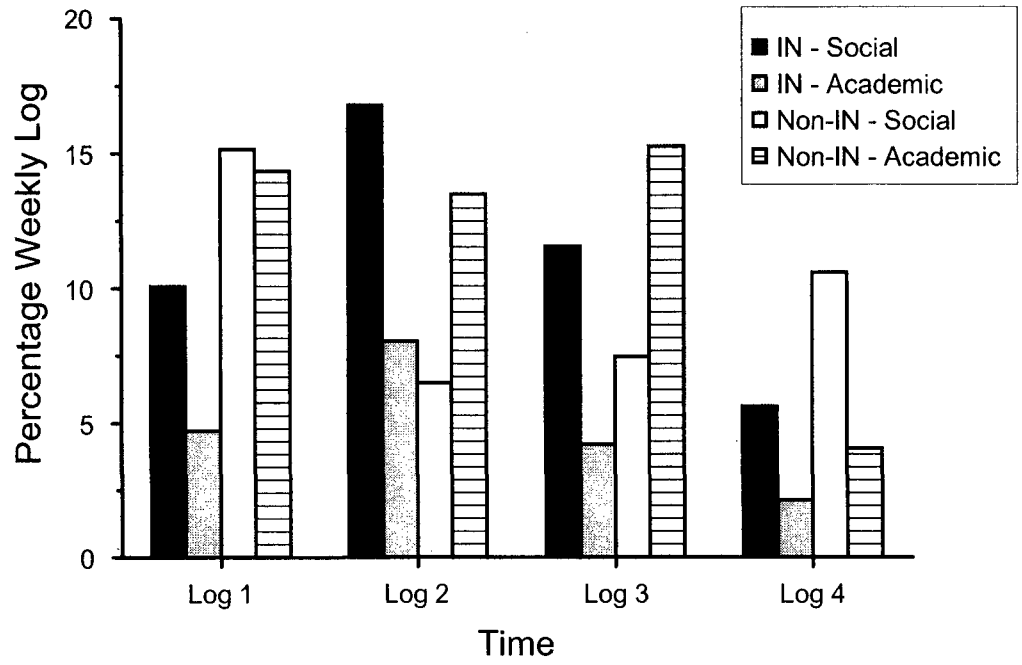

Figure 20. Interactive and non-interactive activities for the treatment group.

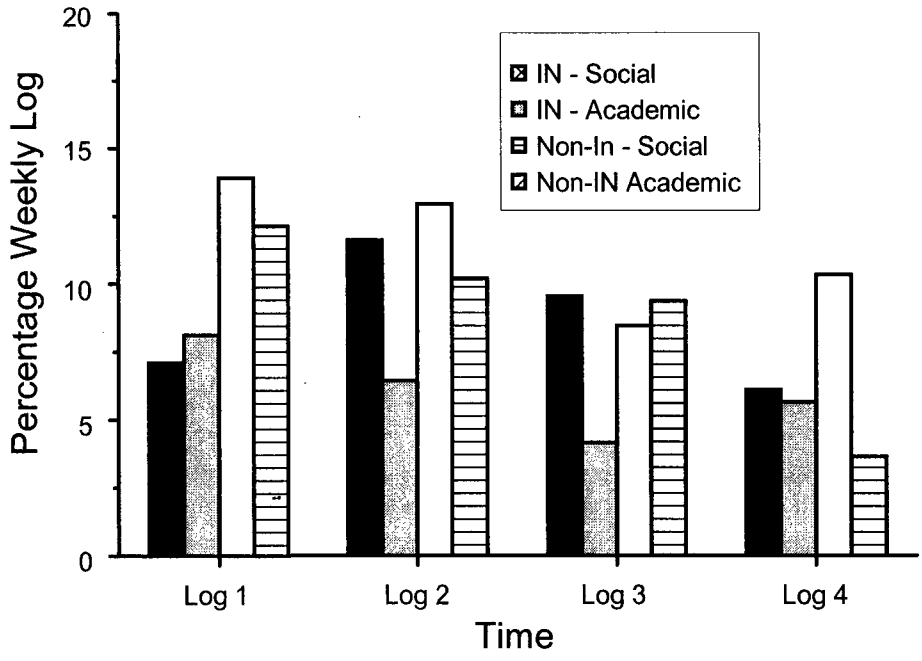

Figure 21. Interactive and non-interactive activities for the control group.

For both groups, the percentage of recorded log time spent doing interactive social activities reached a high in the second log, then dropped, until by the last log, less time was spent on these activities than in the first log. By and large, the treatment group spent somewhat more of their time doing interactive social activities than the control group, usually between 2-5 % more of their time. By contrast, the control group generally spent a greater percentage of their recorded log time than the treatment group doing interactive academic activities, an average of 6.1 % to the treatment group's 4.8 %.

The control group also was more consistent and spent more time on average doing non-interactive social activities than the treatment group, with averages of 11.4 % and 9.9 % of the time, respectively. For non-interactive academic activities, during the first three logs the treatment group spent 1-5 % more time on average on these activities than the control group, but in the last log both groups only spent around 4 % of their time in these types of activities.

Summary: Exposure

The two groups shared some general patterns of exposure and use over the four log periods, but some noticeable differences emerged as well. The total and mean amounts of English exposure and use for each of the log periods were higher for the treatment group. For both groups, academic work was the category with the highest percentage of use/exposure to English relative to overall use/exposure, and for both groups, the percentage increased over time. However, relative to their total English use/exposure, the control group used a higher percentage of English in academic work than the treatment group, both overall and at each log period. Conversely, the treatment

group used more of their total English in attending class and recreation than did the control group.

For both groups, the percentage of time logged doing aural/oral academic activities, both interactive and non-interactive, decreased over time or remained constant. Thus, it appears that the increasing percentage of time both groups spent on the general category of academic work was due more to written than aural/oral work. For both groups, more of their time doing aural/oral academic activities was spent in non-interactive than in interactive ones. Even though the treatment group also spent comparatively more time attending class, however, it was the control group who spent more time in interactive academic activities than the treatment group, while the opposite was true for non-interactive academic activities. The treatment group's higher percentage of interactive social time compared to the control group might have stemmed from the greater amount of time they spent on the general category of recreation. The average percentage of time doing non-interactive social activities was not very different between the groups, but the proportion of time spent by the control group was more consistent.

Non-Native Graduate Students' Views on Language Learning and Language Use

In the results for the three previous research questions, I have examined the nature of the input received by the talker participants, and the output produced by them. However, language learning is not simply an input/output mechanism, but a process which is individual to every learner, each of whom has particular ways of attending to and interpreting their language exposure and use. Learners' own reports about their language learning can serve as evidence of their development; these reports can also

show how learners' thoughts and emotions do not merely *reflect* their decisions and actions in language learning, but can *influence* how learners go about learning a language. The final research question targeted the talker participants' own descriptions and interpretations of their use of and exposure to English.

Research Question 4

What do NN graduate students think about their development in English over a five-month span and during their graduate studies in general?

Procedure

Following the final recording session, the talkers were asked to return for one more session in order to be interviewed, to verify unclear words in the transcripts, and to participate in an individual tutorial. The talkers were interviewed individually in a quiet room. The semi-structured interview covered topics such as the talkers' development in English since arriving at university, their views on the ESL courses they had taken there, language difficulties they had had and continued to have, and techniques and strategies they recommended for progressing in English. Because of the semi-structured nature of the interviews, not all talkers within each group were asked all the same questions, and when talkers raised other subjects which were related to language learning and life as a non-native graduate student, those subjects were also explored. A summary of interview themes is first presented for the control group ($n = 9$), then for the treatment group ($n = 5$). Interviews from Sigman, Christine, and Hui from the treatment group were not conducted, for reasons explained in the Methodology section in Chapter 3.

Control Group

Perceived Improvement

With the exception of one, every talker in the control group judged that their English had improved since they had arrived in Canada. A constant difficulty that almost all had on arrival was understanding what others in their environment were saying. "The things that I was very nervous about is to don't understand what the people say me. Especially when I talk to my supervisor...I was afraid of that maybe he say something and I didn't get it." (Ahmed). Some also had trouble expressing themselves in basic ways. "When I came here...I thought it's difficult to open the mouth to speak with English. It was a big challenge to me. But...I had to open my mouth. No choice." (Ma). However, almost all talkers believed, and some had been told, that their level of English was noticeably higher now than when they first arrived. "Since last year to now, I have improved a lot. There is a professor in [my] department that makes jokes with that. He said me, hey, look at you, you arrived here and ...you weren't able to communicate and now you are speaking an acceptable English." (Lupe).

Some talkers in the control group also stated that their English had improved, either in a stable or temporary way, over the course of the five-month study. Three talkers attributed their improvement in listening or speaking to participating in the study itself because they had engaged in the study tasks. "I just feel every time I come here to stay ...one and a half hours to practise or talking and watch movie and repeat, I think so I can speak a little bit fluently." (Ping). However, three other talkers believed their increased speaking ability or increased confidence was due to their ability to engage in regular conversation and discussion in English with other people, whether at school, at work, or

socially. Xing believed he had not improved in any way from January to May because he had concentrated solely on his studies, but said his abilities in both speaking and listening had greatly improved in the following month, "because in June, my supervisor on vacation," and he had felt relaxed enough to watch movies frequently and talk more to his colleagues in the lab.

Weaknesses and Difficulties

Seven talkers in the control group were asked whether the areas of English they had listed as weak at the beginning of the study were still weak. Three talkers confirmed that their weaknesses remained, but four talkers believed they had improved in their listening, speaking, or presentation skills because of regular exposure to and use of English. "Of course [my listening ability is] a little changed right now because…I heard different people who talks with different accent and different words." (Ahmed).

Talkers were also asked about things that continue to be difficult for them as graduate students. Jiao still found it hard to participate in group meetings or seminars, while two others mentioned the ability to use colloquial rather than formal language. "I just knew that somebody say 'yummy', 'taste good', for me I can say 'taste good', but I think 'taste good' is too formal." Two talkers stated that culturally different thinking styles ("the thinking way and the culture difference") caused problems when they were communicating.

Best Ways of Learning English

All talkers in the control group believed that one of the best ways for them to learn English was to regularly talk to a person or people who spoke it. Two talkers suggested that an English-speaking roommate would be best, though they admitted that

students from their country (China) tended to live and socialize together partly because of shared cultural background. "[L]iving habits is so different." (Ping). Three talkers specified that, ideally, the person should be a native English speaker.

Talkers had taken various steps to maintain or increase their exposure to and use of English. Some talkers watched TV or movies, or listened to the radio to improve their vocabulary and listening comprehension. Other talkers sought out opportunities to interact, by joining a Toastmasters public speaking club (Ahmed), committing to speak English with people at least one hour a day, even with friends of the same L1 (Lupe), or trying to strike up conversations on the bus (Ma). Many talkers, especially those from China, felt that their level of interaction in English was not high enough. "I think I still need ...much more chance to speak in English." (Ma). Esteban, who was very satisfied with his own level of ability in English, maintained that becoming better in English was a matter of practice, "but you also have to be willing to learn, because again, I have another friend at school...I think he has been here longer than me and his English still bad. And that's because he doesn't practise... I think he only surrounds by [people from his country] so there's no way that you can practise your English like that."

One of the principal reasons that control group talkers gave for not interacting as much as they would like was lack of time because of the demands or conditions of their graduate work. "We have to work and I spend...maybe more than eight hours a day in the lab [..].But you know, you cannot keep on talking in a lab." (Jiao). Xing had made a conscious decision to give his program of study priority over work on English, and found that his language ability had actually worsened over his first few years here. As mentioned above, another explanation cited was differences in culture, while at least two

talkers said that sometimes other people were not willing to talk to them. "I'm not attractive to the other people." (Xing).

Communication Breakdowns

Breakdowns in communication were reported by many talkers; Ahmed, who was satisfied with his command of "technical English", wryly noted, "but it doesn't help you when you go for shopping." However, some talkers were at times very distressed by their difficulties in understanding or communicating a message, especially because they thought their difficulties with language obscured their professional competence.

> "For example...you discuss some rather complicated problems with some colleagues, with professors...we obviously have some disadvantage because I cannot express me more clearly and I tend to be not, not say anything before I really organize this thing...Sometimes I just sitting here and looking at some people just talking very fast and I don't know, it seem I don't know this thing totally [because he did not say anything] but actually I know something...It still feel very, very bad." (Jiao).

Even when there was no communication breakdown, Lupe believed that his noticeable accent led "local guys" to view him unfavourably. "Because that puts you in a bad position always, all the time. You can feel that, you can perceive that from Canadians, that they don't like very much immigrants." However, he believed it was his responsibility to learn to speak "with fluidity" so that he could make a good impression.

Oral Communication Course

Five out of the eight control group talkers had previously taken the oral communication course two semesters or more before. They were asked about the course

and what they had learned from it. Three talkers believed that the course had helped, but two talkers were less positive. Ma said that regular interaction was more important than the course, because the course only provided rules which he subsequently forgot. Jiao first stated that the material itself was not helpful, but noted that taking the course pushed him to work on his English, though his subsequent improvement was not necessarily due to the course. However, Jiao later discussed what he had learned about lexical stress and focus words, and said that at least by the end of the course students had improved, even if they may have forgotten these things after the course. For Esteban, the course had helped him realize that his English was not good, and that there were different and better ways of pronouncing words. Xing viewed the course as very useful, and had learned that listeners could understand imperfectly pronounced words if the stress was correctly placed. Finally, Feng believed the course had been very helpful for learning presentation skills, and he had made use of those skills in subsequent classes.

Treatment Group

Perceived Improvement

Everyone in the treatment group believed their level of English had improved from the time they had first arrived at university. Three talkers thought they had improved listening comprehension, with a caveat from Javier that he had understood non-native speakers better when he first arrived, but now found it easier to understand native speakers. Three talkers judged their overall oral communication, both listening and speaking, to be better, with two also mentioning their writing. Four of the talkers also explicitly stated that their greatest difficulties with understanding and being understood

had been in the first three to four months of their university careers, and after that their studies and interactions had become noticeably less difficult.

Over the four months of the oral communication course and one month more, only Xiao stated that her English had not improved in some way, though later she rated herself slightly higher on her listening skills than she had in January. Bao also maintained that she felt her English was better only when she had a lot of practice speaking and listening to a person who spoke fluent English, and in the past one or two months she had not had that experience. Both Javier and Tai Ning felt that their listening comprehension and writing had greatly improved since January, with Javier also mentioning, "and I feel more confident when I talk, but I don't know if I'm doing better than before." In Marie-Pier's opinion, she had improved her presentation skills and her pronunciation was a little better, but she assumed that her problems with pronunciation would remain whatever she did. She stressed that she was now equal to dealing with any situation in English, but attributed her progress to a high level of interaction with English speakers.

Oral Communication Course

The treatment group talkers were asked what they had learned from the oral communication course. Two talkers said that, contrary to what they had thought, they had learned that they actually didn't have good pronunciation. However, most talkers mentioned aspects of suprasegmental pronunciation, such as linking, rhythm and intonation, thought groups, and lexical stress. Piotr and Tai Ning had followed the instructor's advice to slow down their speech, to first think about what they would say and then say it. Piotr mentioned that this had also helped him notice problems with his use of grammatical tenses, since he was paying more attention to his speech. Tai Ning

also said that before the class, people had not understood him because he put stress in the wrong places. After he had learned about proper stress placement, he improved in his ability to communicate. "I kind of correct my pronunciation by stress, and then they will know it." The short-lived effects of the instruction were noted by several talkers. "When I take the course for pronunciation, I think I improve, I learn a lot, but after the course finish... at that time when I speak, I will pay attention to the pronunciation, but now mostly I forgot." (Bao).

Some people in the treatment group also stated that they had had some different expectations of the course, of more discussion between students and instruction in vocabulary or useful phrases. "I was maybe expecting a bit more communication, like more general talk, but that's true that the class is for grad student and we have to do presentation and that kind of stuff." (Marie Pier).

When treatment group talkers discussed working in groups with other non-native speakers during class, opinions were mixed. Two talkers felt that their speech had been adversely affected by some of their partners, or that they had not received accurate feedback on their own production. "In class, you're doing the exercise and the prof is not next to you, and you got...the other student who say, no, no, it's like that. And after...sometime they are wrong also because no one really know." (Marie Pier). Javier had experienced similar problems in one section of the course, but then switched to another section, where he was much happier with his partners. However, Tai Ning believed that group work was helpful throughout his courses, not just the oral communication course. "I think discuss with my classmates is very helpful to understand knowledge and to finish the assignments. It's very useful."

All talkers in the treatment group believed that the exercises on CAN-8 were useful for practising what they had learned. However, most reported that their use of CAN-8 had decreased over the semester, usually because they had to go to the lab in order to use it, or because the lab had been too crowded when they did go. One student suggested, "The best improvement that ...CAN-8 could be...on the WebCT or something like that, then you can do it in your office. Because I'm not going to go all the way down to practise." (Marie Pier).

Difficulties

As with the control group, some talkers spoke about persistent difficulties expressing their message using appropriate vocabulary. "When I want to express myself, I don't know how to begin, which is the way accepted for the native speaker." (Xiao). However, Javier, at least, was becoming accustomed to not being able to speak his thoughts as quickly as in his L1. "Because Spanish is my first language, it's different, I used to think and speak at the same time. But I'm getting used now to, to take time, think what I want to say, and then say it."

Best Ways of Learning English

Unlike the control group, not everyone in the treatment group believed that the best way for them to learn English was to interact with other speakers of English. This may be partly due to their patterns of interaction, as will be discussed below. Xiao thought that another course, one that focused on how she could express herself in a nativelike way, would be ideal. Both Javier and Piotr watched documentaries or movies, and spoke the narration or dialogue back to the TV. Piotr made maximal use of this technique by watching some of his favourite movies over thirty times, to the point that he

could recite whole stretches of dialogue. "This is nice thing, for vocabulary I remember whole sentences or logical phrases...I know when I can apply it."

The other three talkers thought that speaking to someone in English was the best way to learn. Bao preferred talking with someone to watching TV because she could concentrate better when she had to respond. For Marie-Pier and Tai Ning, speaking with other people was the natural way to learn. Tai Ning believed that a native English speaker provided the best opportunity for increasing his knowledge of English. "They will say something and kind of explain themself and chat something. And ...sometime they will explain the same meaning as I did before. But I found that it's a different expression and I know that maybe his way is right, so I correct myself."

Compared to the control group, fewer talkers in the treatment group expressed concern over the level of their exposure to or use of English. Javier reported that because he was completing the language log, he realized that he had to look for ways to use or hear English. Originally, he had modelled his English on that of non-native colleagues because he could understand them, but when he realized that they were making lots of mistakes, he began using TV programs as models, and was quite satisfied with the results. Both Tai Ning and Marie-Pier reported having active social lives in English, and Piotr joked that he was someone who would rather talk than work. Only Bao expressed frustration over her current exposure to English, since she was at an institute where most of her colleagues were francophone and most of the academic meetings were in French. She used English mostly after work in social outings with colleagues and friends, but she was not satisfied with the input she was receiving. "In my life, there are francophone and their English is not very good. If I talk with them, sometimes I follow them, their English

like that, like French English." Bao had had a native English boyfriend and had found it was very good for her English, but the relationship ended and she feared she was losing her ability to speak and understand English.

Communication Breakdowns

Perhaps because they had more regular interactions with other people, talkers from the treatment group put more emphasis in their interviews on their abilities to overcome communication difficulties than on the difficulties themselves. Piotr laughingly reported a professor in his department who made funny faces when he did not understand what Piotr was saying, which reminded Piotr to slow down and speak more clearly. However, he still had problems keeping his sentences short and simple enough for people to understand. Javier described his evolution in lab meetings. "At the beginning, I went to a couple of meetings with the rest of the group. And it was frustrating because first I couldn't understand. After a while I started to understand what they were saying, but I couldn't speak....but now I can give my opinion." Tai Ning as well said that he could communicate much better, both because he had practice talking to friends and because he had learned to think before he spoke. Bao also said she now had few problems understanding people, although her shyness kept her from having frequent long conversations. Marie-Pier provided a striking illustration of how far her communication abilities had progressed. At a departmental retreat in October of her first year, she could not communicate with anyone except the one other francophone there. By the end of her fifth semester, she had served on the departmental student council, she spent all her social time with English-speaking friends, and she had won an award at the departmental banquet, "the life of the party prize". Only Xiao reported continued difficulties with

speaking to others, especially in class discussions, where she was afraid that no one would understand her and she had trouble keeping up with the topic of discussion. "When I thought of the idea, the question has already passed."

Help for Non-Native Graduate Students

Talkers from both groups were asked how non-native graduate students could be helped with their language difficulties at university. Not surprisingly, most suggestions involved formalizing opportunities for interaction. Ma suggested that the government somehow give immigrants more chances to "communicate with the nation," so they could better immerse themselves in society and learn more about its culture. Ping and Jiao proposed organizing specific times for international graduate students to meet each other and discuss some topics, with local English-speaking students also sitting in. Marie-Pier's ideal language-learning situation somewhat resembled the "international student buddy" system at some universities, where the language learner could follow around a proficient speaker. Piotr thought merely "more time" would be helpful, while Xiao hoped for a course where students could better learn how to express themselves appropriately.

Summary: Non-native Graduate Students' Views

It is clear that both groups of students had many of the same concerns. Being able to understand others and to be understood was a significant problem when they first arrived at university, though almost all believed their communication skills had improved since their arrival. Most people believed their skills in English had improved even over the five months of the study, though some in the control group attributed it to participation in the study itself, while for others it was due to the effects of interaction, exposure, and for some, instruction over that time period. The great majority of students

who had taken the oral communication course believed it was useful and a good course for them. Some thought that its effects were not long-lasting, but they appreciated learning more about communication in English, even if they did not apply what they had learned. There were mixed reactions about the level of group work done with other non-native speakers in the class, but everyone asked agreed that the CAN-8 exercises were useful, though not convenient to use.

Overall, talkers in the treatment group expressed fewer concerns about their communication skills and exposure to or use of English than talkers in the control group. They tended to emphasize how far they had come rather than the difficulties they continued to have with communication. In the control group, some talkers believed their communication difficulties were current and at times severe. These same talkers were not happy with their levels of interaction with or exposure to English speakers, but felt they lacked both time and opportunity to increase those levels.

When students were asked how non-native graduate students could be helped, most suggestions addressed conversational interaction between people, and how to increase the opportunity for those types of interactions. Discussion groups and "buddies" were suggested, but a course focusing on appropriate and idiomatic vocabulary and expressions was also proposed.

Summary of Results

The following is a brief summary of results to all four research questions.

Research Question 1

What is the nature of classroom instruction for NN graduate students in an ESL oral communication course emphasizing intelligibility?

The instruction focused on suprasegmental aspects of pronunciation, communication strategies, and presentation structure. Students were encouraged to induce pronunciation patterns from models, and practiced the patterns using form-focused, sentence-length exercises. The presentations also provided opportunities for pronunciation patterns to be applied. Michelle, the instructor, did not expect students' pronunciation to change noticeably over the time period of the course, but did expect students to become aware of pronunciation patterns and to learn strategies so that their pronunciation could develop in the future.

Research Question 2

How does the speech of NN graduate students develop, in terms of its intelligibility and evaluation by listeners:

(a) from the beginning to the end of an ESL course targeting oral communication?

(b) when not taking an ESL oral communication course, but only studying in an English environment?

Intelligibility differences did not appear over time in measures from either of the two tasks. Intelligibility differences between groups were not consistent in measures from the retell task, with the treatment group sometimes as intelligible as the native speaker group. For the on-line task, however, the native speaker group was always more

intelligible than both NN groups, and the treatment group was more intelligible than the control group either at Time 1 or overall.

For listener evaluation, the control group was consistently rated more fluent at later times than at earlier times. The native speaker group was always rated more comprehensible, less accented, and more fluent than both NN groups, while the treatment group was rated more comprehensible and more fluent than the control group.

Research Question 3

In what ways do NN graduate students at an English-medium university come into contact with and use English on a weekly basis?

The treatment group had a higher total and mean English exposure/use over the four log periods than the control NN group. However, for both groups, their highest amount of exposure/use was in academic work, though the control group spent relatively more time on this than the treatment group. The treatment group spent relatively more time in recreation and aural/oral social interaction, while the control group spent more of their aural/oral interaction time in academic activities.

Research Question 4

What do NN graduate students think about their development in English over a five-month span and during their graduate studies in general?

Almost every student in both groups believed their English had improved over the course of the research study and during their graduate studies. Some control group students believed, though, that their recent development was due to participating in the study. Most students who had taken the oral communication course believed that it was useful, though they may not have thought their learning had persisted. Students in the

treatment group expressed fewer concerns about their communication skills and English exposure/use than did the control group students.

In this chapter, I presented results for the four research questions. In the following chapter, I will discuss the significance of and possible explanations for the results.

CHAPTER 5: DISCUSSION OF RESULTS

Overview

As outlined in Chapter 2, a number of studies have described the effects of instruction targeting L2 speech on L2 learners' oral production. However, the design of those studies did not meet at least one of three important conditions: (a) thorough observation of the instruction over time; (b) detailed measurement of L2 learners' production over time; and (c) at least one comparison group of L2 speakers in order to identify changes in L2 speech due to non-instructional factors. The current thesis study was designed to meet these three conditions, leading to the following two research questions: (a) what is the nature of classroom instruction for NN graduate students in an ESL course targeting oral communication?; (b) how does the speech of NN graduate students (treatment group) develop, in terms of its intelligibility and evaluation by listeners, from the beginning to the end of an ESL course targeting oral communication? *and* how does the speech of NN graduate students (control group) develop, in terms of its intelligibility and evaluation by listeners, over the same time period when not taking an ESL oral communication course, but only studying in an English environment?

Researchers studying the longitudinal development of L2 speech have tended to observe and measure *either* classroom instruction *or* non-instructional exposure to the L2. In addition, L2 learners' own views on their development have not often been solicited. This means that researchers have an incomplete picture of the factors which may have influenced L2 speech development. In the current study, in addition to instruction, I investigated two other factors (exposure/use and learners' own views on L2 learning),

which may have influenced the development of L2 learners' speech. These factors were investigated through two research questions: (a) In what ways do NN graduate students at an English-medium university come into contact with and use English on a weekly basis?, and (b) What do NN graduate students think about their development in English over a five month span and during their graduate studies in general?

The first research question, about the nature of classroom instruction, was investigated through observation of 8 classes of an oral communication course for NN graduate students, using Part A of the COLT observation scheme, videorecordings, and field notes, as well as a semi-structured interview with the course instructor. The second question, about the L2 speech development of NN graduate students, was investigated through measurement of listeners' understanding and evaluation of NN graduate students' longitudinal speech samples. The third research question, about NN graduate students' English exposure and use, was investigated through NN graduate students' logged exposure to and use of English over time. The fourth research question, about NN graduate students' views on their English development, was investigated through semi-structured interviews with NN graduate students.

This chapter summarizes and discusses findings for each of the four research questions.

Nature of Classroom Instruction

Research Question 1

What is the nature of classroom instruction for NN graduate students in an ESL course targeting oral communication?

Course Design

The course was designed to help non-native graduate students improve their knowledge of and skills in English communication, with a particular focus on suprasegmental pronunciation and research presentation skills. One notable aspect of the course was the freedom the instructor had in the design and implementation of the course. Basturkmen, Loewen, and Ellis (2004) describe teachers as "people who construct their own personal and workable theories of teaching" (p. 244); Michelle certainly had based the course and her teaching on her understanding of how languages are best learned and taught. Changes to the course were of course restricted by institutional and organizational factors, but Michelle had been able to experiment with topics and activities, choosing the ones which she believed were most effective and/or were supported by research. Sato and Kleinsasser (1999) found that the language teaching knowledge and classroom practices of Japanese as a second language teachers were mostly based on their own beliefs and experiences. In the current study, Michelle did draw on her experiences in the classroom when making pedagogical decisions; however, the choices she made were based not only on her own experiences, but also on research she had read and on her professional contact with teachers and researchers in the field. In this way, she actively worked to articulate and integrate explicit principles of teaching and learning in her course design.

Focus on Suprasegmentals

Michelle's focus on suprasegmental aspects of pronunciation was supported by research cited in Chapter 2, in which more nativelike production of suprasegmentals was an important factor in better intelligibility and rating of L2 speech; in addition, several studies cited in Chapter 2 showed that ESL learners trained in suprasegmental pronunciation improved in both non-spontaneous and spontaneous speech. However, in spite of the training received by the treatment group talkers in this study, there was no clear evidence of substantial improvement in their intelligibility or listener evaluation in spontaneous speech.[5] Interestingly, the instructor herself believed that it was unreasonable to expect that students' communication skills or intelligibility would noticeably improve by the end of the course, since the period of instruction was not sufficient. Instead, the course was meant to provide students with a base of knowledge about suprasegmental pronunciation and learning strategies which they could later use to improve their skills on their own.

As mentioned above, there is no clear evidence that instructed talkers became more intelligible because of instruction. These results may be due to the types of tasks and measures used to assess intelligibility, a factor discussed at the end of this chapter. However, the results may also show that pronunciation teaching which focuses almost completely on suprasegmental aspects is not an unconditional solution to improve L2 learners' intelligibility. Numerous pronunciation researchers, methodologists, and teachers currently advocate a primary focus on suprasegmentals in pronunciation

[5] Both the treatment group and control group also recorded isolated, non-spontaneous sentences, but these were not heard by the listener groups. The treatment group's presentation skills, which were another primary focus of instruction, were not evaluated in this study, so it is not possible to discuss any potential improvement in that area.

instruction (Celce-Murcia, Brinton, & Goodwin, 1996; Morley, 1999; Wong, 1987). However, research-based evidence of its effectiveness (either in itself or compared to other approaches) is far from exhaustive, as seen in the literature review and in the results of the current study.

Time

In Chapter 2, several studies described ESL courses in which learners' intelligibility and/or evaluation by listeners significantly improved (e.g., Derwing et al., 1998; Stevens, 1989). However, one noticeable difference between those courses and the oral communication course in the current study was the class time available. The other courses were intensive, usually having classes every day or every other day for at least three hours each class, for a period of at least a month. This intensive time was not available in the current oral communication course, which had two 90-minute class meetings a week for 14 weeks; the distribution of class time may play some part in explaining why the intelligibility and listener evaluation of the treatment group did not greatly change over the duration of the study.

As Michelle herself noted, the time-course of the instruction was too short for students' oral skills to noticeably develop. She believed that additional time was so important that she developed another follow-up course, not so that students could be taught more, but simply in order to increase students' opportunities for exposure and practice using the guidelines for suprasegmental pronunciation they had already learned. In fact, Michelle believed that the best learning conditions for students included sufficient time to use the guidance they had received about important aspects of pronunciation.

Numerous studies have shown that when adult L2 users arrive in their L2 environment, the degree of their L2 accent is affected by how far away they are from the age of puberty at arrival, and not how long they stay in the L2 environment (Jia, Aaronson & Wu, 2002; Oyama, 1976; Piske, Mackay, & Flege, 2001). This means that, even after years of naturalistic L2 input, some adult L2 learners' pronunciation may be as accented as it was when they arrived (fossilization). Therefore, adult L2 learners may need long and/or concentrated amounts of time for pronunciation instruction and guided practice in order to develop their L2 pronunciation; a semester-long course twice a week may not provide sufficient time for the pronunciation (or intelligibility) of adult L2 learners to develop.

Methodology

In both the teaching methodology typically observed in classes and the instructor's stated preferences for language learning and teaching, language learners were seen primarily as analyzers of language, whose learning was best served if they were allowed to first try to induce patterns of suprasegmental pronunciation from selected models (see also Borg, 1998). This was not as true for the teaching of presentation structures and strategies, where students were often presented a strategy or discourse structure from the outset, with subsequent discussion of the reasons for using it. Because of this difference, the approaches to teaching suprasegmental pronunciation and presentations will be discussed separately, beginning with suprasegmental pronunciation.

Suprasegmental pronunciation. The instructor provided instruction and feedback which focused on producing English in a more nativelike way. This instruction and feedback was meant to give students accurate and relevant information so that they

themselves could analyze and eventually produce English in a more nativelike way. Michelle's stated target for corrective feedback was aspects of pronunciation which had been or were being practised. Typically, the students did receive feedback on suprasegmental aspects which had been taught or were the current focus of the class. In ESL classes observed by Basturkmen, Loewen, and Ellis (2004), teachers' classroom practices in focusing on form often conflicted with their stated beliefs. However, in the current study, the instructor's stated goal for providing corrective feedback was generally quite consistent with her targets for corrective feedback in the classroom.

Students were expected to use their analytic skills to monitor their own production, as well as to provide feedback on the production of other students. Michelle, the instructor, believed that this simultaneous focus on analysis and production, rather than simply mimicking models, helped students to remember and apply the pronunciation patterns better, and allowed them to practice on their own or without the aid of an instructor. This, it must be noted, was what Michelle hoped students would be able to do by the end of the class, even if their oral production had not noticeably changed. Indeed, in assignments and presentations, students were partly evaluated not just on their oral production, but also on how well they could identify and predict in written texts the pronunciation patterns they had learned.

Form versus meaning. Concurrent with the emphasis on analysis of language was the tendency for class practice activities to focus primarily on the form of pronunciation, with much less emphasis on activities where messages had to be communicated. Michelle did discuss with students how nativelike or non-nativelike use of pronunciation patterns could have consequences for listeners' understanding. However, the practice exercises

usually involved reading aloud words, sentences, or paragraphs. When reading the words or sentences, students usually concentrated on producing suprasegmentals in a more nativelike way, and not on successfully communicating a certain message.

The nature of the suprasegmental activities may have been partly a function of the textbook used, whose exercises were primarily focused on the form of pronunciation. In contrast, some textbooks on pronunciation teaching methodology (e.g., Celce-Murcia, Brinton, & Goodwin, 1996), advocate a multi-step process of awareness-raising and practice activities, with initial activities having a predominant focus on form, and subsequent activities demanding that the learners focus on both form and meaning. In the oral communication course, this combined focus on form and meaning did occur in activities related to presentations.

Presentations. The presentations were envisioned as tasks in which the students could apply their knowledge of pronunciation patterns, creating their own message, deciding how to structure and produce it, and conveying this message to the audience. The classroom activities for presentations thus included much more focus on the meaning and use of language, such as selecting appropriate comprehension checks at a given point in a presentation.

Unlike the introduction of suprasegmental aspects, when features of presentation structures or strategies were introduced, students were assumed to have some previous knowledge of them, often leading to discussion of the suitability or meaning of particular behaviours or structures. The structure of presentations was sometimes justified by convention (e.g., "This is how it's done" or "This is how I want you to do it"), but the use

of presentation strategies was always explained by their effect on the audience (e.g., in North American culture, eye contact is a form of connection).

Written activities and exercises. As mentioned by Michelle in the interview, the presentations were included in the course partly because of their relevance to graduate students' academic work, but also as a way of allowing students to put their knowledge about suprasegmentals into practice while communicating a message at the same time. Michelle required students to write out their entire presentation in order to help them figure out how they should apply the suprasegmental guidelines they had learned. They could then use their written-out presentation while they were orally presenting.

The use of these written notes may also go some way towards illuminating why no change was seen in treatment group talkers' intelligibility scores or listener evaluations in spontaneous speech. For students' classwork—the practice exercises, the recorded reading assignments, and the presentations—students did not need to simultaneously generate a message, encode it in some form, and produce the form. This means that when orally producing the sentences, students could concentrate simply on appropriately pronouncing the sentences they read. In contrast, when the treatment group talkers recorded stories for the current study, they could think about the story and prepare notes for up to five minutes before recording the story, but they were not allowed to use their notes while speaking. In other words, in telling the story, they had to attend not only to orally producing their message (pronunciation) but also to the organization and encoding of the message. The processing requirements of the speaking task in the study were therefore different than the requirements of tasks and assignments in the course. The

possible consequences of these different requirements are discussed in the following section.

Transfer-Appropriate Processing

The consequences of processing similarities or differences between cognitive tasks are addressed in the *transfer-appropriate-processing framework*, which postulates that people perform better in a test task when the processing operations required are similar to those required by a (previous) study (i.e., learning) task, than when the operations required are different (Morris, Branford, & Franks, 1977). Since the story-telling tasks in the study required more, processing-wise, from the treatment group talkers than the speaking tasks and assignments in the course did, it is not surprising that any improved performance in the course (learning) tasks would not transfer to the story-telling (test) tasks, for reasons explained below.

Gatbonton and Segalowitz (2005) and Trofimovich and Gatbonton (2006) propose a framework for language teaching which provides learners with repetitive practice while sending and receiving messages and maintaining a focus on the form of production, including pronunciation. They state that while learners are repeatedly producing a range of utterances in a genuinely authentic exchange of information, the meaning-related properties of the utterances become more familiar, which allows the learners to focus more on improving the form of the utterances (Trofimovich & Gatbonton, 2006, p. 531). Thus, when learners are called upon outside the classroom to generate, encode, and orally produce messages, they are able to transfer their learning of pronunciation to these real-world tasks because of their previous practice with classroom tasks which had the same processing requirements.

In the oral communication course, the course tasks encouraged controlled and repetitive practice of the suprasegmental patterns, which allowed learners to focus on the form of their pronunciation. However, messages which were both authentic and spontaneous were not being transmitted, meaning that the learners did not have to attend to both the form and meaning of their speech simultaneously. By contrast, authentic communication tasks outside the classroom require that learners focus on both the form and meaning. Because learners did not have much practice in dividing their attention this way in the classroom, it may help to explain the lack of improvement over the study's duration in the intelligibility of their spontaneous speech.

Summary: Discussion of Nature of Classroom Instruction

The oral communication course was carefully designed and structured, with topics of instruction which were motivated by research findings and learner needs, and teaching methodology which actively involved students in developing their knowledge and analysis of pronunciation patterns, of presentation discourse structure, and of self-monitoring skills and presentation strategies.

Michelle, the instructor, noted that she did not have great hopes of seeing substantial change in students' intelligibility, based on the modest amount of time that the students were exposed to instruction. Another factor which may have worked against changes in intelligibility was the different processing requirements between the course tasks and the study tasks. In many course tasks, the students focused primarily on the form of pronunciation, and for all course tasks, speech was read aloud, not spontaneously produced. For the study tasks, students were required to convey authentic messages in telling their stories, and their speech was spontaneous, not read-aloud.

The NN treatment group talkers received instruction targeting oral communication, motivated by previous research and implemented by an experienced, skilled instructor. It would be expected that the treatment group talkers might show some noticeable improvement in their speech. However, detecting improvement in L2 speech depends on the ways in which L2 speech is measured. Therefore, we next consider the results for the research question on the intelligibility and listener evaluation of NN talker speech, along with the speech of the native speaker group.

Intelligibility and Listener Evaluation

Research Question 2

How does the speech of NN graduate students develop, in terms of its intelligibility and evaluation by listeners:

(a) from the beginning to the end of an ESL course targeting oral communication?

(b) when not taking an ESL oral communication course, but only studying in an English environment?

Three different listener tasks, five different intelligibility measures, and three different ratings measures were used to investigate these questions. I will first discuss the results for intelligibility and the various measures used to explore it, then listener evaluation and the various ratings used to measure it. For each intelligibility measure, the results for the two talker groups and for the individual talkers in each group will be discussed together. For the listener evaluation measures, only the results for the two talker groups will be discussed, since listeners' evaluation was analyzed by group and not by individual talker.

Intelligibility

The results for the intelligibility of groups and individuals differed depending on the tasks and measures used. There are at least two ways of interpreting these differences between measures. One interpretation is that some measures were not sensitive to intelligibility differences between speakers or over time, or were influenced by other factors which interfered in the assessment of intelligibility. Another possibly complementary interpretation is that speech can be intelligible at different levels of discourse, and particular measures were sensitive to intelligibility differences only at particular levels of discourse. Following the discussion of the intelligibility results, these two interpretations will be reviewed.

Accurate Content

Accurate content was measured by a ratio of the number of story elements accurately retold divided by the total number of original story elements. Listeners' retells of the non-native speaker groups' stories were found to have relatively more accurate content than listeners' retells of the native speaker group's stories. Therefore, it seems that this particular measure of intelligibility does not specifically target the characteristics of *L2 speech* which might cause problems for listeners' understanding. Because this measure did not distinguish between native and non-native talkers, it was not used for individual talker analysis.

The result for accurate content might seem counter-intuitive. Why might retells of the NN stories contain more accurate content than retells of the native speaker stories? Together with the nature of the listening task, the length of the stories told by each group

and the speech rate of the talkers may have influenced the amount of accurate content retold.

As seen in Table 27, the native speaker group did not always have the highest mean length of stories (in minutes) at each recording time. Some NN talkers spoke for as long or longer than the native speakers. However, at all recording times, the native speaker group had the highest mean *speech rate*, meaning that generally, a minute of a native speaker's story would have comparatively more information (content) than a minute of an NN talker's story.

In order to retell the story, listeners took notes while listening. With the native speakers' higher speech rate, listeners may have not been able to write down as much content as they had for the stories told by NN talkers, who spoke more slowly. If listeners had written down relatively less content for native speaker stories, they may not have been able to retell as much of the stories, thus leading to lower ratios for accurate content. It was only possible to use the measure of accurate content with the retell task, so this explanation cannot be checked against a similar measure for the on-line intelligibility task.

Table 27

Story Length and Speech Rate

Group and talker	Time 1		Time 3		Time 4	
	Length[a]	Rate[b]	Length[a]	Rate[b]	Length[a]	Rate[b]
Treatment						
Javier	1.48	102	3.40	118	2.77	121
Hui	0.90	121	2.83	118	2.28	126
Jiao	1.00	131	1.67	101	1.67	117
Mean	1.13	118	2.63	112	2.24	121
Control						
Lupe	2.20	130	3.62	128	2.42	133
Ping	1.43	115	4.62	118	6.02	115
Feng	5.20	74	1.93	77	4.48	69
Mean	2.94	106	3.39	108	4.31	106
Native Speakers						
Aileen	1.61	175	1.62	193	1.39	177
Brenda	1.95	191	5.22	179	3.44	192
Daniel	2.62	175	3.05	167	3.72	146
William	0.82	196	1.20	187	1.00	179
Mean	1.75	184	2.77	182	2.39	174

Note. [a]in minutes. [b]measured as words per minute.

Inaccurate Content

Inaccurate content was measured in both the retell task and the on-line intelligibility task. The results for each task will be discussed in turn.

Retells. In measuring inaccurate content in story retells, the results which were obtained depended greatly on the nature of the inaccurate content measured. When the analysis included all retell elements with any inaccurate content at all, even details, no differences appeared over time or between retells from different speaker groups. However, when inaccurate *details* were removed from the analysis, differences appeared. This indicates that listeners retelling stories by native speakers are not inevitably more accurate in all respects than listeners retelling stories by non-native speakers. One reason that listeners may retell details inaccurately is because they do not understand a non-native storyteller's speech; however, details in native speakers' stories (whose speech should be quite intelligible) were also inaccurately retold.

When inaccurate details were removed from the analysis, retells of NN control group stories had more inaccurate content overall than retells of stories from the native speaker group. However, the amount of inaccurate content in NN treatment group retells and native speaker retells was not significantly different overall. These results may mean that measuring inaccurate content in this way does not tap intelligibility differences which are based only on the nativeness or non-nativeness of the speakers. On the other side of the coin, the results might also suggest that for extended and authentic speech, when inaccurate understanding of *content* is the measure of intelligibility, some non-native speakers and their listeners can get beyond non-nativelike pronunciation and other

aspects of non-native production, and reach similar levels of intelligibility as native speakers do for their listeners.

The differences in inaccurate content at specific times between retells of different speaker groups did not seem to be systematic or robust. The differences at Time 3 between the retells for the control group and treatment group, and at Time 4 between the control group and native speaker group seem to have been driven by Ping, in the control group, whose retells had more inaccurate content at Times 3 and 4 than at Time 1. The only other change in inaccurate content for individual talkers was for Hui, in the treatment group, whose retells had more inaccurate content at Time 3 than at Time 1. However, because as a group the control group retells had more inaccurate content than the treatment group retells at Time 3, it appears that the inaccurate content in Ping's (control group) retells increased to a greater degree than the inaccurate content in Hui's (treatment group) retells.

Did the speech of these two talkers actually become less intelligible over time, as it seems from their individual intelligibility results? With regard to the inaccurate content measure, a possible factor is differences in story length over time. Both Ping and Hui's stories were shorter at Time 1 than at later times, so listeners for Time 1 stories would have had to note down comparatively less content, and they possibly understood or noted down content more accurately. However, for other talkers whose stories at Time 1 were also shorter than at later times, listeners' retells were not different over time. It could be that inaccurate content in the retells stemmed not only from story length or the nature of a talker's speech, but also from the nature of an individual talker's storytelling. If a listener did not understand individual words in a talker's story, the listener might still have been

able to use the wider story context to understand the events in the story. However, if the structure of the story itself was confusing, listeners may not have been able to overcome misunderstanding due to non-native pronunciation. It may be that Ping and Hui's speech did not change, but the way they organized their stories at Time 1 may have been different than at later times. This is an area for investigation in future research.

That inaccurate content in retells was not due only to non-native pronunciation can be seen by the results for retells of native speakers, which also contained inaccurate content which fluctuated across recording times. When a post-hoc, two-way ANOVA was conducted with inaccurate content scores, with time as the between-group factor and individual *native* speakers as the within-group factor, a main effect for speaker was revealed, $F(3, 129) = 3.66$, $p = .014$, with a significant interaction, $F(6, 129) = 3.50$, $p = .003$. Tests of simple main effects exploring the interaction showed that both Brenda and William had significantly different scores for inaccurate content between Times 1 and 3 ($ps = .030$ and .004, respectively) and Times 3 and 4 (both $ps = .001$). Changes in pronunciation cannot be the cause of these differences, so the ways in which Brenda and William told the stories may have contributed to this difference. Again, possible effects of story organization are an area to investigate further in future research.

On-line word task. In the on-line word task, there were no differences in listeners' inaccurate word guesses over time, but there were overall differences between groups of talkers. Listeners had the fewest inaccurate word guesses when listening to native speaker stories, followed by treatment group stories, then control group stories. The overall difference between the NN treatment and control groups, combined with the lack of differences between the stories of both NN groups at Time 1, suggest that from Time 1 to

Time 3 to Time 4, treatment group stories elicited fewer inaccurate word guesses (i.e., became more intelligible) than the control group, though the difference between the treatment and control group at any one time may not have been large enough to be significant.

No individual talkers in the treatment group had significantly fewer inaccurate word guesses for their stories over time, though in the control group, Feng's stories did elicit significantly fewer inaccurate guesses at Time 4 than at Time 1. It appears that even with the improvement for Feng's stories in the control group and the lack of significant improvement for individual talkers' treatment group stories over time, the stories from the treatment group still sustained significantly fewer inaccurate guesses overall than the control group, suggesting a slight improvement in intelligibility over time.

Inaccurate content: measures from the two tasks. Compared to the retell task, in the on-line word task there was a more clear-cut difference overall between listeners' performance when listening to native speaker stories and NN stories; this may show that the inaccurate content measure for the on-line task was more sensitive to talkers' pronunciation than the measure for the retell task was. Another difference between the two tasks was the time available to the listeners to decode L2 speech. These explanations are discussed below, beginning with the issue of the intelligibility measures' sensitivity to pronunciation.

First, listeners doing the on-line word task were required to understand individual words within a stream of continuous speech, but not larger story elements. In contrast, retell listeners were required to reproduce story elements, but not necessarily individual words. If a talker's pronunciation of a word was not clear to a listener but the meaning of

the larger story element was clear, the listener would have been able to accurately paraphrase the story element without figuring out the exact word itself.

Second, the retell listeners themselves decided when to start recording their retell, whereas on-line listeners could only interpret what they heard during the time they were listening to the story. Retell listeners thus had the chance to look over their notes and think about unclear words and story elements after hearing the story. In the terms of Anderson and Lynch (1988), when trying to comprehend the story, retell listeners had more time than the on-line listeners to make use of not only their knowledge of language as a system, but also their knowledge of the co-text (what had been said), their background knowledge of what is factually true and socioculturally typical, and their knowledge of how language is usually used in discourse.

If the lower scores for control group stories in the on-line task are indeed an indication of less intelligible pronunciation by the control group, there might be further evidence from the on-line measure for LU statements, and possibly from listeners' ratings of the two NN groups. These measures are discussed further below.

Listener Statements of Lack of Understanding

The measures for listener statements of lack of understanding (LU) were used for both the retell task and the on-line word task. The results for each task will be discussed in turn.

Retells. Fewer LU statements occurred in retells of NS stories than in those of both NN groups. Thus, unlike the inaccurate content measure in the retells, this LU measure differentiated between retells of NS and NN talkers' stories, though not between retells of stories from the two groups of NN talkers. For individual talkers' stories, only

the retells for Javier, in the treatment group, showed changes in the number of LU statements by listeners. The increase in LU statements in Javier's retells from Time 1 to Time 4 was not large enough to generate any differences for the treatment group stories compared to the control group or over time.

The LU measure in the retell task seemed much more linked to the (non)-nativeness of talkers than the measure of inaccurate content was. It may be that the LU retell measure was a purer measure of listeners' understanding of talkers' pronunciation. Listeners may have retold story content *inaccurately* because of talkers' pronunciation, because the story was structured in a confusing way, or because they were not able to note down the content accurately. However, listeners' statements that they could not understand or had trouble understanding a story element (LU statements) were almost entirely restricted to the stories of non-native speakers. This may be because, unlike non-native speakers' stories, there were few elements in native speakers' stories which were completely unintelligible to listeners.

On-line word task. For the on-line word task, NN treatment group and control group stories both elicited more LU statements than stories from the native speaker group. However, the stories of the NN treatment group received fewer LU statements at Time 1 than those of the control group. As well, there were more LU statements for the treatment group stories at Time 4 than at Time 1, but no effect for time for the control group stories. No differences over time appeared for individual talkers' stories except for Feng, in the control group, whose stories had more LU statements at Time 4 than at Time 3. This difference was too small to produce any overall change for the control group stories over time.

The difference in the number of LU statements for the treatment and control group stories at Time 1 meant that it was not possible to analyze whether the control group stories received more LU statements than the treatment group overall. Thus, this measure did not provide *clear* support for or against the possibility that the control group's pronunciation was less intelligible overall than the treatment group's. However, the difference between the two NN groups at Time 1 could possibly be interpreted as another sign that the pronunciation of the treatment group *was* more intelligible than that of the control group, and that this greater intelligibility could be seen even at Time 1.

The LU statements for treatment group stories increased over time. This seems at first to suggest that the treatment group's stories became less understandable to listeners over time. However, the measure for *inaccurate content* in the on-line task suggests that this was not the case, because over time listeners' inaccurate word guesses for treatment group stories *decreased*, though not significantly. Still, listener groups hearing speech samples from different times may have varied in their behaviour when hearing an unclear word. While one group may have been more likely overall to make a(n) (inaccurate) guess (inaccurate content), another may have been more likely to simply state that they did not understand the word (LU statement). This could have led to a listener group having a relatively high ratio of LU statements with a relatively low ratio of inaccurate word guesses.

LU statements: measures from the two tasks. Unlike the measures for inaccurate content, the LU statements measures for both the retell and on-line tasks differentiated between native and non-native speakers. Perhaps these LU measures demonstrate that when listeners are completely unable to decode a word or a story element, this occurs

more often with non-native than with native speech. Even if listeners in a retell task have time to reflect before retelling a story, their knowledge of the story context will not help them if they have no idea of what was said in the part they didn't understand. This lack of understanding may be due to non-native pronunciation.

Summary: Intelligibility

However it was measured, few changes in intelligibility were seen over time for either non-native group. The only unambiguous changes over time were in the on-line word task, when listener comments of lack of understanding for treatment group stories actually increased from Time 1 to Time 4. There appears to have been a decrease over time in inaccurate word guesses in the on-line task for the treatment group stories, but evidence for this decrease is indirect. All in all, for either NN group, any changes in intelligibility over time were not unequivocally for the better or for the worse.

For the stories of individual talkers, few showed any consistent patterns of change for any measure in either direction over time. Most changes involved lower (less favourable) intelligibility scores at later times than at earlier times. For the stories of one talker in the control group, Feng, changes in the two intelligibility measures from the on-line task were actually in different directions.

I return to the two possible explanations, mentioned at the beginning of the section on intelligibility, for the different pattern of results from different measures and different tasks. With regard to intelligibility differences between speaker groups, one task (the on-line word task) provided more straightforward evidence than the other (the retell task). The two measures from the on-line word task showed the native speaker group to be always more intelligible than the NN groups. However, differences between the two

NN groups appeared only for one measure in the on-line word task. Results of measures from the retell task were much more variable, with only one measure showing differences between the native speaker group and both NN groups.

These variable results may spring from several sources. First, other factors may influence intelligibility scores. For example, measures from the retell task may be more open to influence by task effects. Two listeners who understand a story to an equal extent may differ in the intelligibility scores for their retells because of differences in the listeners' note-taking ability. Second, L2 speech may be intelligible to various degrees at different levels of discourse. For example, measures in the retell task may be more reflective of authentic communication than measures in the on-line task, when listeners are trying only to recognize words; therefore, the fact that some talkers pronounce words in a non-nativelike way may not be critical for their intelligibility in the retell task, since listeners are able to go beyond the speech itself and use their knowledge of the context to help them to understand.

Measures from the two tasks may thus be effective at measuring intelligibility at different levels of discourse. Measures from the retell task, especially the inaccurate content measure, may reflect listeners' understanding of speakers in authentic communication. At the level of extended discourse, some non-native speakers can thus be as intelligible as native speakers. In contrast, measures from the on-line word task may reflect listeners' understanding at the level of pronunciation of words. When focusing on word-level intelligibility, differences between native and non-native speakers can be seen.

The differing results between the measures from each task might be accounted for by the two explanations above. However, the intelligibility scores *as a whole* may have been affected by two additional factors. First, the tendency for intelligibility scores to be lower at later times may be a function of the increased length of later stories for most talkers, as longer stories gave listeners more opportunities to misunderstand, and for retells, made greater demands on listeners' note-taking ability. The use of ratios in measuring intelligibility was an attempt to control for story length, but it may not have been entirely successful. Second, the factor of listener group composition must be considered. The different listener groups for different times and tasks could have reacted in different ways to the stories. The distribution of self-rated listener experience with L2 speech was matched across groups, but there may have still been differences between listener groups (e.g., in the *type* of experience with L2 speech) that led to different levels of understanding.

Listener Evaluation

The second part of the discussion of research question 2 deals with listeners' evaluation of talkers' speech. Talkers were evaluated on three different aspects: comprehensibility, accentedness, and fluency.

Rater Reliability

Even with the removal of outliers, the reliability of listeners' ratings was sometimes not high. However, this variability usually occurred with listeners' ratings of the treatment group, demonstrating that listeners did not fully agree in their evaluation of this group. Particular aspects of treatment group speech received variable ratings. For example, fluency ratings were uneven, likely because listeners were not all using the

same criteria for rating. Although listeners had been given examples of particular features of dysfluency, they often differed from each other in their judgements of fluent and dysfluent speech for treatment group talkers. In addition, listeners from the on-line word task at Times 1 and 3 generally had lower rater reliability for treatment group talkers. However, there was a small number of listeners in any given listener group for the on-line word task, meaning that small differences in ratings would have had a large effect on rater reliability coefficients.

Although the reliability of ratings was sometimes not high, the patterns of ratings across listener groups were usually quite similar; this suggests that although some sets of listener evaluations were not high in reliability, overall, listener groups behaved in similar ways. Ratings for comprehensibility, accentedness, and fluency will be discussed in turn.

Comprehensibility

For both groups of listeners (retell task and on-line task), the treatment group's stories were rated easier to understand than the control group's stories at Time 1. The stories of both NN groups were also rated harder to understand than those of the native speaker group. As with most of the intelligibility measures, no changes over time appeared for the speaker groups, meaning that a given speaker group was not rated harder or easier to understand over the three recording sessions.

Although there was no difference between the stories of the two NN groups at Time 1 for most of the *intelligibility* measures, listeners at Time 1 seem to have had at least the *perception* that the stories of the control group were harder to understand. For both groups of listeners, the comprehensibility results are closer to the intelligibility results for the on-line word task than to the intelligibility results for the retell task. For the

measures from the on-line task, the control group stories were less intelligible than those of the treatment group, either at Time 1 or overall. It may be that listeners equated their ease of understanding with their understanding of individual words, rather than their understanding of a talker's overall message.

Accentedness

Results were similar for both groups of listeners. The stories of the native speaker group received the most nativelike accentedness ratings; the stories of the NN groups received significantly less nativelike accentedness ratings. For listeners in the retell task, there was no difference in accentedness ratings for the two NN groups. For listeners in the on-line word task, the treatment group was rated as having more nativelike accents than the control group.

These results provide more support for the proposal discussed in the section on intelligibility above that the on-line task was a purer measure of talkers' pronunciation than the retell task. The pattern of results for the accentedness ratings for the on-line word task mirrored the results for inaccurate word guesses for the on-line word task. The accentedness results for the on-line task also suggest that the treatment group talkers may have become slightly more nativelike in their pronunciation over time, since there was no difference in accentedness ratings at Time 1.

Fluency

For fluency, the pattern was somewhat different for the two groups of listeners. For both groups of listeners, the treatment group talkers were rated more fluent than the control group at Time 1, but the control group talkers had significantly improved in their fluency ratings by Time 4. Interestingly, for the on-line listeners only, the native speaker

talkers were rated less fluent at Times 3 and 4 than at Time 1. These differences in ratings over time are not reflected in the mean speech rates for talker groups in Table 27. It seems that the broad measure of words per minute was not the only criteria that listeners used to make their fluency ratings.

Summary: Listener Evaluation

Although rater reliability was sometimes not very high, listeners' patterns of rating scores were more consistent across listener groups than their pattern of intelligibility scores, which suggests that although the previous listening tasks (retell and on-line word intelligibility) focused on different levels of intelligibility, the different focuses did not seem to affect listener groups differently in their subsequent rating task. The treatment group talkers were at least *perceived* by listeners to be generally better speakers than the control group talkers overall, even though not all intelligibility measures showed this unconditionally.

No unambiguous change in ratings over time appeared, except for the control group's improved fluency ratings, which occurred independently of their ratings on any other measure. The increase in control group talkers' cumulative exposure to English over the five-month study may have been a factor in listeners giving the group higher fluency ratings at Time 4 than at other times. However, it is peculiar that the same increase in fluency ratings did not occur for treatment group talkers, whose cumulative exposure to English also increased over the five months.

Summary: Discussion of Intelligibility and Listener Evaluation

In general, neither non-native group appeared to unequivocally improve or to worsen in their intelligibility or listener evaluation over time. It appears that intelligibility

of speakers can be measured at different levels of discourse, with some speakers equal in intelligibility at one level of discourse, and different at another. Some results suggest that the treatment group did make slight gains over time in their pronunciation, with changes in the ratio of incorrect word guesses by listeners and in listeners' accentedness ratings. However, other measures indicate that at the beginning of the study, after receiving only a few hours of instruction, the treatment group was already better than the control group.

For the two NN talker groups, most or all of the time they were using or exposed to English was outside the oral communication classes. I will therefore discuss the ways in which both groups of NN talkers were exposed to and used English.

English Exposure/Use

Research Question 3

In what ways do NN graduate students at an English-medium university come into contact with and use English on a weekly basis?

Measurement of Exposure and Use

It is worth noting that for almost all previous studies in which learners' self-reported exposure and use was measured, it was measured with questionnaires or surveys administered once or, at most, twice; learners' responses were based on their estimates and/or recollections. Apart from an unpublished study conducted by Leila Ranta at the University of Alberta, the current study is the only one known to the author in which the frequency and nature of learners' self-reported *daily* out-of-classroom L2 exposure and use was recorded longitudinally.

Accent and Use of L2

As discussed previously in Chapter 2, Flege et al. (1999) found that for L1 Korean learners of English who were matched on age of arrival, length of residence, and amount of U.S. education, the learners who reported using English relatively often in their daily lives were rated with significantly more nativelike accents than learners who reported using English relatively infrequently.

In order to weigh the results of this current study against Flege et al.'s, I would have needed measures of exposure and accent from the non-native talkers, as well as biographical data. There are complete biographical data and accent ratings for the six talkers whose stories were heard by listeners, but because one talker in the control group did not complete at least three English exposure logs, her exposure/use data could not be used. Therefore, it was not possible to analyze how the talker participants' longitudinal exposure to or use of English related to how intelligible or accented they were to listeners. However, I will discuss the patterns of English exposure and use from the larger set of treatment and control participants who did complete at least three of the four logs.

Overall Weekly Exposure/Use

One interesting finding was that the treatment group used or was exposed to a higher mean amount of English per talker than the control group (263 minutes more on average, ranging from 48-827 min/week). This might be explained by the fact that the treatment group was attending the oral communication class, which totalled 240 minutes of instruction per week beyond their regular program of study. However, the control group was studying in an English environment as well, which presumably provided them with similar opportunities to use and be exposed to English. The treatment group's

greater amount of mean exposure/use might be explained by the types of activities they engaged in and the characteristics of the talker groups, which are discussed below.

Exposure/Use across Activity Categories

The pattern of English use and exposure between groups was different in some ways. The treatment group's comparatively greater use of English for the category of attending class is just one indicator of differences in the composition of the treatment and control groups. The treatment group contained a greater number of Master's students, who certainly spent more time in classes overall than the mainly PhD students in the control group. The increased class time for the treatment group can also be seen in the comparatively greater time they spent in non-interactive aural/oral academic activities (e.g., lectures, presentations). The control group's comparatively greater use of English in the category of academic work was likely because the PhD students in the group took fewer courses than the treatment group students; consequently, the control group students spent more of their overall academic time on non-course tasks.

The treatment group spent comparatively more time using English not only in their recreation, but also in general non-academic interaction: they used comparatively more English than the control group in the recreation category, and in aural/oral interactive social activities as well. It does seem, therefore, that the treatment group spent comparatively more time in English outside the academic sphere. This may again be due to the higher number of Master's students in the treatment group, who were generally younger and may have had fewer family commitments and perhaps a somewhat lighter academic load; therefore, they may have generally had more time both for recreation and for social activities; in addition, the Master's students in the treatment group may have

been more likely to relax and socialize with friends of a different L1, not just same-L1 family members.

Exposure/Use in Aural/Oral Activities

In looking solely at the domain of the selected aural/oral activities, both groups spent more time in English in social activities than in academic activities. This is true whether the activities were interactive or not. This may indicate simply that much academic work is done through writing and reading rather than listening and speaking. An indication of how the talkers' academic work was predominantly written in nature was the continuing decrease over time for both groups in the use of English for aural/oral academic activities, but a concurrent increase over time in the use of English for academic work overall. No doubt the end of classes and the talkers' associated focus on term papers, exams, and articles had a lot to do with this.

Summary: Discussion of English Exposure/Use

Overall, the control group seemed to engage in less recreation and general aural/oral interaction in English than the treatment group did. This means that much of the control group's exposure to and use of English was limited to the academic sphere or consisted of activities in which English was primarily written, not spoken. According to proponents of the interaction hypothesis, interaction between an L2 learner and another interlocutor is one of the drivers of language learning; through the efforts of both the language learner and interlocutor to understand and to be understood, the learner's attention is drawn to the language forms which convey the message she wants to send or understand (Gass & Mackey, 2006).

Therefore, the comparatively lower amount of time the control group spent overall in aural/oral interaction might suggest that this group had (a) less opportunity to receive aural input which could be adjusted to their capacity for understanding; (b) less opportunity to notice aural language forms which they did not already have control over; and (c) less opportunity to restate their message in a more accurate way. This does not necessarily mean that the control group learned less English than the treatment group over the five-month study period; however, the low level of interaction in English was keenly felt by some graduate students in the control group. In their interviews, they described their frustration with their opportunities for English use and exposure. It is discussion of those interviews that I turn to next.

NN Graduate Students' Views on their Language Learning

Research Question 4

What do NN graduate students think about their development in English over a five-month span and during their graduate studies in general?

Academic and Language Difficulties

All NN graduate student participants had this in common: they were non-native speakers of English at an English-medium university. However, each graduate student had his or her own individual understanding and opinion of his or her academic experiences in English. What came out clearly from the individual interviews was that graduate students who reported having major difficulties with spoken English also believed that these difficulties contributed to problems in their studies and in their

interactions with others. This finding echoes results in other studies on NN graduate students in Canada (Cheng, Myles, & Curtis, 2004; Fletcher & Stren, 1989; Lee, 2005).

Perceived Improvement

In the interviews conducted at the end of the data collection for the talkers, the majority of students reported extensive difficulties when they first began their university programs, suggesting that they had not received adequate preparation in using and understanding English before arriving at university. However, as was similarly reported in Cheng, Myles, and Curtis (2004), many students stated that they became much better at coping in an academic English environment, especially at understanding spoken English, simply with more exposure to and practice with English in their studies. This development was most noticeable to them over the course of the first few months.

Most students in this study also believed that they had improved over the five-month data collection period, though a few students in the control group attributed their improvement partly to the increased input and interaction from participating in the study itself. The control group's logs, showing a relatively lower amount of time interacting socially in English, were corroborated by control group interviews, in which some students worried that they did not spend enough time interacting in English. This might explain why, in the minds of some control group students, meeting me and completing aural/oral research tasks once every six weeks was judged to be a significant driver of language development. "I just feel every time I come here to stay two hours or half an hours or one and a half an hours to practise or talking and watch movie and repeat, I think so I can speak a little bit fluently".

As in Lee (2005), some students' self-confidence in themselves and in their academic competence was jeopardized by their struggles with interaction in English. When students could not understand others' speech or could understand, but were not able to show that they had something to contribute, their confidence was reduced and they were left feeling frustrated and helpless.

Oral Communication Course

As the great majority of students remarked, they had become better in using and understanding English over the five-month study. Did treatment group learners credit the oral communication course for their progress? Several learners mentioned that what they had learned, though useful, had no lasting effect on their speech after the course had ended. However, two learners, Tai Ning and Piotr, believed that the feedback they had received from the teacher had helped them to be better understood (Tai Ning) and more grammatically accurate (Piotr) when they spoke.

Significantly, the teacher herself did not expect students to greatly improve their communicative skills by the end of the course, but expected the students to be able to eventually put into practice the suprasegmental information and independent learning strategies they had learned in the course. It may be that the treatment group students needed more time to both match the pronunciation patterns they had learned with the naturalistic input they received, and to consequently produce those patterns in their own output. It may also be that some students did not fully realize the extent to which they would have to themselves continue to work on their pronunciation after the course by using the independent learning strategies covered in the course.

What did the treatment group students believe they had in fact learned from the course? Most focused on the suprasegmental aspects which had been taught, but a few also remembered advice to slow their speech rate. Some students had expected to learn vocabulary or set expressions, or to do more exercises involving "communication". However, the course description was fairly explicit in specifying that suprasegmental pronunciation and presentations were the main focuses of the course.

Of the control group talkers who had previously taken the course some time before, most believed that it had been helpful for them, and they usually remembered learning about lexical stress and focus words. They did not, however, attribute any improvement in English over the five-month study period to what they had learned in the oral communication course long before. Instead, they credited the opportunities they had had to use English with other people. Nevertheless, it is possible that, post-course, the talkers did use the knowledge and strategies they had learned, but were unaware of using them. If learners began to apply their knowledge and use of strategies more automatically as time went on, they might have been less and less aware that they were drawing on that knowledge and those strategies.

Summary: Discussion

Taken together, the results discussed here suggest a picture of two distinctly different groups of NN graduate students, who had different sets of behaviours and different perspectives on their communication and interaction in English. As seen from the logs and interviews, the treatment group spent comparatively more time interacting and socializing with other people in English, and generally viewed their development and

current abilities in a positive way. The control group, however, spent comparatively more time on academic work, as well as on social activities which did not require much interaction in English. Control group students believed their English had improved, but many were unhappy about their ability to communicate and their level of social contact in English.

Although the two groups seemed distinct in their characteristics and behaviours, no clear-cut difference in intelligibility was seen between the two groups. Moreover, no differences in intelligibility appeared over a five-month period for either group. This was in spite of the fact that the treatment group had received months of theoretically-motivated training in oral communication, provided by a knowledgeable and enthusiastic instructor. This lack of conclusive evidence of change could be explained in several ways. One possibility is that focussing instruction primarily on suprasegmental aspects of pronunciation is too restrictive an approach for learners with different weaknesses in pronunciation. Another possibility is that the classroom and the research tasks had different processing demands. An additional factor is a possible link between listeners' retells and their note-taking skills, which would have affected scores on the two intelligibility measures for the retell task. The use of two tasks, which may have targeted intelligibility at two different levels of discourse, is a final factor.

However, although the intelligibility measures showed no clear differences between groups or over time, the listener evaluation measures did show consistent differences between the two NN groups. For almost all measures of listener evaluation, the treatment group talkers were rated more favourably than the control group talkers at Time 1. This suggests first, that the listeners were able to hear differences between the

two sets of talkers even before the treatment began, and second, that those differences were not clearly brought out by the intelligibility measures.

We might conclude that if listeners perceive one talker to be comparatively harder to understand or comparatively more accented than another talker, it does not necessarily follow that the first talker is actually less intelligible. In past research, the relationship between listener evaluation ratings (comprehensibility, accentedness) and intelligibility scores has not been consistent (cf. Munro & Derwing, 1995; Wijngaarden, Steeneken, & Houtgast, 2002). However, both listener evaluation and intelligibility scores have been consistently linked to goodness of suprasegmental production. In the current study, talkers' suprasegmental production was not assessed in itself, so it is not known whether it changed over time or how it may have contributed to the listener evaluation or intelligibility scores. Further research is clearly called for in order to identify which characteristics of NN speech are linked to more or less favourable listener evaluation, and which characteristics are linked to a talker's intelligibility to a listener.

CHAPTER 6: CONCLUSION

In this chapter, limitations of the study and suggestions for future research are discussed. Then, implications of the study's results for university administrators, faculty, staff, and students are considered. The chapter ends with concluding remarks about universities' investment in non-native graduate students.

Limitations and Future Research

Linguistic Context

This research study was conducted within a certain context, that of an English-medium university in a city with two main languages, French being the dominant one. This linguistic context is unique in North America, so the results of this research are not necessarily generalizeable to other North American linguistic contexts. Therefore, additional research on L2 instructional effects on NN graduate students' intelligibility or pronunciation should be conducted in other North American contexts.

Sample Selection

Due to methodological considerations (keeping listening sessions to a reasonable length), only a subset of talkers from both NN talker groups had their speech samples heard by listeners. These talkers (three from each group) were matched for L1 and length of residence, but in future research, a larger sample of talkers is clearly desirable. There would then be wider representation of the nature of group members' L2 speech; therefore, the results for intelligibility and listener evaluation would be more representative of the group overall. The sample of listeners for the on-line task was also

not large; therefore, atypical listener performances could have excessively affected the overall results for the group.

In another sample selection issue, the samples of some talkers were not included in listening sessions simply because one or two of their stories were too long. In future research, talkers could be asked to limit their stories to around two minutes. In the current study, it was not known how time limits might have affected the nature of the stories and storytelling, and so no time limits were given. However, future research could investigate whether or how time limits on stories affect the nature of the stories told.

Narrative Analysis Scheme

Although the use of a modified scheme of narrative analysis was replicated by an independent analyst, this scheme should be further refined and validated in future research with different sets of narratives.

Intelligibility Measures and Constructs

One of the strengths of this study was that L2 speech was measured and evaluated in various ways, with two tasks and five measures used to measure talker intelligibility. This meant that a more nuanced view of intelligibility was possible. However, the intelligibility construct behind the two tasks was different (construct – the abstract theoretical concept of a given ability or trait), with one task focussing on accurate word identification and the other on accurate retelling of stories. Because of the two tasks' different constructs of intelligibility, measures from the two tasks could show different results while still being valid measures. That is, each intelligibility measure could be compared to other measures stemming from the same task (e.g., inaccurate content compared to lack of understanding in the retell task), to see if results were similar. If the

results were not similar for measures within the task, this might be a clue that there was a validity problem with one or both measures, since the two measures were based on the same intelligibility construct. However, if the results for the measures across tasks were different, this would not give conclusive information about the validity of the measures, since the two tasks were based on different intelligibility constructs. In future research, intelligibility could be measured with several tasks based not just on different intelligibility constructs, but on similar constructs as well. This would allow researchers to more fully evaluate the validity of the intelligibility measures for each task.

Helping NN Graduate Students

Apart from Stevens (1989) and Derwing, Munro, and Wiebe (1998) , described in Chapter 2, the studies which described ESL courses for NN graduate students, particularly ITAs, did not include formal evaluation of the courses' instructional effects. Various approaches were taken in these courses; however, evaluation of the real-world effects of these approaches and materials was generally impressionistic, through students' and instructors' comments and ratings of the course itself. Very little data are available showing NN graduate students' performance in the academic sphere upon completion of such courses.

What do we know, then, about how NN graduate students can be helped in their efforts to use English in academia (and the outside world)? What seems undeniable is that the NN graduate students in the current study who struggled with English felt they needed more opportunities to use English; also, most NN graduate students in this study believed that spoken interaction with others in English was the best way for them to

learn. Interestingly, when Zimmerman (1995) surveyed NN graduate students in a U.S. college, she found that a significant part of their satisfaction with their own communication was predicted by how frequently they interacted with American students. Because some students in the current study had difficulty fully participating in English conversations or even finding opportunities to participate, it appears that "sheltered" interactions would be one means by which NN graduate students could gain more practice with English.

Interactions like this already exist in many forms at many universities – ESL conversation courses or conversation times, classroom or tutorial observations, or peer pairs or "buddies", recruited from departments or from the university at large, who are matched to an incoming student for a year. In research on the effects of formalizing non-native students' contact with native speaking students, Westwood and Barker (1990) showed that international undergraduate students who were paired with a "host national" for their first year were significantly more successful academically and less likely to leave their program than students who were not paired.

The university in this study offered several programs of these sorts. However, not all NN graduate students choose to participate in these opportunities. As mentioned in the interviews, some reasons may be time constraints, cultural differences in thinking and speaking, or simply fear of being rejected by potential interlocutors.

Academic Departments

The concerns about time constraints and anxiety about initiating interactions can be partly addressed at the level of the supervisor and the department. It would seem that if students are actively encouraged by their supervisors and the department administration

to take the time to work on their English, the students might be more likely to do so, even if it meant completing some other academic tasks more slowly. In addition, a department which promotes social interaction between and among students, faculty, and staff would be a less intimidating place to strike up a conversation. For example, in Tai Ning's department, an informal gathering for all graduate students was held every Friday afternoon, with refreshments provided and no set agenda. Students could talk to other people who shared some of the same academic and research interests.

These types of initiatives may involve changes to the culture of an academic department for faculty members, staff, and students. Such changes may not be welcomed by some, who may believe that non-native graduate students should themselves take responsibility for addressing their difficulties with language (S. Jenkins, 2000). However, it is clear from the interviews in this study that graduate students who still struggled with English in academic and social contexts did not wish to deny their responsibility for developing their skills in English. They did, though, desire some assistance and encouragement in becoming more skilled. Therefore, if departments and supervisors decide to provide that assistance and encouragement, struggling graduate students will likely jump at the opportunity and make the most of it (Capraro, 2002; Perucci & Hu, 1995).

ESL Outreach

ESL administrators and instructors can also do their part, both in making opportunities for interaction more accessible and in raising awareness among supervisors, professors, and departmental administrators of how to provide support for NN graduate students. To begin with, both academic and student service departments should be

provided accurate, up-to-date information about ESL services and courses provided to NN graduate students (Sung, 2000). This information can be provided to students when they are admitted and/or at orientation sessions, with periodic reminders throughout the year. Workshops for faculty and NS classmates could cover strategies which encourage NN graduate student comprehension and participation in class discussions or research meetings, such as more frequent reviews of the content discussed to that point, explicit solicitations of NN graduate students' ideas, or simply a slightly more measured pace of discussion.

Planned Interactions

Five of the eight control group talkers, who were not currently taking an ESL course, had previously taken the ESL course in oral communication. These talkers had taken steps to formally work on their communication in English, but some were still struggling after the course to initiate and participate in conversations and discussions. Where such services are available, such graduate students could register to be matched to another student or mentor, but for those who want to have planned interactions, small-scale discussion or conversation groups might meet their needs. International student offices, graduate student associations, or other university bodies often offer these groups, but not all students may know about them; this underlines the need for ESL administrators, departmental administrators, and faculty to be informed about the options which are available, and to inform students in turn.

Early, Intensive Intervention

Such opportunities for instruction and interaction should of course be available to NN graduate students throughout their studies, but the most fruitful time for students to

receive targeted language support might be during their first semester or even before beginning their graduate studies proper. Several participants stated that their struggles with using English in academia were most acute for the first three or four months, after which they seemed to reach some level which was adequate for the academic tasks they had to perform. If NN graduate students often have the most difficulty with English at the beginning of their studies, it would seem sensible to either ensure that they are better prepared beforehand or that they are getting assistance with English at their most vulnerable time (Sung, 2000).

The education systems in some countries, such as the Netherlands, require one-year preparatory courses for non-native students who wish to study in undergraduate or graduate programs in the national language but who do not or cannot pass the state language exam. These courses include (Dutch) language training, study skills, and field-specific courses. Non-native students who complete the highest-level courses and are admitted to university are more successful in their academic programs than those who simply pass the state exam (InHolland University, 2006).

At some universities in Canada, there are similar one-year bridge programs for those graduate students who meet academic requirements but whose scores on language tests are slightly below the minimum. However, these types of programs are not the norm at most universities, and NN graduate students are often forced to acclimate simultaneously to the demands of their graduate program and the demands of a new social and linguistic environment. Many Canadian universities are more likely to offer non-credit intensive English programs, designed to prepare NN students to meet the

university's English language requirements and to succeed in an English-language university environment.

The success of either of these types of programs in preparing NN graduate students to succeed at university is unclear. Most journal articles published about these programs either (a) describe program curricula, but not the success of students who have followed the curricula (e.g., Krasnick, 1990), or (b) describe the effect of program curricula on students' success in meeting language requirements for the program itself or for entering university (e.g., Green, 2005). Because higher-level courses in these programs are designed to help NN students not simply to enter university, but to succeed in their studies, it is vital for researchers to track, survey, and interview the graduates of these programs who enter universities. In this way, the program curricula can be modified to reflect and address the linguistic challenges which will be faced by NN university entrants.

Conclusion

According to policy statements from universities and governments in Canada, there is a serious need for increased numbers of students and workers with specialized knowledge and skills. Both universities and government see international or immigrant graduate students as an important resource in meeting this need. However, simply bringing such people to Canada and Canadian universities should not be the end of the story. The short- and long-term communication difficulties faced by non-native graduate students cannot be glossed over or taken for granted.

Administrators, departments, faculty supervisors, and ESL instructors need to develop creative and diverse approaches in identifying and assisting non-native students who are struggling to succeed because of their language level. Students who are confident and proficient in the language of their environment will be better able to succeed in their studies and to integrate into their environment. If it is worth our while to recruit expertise from outside Canada, then it must certainly be worth our while to help those experts succeed when they arrive.

REFERENCES

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly, 18,* 71-85.

Anderson, A. & Lynch, T. (1988). *Listening.* Cambridge: Cambridge University Press.

Anderson-Hsieh, J., Johnson, J. & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529-555.

Anderson-Hsieh, J. & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38,* 561-613.

Association of Universities and Colleges of Canada. (2007a). Enrolment. *Trends in higher education.* Ottawa: Association of Universities and Colleges of Canada.

Association of Universities and Colleges of Canada. (2007b). *Creating a knowledge advantage through international education.* Ottawa: Association of Universities and Colleges of Canada.

Barker, C. (2001). *Strange invaders.* (Available from National Film Board, PO Box 6100, Centre-ville Station, Montreal, QC, Canada, H3C 3H5).

Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics, 25,* 243-272.

Bent, T. & Bradlow., T. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114,* 1600-1610.

Berman, R. A. & Slobin, D. A. (1994). *Relating events in narrative: A crosslinguistic developmental study.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Berman, R. & Cheng, L. (2001). English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics, 4,* 25-40.

Borg, S. (1998). Talking about grammar in the foreign language classroom. *Language Awareness, 7,* 159-175.

Bradlow, A. & Bent., T. (2003). *Listener adaptation to foreign-accented English.* Proceedings of the International Congress of Phonetic Sciences [CD-ROM].

Bradlow, A. & Pisoni, D. (1999). Recognition of spoken words by native and non-native listeners. *Journal of the Acoustical Society of America, 106,* 2075-2085.

Breck, E. (1998). *Soundscriber.* English Language Institute: University of Michigan.

Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *Modern Language Journal, 91*, 5, 923-939.

Capraro, F. (2002). *A journal study of the spoken English learning experience of prospective international teaching assistants*. Unpublished doctoral dissertation, The Ohio State University, Columbus.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Cheng, L., Myles, J. and Curtis, A. (2004). Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal, 21,* 50-71.

Cobb, T. (n.d.). *The compleat lexical tutor*. Retrieved June 12, 2006, from http://www.lextutor.ca

Damron, J. A. (2001). *Chinese 101, a prerequisite to Math 100? A look at undergraduate students' beliefs about their role in communication with international teaching assistants*. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN

Dauer, R. M. (1993). *Accurate English: A complete Course in Pronunciation*. Upper Saddle River, NJ: Prentice Hall Regents.

Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon, UK: Multilingual Matters.

Derwing, T., Munro, M. & Wiebe, G. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning, 8,* 217-235.

Derwing, T., Munro, M. & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning, 48,* 393-410.

Derwing, T. & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19,* 1-16.

Educational Testing Service. (1996). *Speaking Proficiency English Assessment Kit (SPEAK) Test*. Princeton, NJ: Educational Testing Service.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal, 79,* 530-542.

Flege, J. E. (1988). Factors affecting degree of foreign accent in English sentences. *Journal of the Acoustical Society of America, 84,* 70-79.

Flege, J. E., Munro, M. & Mackay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97,* 3125-3124.

Flege, J. E., Frieda, E. M. & Nozawa, T. (1997). Amount of native language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25,* 169-186.

Flege, J. E., Yeni-Komshian, G. & Liu, H. (1999). Age constraints on second language acquisition. *Journal of Memory and Language, 41,* 78-104.

Flege, J. E. & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America, 91,* 370-389.

Fletcher, J. F. & Stren, R. E. (1989). Language skills and adaptation: A study of foreign students in a Canadian university. *Curriculum Inquiry, 19,* 293-308.

Gallego, J. C. (1990). The intelligibility of three non-native English-speaking teaching assistants. *Issues in Applied Linguistics, 1,* 219-237.

Gatbonton, E. & Segalowitz, N. (2005). Rethinking communicative language teaching: A focuss on access to fluency. *Canadian Modern Language Review, 61,* 325-353.

Green, A. (2005). EAP study recommendations and score gains on the IELTS academic writing test. *Assessing Writing, 10,* 44-60.

Guion, S. G., Flege, J. E. & Loftin, J. D. (1999). The effect of L1 use on foreign accent ratings in Quichua-Spanish bilinguals. *Proceedings of the 14th International Congress of Phonetic Sciences, 2,* 1471-1474.

Guion, S. G., Flege, J. E. & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics, 28,* 27-42.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38,* 201-223.

Hinofitis, B. & Bailey, K. M. (1981). American undergraduates' reactions to the communication skills of foreign teaching assistants. In J. Fisher, M. Clarke, & J. Schachter (Eds.). *On TESOL '80: Building Bridges* (pp. 120-133). Washington: TESOL.

InHolland University (n.d.). English Summary. Retrieved August 8, 2007, from http://www.inholland.nl/NR/rdonlyres/844461D0-F409-4757-B2D1-5CD234C71F0E/1674/TS_englishsummary.pdf

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford: Oxford University Press.

Jenkins, S. (2000). Cultural and linguistic miscues. *International Journal of Intercultural Relations, 24,* 477-501.

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants. *Applied Psycholinguistics, 23,* 599-621.

Johnston, D. (2000). *CoolEdit 2000.* Phoenix, Syntrillium Software Corporation.

Kennedy, S. (2006, June). *Task effects on ratings of non-native speech.* Poster session presented at the American Association of Applied Linguistics, Montreal, QC.

Kerr (2003). The rest of the iceberg; Articulatory setting and intelligibility. In J. Burton & C. Clennell. (Eds.) *Interaction and Language Learning* (pp. 99-114). Alexandria, VA, TESOL

Krasnick, H. (1990). Preparing Indonesians for graduate study in Canada. *TESL Reporter, 23,* 33-36.

Labov, W. (1972). *Language in the inner city: Studies in the black English vernacular.* Philadelphia: University of Pennsylvania.

Lang, T. (1990). *Licence to kill, part MCMXC.* Canada: 3 min.

Lee, K.-H. (2006). *A traveler's tale: The experience of study in a foreign language.* Unpublished doctoral dissertation, University of Alberta, Edmonton.

Li, P., Sepanski, S. & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behaviour Research Methods, 38,* 202-210.

Liu, T. H. (2001). *Learners' attitudes, contact, and oral ESL proficiency: A study of Chinese graduate students at U. S. universities.* Unpublished doctoral dissertation. Purdue University, West Lafayette, IN.

Macdonald, D., Yule, G. & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning, 44,* 75-100.

Matsuura, H., Chiba, R. & Fujieda, M. (1999). Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes, 18,* 49-62.

Morley, J. (1999). Trends in speech/pronunciation instructional theory and practice. *TESOL Matters,* August/September.

Morris, C., Branford, D., & Franks, J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behaviour, 16, 519-533.*

Munro, M. & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45,* 159-182.

Munro, M. & Derwing, T. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48,* 159-182.

Munro, M. & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition, 23,* 451-468.

Myles, J. & Cheng, L. (2003). The social and cultural life of non-native English-speaking international graduate students at a Canadian university. *Journal of English for Academic Purposes, 2,* 247-263.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5,* 261-263.

Perucci, R. & Hu, H. (1995). Satisfaction with social and educational experiences among international graduate students. *Research in Higher Education, 36,* 491-508.

Piper, T. & Cansin, D. (1988). Factors influencing the foreign accent. *The Canadian Modern Language Review, 44,* 334-342.

Piske, T., Mackay, I. & Flege. J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29,* 191-215.

Purcell, E. T. & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30,* 271-287.

Ranta, L. & Meckleborg, A. (2002). *The Language Activity Log,* Department of Educational Psychology, University of Alberta.

Riney, T., Takada, M. & Ota, M. (2000). Segmentals and global foreign accent. *TESOL Quarterly, 34,* 711-737.

Riney, T. & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition, 20,* 213-243.

Riney, T. & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese ESL speakers. *Language Learning, 49,* 275-302.

Sato, K. & Kleinsasser, R. C. (1999). Communicative language teaching: Practical understandings. *Modern Language Journal, 83,* 494-517.

Schutz, A. & Richards, M. (2003). International students' experience of graduate study in Canada. *Journal of the International Society for Teacher Education, 7,* 56-63.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The Other Tongue: English Across Cultures* (1st ed., pp. 75-90). Urbana, IL: University of Illinois Press.

Snowden, A. (1987). *George and Rosemary*, (Available from National Film Board, PO Box 6100, Centre-ville Station, Montreal, QC, Canada, H3C 3H5).

Spada, N. & Fröhlich, M. (1995). *Communicative Orientation of Language Teaching (COLT) Observation Scheme: Applications for Research.* Sydney: National Centre for English Language Teaching and Research.

Stein, N. L. & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle. (Ed.), *New Directions in Discourse Processing* (pp. 53-119). Norwood, NJ: Ablex, Inc.

Stevens, S. G. (1989). A dramatic approach to improving the intelligibility of ITAs. *English for Specific Purposes, 8,* 181-194.

Sun, W. & Chen, G-M. (1997, March). *Dimensions of difficulties Mainland Chinese students encounter in the United States.* Paper presented at the International Conference in Cross-Cultural Communication, Tempe, AZ.

Sun, Y. (1987). An EFL needs assessment: Chinese students at a Canadian university. *TESL Canada Journal, 5,* 27-44.

Tahta, S., Wood, M. & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech, 24,* 265-272.

Thompson, I. (1991). Foreign accents revisited. *Language Learning, 41,* 177-204.

Thornbury, S. (1993). Having a good jaw: Boice-setting phonology. *ELT Journal, 47,* 126-31.

Towell, R. & Hawkins, R. (1994). *Approaches to Second Language Acquisition.* Clevedon, Multilingual Matters.

Trabasso, T., Van den Broek, P. & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes, 12,* 1-25.

Trofimovich, P. & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Trofimovich, P. & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics, 28,* 251-276.

Trofimovich, P. & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal, 90,* 519-535.

van Dijk, T. A. (1976). Philosophy of action and theory of narrative. *Poetics, 5,* 287-338.

Wajnryb, R., Coan, J. & McCabe, L. (1997). 'What you put into it is what you get out of it': A report on a remedial phonology clinic. *EA Journal, 15,* 38-51.

Wang, Y. (2003). *The contextual knowledge of language and culture in education: Exploring the American university experience of Chinese graduate students.*, Unpublished doctoral dissertation. University of Southern Mississippi, Hattiesburg, MS.

Weldon, J. (1991). *The Lump.* (Available from National Film Board, PO Box 6100, Centre-ville Station, Montreal, QC, Canada, H3C 3H5).

West, M. (1953). *A General Service List of English Words.* London: Longman, Green and Co.

Wijngaarden, S. J., Steeneken, H. J. M. & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *Journal of the Acoustical Society of America, 112,* 3004-3013.

Xu, M. (1991). The impact of English-language proficiency on international graduate students' perceived academic difficulty. *Research in Higher Education, 32,* 557-570.

APPENDIX A

RESEARCH ETHICS CERTIFICATE

McGill

Faculty of Education – Ethics Review Board
McGill University
Faculty of Education
3700 McTavish, Room 230
Montreal H3A 1Y2

Tel: (514) 398-7039
Fax: (514) 398-1527
Ethics website: www.mcgill.ca/rgo/ethics/humans

Faculty of Education – Review Ethics Board
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 623-1205

Project Title : ***Spet***

Applicant's Name: Sara Kennedy **Department**: DISE

Status: PhD student **Supervisor's Name: Lise W**

Granting Agency and Title (if applicable): SSHRC

Type of Review: Expedited ✓ Full ___

This project was reviewed by: Derevensky/Starke-Meyerring

Approved by

R. Bracewell Jan 16, 2006
Signature/Date
Robert Bracewell, Ph.D.
Chair, Education Ethics Review Board

Approval Period: Jan 16/06 **to** Jan 16/07

All research involving human subjects requires review on an annual basis. An Annual Report/Request for Renewal form should be submitted at least one month before the above expiry date. If a project has been completed or terminated for any reason before the expiry date, a Final Report form must be submitted. Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received. This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement on the Ethical Conduct for Research Involving Human Subjects.

1/11/06

APPENDIX B
CONSENT FORMS FOR TALKERS

Control Group Talkers

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in the research project entitled:

Spet

And conducted by: _____Sara Kennedy, McGill University_____

1. Purpose – *The purpose of this research is to see how non-native English speakers speak English.*

2. Procedures –

You will participate in four recording sessions. One will take place in the third week of January, the next in the middle of the Winter term, the next at the end of the Winter term, and the last one month after the end of the Winter term. Each session will take approximately one hour and a half, except for the first, which will take about two hours. During each session, you will read a list of sentences, watch a short film and retell the story, and tell a story from your own life.
During the five months of the study, once every four weeks for a week you will also make entries in a log of your daily exposure to and use of English. You will mark the relevant parts of this log every day, and hand in the log at every recording session. This is an important part of the research and has to be done ***every day of the week, once every four weeks****.*

After the four recording sessions, you will be asked to verify the analyses which have been made of your story recordings, and to talk about your progress. This will take about two hours.

At the end of the five-month study, *you will have the chance to listen to your recordings, and to discuss your progress with an experienced ESL teacher and talk about ways to improve your speech.*

During the five-month study, *the researcher, Sara Kennedy, will not give you any advice or suggestions on improving your English speech. This will be done only* **after** *the five months.*

In all the video and audio recordings, you will be identified with a code so that no one knows your name. The results of this study and the data collected may be presented in a workshop, a class, a research presentation, a Ph.D dissertation, or a journal article. Your name will not be mentioned in any presentation of the results.

3. Conditions of Participation –

You will be recorded in four separate recording sessions, each lasting about one hour and a half, except for the first, which will last about two hours.

Once every four weeks, for one full week, *you will make log entries on your exposure to English and hand in the log at every recording session.*

You will verify the analyses of your recordings after the four recording sessions.

You will be free to listen to your recordings at the end of the data collection, and to then discuss your progress and work on your speech with an ESL teacher in a one-on-one tutorial session.
Participating or not participating in this study will not affect your grades in your courses.
You will receive $75 (CAN) for your participation. One-third of this money will be paid at the beginning of the study, one-third in the middle, and one-third at the end. You must fully participate in the study to receive the full amount.

- I understand the purpose of this study and know about the risks, benefits, and inconveniences involved in this research project.
- I understand that I am free to withdraw at any time from the study without any penalty or prejudice.
- I understand that this research will not affect my grades or evaluation of my work.
- I understand that my name will not be used during and after this research project.
- I understand how the data in this project may be used, especially with respect to publication, communication, and diffusion of results.

I have read the above and I understand all of the above conditions. I freely consent and voluntarily agree to participate in this study.

Name (please print) _____

Signature _____ Date _____

Treatment Group Talkers

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in the research project entitled:

Spet

And conducted by: _____Sara Kennedy, McGill University_____

1. **Purpose** – *The purpose of this research is to see how non-native English speakers speak English.*

2. **Procedures** –

You will participate in four recording sessions. One will take place at the beginning of January, the next in the middle of the Winter term, the next at the end of the Winter term, and the last one month after the end of the Winter term. Each session will take approximately one hour and a half, except for the first, which will take about two hours. During each session, you will read a list of sentences, watch a short film and retell the story, then tell a story from your own life.
During the five months of the study, one week every four weeks, you will also make entries in a log of your daily exposure to English. You will mark the relevant parts of this log every day, and hand in the log at every recording session. This is an important part of the research and has to be done **every day for seven days.**

In your **** course, you, your classmates, and your teacher will be videotaped. The videotape will not be used for any part of your grades and your teacher will only see it (with your permission) after the course grades are submitted.

After the four recording sessions, you will be asked to verify the analyses which have been made of your story recordings, and to talk about your progress. This will take about two hours.

At the end of the five-month study, *you will have the chance to listen to your recordings, and to discuss your progress with an ESL teacher and talk about ways to improve your speech*

During the five-month study, *the researcher, Sara Kennedy, will not give you any advice or suggestions on improving your English speech. This will be done only* **after** *the five months.*

In all the video and audio recordings, you will be identified with a code so that no one knows your name. The results of this study and the data collected may be presented in a workshop, a class, a research presentation, a Ph.D dissertation, or a journal article. Your name will not be mentioned in any presentation of the results.

Conditions of Participation –

You will be recorded in four separate recording sessions, each lasting about one hour and a half, except for the first, which will last about two hours.

You will make **daily** *log entries on your exposure to English and hand in the log at every recording session.*
*You, your classmates, and your teacher will be recorded on videotape in your **** course.*
You will verify the analyses of your recordings after the four recording sessions.
You will be free to listen to your recordings at the end of the data collection, and to discuss your progress and work on your speech with an ESL teacher in a one-on-one tutorial session.
You will not tell your classmates or instructor about the activities you do in this research project.

*Participating or not participating in this study will not affect your grades in ***** in any way.*
You will receive $75 for your participation. One-third of this money will be paid at the beginning of the study, one-third in the middle, and one-third at the end. You must fully participate in the study to receive the full amount.

- I understand the purpose of this study and know about the risks, benefits, and inconveniences involved in this research project.
- I understand that I am free to withdraw at any time from the study without any penalty or prejudice.
- I understand that this research will not affect my grades or evaluation of my work.

- I understand that my name will not be used during and after this research project.
- I understand how the data in this project may be used, especially with respect to publication, communication, and dissemination of results.

I have read the above and I understand all of the above conditions. I freely consent and voluntarily agree to participate in this study.

Name (please print) _____

Signature _____ Date _____

All Students in Oral Communication Course

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in the research project entitled:

Spet

And conducted by: _____Sara Kennedy, McGill University_____

1. Purpose – *The purpose of this research is to see how non-native English speakers speak English.*

2. Procedures –

*In your ***** course, you, your classmates, and your teacher will be videotaped once every week. The videotape will not be used for any part of your grades and your teacher will only see it (with your permission) after the course grades are submitted.*

In all the video recordings, you will be identified with a code so that no one knows your name. The results of this study may be presented in a workshop, a class, a research presentation, a Ph.D dissertation, or a journal article. Your name will not be mentioned in any presentation of the results.

If the researcher wishes to show in a public forum a segment of video in which your image or your voice will be shown, the researcher will need your written consent (in addition to this paper). Without your written consent, the segment will not be shown).

3. Conditions of Participation –

*You, your classmates, and your teacher will be recorded on videotape in your **** course once a week.*
*Participating or not participating in this study will not affect your grades in **** in any way.*

- I understand the purpose of this study and know about the risks, benefits, and inconveniences involved in this research project.
- I understand that I am free to withdraw at any time from the study without any penalty or prejudice.
- I understand that this research will not affect my grades or evaluation of my work.

- I understand that my name will not be used during and after this research project.
- I understand how the data in this project may be used, especially with respect to publication, communication, and dissemination of results.

I have read the above and I understand all of the above conditions. I freely consent and voluntarily agree to participate in this study.

Name (please print) _____

Signature _____Date _____

APPENDIX C

INFORMATION ON NON-NATIVE TALKERS

Language Background Questionnaire

Name: _____ Gender: Male____ Female _____

Phone number: _____ Email address: _____

Is your hearing normal as far as you know? Yes: _____ No: _____

Program at McGill (including year): _____

Date of birth: _____ Birthplace (City, Country) : _____

What do you consider to be your native language?

Were you exposed to this language since birth? Yes: _____ No: _____

What do you consider to be your second language? English:_____ Other:_____

At what age did you start learning your second language? _____

What language do you speak at home now? _____

What is the native language of your mother? _____ Your father? _____

In what language did you attend school in your country? Please circle the appropriate one

-elementary school: native language only English only Other: _____

-high school: native language only English only Other _____

-university: native language only English only Other _____

Please rate your ability to speak, listen to, read and write **your native language** by using the scales in the box below. Please note that 1= extremely poor and 9= extremely fluent

Native Language **1=Extremely Poor** **9= Extremely Fluent**

speaking	listening	Reading	Writing
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

Please rate your ability to speak, listen to, read and write **French** by using the scales in the box below. Please note that 1= extremely poor and 9= extremely fluent

French	**1=Extremely Poor**	**9= Extremely Fluent**	
speaking	listening	Reading	Writing
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

Please rate your ability to speak, listen to, read and write **English** by using the scales in the box below. Please note that 1= extremely poor and 9= extremely fluent

French	**1=Extremely Poor**	**9= Extremely Fluent**	
speaking	listening	Reading	Writing
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

What other languages do you speak? How well do you speak them?

At what age did you start learning English? _____

Please list amount of time (per week) that you learned English in:
Primary school (time) _____ for (number of years) _____
Middle school (time) _____ for (number of years) _____
High school (time) _____ for (number of years) _____
University (time) _____ for (number of years) _____

In your country, did you learn any English outside of school? How?

_____

Have you ever lived in any country other than your native country and Canada? For how long?
For what purpose? _____

When did you arrive in Canada ? _____
Have you ever worked in a place where only English was used? Yes___ No ___
Where (City, country)_____ For how long? _____
What institution did you work in? _____

Have you ever taken an English course that focused on speaking? Yes ____ No _____
If yes, where? _____ For how long? _____

What are your strengths in speaking English?

_____

What are your weaknesses?

_____

What do you want to improve the most in your English speaking?

_____

On average, how many hours/minutes do you hear or speak English every day:

At school _____

At home _____

In your social life _____

APPENDIX D

INFORMATION ON LISTENERS

Language Background Questionnaire

Name: _____ Gender: Male____ Female _____

Phone number: _____ Email address: _____

Is your hearing normal as far as you know? Yes: _____ No: _____

Date of birth: _____ Birthplace (City, Province/State) : _____

What do you consider to be your native language? English: ____
Other: _____

Were you exposed to this language since birth? Yes: _____ No: _____

What do you consider to be your second language? French:_____ Other:_____

At what age did you start learning your second language? _____

How well do you speak your second language? 1 2 3 4 5 6
Poor Excellent

What language do you speak at home now? _____

What is the native language of your mother? _____ Your father? _____

What is your program of study and year? _____

Please rate how frequently you have contact with people who speak English to you, but whose English has a noticeable non-native accent.

Never rarely occasionally fairly frequently frequently very often

Please describe in what context you meet the people above and the mother tongues of the people.

Table D1

Listener Information from the Retell Task

	Name code	Sex	Age at listening	Program of study	Accent exposure
			Time 1		
1	MiLa	M	25	Science, Post-doc	Rarely
2	ElEa	F	19	Arts	Occasionally
3	CaLa	M	22	Science	Occasionally
4	DaLe	F	23	Arts	Occasionally
5	EmBl	F	19	Arts	occasionally
6	SiHl	M	18	Engineering	Fairly frequently
7	ErFr	F	22	Education	Fairly frequently
8	JeTu	F	21	Fine Arts	Fairly frequently
9	DaFr	M	18	Arts	Fairly frequently
10	FrSl	F	23	Management	Fairly frequently
11	MaHu	F	19	Education	Fairly frequently
12	FoBe	M	18	Engineering	Fairly frequently
13	JeAs	F	19	Science	Frequently
14	DaMu	F	21	Science	Frequently
15	MeTh	F	22	Education	Very often
16	ElRh	M	19	Management	Very often
			Time 3		
1	CaGo	F	18	Education	Never
2	LiSh	F	20	Engineering	Occasionally
3	AnSh	F	20	Arts	Occasionally
4	ToJo	M	43	Computer technician	Occasionally
5	EmFr	F	19	Arts	Occasionally
6	MoTe	F	19	Arts	Fairly frequently
7	ErSt	F	23	Arts	Fairly frequently
8	AmMu	F	21	Arts	Fairly frequently
9	OrFa	M	20	Arts	Fairly frequently
10	CaOs	F	21	Arts	Fairly frequently
11	KrMo	F	16	Science	Frequently
12	MaSp	F	19	Science	Very often
13	MaWa	F	19	Fine Arts	Very often
14	AlHa	M	19	Music	Very often
15	StBa	F	20	Education	??
			Time 4		
1	HaDe	F	19	Science	Rarely
2	TaPi	F	17	Arts	Occasionally
3	AvBa	F	29	Arts	Occasionally
4	SaMu	F	19	Arts	Occasionally
5	NiSh	F	20	Arts	Occasionally
6	MaDu	F	20	Arts	Fairly frequently
7	KaSc	F	22	Arts	Fairly frequently

Table D1 (*continued*)

	Name code	Sex	Age at listening	Program of study	Accent exposure
8	LaHa	F	20	Arts	Fairly frequently
9	PaCa	F	30	Education, Master's	Fairly frequently
10	LeMa	F	27	Science	Fairly frequently
11	ReSh	F	20	Education	Fairly frequently
12	AmFr	F	18	Science	Frequently
13	JoAl	F	29	Staff	Very often
14	JuHa	F	25	Education	Very often
15	ViBe	F	18	Arts	Very often

Table D2

Listener Information from the On-Line Word Intelligibility Task

	Code	Sex	Age at listening	Program of study	Accent exposure
			Time 1		
1	TyKr	M	19	Arts	Occasionally
2	MaDa	F	23	Science	Occasionally
3	GaLe	M	???	Science	Occasionally
4	SoBu	F	19	Arts	Occasionally
5	MeZe	F	20	Science	Occasionally
6	BlGr	M	20	Arts	Fairly frequently
7	ChMa	F	26	Science	Fairly frequently
8	AlRi	F	20	Science	frequently
9	CoTh	F	20	Science	Very often
10	MePe	F	21	Arts	Very often
			Time 3		
1	CoTs	F	23	Science	rarely
2	SaGi	F	20	Arts	rarely
3	AnDa	F	25	Science	Occasionally
4	AnHe	F	23	Education	Fairly frequently
5	RePa	F	21	Science	Fairly frequently
6	SaOr	F	23	Arts	Fairly frequently
7	EtTr	M	22	Arts	frequently
8	ShCu	F	20	Arts	frequently
9	FeMa	F	22	Arts	Very often
10	MaGo	M	24	Arts	Very often
			Time 4		
1	LaSu	F	19	Science	Rarely
2	DaBr	F	22	Education	Rarely
3	LaMo	F	22	Education	Occasionally
4	EmGr	F	19	Arts	Occasionally
5	AlSe	M	20	Management	Fairly frequently
6	AlNe	F	19	Arts	Fairly frequently

Table D2 (*continued*)

	Code	Sex	Age at listening	Program of study	Accent exposure
7	EmBa	F	24	Arts	Fairly frequently
8	JaRi	M	25	Architecture, Master's	frequently
9	PaMa	F	18	Science	frequently
10	ScBu	M	36	Research Associate	Very often

APPENDIX E

COLT, PART A

Research Instruments for Classroom Observation

FIGURE 1

Communicative Orientation of Language Teaching (COLT) Observation Scheme

COLT PART A

Communicative Orientation of Language Teaching Observation Scheme

School ..... Grade(s) ..... Observer .....

Teacher ..... Lesson (min.) ..... Visit No .....

Subject ..... Date ..... Page .....

TIME	ACTIVITIES & EPISODES	PARTICIPANT ORGANISATION							CONTENT								CONTENT CONTROL			STUDENT MODALITY					MATERIALS							
		Class			Group		Indiv.		Manag.		Language				Other topics										Type				Source			
																									Text						
		T→S/C	S→S/C	Choral	Same task	Different tasks	Same task	Different tasks	Procedure	Discipline	Form	Function	Discourse	Sociolling.	Narrow	Broad	Teacher/Text	Teacher/Text/Stud.	Student	Listening	Speaking	Reading	Writing	Other	Minimal	Extended	Audio	Visual	L2-NNS	L2-NS	L2-NSA	Student-made
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33

APPENDIX F

TRUE-FALSE SENTENCES

Practice

Most adults can run a mile in two seconds.
Cats are generally good drivers.
Many people go swimming in the ocean.
Tables and chairs can sing and dance.
You can tell time with a shoe.

First Recording Session – Time 1

Keys can open locked doors.
You can see animals at the zoo.
Most animals need air to breathe.
Many people enjoy looking at paintings.
Most people love to go to the dentist.
Doctors often work in hospitals.
Spaceships use hot dogs as fuel.
You can have dinner in a restaurant.
Some people like to read poetry.
Most sailors keep their boats at the airport.

Second Recording Session – Time 2

Opera singers have good voices.
You can borrow a bicycle from the library.
Most mothers think their children are ugly.
Many children's books have pictures.
There are snakes in the jungle.
Many people have telephones.
A raincoat makes an excellent bathing suit.
Refrigerators keep food extremely hot.
You can write with a pen or a pencil.
The United States is a small island.

Third Recording Session – Time 3

Some countries have large armies.
Some babies enjoy reading novels.
The sun always sets in the north.
Some people find music relaxing.
Most cowboys like to ride horses.
Students often use notebooks.
Spaghetti grows on tall trees.
Most people take photographs with their toes.
Most people wear hats on their feet.

Fourth Recording Session – Time 4

A monkey is a kind of bird.
There are many cities on the moon.
Many people drink coffee for breakfast.
All men can have babies.
Hungry cats like to chase mice.
Ships travel on the water.
Gasoline is an excellent drink.
You can start a fire with a match.
People eat through their noses.
Gold is a valuable metal.

APPENDIX G

PROMPTS FOR EXTENDED SPEECH

Warm-up Prompts:

Hometown
Think of your hometown, the place where you were born and/or lived as a child. What is it like? What is it famous for? What kinds of activities can people do there? Are there bad things about your hometown? What is your favourite thing about your hometown?

Film retell
Licence to Kill - a reversal of the typical hunting scenario. A bear takes his gun and goes into the city to try his luck

Place
Think of a place that you remember well. What was it like? Why do you remember it? You will have five minutes to think of what you will say. You can write notes, but you will not be able to use them when you talk.

Person
Think of a person who has made a good impression on you. Why did s/he make that impression on you? What is s/he like?

Personal Anecdote Prompts

Purchase
Think of a time when you made a very good purchase or a very bad purchase. Where were you? What did you buy? You will have five minutes to think of what you will say. You can write notes, but you will not be able to use them when you talk.

Surprise
Think of a time when you were surprised by the behaviour of someone else (he or she didn't behave as you expected). What did this person do? Why were you surprised? Did this person surprise you again later on? You will have five minutes to think of what you will say. You can write notes, but you will not be able to use them when you talk.

Job/Course
Think of a job or a course you were in which was really bad or really good. What was the job/course about? What was the boss or instructor like? Why was it so good or so bad? What did you get out of the experience?

Decision

Think of a time you made a very good or very bad decision. What did you decide? What were your reasons for making the decision that you did? What was the result of your decision?

Film Prompts

George and Rosemary – a retired man fantasizes about romance with his neighbour of the same age, but lacks the courage to talk to her

Strange Invaders – a childless couple welcome a little creature into their home, turning their lives upside down and revealing the little one's true origins

Onions and Garlic – two peasants try to make their fortune in a far-away land, learning that value is in the eye of the beholder

The Lump – an unattractive, friendless man grows a lump on his head which changes his status and his life

APPENDIX H

LANGUAGE LOG SAMPLES AND INSTRUCTIONS

Name:
Date:

1 – a little of the language I used/heard/read
2 – some of the language I used/heard/read
3 – most of the language I used/heard/read
4 – all of the language I used/heard/read

How much time *in minutes* did you spend using/hearing/ reading English?

Put any comments about the activity below (optional)

Table H1
Language Log Completion: English Use/Exposure

Category	***Sub-Category***	English Use/Exposure during Activity	Time	Comments

Table H2
Log Categories: Exposure to and Use of English Daily

Category	***Sub-Category***
1. Daily Living	a. Appointments (*medical, dental, hair, immigration*) b. Child care (*picking up or dropping off children, making or receiving phone calls about children*) c. Paperwork *(writing cheques, doing accounts, filling in forms)* d. Running errands/shopping *(short trips to buy or deliver something, going to bank, post office, paying bills, grocery shopping)* e. Chores and daily tasks (*reading recipes while cooking, organizing material)* f. Thinking (*planning, praying, problem solving, remembering*) g. Transportation *(going from one place to another)*
2. Social Interaction	a. Attending a meeting b. Electronic chat (*involves writing such as ICQ, chat rooms)* c. Face-to-face conversation (general) d. Face-to-face conversation (personal) e. Face-to-face discussion (*specific topic*) f. Meeting with academic advisor/professor g. Online conversation (*involves speaking such as NetMeeting,*) h. Reading/writing email i. Reading/writing formal correspondence (*business letters)* j. Reading/writing personal correspondence *(personal letters)* k. Telephone conversation
3. Academic Work	a. Borrowing resources from the library (*going to library, taking out books*) b. Collecting data/doing an experiment c. Computing d. Doing language log e. Face-to-face discussion (*specific topic*)

	f. Listening to a presentation/lecture (*by professor, lecturer, classmate*) g. Making a presentation *(to a class, a group or the public)* k. Photocopying l. Preparing for a presentation m. Reading an academic article/text n. Reading instructions o. Solving problems p. Studying for an exam/test q. Surfing the Web/library searches r. Telephone conversation s. Thinking (analyzing, problem solving, planning) t. Writing a memo/report u. Writing an assignment or paper
4. Attending Class	a. Collecting data/doing an experiment b. Computing c. Face-to-face discussion d. Listening to a presentation/lecture e. Making a presentation f. Reading academic article/text g. Reading instructions h. Solving problems i. Surfing the Web/Library searches j. taking notes k. Thinking l. Writing a memo/report m. Writing an assignment or a paper n. Writing an exam
5. RA/TA	a. Attending a meeting b. Borrowing material from library c. Chores / daily tasks d. Course planning and preparation

	e. Face to face conversation f. Face to face discussion g. Helping students h. Listening to a lecture or speaker i. Making a speech or presentation j. Marking k. Reading a reference book l. Reading academic articles/texts m. Reading instructions n. Reading-writing email / messages o. Surfing the web and library searches p. Taking notes/making lists q. Talking on the telephone r. Writing an academic paper
6. Recreation	a. Attending a party b. Eating at restaurant or someone's house c. Exercising/sports (*going to the gym, playing team sports)* d. Going on a trip e. Going to a dance or nightclub f. Hobby (*arts and crafts, bird watching, collecting, etc.)* g. Personal writing/journal (*such as a diary*) h. Playing computer games i. Reading comics j. Reading fiction (*novels, short stories*) k. Reading non-fiction (*biographies, self-help, newspapers, historical)* l. Surfing the Web/computing m. Watching TV/movie

To complete the daily language log:

1. Find a moment when you have about 10 minutes of quiet time, either at the end of the day or at the beginning of the morning.

Date: February 31, 2005.

1 – a little of the language I used/heard/read
2 – some of the language I used/heard/read
3 – most of the language I used/heard/read
4 – all of the language I used/heard/read

How much time *in minutes* did you spend using/hearing/ reading English?

Put any comments about the activity below (optional)

Table H3

Sample Log Completion

Category	***Sub-Category***	English Use/Exposure during Activity	Time	Comments
1. Daily Living	a. Appointments	4	15 min.	dentist
	d. Running errands/shopping	2	10 min.	drugstore
	e. Chores and daily tasks	1	2 min.	giving instructions on using washing machine
	g. Transportation	4	3 min.	Buying ticket and reading metro map
2. Social Interaction	c. Face-to-face conversation (general)	4	3 min.	Talk about new professor with Jorge
	d. Face-to-face conversation (personal)	4	15. min	Lunch with Diane
	f. Meeting with academic advisor/professor	4	25 min.	Discussed proposal
	h. reading/writing email	1	10 min.	Personal emails

	i. Reading/writing formal correspondence	3	15 min.	Reading letters from X university and Y granting agency
	k. Telephone conversation	4	5 min.	With Stan
3. Academic Work	b. Collecting data/doing an experiment	1	4 min.	Ask technician questions
	e. face to face discussion	4	15 min.	Biochemistry presentation
	d. Doing language log)	4	15 min.	Looked at agenda
	f. Listening to a presentation/lecture	3	20 min.	Biochemistry/asked friends to explain something in my language
	l. Preparing for a presentation	3	20 min.	practised with help from a friend.
	m. Reading an academic article/text	3	35 min.	Used translator sometimes
	q. Surfing the Web/library searches	4	20 min.	For presentation
4. Attending Class	c. Face-to-face discussion	3	20 min.	Discussion in seminar about enzymes
	d. Listening to a presentation/lecture	4	50	Prof. and student presenter
	f. Reading academic article/text	4	5	Prof in lecture said look page xx
	h. Solving problems	2	5	Discuss with classmates
	i. taking notes	3	50 min.	
5. RA/TA	d. Course planning and preparation	4	40 min.	tutorial

	f. Face to face discussion	4	10	With student about grade
	j. Marking	4	50 min.	Lab reports
	n. Reading-writing email /messages	4	10 min.	From/to students
	q. Talking on the telephone	4	5 min.	With student
	r. Writing an academic paper	4	30 min.	Methodology section
6. Recreation	c. Exercising/sports	2	20 min.	Basketball game
	m. Watching TV/movie	4	20 min.	news

APPENDIX I

RATING SCALES

Based on Munro and Derwing (1995)

Instructions for Listening and Rating

You will hear excerpts from the same 11 short stories.

- After each story, you are going to rate the speaker on scales of 1 to 9 for accentedness, comprehensibility, and fluency. Please listen to the **complete** excerpt **before** rating it.
- Put an X in one of the 9 boxes for each scale. *Please try to use the entire scale.*
- Please put the X in one box or another, and not in between. Remember, for comprehensibility, you are **not** rating speakers on whether they tell interesting stories. You are rating speakers on how easy it is for you to understand the speaker.

PRACTICE

Comprehensibility

1	2	3	4	5	6	7	8	9
extremely easy to understand								extremely difficult to understand

Accent

1	2	3	4	5	6	7	8	9
no foreign accent								very strong foreign accent

Fluency

1	2	3	4	5	6	7	8	9
Extremely fluent BEGIN								Extremely dysfluent

APPENDIX J

TIMELINE AND TALKER COUNTERBALANCING

Table J1
Timeline for Talker Recording Sessions

Recording session	Tasks
1 (January 16-24)	1) Consent form signing 2) Questionnaire completion 3) Warm-up - read aloud 15 T-F sents, incl. five warm-up 4) T-F sentence recording 5) Warm-up – hometown description recording, 1 min. preparation 6) George and Rosemary watch 2X and retell, purchase story tell, 5 min. preparation each (counterbalanced: half movie then story; half story then movie) 7) Log explanation and practice
2 (Feb. 27 – Mar. 7)	3) Warm-up - read aloud 15 T-F sents, incl. five warm-up 4) T-F sentence recording 5) Warm-up – watch bear film 2X, retell recording, 1 min. preparation 6) Strange Invader watch 2X and retell, surprise story tell, 5 min. preparation each (counterbalanced: half movie then story; half story then movie) 7) Log debriefing
3 (Apr. 10-21)	3) Warm-up - read aloud 15 T-F sents, incl. five warm-up 4) T-F sentence recording 5) Warm-up – place description recording, 1 min. preparation 6) Onions and Garlic watch 2X and retell, school/job story tell, 5 min. preparation each (counterbalanced: half movie then story; half story then movie) 7) Log debriefing
4 (May 8-18)	3) Warm-up - read aloud 15 T-F sents, incl. five warm-up 4) T-F sentence recording 5) Warm-up – person recording, 1 min. preparation 6) The Lump watch 2X and retell, decision story tell, 5 min. preparation each (counterbalanced: half movie then story; half story then movie) 7) Log debriefing

Table J2

Counterbalancing of Extended Speech Tasks for Treatment Group

Code	L1	First Recording Session	Second Recording Session	Third Recording Session	Fourth Recording Session
Tai Ning	Mandarin	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Javier	Spanish	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Sigman	Tamil	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Marie-Pier	French	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Hui	Mandarin	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Piotr	Russian	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Bao	Mandarin	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Xiao	Mandarin	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Christine	French	1)film 2) personal	1) personal 2) film	1) personal 2) film	1)film 2) personal

Table J3

Counterbalancing of Extended Speech Tasks for Control Group

Code	L1	First Recording Session	Second Recording Session	Third Recording Session	Fourth Recording Session
Ahmed	Farsi	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Lupe	Spanish	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Esteban	Spanish	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Ma	Mandarin	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Ping	Mandarin	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Feng	Mandarin	1) film 2) personal	1) personal 2) film	1) personal 2) film	1) film 2) personal
Jiao	Mandarin	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film
Xing	Mandarin	1) personal 2) film	1) film 2) personal	1) film 2) personal	1) personal 2) film

APPENDIX K

SEMI-STRUCTURED INTERVIEW QUESTIONS

Instructor

Background

What are origins of the course? For what reason was it developed? When it was developed, how did you view it ideally?

You've said that course has changed some since its beginning. How has it changed? Why did you add things on looking up and reading, etc.?

Class

What are your goals/expectations for the students in this class?
How do you see this class as fitting into the life or demands of a graduate student, what they need?
What are your goals/expectations for your own teaching in this class?

How did you feel about how the class went this term? What went well?
What sorts of things did people improve on in the course and what sorts of things did they generally not improve on?

What would you do differently in this class if you had another chance w/ same people?

Why did you focus on the things that you did over the course? What informed your choice of the things that you taught and evaluated?

Materials

How long have you been using *Accurate English* and other materials? What do you like about them, why use those ones in particular?

Methodology

How do you see yourself in the class? What is your role in preparing and conducting the class?

To present new stuff, you often use discovery approach. Why do that?

Students did a lot of work in pairs. Why do that? Are people worried about others' accents? How did class get along w/ each other? Did some really help others a lot?

I noticed you started to correct people more often in whole class near the end of course. What's your policy on giving feedback in class? When and why do you do it?

What is your teacher training and experience teaching? How did your teacher training and experience prepare you to teach this class?

You are very organized in your classes. Are there ever times when you would depart from your lesson plan, and what are those times?

You did more feedback on segments, especially final consonants, in later part of course. Was that planned?

You called individual students' attention to things they need to work on that they will be doing in class. Done to make sure students' pay attention?

Are there any things that you do in your class that might be considered unorthodox or "wrong" according to common practice in pronunciation teaching? Why do you do it?

There is emphasis on conscious awareness of patterns in your class. Do you do this because of the students you have, your own learning style, or your pedagogical approach?

How do you think adults learn pronunciation?

What do you think are biggest mistakes students make about trying to learn pronunciation and oral communication?

NN Talker Participants

How did you do with your graduate work this term? What were you working on?

How do you think your English developed over the last 5 months (Jan. to May)?

How have your goals and your estimations of your strengths and weaknesses changed?

What do you find hardest about learning/speaking English?

How did you interact with your classmates? Did you make any friends? Did anyone help you?

Have your goals changed over past 5 months w.r.t your English development?

What was the most difficult thing for you here as a non-native English speaking grad student?

What do you think is the best way for you to learn speaking?

APPENDIX L

LENGTH OF PERSONAL ANECDOTES

Table L1
Anecdote Lengths of Treatment Group

Code	L1	First Recording Session (minutes)	Second Recording Session (minutes)	Third Recording Session (minutes)	Fourth Recording Session (minutes)
Tai Ning	Mandarin	2.45	3.22	13.22	4.52
Javier	Spanish	1.48	1.87	3.4	2.77
Sigman	Tamil	2.78	2.5	2.95	3.0
Marie-Pier	French	1.25	1.95	2.06	2.1
Hui	Mandarin	0.9	1.83	2.83	2.28
Piotr	Russian	5.5	N/A	5.48	0.51
Bao	Mandarin	1.93	2.4	3.2	2.63
Xiao	Mandarin	1.00	1.1	1.67	1.67
Christine	French	2.05	1.97	2.52	2.83

Table L2
Anecdote Lengths of Control Group

Code	L1	First Recording Session (minutes)	Second Recording Session (minutes)	Third Recording Session (minutes)	Fourth Recording Session (minutes)
Ahmed	Farsi	1.47	0.97	1.72	1.27
Lupe	Spanish	2.2	1.77	3.62	2.42
Esteban	Spanish	2.07	3.03	2.08	1.53
Ma	Mandarin	4.1	2.65	10.12	4.4
Ping	Mandarin	1.43	6.15	4.62	6.02
Feng	Mandarin	5.2	3.13	1.93	4.48
Jiao	Mandarin	8.28	11.12	18.67	10.8
Xing	Mandarin	4.42	1.83	11.58	5.8

APPENDIX M

LISTENER CONSENT FORMS

Retell Task

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in the research project entitled:
SPET
And conducted by: Sara Kennedy, McGill University

1. ***Purpose*** *– The purpose of this research is to investigate speaking and listening in English.*

2. ***Procedures*** *–*

   You will hear eleven short stories. You will listen to each story and take notes on the story. After the story is finished, you will use your notes, and be recorded telling the story again. After re-telling all the stories, you will then listen to excerpts from the original stories and rate them on three different scales. Your notes, recordings, and ratings will receive a code so that no one knows your name. The results of this study may be presented in a workshop, a research presentation, or a Ph.D dissertation and subsequently published in a journal. Your name will not be mentioned in any presentation of the results.

3. ***Conditions of Participation*** *–*

   You will hear stories, take notes, re-tell the story again while being recorded, then listen to excerpts of the stories again and rate them. This will take about one hour and a half.
   You will then receive $15 (CAN).
   Participating in this study will not affect your grades in any course.

- I understand the purpose of this study and know about the risks, benefits and inconveniences involved in this research project.
- I understand that I am free to withdraw at any time from the study without any penalty or prejudice.
- I understand that this research will not affect my grades or evaluation of my work.
- I understand that my identity will be kept secret for this research project.
- I understand how the data will be used, especially with respect to publication, communication and dissemination of results.

I have read the above and I understand all of the above conditions. I freely consent and voluntarily agree to participate in this study.

Name (please print) _____

Signature _____ Date _____

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in the research project entitled:
SPET
And conducted by: Sara Kennedy, McGill University

1. ***Purpose*** *– The purpose of this research is to investigate the comprehension of speaking in English.*

2. ***Procedures*** *–*

   You will hear eleven short stories. You will listen to each story and pause the recording when you do not understand a word. When you pause the recording, you will be recorded talking about what you did not understand and why you think you did not understand it. The recording of your voice will receive a code so that no one knows your name. After hearing all the stories, you will then hear excerpts of the stories and rate them on three aspects. The results of this study may be presented in a workshop, a research presentation, or a Ph.D dissertation and subsequently published in a journal. Your name will not be mentioned in any presentation of the results.

3. ***Conditions of Participation*** *–*

   You will hear stories, pause the recording when you don't understand a word, then be recorded explaining what and why you didn't understand. You will then hear excerpts of the stories and rate them.This will take about one hour and a half.
   You will then receive $15 (CAN).
   Participating in this study will not affect your grades in any course.

- I understand the purpose of this study and know about the risks, benefits and inconveniences involved in this research project.
- I understand that I am free to withdraw at any time from the study without any penalty or prejudice.
- I understand that this research will not affect my grades or evaluation of my work.
- I understand that my identity will be kept secret for this research project.
- I understand how the data will be used, especially with respect to publication, communication and dissemination of results.

I have read the above and I understand all of the above conditions. I freely consent and voluntarily agree to participate in this study.

Name (please print) _____

Signature _____ Date _____

APPENDIX N

INSTRUCTIONS FOR LISTENERS

Story Retell Task

Instructions for Listening and Note-Taking

You will listen to 11 short stories. After each story, you are going to tell the story again in as much detail as possible.

- While you listen to each story, take detailed notes on the sheets provided.
- At the end of each story, you will hear "Stop." Pause the CD player. Look at your notes and organize your thoughts.
- When you are ready, start the digital recorder and record your story. Begin by saying the story number, then use your notes to tell the story. If there is something in the story you did not understand, try to tell the story as you understand it, and also mention what you did not understand.
- Stop the recorder when you are finished telling the story. Take a little break, then when you are ready, press pause again on the CD player and listen to the next story, taking detailed notes, etc..

PLEASE DO NOT STOP THE CD PLAYER UNTIL EACH STORY IS FINISHED.

The first story allows you to practice taking notes and telling a story. The speaker was asked to tell a story about a time when they were surprised by someone else's behaviour.

(*Time 1*) For the other 10 stories, the speakers were asked to tell a story about a time they had made a very good or very bad purchase.

(*Time 3*) For the other 10 stories, the speakers were asked to tell a story about a job or a course they had had which was really bad or really good.

(*Time 4*) For the other 10 stories, the speakers were asked to tell a story about a time that they had made a very good or very bad decision.

Instructions for Listening

You will listen to 11 short stories, one at a time. While you are listening to each story, every time you don't understand a word, you will make a signal. The recording will be paused, and you will talk about the word you didn't understand, and why you think you didn't understand it.
If you are listening and at **first** you don't understand a word, **then** you understand it, make a signal. You will talk about which word you didn't understand at first, why you think you didn't understand it, and how you finally came to understand it.

The first story allows you to practice signalling when you don't understand. The speaker was asked to tell a story about a time when they were surprised by someone else's behaviour.

(*Time* 1) For the other 10 stories, the speakers were asked to tell a story about a time they had made a very good or very bad purchase.

(*Time* 3) For the other 10 stories, the speakers were asked to tell a story about a job or a course they had had which was really bad or really good.

(*Time* 4) For the other 10 stories, the speakers were asked to tell a story about a time that they had made a very good or very bad decision.

APPENDIX O

SUMMARY OF CLASSES

Class	Interaction	Instructional focus of class	Instructional activities	Content
1	65% T/C 29% group 6% S-C	56% focus words (including 22% pitch changes) 40% pausing	T and Ss deciding where to put pauses in sentences. T and Ss reading sentences aloud with pauses. T and Ss reading sentences aloud with last word as focus word. T and Ss practicing bobbing head on focus word. T and Ss reading aloud sentences with pitch change on focus word.	Form – 100% Procedure – 22% Function – 9% Sociolinguistics – 4%
2	64% T-C 22% group 12% S-C 2% indiv	82% linking 9% focus words	T and Ss marking focus words in sentences with upward arrows. T and Ss marking links between words in thought groups in sentences. T and Ss reading sentences aloud with linking. Ss identifying C-V and V-V links in sentences.	Form – 89% Procedure – 19%

Class	Interaction	Instructional focus of class	Instructional activities	Content
3	51% group 49% T-C	91% lexical stress (including 69% suffixes)	T and Ss counting syllables in words. T and Ss writing down suffixes. T and Ss listing words ending in certain suffixes. T and Ss predicting stress placement based on suffixes. Ss dividing word root from suffix and marking stress placement in root. Ss suggesting guidelines for stress placement for specific suffixes.	Form – 96% Procedure – 26%
4	58% T-C 42% group	89% lexical stress	Ss predicting stress placement for words with prefixes. T and Ss reading aloud words with changeable stress depending on part of speech, both in isolation and in dialogues. T and Ss reading aloud questions with changeable words. Ss identifying secondary stress in words in read-aloud sentences. Ss identifying reduced and unreduced vowels in syllables. Ss suggesting guideline for reducing vowels in certain suffixes.	Form – 86% Procedure – 21% Discourse – 1%

Class	Interaction	Instructional focus of class	Instructional activities	Content
5	59% T-C 41% group	47% reading looking up 45% lexical stress	T and Ss practicing reading from a text while periodically looking up. Ss reading paragraph along with T or marking lexical stress. Ss identifying number of syllables and stress placement in words. Ss practicing reading words with reduced vowels.	Form – 85% Procedure – 47% Discourse – 2%
6	49% group 45% T-C 6% indiv	49% *-ed* endings 35% - *es* endings 18% lexical stress/linking	T and Ss identifying number of syllables in words with and without *–ed* endings. Ss suggesting 3 rules of pronunciation of words with *–ed.* T and Ss classifying words with *–ed* into 3 categories. Ss identifying stress placement and internal links within words. Ss suggesting appropriate pronunciation of words ending in *–es*. Ss suggesting pronunciation rules for words with *–es.* T and Ss reading aloud text with *–ed* and *–es* words.	Form – 88% Procedure – 13%
	mid-term presentations			

Class	Interaction	Instructional focus of class	Instructional activities	Content
7	63% T-C 37% group	41% effective body language 32% presentation conclusions 21% comprehension checks	Ss brainstorming possible conclusion for presentation transcript. Ss choosing best conclusion from a selection and discussing choice. Ss suggesting possible questions for comprehension checks. Ss suggesting positive and negative uses of body language in presentations. Ss completing worksheet on interpreting body language.	Function – 49% Discourse – 31% Form – 23% Procedure – 22%
8	63% T-C 37% group	74% combination of lexical stress in compound and atypical words, linking, and thought groups 14% problem-type presentation structure.	T and Ss read aloud number, compound, or acronym words, or sentences with such words, focusing on lexical stress, pausing, and linking. Ss fill in outline for problem-type presentation.	Form – 76% Discourse – 13% Procedure – 12%
Total	58% T-C 39% group 2% S-C 1% indiv			78% - Form 24% - Procedure 7% - Function 6% - Discourse 0.5% - Sociolinguistics

APPENDIX P

SELECTED AURAL/ORAL ACTIVITY SUBCATEGORIES

FOR TWO DOMAINS AND MODES

Academic domain	Interactive	3e – face-to-face discussion 3r – telephone conversation 4c – face-to-face discussion (class) 5a – attending a meeting 5e – face-to-face conversation 5f – face-to-face discussion 5g – helping students 5q – talking on the telephone
	Non-interactive	3f –listening to a presentation/lecture 3g - making a presentation 4d – listening to presentation/lecture 4e – making a presentation 5h – listening to a lecture or speaker 5i – making a speech or presentation
Social domain	Interactive	2a – attending a meeting 2c – face-to-face conversation (general) 2d – face-to-face conversation (personal) 2e – face-to-face discussion 2f – meeting with academic advisor/professor 2k – telephone conversation 6a – attending a party 6b – eating at restaurant or someone's house
	Non-interactive	6m – watching TV/movie

APPENDIX Q

EXAMPLE OF ORIGINAL AND RETELL CODING

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
1a	First of all, I don't think there're bad decisions, first.	evaluation	1a	The third story is about a man who doesn't believe that there are bad decisions	Y	evaluation
1b	Because even if they, they seem bad,		1c	because he thinks that	Y	
1c	at the end, all of them contribute to the, my experience.			no matter what, you'll learn from your experiences		
2a	And when I, when somebody asks me,	evaluation				
	would you change something		4a, 4b	Anyways, he decided to get his master's at university in chemical engineering	Y, W	IE
	if you can go back,		7	and there's a possibility	Y	goal
2b	I said no.			of him getting a position at a mining, in a mining department		
2c	no, because I am who I am because of my experiences.		8a	but he had more of a chance of getting this position	Y	attempt
3r1	But I have a very good one, decision.	abstract	8b	if he is part of the engineering department	PM	

Appendix Q (*continued*)

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
4a	I, I decided to start studies as, to, to get a Master in my university.	IE	10	so in the end he got his master's	Y	outcome
4br1	And that Master was in the, given in the department of, mining department.		9	and got the job at McGill in the mining department	PW	
5r1	I was finishing my studies of chemical engineering, in the chemical engineering department.	setting	13a, 13b	And he's now part of a program called Jong academics in the engineering department	DW, Y	outcome
4br2	And I went to the mining engineering department to get my Master	attempt	14	and he says he has to work for twice the time he was there	Y	outcome
6	And what were the reasons?	discourse	3r1	but he thinks	Y	evaluation
7	The reason were that,	goal		he made a good decision.		
	there was a possibility of getting a position there in the mining department					
8a	And the only way that I could try to get the position,	attempt				
8b	it was being part of the department.					

Appendix Q (*continued*)

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
5r2	I mean, I was before part of the, the chemical engineering department	attempt				
4br3	and I moved to the mining department.					
9r1	And well, the, the results, I got the job at, at the, at the end.	outcome				
10r1	I got my master	outcome				
11	and there I met then, I met there one person that works here that was my contact to came here.	IE				
10r2	And once I got, I, I got my Master	outcome				
12	and I, I was applying to McGill,	attempt				
9r2	they decided that	outcome				
	I was the right person to, for the, for the position					
13a	and now I am part of a, a program,	outcome				
	it's called, Young Academics or something like that,					

Appendix Q (*continued*)

Original story			Retold story			
Element id	Clause	Element	Element id	Clause	Category	Element
13b	in which you are an academic in, in, in, in the engineering school,	outcome				
	which, in which is the, the department, mining department					
13c	and they, they keep your, your salary each month, the end.					
13d	and while I, I'm, while you're, you're obtaining your, your PhD					
14	But at the end I have to work for twice, twice as much the time that I was here.	outcome				
15	But doesn't matter.	evaluation				
3r2	And I think that was a good decision.	evaluation				
13a	and now I am part of a, a program,	outcome				
	it's called, Young Academics or something like that,					

APPENDIX R

SITE MAP FOR ORAL COMMUNICATION COURSE WEBSITE

- Home Page
  - Course outline and schedule
  - Course Material
    - All assignments
      - Detailed assignment descriptions
      - Assessment grids
      - Example assessment grids
    - Classwork and more
      - Class handouts
      - Activities
      - Follow-up info re. class discussion
    - Model presentations
      - Videos of student presentations (students who took the course previously)
  - Resources
    - Pronunciation
    - Listening
    - Vocabulary
    - Dictionaries
    - Presentation Skills
    - Relevant articles
  - Interview sign-up sheet
  - Presentation schedules