THE UNIVERSITY OF MANITOBA

A STUDY OF LEXICAL CHOICE IN ESL FOREIGNER TALK

BY

SANDI HOWELL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF EDUCATION

DEPARTMENT OF CURRICULUM:

HUMANITIES AND SOCIAL SCIENCES

FACULTY OF EDUCATION

WINNIPEG, MANITOBA

JANUARY 1994

ISBN 0-315-92238-9

Name ___

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

LINGUISTICS

SUBJECT TERM

0290

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom _____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

_____ ☐☐☐ U·M·I

SUJET — CODE DE SUJET

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION
Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Médiévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES
Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-être social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes. 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des pâturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE
Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hôpitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES
Sciences Pures
Chimie
- Généralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÈres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Électronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie
Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et thermodynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systèmes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

A STUDY OF LEXICAL CHOICE IN ESL FOREIGNER TALK

BY

SANDI HOWELL

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF EDUCATION

ABSTRACT

This is an exploratory study which is intended to add to the current body of knowledge regarding the language adjustments teachers make when communicating with Non-Native Speaker(NNS) students. This language adjustment is called teacher talk, a sub-category of a phenomenon known as foreigner talk. It is important to understand language input and its impact on the second language acquisition process. Such an understanding will allow teachers to make more informed pedagogic decisions about the type of language input used with NNS students, focusing on the idea of linguistic simplification and its impact on cognitive simplification.

Specifically, this study is about vocabulary form and commonness or word frequency. Vocabulary form features include grammatical category (e.g., noun, verb), general classification (i.e., function or content word), inflectional morphemes(e.g., endings like "ed"), productive derivational morphemes (e.g., prefixes and suffixes), verb form, pronoun form and contractions. A computer program called PARSAID was developed and used to count the forms in the corpora of ten English as a Second Language (ESL) teachers (Group A) and ten Non-ESL teachers (Group B) of adults, using a two-minute sample corpus from each teacher. The samples were transcribed and each word was classified and coded according to the pre-selected vocabulary features. The corpora of Group A and Group B were compared for significant differences using chi

square. In addition, commonness or word frequency was studied by comparing the corpora with an established word frequency dictionary (Carroll, Davies & Richman, 1971). A dictionary based on this book was entered into PARSAID which then parsed each word in each teacher's corpus according to frequency of use per million words. The results were described with a cumulative frequency table and polygon. In addition, the total mean frequency of use per million and the total mean number of words were compared using a *t* test with each.

This study found that there was a significant difference between Group A and Group B regarding the use of general classification of words, grammatical form, pronoun form, contractions and derivational morphemes. For commonness or word frequency per million, Group B used a slightly higher percentage in the most frequent range of 3,351-73,123 words per million. Group A used a higher percentage in the second highest range. Group B had a significantly higher frequency per million on average than did Group A. Also, Group A used 72 per cent of the total number of words used by Group B, a significant difference.

This study resulted in one profile of teacher talk which focuses on vocabulary form and commonness, and may contribute to further research in the area. Such a profile can assist teachers in their reflections on the simplification of language input and its impact on the cognitive domain, and in making relevant pedagogic decisions.

ACKNOWLEDGEMENTS

I would like to express my appreciation to the members of my thesis committee as follows:

Dr. Richard Kidd, as the committee chairperson and linguistic advisor, for his ongoing feedback on my ideas and writing of the thesis.

Dr. Eric MacPherson, as a committee member and experimental and statistical design advisor, for his experience and creative input regarding this aspect of the thesis. I would also like to thank Dr. MacPherson, Jim MacPherson and Bruce MacPherson for their programming and refinement of PARSAID.

Dr. Bev Zakaluk, as the final committee member, for her comments, suggestions and background knowledge.

A great deal of thanks goes to the 20 teachers who so willingly allowed me to come into their classrooms to tape their speech.

Also, I would like to thank my husband, David, whose patience, understanding and encouragement greatly helped me to complete this study.

Thanks goes to my father, whose clear - headed comments helped me to refine my communication, and whose ongoing support helped to make this possible.

I would also like to thank my friend, Arlene, whose own experience with thesis work helped me enormously in gaining perspective. Along with her never - ending patience, hers was a great contribution.

Finally, thank you to the many other people in my life who continued to support me throughout this endeavour.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Statement of Problem and Its Significance

Statement of Problem

The purpose of this study is to add to the current body of research concerning the language adjustments teachers make when communicating with Non-Native Speaker (NNS) students. This is a phenomenon known as Teacher Talk (TT), which is a sub-category of a speech variant called Foreigner Talk (FT). Studies of FT examine the way Native Speakers (NSs) of a language speak to people whom they perceive as lacking full proficiency in the language.

In a general sense, this study is about the second language acquisition process and, specifically, it focuses on the vocabulary form of the language input aspect of the process. It is an exploratory study intended to determine similarities in the speech of ESL teachers with respect to pre-selected lexical form features, thereby creating a profile of TT with respect to vocabulary. It is important to establish how the language used in the instruction of ESL students differs from that used with NS students. Information about the way language is modified for the purpose of facilitating communication can be used by teachers to reflect upon the notion of simplification and its impact on the cognitive domain in the classroom. Ultimately, this should help

teachers to make more informed pedagogic decisions about second language instruction.

The basic question asked is whether or not there are any significant differences between the following two groups of teachers regarding the frequency of vocabulary forms occurring in selected corpora:

1. Group A. Ten teachers of intermediate or advanced language proficiency adult ESL students.
2. Group B. Ten teachers of adult non-ESL students.

The teachers involved in the study were chosen from a variety of class content areas so as to provide a wide spectrum of vocabulary. They were audiotaped during a time when they were explaining something to their students. The resulting corpora were then transcribed, coded and analyzed.

Although vocabulary can be investigated in terms of semantic features (i.e., word meaning), this study focused exclusively on vocabulary forms (Robinett, 1978). Form includes:

- general classification of words; i.e., function, content.
- grammatical category; e.g., noun, verb, and differences within each category, e.g., pronoun and verb forms.
- inflectional morphemes; e.g., grammatical morphemes like "ed", "s".
- productive derivational morphemes; e.g., prefixes like "anti-" and suffixes like "-able".

- contraction forms; e.g., 'm for am.

A related aspect of form is vocabulary commonness or word frequency, which is a useful feature for teachers to understand especially as it relates to the simplicity of language input (Robinett, 1978). Commonness can be expressed as the number of times the word appears per million calculated according to a lognormal model (Carroll, Davies & Richman, 1971). A word such as "the" occurs 73,123 times per million and is the most common word in English, as listed in *The Word Frequency Book* (Carroll, Davies & Richman, 1971). A less common word, for example, is "credentials" which occurs approximately one time per million. So we would say that "the" is a more common word than "credentials". Impressionistically, commonness may be apparent. This study attempts to quantify the idea of commonness, again to contribute to teachers' discussions about the nature of simplification and pedagogic decisions.

Another aspect of vocabulary related to the phenomenon of TT is the total number of words in a corpus. Using less words in a given time frame is a common way of simplifying language input (Ferguson, 1981; Blau, 1982; Wesche & Ready, 1985; Kliefgan, 1985). However, Chaudron (1983) points out that simplification can take the form of confusingly redundant overelaboration. This study attempts to add information to the profile of TT about the number of words used.

To assist in the calculation of frequencies from the corpora, a computer program called PARSAID was created. PARSAID makes it possible to classify individual words easily according to user designation, and to determine total frequencies in large corpora. It also allows a determination of vocabulary commonness with respect to a user-entered frequency dictionary. For this study, *The Word Frequency Book* (Carroll,Davies & Richman, 1971) was used to create a frequency dictionary for PARSAID.

Group A and Group B were compared for vocabulary forms using chi square, $p < .05$, to detect any significant differences. The two groups were then compared for vocabulary commonness by using a cumulative frequency distribution with respect to mean per cent and grouped frequency intervals. The total number of words used by each group and the cumulative mean frequency of commonness were also compared.

Previous research (Chaudron, 1983; Wesche & Ready, 1985; Early, 1987; Kleifgan, 1985) has indicated that teachers do in fact modify their speech to NNS students. This study focuses on a specific aspect of these adjustments, vocabulary form and commonness, for the purpose of creating a profile of TT.

Significance of the Problem

Modern second language (L2) acquisition theory classifies the input of language to the learners as a distinct phenomenon. Furthermore, Krashen (1982) maintains that an

appropriate and plentiful supply of comprehensible input (i.e., language which the learner understands) is a necessary condition for L2 acquisition. It has long been recognized that the comprehensibility of language can be increased if it is modified in various ways. Ferguson (1971), for example, identified some of the ways we address persons whose first language is not the same as the speaker's and who do not have full native competence. He called the resulting type of speech "FT", claiming that it varies structurally from NS speech with respect to phonological, grammatical, semantic and discoursal elements.

Opinions differ about the reason, conscious or subconscious, for the use of FT as a communication device. Interactionists like Freed (1981) claim that variations are shaped depending on the speaker's perceptions of age, purpose, cognitive ability, relative status, topic, situation, linguistic sufficiency, familiarity with NNSs' mistakes, and degree of personal acquaintanceship. Other theorists like Ferguson (1971) argue that FT reflects the NNSs' mistakes in a kind of imitation and is an attempt to facilitate communication. Another view (Valdman, 1981) holds that FT is a way of maintaining social distance, denying NNSs the opportunity to hear fully-formed samples of the target language and preventing the development of communicative competence. Granting at least partial truth to all of these assumptions, it follows that TT includes a range of

communicative intentions which may be positive or negative for the L2 learner. Since language input from the teacher almost surely has a significant effect upon the L2 acquisition process in the classroom, it is important to understand the nature of FT and TT, ultimately to be able to use it in a teaching situation effectively. A useful model of FT should consist of a framework which is a continuum, encompassing different styles of FT, such as TT, and variations within these styles. This is really the fundamental rationale for the research undertaken here, as it is expected that further study of the impact of TT on the learning process should provide a clearer picture about its optimum use in pedagogic situations.

One element of FT which has been marginally studied is lexical choice. It has been found (Henzl, 1979; Wesche & Ready, 1985; Hatch, 1979; Long, 1983) that NSs tend to choose "simplified" vocabulary when speaking to NNSs. Speech that is lexically simplified:

- contains fewer compound formations, idioms, or slang,
- features higher frequency vocabulary, and
- has greater repetition of known words.

However, Chaudron (1983) claims that lexical simplification may result in important nuances being lost, thus fostering cognitive simplicity. Most studies about vocabulary which have gone beyond casual observation have focused on type/token

analysis of vocabulary simplicity, wherein the final ratio represents the proportion of different words to the total number of words produced. The smaller the ratio, the less diverse or more simplified the language is thought to be. Impressionistic observations (Long, 1983; Chaudron, 1983) conclude that there is a difference in vocabulary choice between teachers of ESL students and those of non-ESL students. However, it appears that few studies have focused on vocabulary in a comprehensive and quantitative manner. Chaudron (1983) claims:

> Short of an accumulated measure of commonness of all words used in a given lesson, it is difficult to determine the simplicity of vocabulary use in that entire classroom.

Following his cue, this study aims to increase our understanding of lexical simplification in TT by examining the commonness of words in a given text in an exploratory but quantitative way. By counting and categorizing every word in selected TT corpora produced by ten ESL teachers and comparing the resulting data with that derived from similar selected corpora of ten Non-ESL teachers according to a number of form classifications, it should be possible to draw a profile of TT with respect to features of vocabulary. This research will thus add to our current understanding of FT and TT, hopefully leading to more informed and reflective decisions about the choices made for language input in pedagogic situations.

Definitions

Commonness refers to how often a word is used in the English language. This can be stated in quantitative terms as the frequency of occurrence per million words (Carroll, Davies & Richman, 1971). The more frequently occurring a word is, the more common it is said to be.

Foreigner Talk (FT) was first termed by Ferguson (1971) who used it to describe a register of simplified speech which is used by speakers of a language to outsiders who are felt to have a very limited command of the language or no knowledge of it at all. It can be said to occur in varying degrees, often in relation to the speaker's perceived proficiency of the listener. Teacher Talk (TT) is a subset of FT. Specifically, it is the way teachers modify their speech to NNS students.

Frequency is the number of times a word appears within a given interval. Within this study, frequencies are expressed in terms of time, e.g., the number of words in two minutes or in terms of quantity, e.g., the number of words per million words.

Second Language Acquisition results in the ability to produce a second language in its spoken and/or written forms and, therefore, requires the acquisition of some or all of the skills of reading, writing, listening and speaking (Robinett, 1978). **Language input** is part of the acquisition process as the learner is receiving an input of the second language in either a deliberate manner, such as in a classroom, or in an

accidental manner as occurs in a natural setting such as interaction with a store clerk.

Simplification is the process by which language input is changed to make it more salient to the listener. FT and TT are profiles of the language which result from simplification. Simplification should not be thought of as only reductionism, but can also include types of expansion such as elaboration.

Vocabulary is one aspect of FT which can be profiled. **Lexical characteristics** can be of two types, including those related to the forms of words and those related to meaning categories (Robinett, 1978). **Vocabulary form** can be defined as the different ways a word can appear. The form of a word can be described by its classification as function or content word, grammatical category, type of grammatical morpheme such as "ed" endings or plural "s" endings, contractions, and prefix or suffix addition. More detailed definitions are contained in Appendix A.

Summary and Research Questions

This study explores the vocabulary form features in the corpora of two groups of teachers, 10 ESL and 10 Non-ESL, for the purpose of creating a profile of TT with respect to vocabulary. PARSAID enables an efficient and quantitative study of user - selected features, resulting in a reliable characterization of TT.

The research questions follow, accompanied by the corresponding null hypothesis and some preliminary information about how these hypothesis will be tested.

Research Question 1. Is there any significant difference between Group A, ESL teachers and Group B, Non-ESL teachers, in selected corpora with respect to a number of vocabulary forms?

> **Null Hypothesis 1.** There is no significant difference between Group A and Group B for selected vocabulary forms.
>
> The test of significance will be chi square, $p < .05$.

Research Question 2. Is there any difference in vocabulary commonness between Group A and Group B?

> **Null Hypothesis 2.** There is no significant difference between Group A and Group B for vocabulary commonness.
>
> The frequencies will be expressed in a cumulative frequency distribution table and polygon.

Research Question 3. Is there any difference between Group A and Group B for the mean frequency of occurrence per million words?

> **Null Hypothesis 3.** There is no difference between Group A and Group B for the mean frequency of occurrence per million words.
>
> The test of significance will be a two-tailed *t* test, df 18, $p < .01$.

Research Question 4. Is there any difference between Group A and Group B for the mean total number of words?

> **Null Hypothesis 4.** There is no significant difference between Group A and Group B for the mean total number of words.
>
> The test of significance will be a two-tailed *t* test , df 18, $p < .01$.

Chapter Two of this study contains a discussion of relevant literature. Chapter Three outlines the methods, the instrument and the statistical procedures. Chapter Four contains a summary of the data as related to the research questions asked and the null hypotheses posed. Chapter Five contains an examination of these results and discussion of their relationship to previous studies as outlined in Chapter 2. This chapter also has a discussion of pedagogic implications and suggestions for further research. The Appendices contain raw data, expanded grammatical definitions, a sample corpus, and a sample coded corpus.

Chapter Two

Literature Review

Foreigner Talk

Many researchers (Brulhart, 1986; Long, 1982; Chaudron, 1988), theorists (Krashen, 1982) and methodologists (Krashen & Terrell, 1983) hold the view that language input is an independent element factoring in second language acquisition, which is primarily a subconscious process resulting in an intuitive knowledge of the language. Krashen (1982) argues that a large part of the teacher's role is the provision of comprehensible input (i.e., language which is understood) to learners at the beginner level. Satisfactory provision of such input, he claims, will result in an implicit knowledge of the rules of grammar. Subsequently, learners will gain a large part of their interactional and socio-linguistic competence from acquisition experiences which take place in a more natural environment such as social situations. This theory focuses interest on input as an important element of the language acquisition process. While formal, explicit study of a language also has a place in L2 curriculum, one cannot doubt the importance of comprehensible input in promoting acquisition.

One variety of input which has been studied is FT. Ferguson (1971) defines FT as a register of simplified speech used by speakers of a language to outsiders who are felt to

have a limited command of the language or no knowledge of it at all. FT exists on a continuum from the way in which NSs interact with each other to the modified input of the second language teacher in the classroom. Speech modification of this sort is not restricted to NS-NNS discourse. Even NSs adjust their speech deliberately to other NSs in order to clarify messages. Ferguson feels that NSs will adjust their speech according to what they feel will provide the best comprehensible input to NNSs. Long (1982) claims that FT is a natural phenomenon and that every NS will readily use it to lighten the interactional burden for a NNS no matter what the environment. However, some (Carty, n.d.) doubt the existence of FT in a natural environment and its beneficial use in the classroom. Chaudron (1983) even suggests that FT may be detrimental to students in the areas of language and general knowledge attainment (see below).

FT Vocabulary Studies

One feature of FT which has been marginally studied is vocabulary. Hatch (1978) stresses the importance of vocabulary to L2 learners who tend to regard lexical acquisition as a key to improving communication skills. She says that second language learners constantly stress the importance of development in this area for further understanding and that it is vocabulary which forms the framework for grammatical constructs. She quotes Krashen as

noting that learners can often be observed carrying dictionaries but rarely grammar books.

Chaudron (1983) classified FT in this respect as being either ambiguous oversimplification or confusingly redundant overelaboration. He claims that linguistic simplification can lead to simplification in the cognitive sense. For example, the use of simpler paraphrases for words such as "ironic" may lead to the exclusion of important nuances, and the omission of these can be detrimental to the learner. On the other hand, overelaboration can lead to confusion (Chaudron, 1982), as several of the structures used for elaboration (e.g., opposition and conjunction) can lead to ambiguity as to whether new meanings are being added or whether a particular word is being elaborated. Chaudron (1983) claims that informed decisions about the degree of simplification or elaboration of vocabulary should improve teaching methods.

Miesel (1976), in a study of German ESL learners, explores the notion of simplification with respect to FT and the interlanguage of second language learners. He examines the linguistic theory that imitation as a method of FT is a circular process implying a contemptuous mimicking of a foreigner's desperate attempts at the language followed by the foreigner's imitation of the FT. He rejects the theory of imitation on the basis that:

1. there are common characteristics of simplified register in learners across languages;

2. FT is actually used in a limited number of cases; and
3. foreigners tend to associate with each other rather than NSs, relying on their first language for a lot of communication so their opportunity to engage in this behaviour is limited.

Miesel concludes that FT is a psychological process and that simplification is a complex strategy for communication. With regards to vocabulary, it includes such things as:

1. less marked lexicon, e.g., to do vs. does;
2. analytic phrases replacing what the speaker intuitively judges to be "complex" expressions;
3. missing articles, prepositions, verbs and personal pronouns;
4. deletion of inflectional endings on verbs, adjectives and nouns; and
5. insertion of subject pronouns in the imperative.

A teacher will engage in FT to adapt communicative needs to the learner's linguistic competence. In other words, FT is a way of optimizing communication through reduction. This Meisel calls restrictive simplification, claiming that it is a feature of the early stages of language learning.

With regards to the idea of simplification, Ferguson (1971) says that many languages have particular features of pronunciation, grammar and lexicon which are characteristically used in NS - NNS interactions. He feels that FT is an attempt by NSs to simplify the language by imitating NNSs' speech features. There is a general tendency

to omit the copula, prepositions, articles and inflectional endings. These omissions are viewed by NSs as being a simpler version of the language even though a linguistic analysis may show that, for example, the absence of the copula is regarded as a deletion and, hence, grammatically more complex. Also, compensation for the absence of the copula may include complicated patterns of allomorphy(alternate phonetic form) and synonym use. In a later publication, Ferguson (1981) says that there is general agreement that language is considered to be simpler if there is a much smaller total lexicon, strikingly fewer grammatical categories as expressed by inflectional and derivational morphology, or by much less allomorphy. However, he claims, it is unclear whether, or to what extent, these types of simplicities are compensated for by complexities elsewhere in the language.

Further to the discussion of simplification, Larsen-Freeman (1983) states that strategies for communication do not always result in linguistic simplification. She claims that these strategies result in sentences with a deviant word order. An example would be, "Friday, Saturday, did you have a nice weekend?", which is a left dislocation of the topic. Larsen-Freeman claims that it is not necessarily simplification, but comprehensibility that is negotiated between speaker and listener.

Finally, Terrell (1982) points out that although the input of speech must be understood, lexical simplification is

not always a necessary feature for this to occur. Contextual supports, for example, may overcome the complexity of vocabulary.

Hatch, Shapira and Gough (1978) studied simplification strategies within the discourse of two groups of speakers. The first group consisted of NSs who were very familiar with NNSs, such as ESL teachers. There was a tendency for these NSs to imitate the NNSs. Their FT included "it" deletion, lack of verb tense, "do" deletion in negatives and questions, copula deletion, non-inflected plurals, no possessives and no possessive pronouns. The occurrence of the above varied according to the individual speaker and was not a consistent phenomenon. The second group contained speakers who were not accustomed to NNSs, such as persons giving information over the phone to the public. Their strategies were different. They tended to use less grammatical reduction, relying more on discourse tactics such as repetition, restatements, elaboration with synonyms, definition and finishing incomplete sentences for the NNS. Again, there was great individual variation, and the researchers concluded that future studies could focus on who chose what type of FT and for what reasons.

Blau (1982) studied the effects of modified syntax, speech rate and pausing on the comprehension of second language learners. She systematically varied taped passages in terms of sentence complexity and speed in Study 1, and pause times in Study 2. In Study 1, reduced velocity yielded

slightly higher comprehension scores with one of two groups who were at the lowest level of language proficiency. Variations in sentence structure yielded no significant results. In Study 2, the version with increased pause times yielded significantly higher comprehension scores. Blau concludes that pause times at constituent boundaries significantly enhance the comprehension of aural input, and that too slow an input rate potentially can impair comprehension by prolonging the time a pattern must be held in short-term memory and allowing more time for memory traces to fade.

Henzl (1973) studied Czech speakers relaying stories to NSs and NNSs. She found that as a group, NSs talking to NNSs tended to make certain types of lexical choices that were different from the choices used when talking to NSs. One major difference was the use of basic vocabulary in NS-NNS discourse versus extremely rich diversity in NS-NS discourse including elaboration and the use of synonyms. The following are other features of NSs talking to NNSs:

1. Little elaboration of key words with a tendency to utilize these words maximally.
2. An effort to be concrete by avoiding indefinite pronouns and adjectives and using more personal pronouns or concrete names.
3. Omitting or replacing with standard equivalents the following features:

    a) colloquial expressions
    b) defamatory expressions

c) compound words
d) invented names
e) idiomatic expressions

A type-token analysis of verb tense indicates less use of past tense forms with NNSs. Henzl's conclusion was that NSs made linguistic choices according to their judgement of what is easy, simple and clear. Speech is somehow systematically patterned to match the NNS's competence.

Early (1987) studied FT by investigating discourse interaction features and linguistic input features for eight ESL teachers with NNS students compared to eight regular classroom teachers with NS students using chi square and Z scores as tests of significance. All taught social studies to adolescents. She also compared her findings with Long's (1980) findings regarding input and interaction outside of classrooms. Early found that for interaction features there were differences between the two types of teachers. In comparison to the TT of teachers of NS students, the speech of ESL teachers tended to contain:

1. Different frequencies of questions, statements and imperatives.
2. More comprehension checks, self-repetitions, other repetitions and expansions.
3. Shorter average length of t-units in words.
4. Lower average number of tensed verbs per t-unit.

Early concludes that there were many adjustments to the interactional structure of classroom conversations by ESL teachers with NNS students and that clearly TT exists in these classrooms. ESL teachers also tended to use shorter, syntactically less complex utterances, but there was no ungrammaticality. Type-token lexical ratios indicate that there were no significant differences between teachers and, therefore, the lexicon was not restricted. Early's comparison with Long's 1980 study indicates agreement regarding type-token ratio results. Early concludes that there are clear and strongly marked differences between teachers' linguistic input to NNS and NS students. Teachers adjust their speech downwards and modify their interactions when talking to students with limited proficiency in English. While TT and FT are similar in some ways, they also differ in important ways.

Wesche and Ready (1985) studied the language of two psychology professors. One was a NS of English lecturing in English to both NS and NNS classes. The other was a NS of French lecturing in French to both NS and NNS classes with parallel subject matter to the English professor. The NNS students were considered to be of high intermediate proficiency. The language of the two professors was compared with respect to the way they spoke to the NNS students versus the NS students.

Part of the study focused on analyzing vocabulary features, including type-token ratio, grammatical word form (e.g., noun), content vs. function words and verb forms, using chi square as a test of significance. The only significant differences for the English professor with regard to verb forms were the use of more tensed non-auxiliary verbs with the NNS students than the NS students and the use of more non-present tensed copula with the NS students. A type-token ratio was calculated on vocabulary and no significant difference was found. A type-token ratio calculated for verb forms also resulted in no significant difference. There were no significant differences found for grammatical form (e.g., noun, verb, adjective). The English professor used significantly more words when speaking to NNS students. There was also a significant difference in the total number of words used by both professors, as both used more words when speaking to NS students. The use of more words was attributed to extra illustrative examples being given.

However, Wesche and Ready found that a more detailed analysis of specific grammatical forms showed an even greater difference. The English professor used more "will" and "so" while avoiding conditionals with the NNS group. At the same time, he used the word "would" more with NS students. He also made greater use of the infinitive forms, non-present "be", and a variety of tenses with NSs, in contrast to the almost exclusive use of present tense with NNSs. Wesche and Ready

concluded that the most important predictor of the speech characteristics of a NS addressing NNSs are the individual characteristics of the NS's speech rather than a given set of FT discourse characteristics. However, notwithstanding this conclusion, both speakers in the study showed systematic deviations from speech directed to NSs when addressing NNSs in parallel situations. Wesche and Ready attributed the fact that no quantitative differences were found for overall vocabulary characteristics to the high proficiency level of the students as well as the professors' awareness that the NNS students were familiar with the content of the lectures through readings and through language instruction sessions.

Kliefgan (1982) focused on a kindergarten teacher's speech with a group of students whose English competence ranged from zero proficiency to full native competency. The language was characterized by deletions of determiners, copulas, auxiliaries, pronouns and verbs, which diminished with children of greater linguistic proficiency. Type-token results indicated that the greatest number of words were directed at the least proficient child with less lexical variety (total corpus 933 words). A subsequent study (1983) revealed that this teacher increased lexical variety as the students' language ability increased. However, the length of speech directed at NNSs was generally shorter than that intended for NSs.

Long and Sato (1983) conducted an exploratory study of ESL teachers' classroom questions, testing for the following:

1. question type
2. frequency of questions, statements and imperatives
3. present versus non-present verb forms
4. rank order of grammatical morphemes for ESL teachers as compared to rank order for NS-NNS outside classrooms

The comparison group was NSs and NNSs in informal conversations outside the classroom and the test of significance was chi square. Spearman rank order correlation co-efficient was also calculated. The ESL teachers were found to use significantly more present tense verb forms, ask more display and fewer referential questions, and use fewer questions and more statements and imperatives in T-units. The rank order of grammatical morpheme use was significantly positively correlated with the frequency order for these items in the comparison group. From these results, Long and Sato conclude that ESL teachers continue to emphasize form over meaning and accuracy over communication in spite of the trend in language teaching methodology towards focusing on communication.

Derwing (1991) performed a study which compared the speech of eight ESL instructors and that of eight persons who had no regular contact with NNSs, with respect to word frequency. The subjects were asked to explain something to

NNSs considered to be at a high beginner language proficiency level. The two groups did not differ for word frequency or rate. However, when the subjects were grouped according to the personality traits of interpersonal affect and social participation, there were significant differences in word frequency and speech rate. Derwing relied on the Thorndike and Lorge (1944) word count for frequency counts which were classified as follows:

1. The 500 most frequent words.
2. The 1000 most frequent words.
3. Words occurring 100 or more times per million.
4. Words occurring between 50 and 99 times per million.
5. Words occurring 26-49 times per million.
6. Words occurring 0-25 times per million.

Derwing found that narrators used more words in frequent categories and fewer in less frequent categories when speaking to NNSs. Those subjects in the high interpersonal affect/social group (as determined by the Jackson personality inventory) used significantly more of the top 500 words with NNSs and significantly fewer of the least frequent. In the low interpersonal/social participation group, there were no significant adjustments of word frequency.

There were no group (experienced vs. inexperienced narrator) differences in overall rate. However, there was a reduction in rate when addressing NNSs by both groups. In the high versus low interpersonal affect/social participation

groups, the high group did not slow down while the low group did in terms of increasing pause times. Derwing concluded that consideration of how to talk to low proficiency NNSs is excluded from TESL education programs and that there is a need for an emphasis on communication skills in teacher training.

There have been a number of other impressionistic observations about vocabulary in FT. Long's (1983) overall observations characterize FT as consisting of more restricted vocabulary with greater repetition and more concrete versus abstract word forms. Chaudron (1983) notes that vocabulary used with lower level learners looks less complex lexically. He found that vocabulary can be simplified by employing more frequently used words or simple circumlocutions (e.g., "clinging" changed to "hold on very tightly"). He also notes that more difficult vocabulary was clarified, elaborated upon or made simpler by redundancy. Chaudron (1983) argues that short of an accumulated measure of the commonness of all words in a given lesson, it is difficult to determine simplicity. In order to investigate this claim with sufficient rigor, we need to clarify the concept of commonness.

Vocabulary Commonness

Several word frequency books have attempted to establish the commonness of vocabulary (West, 1967; Kucera & Francis, 1967; Jones & Wepman, 1966; Dahl, 1979; Carroll, Davies & Richman, 1971). West's list is based on written materials and

reflects the frequency of occurrence of the various meanings and uses of words in a study of five million words. He claims the chief value of the list is in cutting out that which is not essential to the learning of "major words" or "heavy words" and that anything which seems in the least degree unusual or doubtful should certainly be excluded from the teaching plan. Besides frequency, factors to be considered for the inclusion of vocabulary in the teaching plan are ease of learning, necessity, cover (word nearest the root sense of an idea versus tricky idioms, for example), stylistic level and intensive or emotional words. West maintains that educators should stick to words that express ideas versus emotions in the classroom.

Jones and Wepman's list represents the frequency of adult spoken vocabulary based on the telling of a story from twenty pictures of the Thematic Apperception Test. Their purpose for compiling the list was to provide a baseline for the linguistic analysis of aphasic patients. Data were analyzed according to frequency of use, grammatical class (e.g., noun), inflections and contractions.

Kucera and Francis's list is based on the written Corpus of Present-Day Edited American English. This consisted of 1,014,232 words divided into fifteen categories such as:

1. Press: Reportage
2. Fiction: Science

There are 500 samples of 2,000 words each. Their analysis is of word frequency, word length, sentence length and a lognormal model of frequency distribution (Carroll, Davies & Richman, 1971).

Dahl's list is generated from psychoanalytical cases including patients and psychoanalysts. The total word count was 1,058,888 words with 17,781 different words and the analysis was of frequency, number of speakers to use a word and the rate per million based on Carroll's lognormal model.

Carroll, Davies and Richman's list, which was used in the present study, was taken from the textual samples of published materials for Grade 3 to 9 students. They used the American Heritage Intermediate Corpus with over five million words in 500 word samples and 86,741 different types. The analysis is based on a statistical model which postulates that the total vocabulary underlying a corpus is distributed according to the familiar normal distribution as logarithms of frequency are used. Statistics presented include word frequencies, dispersion of frequencies over 17 subject categories and frequency per million adjusted for dispersion.

Nature of Vocabulary Features

Vocabulary can be said to have characteristics related to meaning, usage or form. Robinett (1978) distinguishes the semantics or meaning in language as being of three types: lexical, experiential and grammatical. She states that words

are the labels of a categorization process. Words relate to each other and fall into classifications. A study of vocabulary from this perspective includes collocation, synonyms, antonyms, idioms, proverbs and cliches. The experiential component adds meaning to a word based on personal or cultural influences and colours vocabulary subjectively. Grammatical relationships also add meaning. For example, a noun in either the subject or object position takes on slightly different meaning as the agent or receiver of an action.

The characteristics of usage include variables which influence an individual's word choice such as age, dialect, sex, social position, purpose, discourse manner and formality. These characteristics vary according to personality and situation.

Finally, the lexicon can be understood in terms of the characteristics of form. These include morphological variants (word changes according to grammatical shift, e.g., "give-gave"), derivational forms (prefixes and suffixes), functional shift (same word but different grammatical form, e.g., "record" as noun and "record" as verb). Robinett also includes frequency as a form characteristic. It is the commonness of vocabulary according to form characteristics to which this study addresses itself.

Conclusions

Krashen and Terrell's (1983) theory states that comprehensible input in a pedagogic situation is an important element in the second language acquisition process. One means of creating comprehensible input is by using FT, defined by Ferguson (1971) as a register of simplified speech which exists on a continuum from the way NSs interact with each other to the deliberately modified speech of second language teachers in the classroom. However, at this point in the discussion, there is conflict over whether or not FT is beneficial for second language learners. Long (1982) claims that FT is a natural phenomenon intended by the NS to lighten the interactional burden for NNSs in any environment. Others (Carty, n.d.; Chaudron, 1983) say that FT may be detrimental in terms of language and knowledge attainment. Terrell (1982) states that although input must be understood, lexical simplification is not a necessary feature. These perspectives raise valid questions related to pedagogy and the definition and use of FT in the classroom. Research into the nature of FT, such as that encompassed in this study, would seem to be essential to making informed decisions about the way teachers communicate with students.

In addition to the issue of the wisdom of using FT as a pedagogic aid, there exist valid questions about the definition of simplification. The general definition of FT as a register of simplified speech is insufficient in

characterizing simplification with any degree of exactness. Meisel (1976) explores the issue of restrictive simplification, which is a way of optimizing communication through reduction. Conversely, Chaudron (1983) points out that attempts to simplify can end up as confusingly redundant overelaboration. Therefore, simplification also seems to exist on a continuum from reduction to expansion. Intuitively we seem to be able to identify and even produce "simple" language. In terms of systematic identification, there have been few studies to date which deal with this topic in a systematic and rigorous manner. It would seem critical that teachers be aware of what simplification is and how it affects students. This study, by creating a profile of vocabulary choice, will add to knowledge about the continuum.

With respect to FT itself, there are few studies which provide a systematic and in-depth profile of vocabulary choice. Impressionistic observations (Long, 1983; Chaudron, 1983), included with other studies, indicate that simpler words tend to be chosen. However, the nature of "simpler" is never really made completely clear in these studies. Derwing (1991) attempted to profile the vocabulary choice of two groups of NSs speaking to NNSs in terms of word frequency or occurrence per million as determined by the Thorndike and Lorge (1944) word count. The assumption is that the more frequently words appear, the more salient they are to the listener. This seems to be one way of systematically defining

word simplicity and is also a way to quantify. This study also attempts to profile vocabulary from this perspective.

Frequency as discussed above is only one aspect of describing vocabulary. Robinett (1978) outlines the various ways that we can analyze language, including the examination of discoursal, grammatical and lexical perspectives. To adequately profile the linguistic nature of FT, there needs to be systematic definition of all aspects. Further to this, there must be consideration of usage factors which can affect communicative situations and which may influence the nature of FT, for example, personality type, learner's language proficiency, and formality to name just a few. In other words, the phenomenon of FT needs to be examined from a holistic perspective.

Both Chaudron (1983) and Derwing (1991) clearly point out that there should be more informed decisions on the part of teachers in terms of the way they communicate with students. Derwing feels that this should be a part of teacher education. To achieve this, it will be necessary to create a more complete profile of FT and its effects, as it would be difficult to accurately describe the phenomenon based on the information collected to date. This is a large area to explore, and related research so far has been inadequate. Much more work needs to be done in this area, and this study is intended to add to the profile by more systematically characterizing vocabulary use by teachers.

Chapter Three

Research Design

Questions

This is an exploratory and quantitative study of vocabulary present in the corpora of both ESL and Non-ESL teachers of adults, intended to add to the profile of FT. The research to date provides little reliable information on vocabulary as an aspect of FT. The present study follows a direction suggested by Chaudron (1983), who claims that short of an accumulated measure of the commonness of all words in a given text, it is difficult to determine the simplicity of vocabulary.

This study is intended to determine simplicity by first addressing the following question:

> Research Question 1. Is there any significant difference between Group A, ESL teachers and Group B, Non-ESL teachers, in selected corpora with respect to a number of vocabulary forms?

The form features measured are as follows. These features are explained briefly in the Procedures section below. Detailed information about the following categories and sub-categories is included in Appendix A.

A. General Classification of Words

1. Function Word
2. Content Word
3. Interjection

B. Grammatical Form

1. Noun
2. Verb
3. Pronoun
4. Adjective
5. Adverb
6. Preposition
7. Co-ordinating Conjunction
8. Subordinating Conjunction
9. Determiner

C. Pronoun Form

1. Personal
2. Reflexive
3. Possessive
4. Reciprocal
5. Interrogative
6. Demonstrative
7. Universal
8. Partitive
9. Relative

D. Verb Form

1. Auxiliary in Finite Form
2. Main Verb in Finite Form
3. Auxiliary in Non-finite Form
4. Main Verb in Non-finite Form
5. Infinitive with "To"
6. Bare Infinitive
7. Copula "Be"
8. Imperative

E. Inflectional Morphemes

1. Plural
2. Possessive
3. 3rd Person Singular
4. Past Tense
5. Past Participle
6. "ing" Form
7. Comparative
8. Superlative

F. Contractions

1. is - 's
2. not - n't
3. will - 'll
4. am - 'm
5. has - 's
6. would - 'd
7. had - 'd
8. have - 've
9. are - 're

G. Contractions Cont'd

1. us - 's

H. Productive Derivational Morphemes

1. Prefix
2. Suffix
3. Prefix and Suffix

Also, an exploratory measure of vocabulary commonness will be taken. The second research question is as follows:

Research Question 2. Is there is any difference in vocabulary commonness between Group A and Group B?

The following questions will also be answered:

Research Question 3. Is there any difference between Group A and Group B for the mean frequency of occurrence per million words?

Research Question 4. Is there any difference between Group A and Group B for the mean total number of words?

Procedure

Subjects.

The Oxford dictionary defines "teacher" as one who explains, shows or states by way of instruction. A "student" is defined as a person studying in order to qualify himself/herself for some occupation or devoting himself/herself to some branch of learning or investigation. The teachers selected for this study were chosen with these definitions in mind. All subjects had to be engaged in instructing students who were devoted to learning for a purpose, such as an occupation or a branch of learning. These definitions would exclude presenters and their general audience but include teachers giving a workshop to their peers, three of whom participated in this study.

Group A consisted of ten teachers of adult ESL students. These teachers classified their students as having either intermediate or advanced English proficiency. The teachers were employed at a variety of locations including specific purpose programs (English in the Workplace) and generic

programs. The content of the classes varied from grammatical to cultural emphasis.

Group B consisted of ten teachers of adult Non-ESL students or students identified as being of NS language proficiency. These teachers were also employed in a variety of situations, including public school system upgrading, technical colleges and workplace programs. Again, the content of the classes was varied. The topics ranged from English to Graphic Arts and Business Law.

All teachers were asked if their students fit the above language proficiency descriptors. Subjects were included in the research based on the teachers' identification and the definitions of teacher and student.

Taping Procedures.

The classes met at a variety of times and locations. The teachers were asked to pick a time when they would primarily be explaining something to the students. In other words, the teachers would be the ones speaking.

The researcher went to one class of each subject and audiotaped approximately 20 minutes of the class. In a few instances (ESL 1, NESL 3), lengthier taping was required because of unforeseen interruptions. The intent was to tape 20 minutes and then to transcribe two minutes, which were minutes seven and eight. The first five minutes were intended to allow the teachers to relax. The tape recorder was placed

as unobtrusively as possible and the researcher remained in the class during the taping. The teachers were asked to sign a release form and were told that the tapes were confidential and would be erased after the research project was completed. The teachers were told that the research was simply counting certain vocabulary features such as nouns or verbs. Each sample was given a number to protect confidentiality.

Transcription and Coding.

Each audiotaped sample was timed and two minutes were transcribed. The transcribed samples were coded by the researcher using PARSAID, which is described in detail at the end of this chapter.

Each word in a corpus was coded according to the categories listed in the question section under Question 1. What follows is a brief definition of each. An expanded and detailed version of coding categories appears in Appendix A.

A. *General Classification of Words*

1. Function words are those with meaning which is related to their grammatical function.
2. Content words are those with lexical meaning correlated with experience in the outside world (Robinett, 1978).
3. Interjections are words which are used primarily as a pausing device and are not intended to convey any grammatical or referential meaning.

B. *Grammatical Form*

1. A noun is a word which names a person, place or thing. E.g., *dog, Elizabeth, chair.*
2. A verb is a word or series of words that describe an action or state of being. E.g., run, should go, to allow, used to..... A verb form can function in another grammatical form such as a noun. E.g., *Swimming* is my favourite sport.
3. Pronouns are words which substitute for a noun. E.g., *she, someone, those.*
4. An adjective is a word which describes or limits a noun or pronoun. E.g., *pretty* girl.
5. An adverb modifies the meaning of a verb, adverb or adjective. E.g., ran *quickly, very* pretty girl.
6. A preposition shows a relationship between a noun and another word in the sentence. E.g., *in* the car.
7. A co-ordinating conjunction is a word which links two units. E.g., I ate at six *and* he ate at five.
8. A sub-ordinating conjunction is a word which links two units, one of which is sub-ordinate in meaning to the other. E.g., I ate *because* I was hungry again at seven.
9. Determiners come at the beginning of a noun phrase but are not adjectives. They add further definition to the noun. E.g., *the, some, that.*

C. *Pronoun Form (Quirk & Greenbaum, 1973)*

1. Personal pronouns function as replacements for noun phrases. E.g., *she.*
2. Reflexive pronouns replace co-referential noun phrases, usually within the same finite verb clause. E.g., *himself.*

3. Possessive pronouns are pronouns which indicate possession. E.g., *her* dog, *hers*.
4. Reciprocal pronouns replace nouns expressing a relationship. E.g., They like *each other*.
5. Interrogative pronouns are obviously used in questions. E.g., *Which* do you....
6. Demonstrative pronouns show near or distant reference. E.g., I like *these*.
7. Universal pronouns are the *every* compounds and are collective. E.g., *everything*.
8. Partitive pronouns are parallel to universals. E.g., *something, anything*.
9. Relative pronouns are connected with relative clauses. E.g., The girl, *who* is on the left.....

D. *Verb Form*

A finite verb form shows tense, aspect, mood and voice, and is tied to the subject. E.g., She *ran* home.

A non-finite verb form does not show tense or mood but can still show aspect and voice. E.g., The *contented* dog slept. *Having been offended* before, he was sensitive.

1. Auxiliary in the finite form includes be, have, do, modals and periphrastic modals. E.g., I *should* go soon.
2. Main verb in the finite form is a verb in a main clause linked to a subject and showing tense, mood etc. E.g., She *ran* home quickly.
3. Auxiliary in the non-finite would be an auxiliary in a non-finite phrase. E.g., *Having been* offended before, he was sensitive.
4. Main verb in a non-finite form includes infinitives, "ing" and "ed" participles. E.g., *Swimming* is my favourite sport.

5. Infinitive with "to". E.g., This will allow us *to go* at five.
6. Bare infinite. E.g., He made me *go*.
7. When "be" is used to describe a state of being it is called the Copula and is not an auxiliary. E.g., She *is* pretty.
8. The imperative is a command and usually the subject is the implied you. E.g., *Go* to the corner.

E. *Inflectional Morphemes (Jones & Wepman, 1966)*

These are morphemes attached to a word to function as grammatical markers.

1. Plural of nouns E.g., boy<u>s</u>
2. Possessive of nouns E.g., Girl<u>'s</u>
3. 3rd Person Singular of verbs E.g., She eat<u>s</u>
4. Past Tense of verbs E.g., walk<u>ed</u>
5. Past Participle of verbs E.g., Has walk<u>ed</u>
6. "ing" Form of verbs E.g., swimm<u>ing</u>
7. Comparative of adjectives E.g., smart<u>er</u>
8. Superlative of adjectives E.g., smart<u>est</u>

F. and G. *Contractions (Jones & Wepman, 1966)*

Contractions are a shortened form of a word which are more characteristic of spoken than written language. As there were 10 contractions and each column in PARSAID allows a maximum of nine sub-categories, an additional column, G. Contractions, was created. However, for the purpose of testing the hypothesis and answering the research question, Columns "F" and "G" were combined.

1. is = 's
2. not = n't

3. will = 'll
4. am = 'm
5. has = 's
6. would = 'd
7. had = 'd
8. have = 've
9. are = 're

1. us = 's

H. *Productive Derivational Morphemes (Fromkin & Rodman, 1983)*

These are morphemes which can be freely added to a word and can result in a change of grammatical form.

1. Prefix E.g., "un" unheard of.
2. Suffix E.g., "ly" quickly.
3. Prefix and Suffix E.g., "pre" and "ion" predetermination.

The coding was done as consistently as possible. After each group was coded, the samples were reviewed once for consistency and accuracy. It is presumed that the coding is reliable within the one to two per cent margin of error associated with this type of study. A random sample of 30 words showed this to be true.

An example of the coding would be as follows:

"I will go home quickly."

I[231] will[1201] go[2202] home[21] quickly[25000002].

Using PARSAID, the sub-categories were totalled for each subject. Then each sample was parsed for word commonness as

compared to the *Word Frequency Book* (1971). The following section on the Instrument contains details of this procedure. Names, technical words or jargon such as "Medicare", interjections such as "um" and slang such as "worrywart" were excluded from this measure, i.e., the words were not counted.

The Instrument

PARSAID was created in response to the challenge of examining every word in a corpus. The researcher outlined the potential uses of a computer program that could measure language and Dr. Eric MacPherson and Jim MacPherson created PARSAID. It has the ability to do several useful things as described in the following four sections.

Dictionary.

A dictionary of a maximum of 28,886 words and each associated frequency can be entered into the program. For this study, 5,336 words were entered with the frequency per million (U score) from The *Word Frequency Book* (Carroll, Davies & Richman, 1971). To begin, the 4500 most frequent words were entered directly from the dictionary. Each corpus was word searched and unknown words were indicated by PARSAID. These words and their frequencies were then entered. It is possible to remove or amend words if required.

Text.

Text can be entered into PARSAID basically like any other word processing package. The text can be edited, saved, printed, recalled and cleared. One thing that the program does not do is automatically push the text forward if there are additions. Therefore, it became necessary to leave some space at the end of each line for additions or errors. For purposes of later reference and discussion, each line was numbered.

The text can be coded in three different ways as it is being entered. By using the brackets [], {}, and/or <>, each word or series of words, including sentences, can be coded. There is the potential for 10 different bracket designations for each type of bracket, which are referred to in this study as the columns. There can be nine sub-categories in each of the 10 columns. Therefore, a full utilization of one type of bracket would result in 90 categorizations. Using all three types of brackets, this would mean a potential 270 categories. The columns and sub-categories are user-designated, allowing for studies of any nature.

This study used one type of bracket, [], with eight columns out of 10 being utilized. The sub-categories of each are listed in the Question Section above. A total of 50 sub-categories were designated. To enter text and coding, the researcher simply types a word followed by the first bracket, a number placed in the appropriate sub-category position and

the second or ending bracket. It is essential to the frequency totals that PARSAID reads the brackets, so both are necessary for each word. An example, as cited above, would be:

"I will go home quickly."

I[231] will[1201] go[2202] home[21] quickly[25000002].

The word "I" was coded as being a content word, pronoun and personal pronoun. Notice that "will" has a zero within its bracket. This indicates that column number three, Pronoun Form, does not relate to this word. If there is to be a column entered after the one which does not relate, then a zero must be entered for the unrelated column. If, for example, there are no columns entered after the third column as in the "I" above, the computer automatically reads the remaining columns after the last number keyed as being zero

For convenience, the most frequently used columns were labelled as the first in the series. The first two columns in this study were ones that apply to every word. The order of the rest reflect the frequency of use for coding purposes.

After the coding is complete, the text can be named and stored, and/or printed in a traditional manner. Editing can be done at any time. Appendix B contains a sample corpus and Appendix C is a sample of the same corpus but coded.

Parse.

Before the parsing begins, the user must request that the text be checked for brackets. PARSAID will indicate if there are any incomplete or backward brackets, allowing the user to edit before parsing. This is an important step as incorrect brackets will result in PARSAID being unable to total columns. The text can be parsed in a number of different ways. PARSAID will total each sub-category in each column for a selected text. By indicating the user-designated number of each sub-category in the graphic of the brackets while in Parse, and then directing the computer to count, the total of this sub-category, an associated percentage of the corpus and the total number of words in the corpus will be calculated. Also, PARSAID will count an intersection of sub-categories as it is possible to enter several sub-categories in different columns at the same time for a count. For example, it is possible to determine how many main verbs in the non-finite form were also in the "ing" form. This study did not require this feature but PARSAID was created with potential future use in mind. Also, this study only required the use of one type of bracket.

Certain words, as indicated in the Transcription Coding Section above, were not used in the commonness count because of their specialness or uniqueness. When these words came up in the search, PARSAID allowed the user to enter "ignore", and the program just passed over them. The number of words ignored in a text can be determined by subtracting the number

of parsed words from the total number of words in a corpus. At this point, if there are unknown words in the text the program will indicate the number unknown. In this study, there should not be any unknown, as all words are searched and a frequency or "ignore" is entered for each.

PARSAID will also calculate commonness with respect to the entered dictionary. The frequency function includes a calculation of mean frequency, frequency range measure and total number of words used. The frequency range is user designated. An example of frequency ranges would be:

1. Words occurring 74,000-2001 times per million.
2. 2000-1501 words per million.
3. 1500-501 words per million.
4. 500-101 words per million.
5. 100-51 words per million.
6. 50-11 words per million.
7. 10-0 words per million.

With the ranges entered, PARSAID will then show the numbers and percentages for each range. For example, in the range 51-100 words per million, a total of 50 words representing 20 per cent of the corpora may be calculated. In this study, the number of words parsed for commonness was lower than the total used in the column counts as there were a number of words ignored. The number ignored per subject was relatively consistent for all subjects.

Equivalences.

PARSAID also has an equivalences function. For a given text, a user designated list of equivalent words can be entered and the text can be parsed as above. For example, the user could designate all forms of one verb to be the same (give=gave=giving=given). This count was not included in this study.

Statistical Procedures

The statistical procedures used in this study will be discussed in terms of the hypotheses they were designed to test. The four null hypotheses which correspond to the four research questions in the Question Section above, are as follows:

Null Hypothesis 1. There is no significant difference between Group A and Group B for selected vocabulary forms.

Null Hypothesis 2. There is no significant difference between Group A and Group B for vocabulary commonness.

Null Hypothesis 3. There is no significant difference between Group A and Group B for the mean frequency of occurrence per million words.

Null Hypothesis 4. There is no significant difference between Group A and Group B for the mean total number of words.

The data was collected and put into Table 1, (for Group A) and Table 2, (for Group B) in Appendix D. For each subject, the tables show the raw frequencies for each sub-category, the total number of words in each subject's corpus, the total number of parsed words and the average frequency of occurrence per million words for each subject. This is all data which PARSAID has totalled. In addition, each group's total frequencies for each sub-category, total number of words, total number of parsed words, mean totals for each group($\bar{x}$) and mean occurrence per million were calculated and added to the table.

Testing Null Hypothesis 1.

For this hypothesis, the group-total raw frequencies for each sub-category were used. To determine if the distribution of responses is similar in Group A and Group B, a chi square test was carried out for each main category or column.

Early (1987), Wesche and Ready (1985), Long and Sato (1983) and Long (1981-82) have also conducted research using the testing of hypotheses with chi square. Like the present study, their studies compared a number of language characteristics as exhibited by a group speaking to an ESL population and a group speaking to a Non-ESL population.

In this study, a chi square test was conducted, for example, on Form Feature A (General Classification of Words), to see if Group A differed from Group B. A chi square value

which exceeds the critical chi square value at $p < .05$ with the applicable degrees of freedom would mean a rejection of the null hypothesis. This would mean that there is a significant difference between ESL teachers and Non-ESL teachers with respect to the vocabulary forms used. Since this is an exploratory study, some risks in analyzing the data were taken which may be optimizing on chance. In an attempt to reduce this risk, the sub-categories in each main category were grouped so that each chi square would involve a two up to five factor by two factor analysis. The groupings are as follows:

A. General Classification of Words
   1. Function Words
   2. Content Words
   3. Interjections

B. Grammatical Form
   1. Content words related to Noun Phrases - noun, pronoun, adjective.
   2. Content Words related to Verb Phrases - verb, adverb.
   3. Function words - preposition, co-ordinating conjunction, subordinating conjunction, determiner.

C. Pronoun Form
   1. Content Word Pronouns - personal, reflexive, possessive, reciprocal, demonstrative, partitive, universal.
   2. Function Word Pronouns - interrogative, relative.

D. Verb Form
   1. Finite Verbs - auxiliary and main.
   2. Non-finite Verbs - auxiliary and main.

3. Infinitives - bare and with "to".
4. Copula.
5. Imperative.

E. Inflectional Morphemes

1. 3rd person singular, past tense, and "ing" form.
2. Plural, and possessive.
3. Comparative and superlative.

F. & G. Contractions

1. Not, and us.
2. Is, am, and are.
3. Will, and would.
4. Has, have, and had.

H. Derivational Morphemes

1. Prefix.
2. Suffix.
3. Prefix and suffix.

The chi square results are reported in Chapter 4 in Tables 1 - 7, with a summary of the results in Table 8.

Testing Null Hypothesis 2.

Each subject's two-minute corpus was word searched. The necessary additions were made to PARSAID's dictionary, and then each corpus was parsed for word commonness according to a frequency range. *The Word Frequency Book* (Carroll, Davies & Richman, 1971) has approximately 87,000 words in its dictionary. The top 100 ranked words have frequencies ranging from 73,123 occurrences per million words (for the most common word) to 1072 occurrences per million words (for the 100th ranked word). These words (shown in Appendix E, Table 1)

represent the most frequently occurring words in the general population. The words ranked 4,950 to 86,740 occur 0 to 10 times per million and comprise the majority of the dictionary with the least frequency of occurrence. The remaining 4,850 words occur 1071 to 4949 times per million. This means that there is an uneven spread of occurrence versus rank. The frequency range chosen for this study attempted to take this into consideration and is as follows:

1. 3351 - 73,123 words per million (the 31 most frequent words)
2. 1601 - 3350 (the next 40 most frequent words)
3. 1072 - 1600 (the next 29 most frequent words)
4. 11 - 1071 (the next 4885 most frequent words)
5. 0 - 10 (the next 81,755 most frequent words)

The raw data for Groups A and B are shown in Appendix F, in Tables 1 and 2 respectively. The results are expressed in Chapter 4 as a cumulative frequency distribution of mean per cent of the corpus of Group A, Table 9 and Group B, Table 10. Figure 1 of this chapter expresses the relationship of Group A and Group B for this distribution in a Cumulative Frequency Distribution Polygon.

Testing Null Hypothesis 3.

The mean frequency of occurrence per million words is shown in Appendix F, Tables 1 and 2 and Chapter 4, Table 12. The means were tested for significant differences by using a two-tailed *t* test, df 18, $p < .05$.

Testing Null Hypothesis 4.

The mean total number of words as shown in Appendix F, Table 1 and 2 and Chapter 4, Table 11 were compared using a two-tailed *t* test, df 18, $p < .05$.

Summary

The frequency measures of the vocabulary forms described above constitute the basis for analysis in this study. From the analysis of these forms, a profile of vocabulary with respect to FT has emerged, based on the differences and similarities between the two groups studied.

Chapter 4

Results

Summary

This study focused on the vocabulary forms and vocabulary commonness of the speech input of teachers to L2 learners, thus adding to the profile of FT and TT. The instrument used to code the corpora and calculate frequencies was a computer program called PARSAID. The program permitted the efficient coding and frequency calculations of large corpora, and also allowed the entry of a reference frequency dictionary for the purpose of analysing word frequency per million.

The corpora of 10 ESL and 10 Non-ESL teachers were recorded, transcribed, coded and analyzed for vocabulary form features, word commonness, mean frequency per million and mean total word count. PARSAID allowed the entry of a coding system to characterize the vocabulary forms in selected corpora and the calculation of frequency totals. In addition, PARSAID calculated the frequency per million of each word.

In order to create a profile of TT by determining if there were any significant differences between Group A and Group B regarding vocabulary forms and vocabulary commonness, four research questions and four corresponding null hypotheses were formulated. This chapter presents the results of the tests of the four null hypotheses. These results provide answers to the corresponding research questions.

Research Question 1

The first research question was:

Is there any significant difference between Group A, ESL teachers, and Group B, Non-ESL teachers, in selected corpora with respect to a number of vocabulary forms?

The null hypothesis was:

There is no significant difference between Group A and Group B for selected vocabulary forms.

The designated test of significance was chi square, $p < .05$. Since this was an exploratory study, some risks were taken which may be optimizing on chance. For the purpose of reducing this risk, the sub-categories in each form classification were grouped into a smaller number of sets. Tables 1 - 7 below show the chi square test for each classification. Table 8 is a summary of these chi square results.

Table 1

A. Chi Square Test for General Classification of Words

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
Function Word	697	1052	758	991
Content Word	1613	1988	1561	2040
Interjection	74	77	65	86
TOTAL	2384	3117	2384	3117

$x^2 = 13.9$, df = 2, significant for $p < .05$

The chi square test for the figures in Table 1 indicates that there is a significant difference between Group A and Group B for the frequency of occurrence of the general classification of words.

Table 2

B. Chi Square Test for Grammatical Form

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
Noun,Pronoun, Adjective	953	1287	937	1303
Verb,Adverb	788	995	746	1037
Determiner,Prep., Co-ord. & Subordin. Conjunction	500	833	558	775
TOTAL	2241	3115	2241	3115

$x^2 = 14.9$, df = 2, significant for $p < .05$

The chi square test for the figures in Table 2 indicates that there is a significant difference between Group A and Group B for the frequency of occurrence of grammatical forms.

Table 3

C. Chi Square Test for Pronoun Form

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
Personal	298	325	289	334
General Content*	112	118	107	123
General Function**	42	78	56	64
TOTAL	452	521	452	521

$x^2 = 12.01$, df = 2, significant for $p < .05$

* Note: General Content Pronouns include Possessive, Reciprocal, Demonstrative, Partitive, and Universal.
** Function Pronouns include Interrogative and Relative.

The chi square test for the figures in Table 3 indicates that there is a significant difference between Group A and Group B for the frequency of occurrence of pronoun forms.

Table 4

D. Chi Square for Verb Forms

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
Finite Verbs	351	454	357	448
Non-finite Verbs	56	63	53	66
Infinitive Forms	56	72	57	71
Copula	83	114	87	110
Imperative	26	15	18	23
TOTAL	572	718	572	718

$x^2 = 7.19$, df = 4, not significant for $p < .05$

The chi square test for the figures in Table 4 indicates that there is no significant difference between Group A and Group B for the frequency of occurrence of verb forms.

Table 5

E. Chi Square Test for Inflectional Morphemes

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
3rd Person Sing., Past Tense,"ing"	102	210	114	198
Plural, Poss.	87	119	75	131
Compar.,Super.	1	3	1	3
TOTAL	190	332	190	332

$x^2 = 5.02$, df = 2, not significant for $p < .05$

The chi square test for the figures in Table 5 indicate that there is no significant difference between Group A and Group B for the frequency of occurrence of inflectional morphemes.

Table 6

F. & G. Chi Square Test for Contractions

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
not,us	26	19	20	25
is,am,are	48	81	59	70
will,would	18	10	13	15
has,have,had	5	7	5	7
TOTAL	97	117	97	117

$x^2 = 10.61$, df = 3, $p < .05$

The chi square for figures in Table 5 indicates that there is a significant difference between Group A and Group B for the frequency of occurrence of contractions.

Table 7

H. Chi Square Test for Derivational Morphemes

Classification	Observed		Expected	
	Group A	Group B	Group A	Group B
Prefix	7	1	3	5
Suffix	29	63	32	60
Prefix & Suffix	0	2	1	1
TOTAL	36	66	36	66

$x^2 = 10.96$, df = 2, $p < .05$

The chi square test for the figures in Table 7 indicates that there is a significant difference between Group A and Group B for the frequency of occurrence of derivational morphemes.

Table 8

Summary of Chi Square Test Results for Vocabulary Form Features

Vocabulary Form Features	Chi Square
A. General Classification of Words	13.9, df = 2, p < .05*
B. Grammatical Form	14.9, df = 2, p < .05*
C. Pronoun Form	6.9, df = 1, p < .05*
D. Verb Form	7.2, df = 2, n.s.
E. Inflectional Morphemes	5.0, df = 2, n.s.
F. & G. Contractions	10.6, df = 3, p < .05*
H. Derivational Morphemes	11.0, df = 2, p < .05*

Note: Reject H_o when p < .05 * = significant difference
n.s.= not significant

Table 8 shows a summary of the vocabulary forms tested for Hypothesis 1 and the chi square results for each form. With the exception of Verb Forms and Inflectional Morphemes, the null hypothesis can be rejected for all forms as indicated in the table. This means that there is a significant difference between Group A, ESL Teachers, and Group B, Non-ESL Teachers, for the use of the following vocabulary forms:

- A. General Classification of Words
- B. Grammatical Form
- C. Pronoun Form
- F. & G. Contractions
- H. Derivational Morphemes

Research Question 2

The second research question is:

Is there any difference in vocabulary commonness between Group A and Group B?

The null hypothesis is:

There is no significant difference between Group A and Group B for vocabulary commonness.

The frequencies will be expressed in a cumulative frequency distribution table for Group A (Table 9) and Group B (Table 10), and a cumulative frequency distribution polygon (Figure 1).

Table 9

Cumulative Frequency Distribution Table for Vocabulary Commonness - Group A

Word Commonness Rank	Frequency of Occurrence/ Million Words	Frequency as Mean % of Corpora	Cumulative Frequency as Mean % of Corpora
1-31	73123 - 3351	35	100
32-71	3350 - 1601	17	65
72-100	1600 - 1072	4	48
101-4984	1071 - 11	36	44
4985-86740	10 - 0	8	8

Table 10

Cumulative Frequency Distribution Table for Vocabulary Commonness - Group B

Word Commonness Rank	Frequency of Occurrence/ Million Words	Frequency as Mean % of Corpora	Cumulative Frequency as Mean % of Corpora
1-31	73123 - 3351	40	100
32-71	3350 - 1601	13	60
72-100	1600 - 1720	6	47
101-4984	1071 - 11	32	41
4985-86740	10 - 0	9	9

From the figures in Table 9 and Table 10, it appears that Group B uses more vocabulary, 35 per cent, in the most frequent occurrence range of 3,351 - 73,123 words per million than Group A who uses 35 per cent. However, Group A uses a slightly higher percentage, 17 per cent, in the second range of 1,601 - 3,350 words per million than Group B who uses 13 per cent. In the fourth range of 11 - 1,071 words per million, Group A has 36 percent versus Group B who has 32 per cent. Figure 1 below provides a graphic representation of these numbers according to the cumulative frequency percentages in each range of occurrence per million and for each group.

Figure 1

Cumulative Frequency Distribution Polygon for Vocabulary Commonness

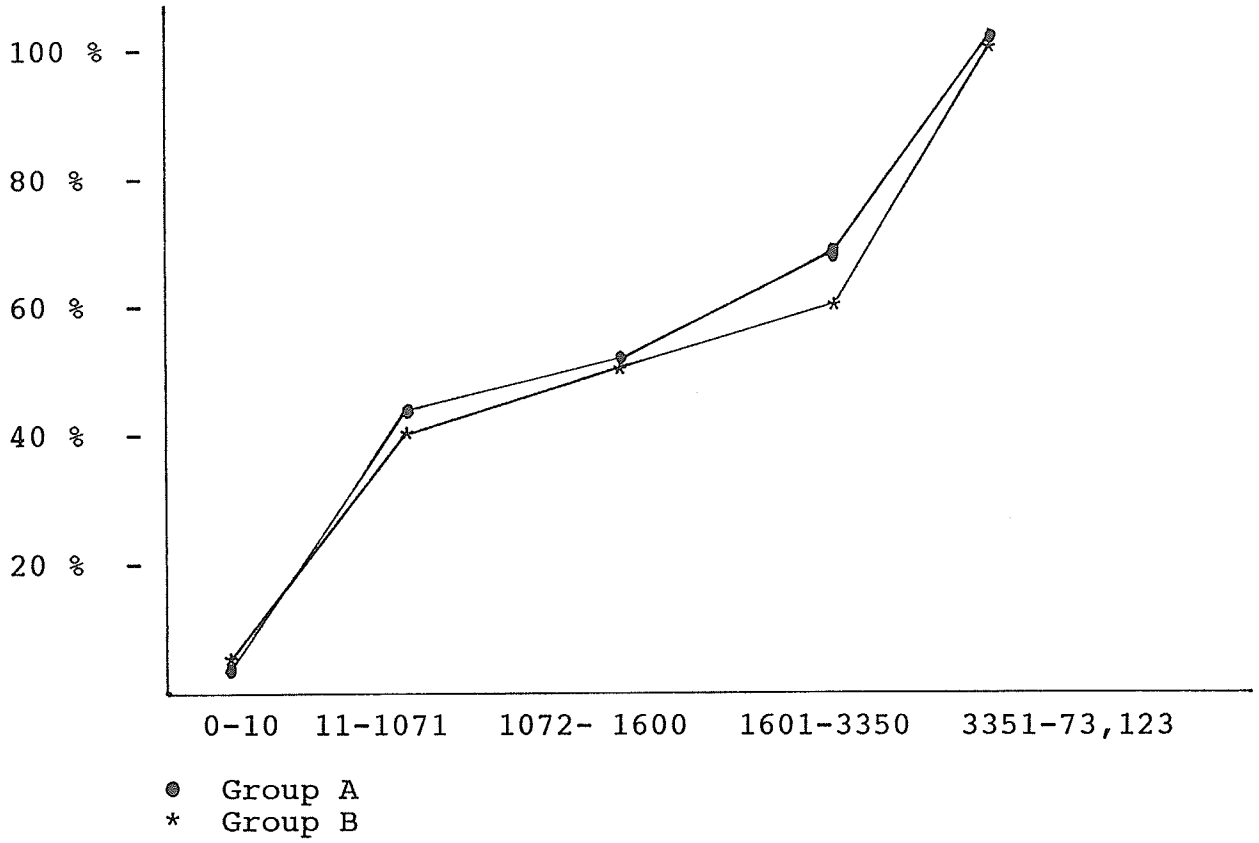

Figure 1 shows the difference between Group A and Group B for the cumulative frequency to five ranges of frequency of word occurrence per million. From these results it appears that there is no significant difference between the two groups for vocabulary commonness.

Research Question 3

The third research question is:

Is there any difference between Group A and Group B for the mean frequency of occurrence per million words?

The null hypothesis is:

There is no significant difference between Group A and Group B for the mean frequency of occurrence per million words.

Table 12 shows the mean frequency for each group.

Table 11

Mean Frequency of Occurrence per Million Words

	$\bar{x}$	s
Group A	7269.4	1932.18
Group B	8831.29	560.74

t = 15.79, df 18, $p < .01$

The t test for the figures in Table 11 indicates that there is a significant difference between Group A and Group B for the mean frequency of occurrence per million words. Specifically, the corpora of Group B (teachers speaking to NS listeners) contained significantly higher frequency per million words than the corpora of Group A (teachers speaking to NNS listeners).

Research Question 4

The fourth research question is:

Is there any difference between Group A and Group B for the mean total number of words?

The null hypothesis is:

There is no significant difference between Group A and Group B for the mean total number of words.

Table 12 shows the mean number of words per two minute corpus for Group A and Group B.

Table 12

Mean Total Number of Words per Two Minutes

	$\bar{x}$	s
Group A	240.1	68.52
Group B	331.4	58.54

t = 25.65, df 18, significant for $p < .01$.

The t test For the figures in Table 12 indicates that there is a significant difference between Group A and Group B for the mean total number of words used. Specifically, the corpora of Group B (teachers speaking to NS listeners)

contained significantly more words (about 38 per cent more on average) than the corpora of Group A (teachers speaking to NNS listeners).

Chapter 5

Discussion, Conclusion and Implications

The purpose of this study is to add to the current body of research concerning the language adjustments teachers make when communicating with NNS students. It is an exploratory study intended to determine similarities in the speech of ESL teachers with respect to pre-selected vocabulary form features, thereby creating a profile of TT with respect to vocabulary. As TT is a sub-category of FT, the profile will also add to a more general description of FT. Hopefully, this type of information will lead to more informed and reflective pedagogic decisions in ESL classrooms.

Discussion of Results

Research Question 1.

Is there any significant difference between Group A, ESL teachers and Group B, Non-ESL teachers, in selected corpora with respect to a number of vocabulary form features?

This study found that there was a significant difference between Group A and Group B for the following vocabulary forms:

- general classification of words
- grammatical form
- pronoun form

- contractions
- derivational morphemes

No significant difference was found between the two groups for the following vocabulary forms:

- verb form
- inflectional morphemes

These results including possible underlying causes of significant differences, are explored in the following sections.

General Classification of Words

Wesche and Ready (1985) found that there was no significant difference in the number of content words versus the number of function words for NSs when speaking to either NS or NNS students. It should be noted that their study did not include adverbs or pronouns as content words. The findings of the present study indicate that there is a significant difference between Group A and Group B regarding the use of function versus content words, with a $p < .05$, $x^2 = 13.9$ (see Table 1). The difference in the findings of the two studies can perhaps be explained on the basis of differing definitions of content and function words.

In terms of the relative percentages of the each group's total corpus, Group A (ESL teachers) used 29 per cent function words, 68 per cent content words and 3 per cent interjections.

Group B (Non-ESL) teachers used 34 per cent function words, 64 per cent content words and 2 per cent interjection. It is interesting to note that function words include forms like co-ordinating conjunctions, relative pronouns, prepositions and the complementizer "that". These types of words often signal the introduction of linguistically more complex clauses. We may speculate that the difference in usage between Group A and Group B is due to more complicated types of sentence structures being used by the Non-ESL teachers, thus requiring the use of more function words. It is beyond the scope of this study to proceed with an analysis of this type; however, with regard to vocabulary choice, it can be concluded that there is a difference and that it is likely of this nature.

Grammatical Form

Wesche and Ready's (1985) study found no significant difference for grammatical form. Data from the present research indicated a significant difference between Group A (ESL teachers) and Group B (Non-ESL) teachers, $p < .05$, $x^2 = 14.9$ (see Table 2). ESL teachers thus appeared to use different grammatical forms in their speech, indicating a further characteristic of TT.

Reflecting on the possible source of the difference by looking at the percentage of use in each group's total corpus, the ESL teachers used 43 per cent noun/pronoun/adjective

versus 41 per cent for Non-ESL teachers, 35 per cent verb/ adverb versus 32 per cent for Non-ESL teachers and 22 per cent determiner/preposition/co-ordinating and subordinating conjunction versus 27 per cent for Non-ESL teachers. The largest difference seems to be in the last sub-group which comprises the function words. This is consistent with the findings for general classification of words. The conclusion could be the same in that Non-ESL teachers used more vocabulary forms which signalled the introduction of additional or more complex clauses.

Pronoun Form

Meisel (1976) concluded that one characteristic of FT was the deletion of personal pronouns. Hatch, Shapira & Gough (1975) found that there was a tendency not to use possessive pronouns in FT. Henzl (1974) observed fewer indefinite pronouns (partitive and universal) and a greater number of personal pronouns in a speaker's effort to be more concrete in his/her word choice. Kliefgan (1985) found a tendency to delete pronouns if possible. However, Wesche & Ready (1985) found no significant difference in pronoun use. This study found that there is a significant difference in the use of pronouns by Group A and Group B teachers (see Table 3). By impression, it would appear that Group A (ESL teachers) use slightly more personal pronouns (66 per cent of the total pronouns) and fewer function word types of pronouns (9 per

cent of the total pronouns) than Group B (Non-ESL teachers) whose personal pronouns were 62 per cent of the corpus and function word types were 15 per cent. Again, this seems to indicate that in the case of function word types of pronouns there is a more complex type of syntax being used by Group B teachers than Group A. The use of slightly more personal pronouns is probably an attempt by the ESL teachers to be more concrete and less ambiguous in their vocabulary choice, as Henzl (1974) points out.

Although it may be reasonable to assume that ESL teachers avoid personal pronouns for this reason, this suspicion is not borne out by the data on the total number of pronoun forms in each group's corpus. For Group A (ESL teachers), pronouns, constitute 19 per cent of the total corpus. For Group B (Non-ESL teachers) pronouns are 17 per cent of the total corpus. Clearly then, pronoun deletion of any nature does not appear to be a characteristic of TT and the opposite tendency, that being increased use, may be happening. This may be related to an attempt by ESL teachers to avoid ambiguous circumlocutions and ambiguous deletion as a means of simplification and to use more concrete and direct language involving the use of pronouns as a way of being specific. For example, ESL Teacher 1 said "Anything more about this one? This conversation? No? You understand all those words?" There seems to be an attempt by the teacher to be very specific about which conversation and which words he/she is referring to. In contrast, he/she

could have said, "Anything more? Do you understand?" This would have been simplification by deletion and the message is much less specific.

Contractions

There has been no research which has specifically studied contracted forms. In terms of L2 acquisition, contractions are generally viewed as a more advanced way of speaking and reflect the use of stress and rhythm in the target language. Contractions are often not heard when lower proficiency NNSs are listening to Nss. Therefore, logically, an attempt to simplify speech would probably involve less use of contractions. This research found that there is a significant difference in the use of contracted vocabulary forms, $p < .05$, $x^2 = 10.6$ (see Table 6) with Non-ESL teachers using more contracted forms of the verb "to be" (69 per cent of the total contractions) versus the ESL teachers (49 per cent). However, the ESL teachers used more contracted "not" and "us" (Group A, 27 per cent versus Group B, 16 per cent) and more "will" and "would" (Group A, 19 per cent and Group B, 9 per cent). One explanation for a greater use of the "not" and "us" contraction sub-category is the discourse style of the speaker. For example, ESL Teacher 1 in Group A used statements like "Let's do one more on the same theme". In contrast, ESL Teacher 2 gave directions in a different manner,

for example, "Practise reading along with me", whereas he/she could have said "Let's read together".

The results of this study indicate that ESL teachers do not merely use fewer contracted forms; they also tend to use contractions of different types. Further research would be necessary, however, to clarify these tendencies.

Derivational Morphemes

Henzl (1974) characterized FT as having more basic vocabulary, i.e., root forms. Ferguson (1971) also claims that there would be fewer derivational forms found in FT. Chaudron (1983) observed that the nature of vocabulary tends to be simpler with more frequently used words. In the present study, Group B (Non-ESL teachers) used significantly more derivational morphemes than Group A (ESL teachers), $p < .05$, $x^2 = 11.0$ (see Table 7). It can be concluded that the absence of derivational prefix and suffix use is a characteristic of TT.

Verb Forms

Kliefgan (1985) found that TT could be characterized by a deletion of copulas, auxiliaries and verbs. Meisel's (1976) characteristics of FT include an absence of verb forms. Ferguson (1971) claims that FT involves an omission of the copula. The data from the present study (see Table 4) indicates that there is no significant difference in the use

of verb forms between Group A (ESL teachers) and Group B (Non-ESL teachers). Even speculating by comparing percentages of use in the two groups suggests that there is no difference in the way verb forms are employed by either group. Without further study, the only explanation to be offered for this in the light of previous research findings is the proficiency level of the learners. The learners chosen for this study were considered to be at an intermediate or advanced level. Therefore, the learners' speech likely contains many of the verb forms, perhaps without the appropriate inflectional morphemes. If the NS is indeed mimicking the NNS, then he/she would be imitating the characteristics of the proficiency level.

Inflectional Morphemes

Henzl (1974) found that there are fewer past tense verbs in FT. Early (1987) and Long & Sato (1983) claim that one characteristic of FT is the use of fewer tensed verbs. Wesche & Ready (1985) found that the English professor used present tense verb forms almost exclusively with NNS students and more non-present tensed copula with NS students. Miesel (1976) concludes that in FT there is a less marked lexicon, for example, more use of "to do" versus "does" and that, generally speaking, there are fewer inflections on verbs, adjectives and nouns. Ferguson (1981) says that there are fewer inflections in FT. Hatch, Shapira & Gough (1975) found a lack of verb

tense and no possessive forms in FT. However, this study found that there was no significant difference for inflectional morpheme use by the two groups of teachers. By speculating with percentages, it appears that Non-ESL teachers do use a higher per cent (63 per cent of the total inflected forms) of inflected forms for 3rd person, past tense and "ing" versus ESL teachers (54 per cent). However, the ESL teachers have a higher percentage of inflected forms (45 per cent of the total inflected forms) as plural or possessives than Non-ESL teachers (36 per cent).

On the basis of previous research, we might suspect that a lower percentage of inflected verb forms would be characteristic of FT, but verbs are not the only inflected forms to be investigated in this study. Speculating further into noun and possessive inflections, it appears that for Group A (ESL teachers), 41 per cent of the inflectional morphemes are plural nouns and 5 per cent are possessives. For Group B (Non-ESL teachers) 34 per cent of the inflections are plural nouns and 1 per cent are possessives. Although the results of the present research indicated that there was no difference in the use of inflections, impressionistic observation indicates that additional research could add interesting conclusions about the differing use of inflections as a possible aspect of FT.

Research Question 2.

Is there any significant difference in vocabulary commonness between Group A and Group B?

Derwing (1991) found that Nss used more words in the most frequent categories when speaking to NNSs. The high interpersonal affective speakers used significantly more of the top 500 words with NNSs and significantly fewer of the least frequent. Chaudron (1983) makes an impressionistic observation and says that the vocabulary choice in FT appears to be simpler in that more frequently occurring words are used. Henzl (1974) claims that in FT,vocabulary choice is less diversified and more basic in nature.

This study found that if there was any difference, Group B (Non-ESL teachers) used a slightly higher percentage of the corpus (40 per cent) in the 31 most frequently used words, versus Group A (35 per cent). Comparatively, Group A used a higher percentage (17 per cent) in the second category with Group B using slightly fewer (13 per cent). In terms of cumulative frequency, the largest difference is in this second category, with Group A having 65 per cent of the corpora below this range, with Group B having 60 per cent below. Again, this means that Group B actually had a slightly higher percentage in the most frequent range. Appendix E contains a list of the 100 most frequently occurring words per million. In ranking 1 - 31, seven words are pronouns and the rest are function words. The majority of words in this range are

function words. Therefore, consistent with the Research Question 1 findings that Group B consistently used more function word forms than Group A, this would seem to make some sense. As function words signal grammatical relationships, and the grammatical structures used by Non-ESL teachers appear to be more complex, we might expect that Non-ESL teachers would use a higher frequency of non-content words. In contrast, ESL teacher discourse tends to be more concrete and fewer of these types of words would be expected.

Research Question 3.

Is there any difference between Group A and Group B for the mean frequency of occurrence per million words?

It appears from the results of this study that Group B teachers used significantly higher frequency per million words than Group A. This means that Group B's lexicon was more common in nature than Group A. As in Research Question 2, this could be attributed to the greater frequency of function words.

Research Question 4.

Is there any difference between Group A and Group B for the mean total number of words?

Wesche & Ready (1985) found that the English speaking professor used more words when speaking to NS students. Kliefgan (1985) concluded that teachers used fewer words with

less variety when speaking to NNS students. Ferguson (1981) feels that there is the use of less words with FT. This study found a significant difference in the total number of words used. Group A (ESL teachers) used 72 per cent of the total number of words used by Group B (Non-ESL teachers), a difference of 28 per cent. Obviously, the use of fewer words in total is a characteristic of TT. Whether this is due to decreased velocity or increased pause time is an area for further study.

Summary and Significance

This study was intended to add information to the current body of knowledge about the language adjustments teachers make to vocabulary forms and vocabulary commonness when speaking to NNS students. Ten ESL teachers speaking to NNS students with at least an intermediate level of language proficiency were compared to ten Non-ESL teachers speaking to NNS students. The resulting profile of TT lends support to some findings of previous research, but is also contradicting.

The significance of the present study lies in the fact that it focuses on vocabulary forms and vocabulary commonness in a quantitative and comprehensive manner. It is also the first study which used the computer program PARSAID to assist with the classification of language forms. This program allows a large corpus to be categorized in a multitude of ways simultaneously and efficiently. In the future, additional

studies of this design may allow a more definitive examination of language features.

The results of this study indicate that TT directed at NNS contains a lesser proportion of function words than NS-NS types of interaction. The grammatical forms present in NS-NNS interactions are also different in some respects. This difference appears to lie in the direction of using fewer determiners, prepositions, co-ordinating and subordinating conjunctions, with slightly more predominance of nouns, adjectives, pronouns, adverbs and verbs. This result confirms the finding that fewer function words are found in TT. In terms of the pronoun forms, there is a difference in the forms used by ESL teachers versus Non-ESL teachers, with ESL teachers employing more personal pronouns and fewer interrogative and relative pronouns than Non-ESL teachers. In contrast to previous research, deletion of pronoun forms was not found to be a characteristic of TT and, if anything, there was a slight increase in the use of pronouns by ESL teachers. ESL teachers used fewer contractions of the verb "to be" than Non-ESL teachers. However, ESL teachers used more contractions of "not", "us", "will", and "would". Therefore, a profile of TT would have to include the differential use of contractions versus a deletion. Further research is needed to be more specific.

There is a reduction in the use of derivational morphemes in the speech of ESL teachers. Vocabulary forms in TT appear to lack derivational prefixes and suffixes.

Surprisingly, this study found that, in general, verb forms in TT are characterized in the same manner as NS-NS interactions. This seems contrary to a great deal of the previous research. Also in contrast to much of the previous research, there was no real difference in the use of inflectional morphemes. There appear to be slightly fewer inflections of 3rd person, past tense and "ing" forms and a slightly higher occurrence of plural nouns and possessive inflections in TT. This finding would have to be confirmed by further in-depth study.

The data from the present study indicates that NS-NS discourse contains a higher percentage of words which are in the most frequently occurring per million range, 3351 to 73,123 words per million. An examination of the *Word Frequency Book* (Carroll, Davies & Richman, 1971) shows that most of these 31 most frequently occurring words are function types of words. Again, this finding is consistent with other findings in this study. In the next 32 - 71 most frequently occurring words, Group A teachers had a slightly higher percentage than Group B. There may be a tendency for ESL teachers to use more commonly occurring content types of words than Non-ESL teachers, but this would have to be examined in greater depth.

Limitations

Ferguson (1981) says:

> "Perhaps the most serious disadvantage of all these names for simplified registers is that they tend to focus research on particular registers or styles rather than on the dimensions of variation which cut across them".

Certainly, that is one of the limitations to generalizability of this study, in that it is limited to TT situations involving NNSs who are at an intermediate language proficiency level. This means that the TT profile may not reflect other teaching situations involving, for example, learners at beginner proficiency levels. Also in terms of dimensions of variation, this study's results represent an average of a group of ten teachers and do not account for the individual language style differences between each teacher.

Furthermore, this study has not accounted for the interactional nature of the students and teachers in the classroom. Communication is, to a greater or lesser degree, a two-way process in that often what one individual says elicits a certain type of response from the second. This study does not account for language choice, in part, being formed by the demands of the communicative situation.

Additionally, it is unclear to what extent vocabulary choice is more determined by other factors such as clause or sentence type or discourse style, for example. Vocabulary may

be the result of another causal factor instead of an independent choice. There is no control for this in the present study.

Also, the language proficiency of the students was not formally tested but left to identification by the teacher. The leaves an element of subjectivity which could influence the findings.

Because of the large number of categories and sub-categories of vocabulary forms, more powerful statistical tests could not be used because of the risk of optimizing on chance. As a result, specific comparisons between Group A (ESL teachers) and Group B (Non-ESL teachers) for each feature could not be made. Also, because of the ranges chosen for the frequency per million comparison, more exacting conclusions could not be drawn about the nature of the vocabulary differences in this respect.

Further Research

This study was an exploratory study of a large number of features specific to vocabulary form. Any of the sub-categories could be studied more extensively. For example, a study could be conducted on pronoun forms using the same research design. Because of the more limited number of categories, the statistical test in this case could be a *t* or *F* test, thus also allowing for individual variation of the scores.

The results of this study point to the idea that vocabulary forms are indicators of syntactical differences. PARSAID would allow for research into this type of difference using the same research design as the present study.

One area that this study did not account for was the individual discourse style of the speakers. It is not known how heavily this influences the findings; however, averaging of the data accounted for this to a certain extent. This study did include three teachers who were in both Group A (ESL teachers) and Group B (Non-ESL teachers). It would be an interesting study to use the same research design and vocabulary form features but use the same individuals in two different speaking situations as a basis for creating a profile which could be compared to those created in this study.

An extension of this idea would be to use the same teacher or the same group of teachers speaking to NNSs with differing proficiency levels. This would help to identify if FT is a way of mimicking the NNSs. It would also create a continuum of FT from the way in which NS speak to each other to the way in which they speak to those they perceive to have little or no proficiency in their language.

As there was a significant difference in the total number of words used by ESL teachers, a study of this difference would also be valuable. Such an investigation would involve a research design intended to discover whether it is the

velocity, pause time or a combination of both causing the difference in the total number of words.

Further research into the frequency of occurrence per million words could yield some valuable information about FT. More data and further sub-divisions of ranges of occurrence could give us more information about lexical choice. The notion of simplicity in this respect is still largely undefined. It would be interesting to have a theory proposed about the notion of simplicity and word choice. So far, simplicity and vocabulary choice in FT seems to be based on intuitive choices and explanations.

Another interesting study would be a survey or quantitative study on whether or not teachers are making conscious choices about their language. If so, what do they think they are doing or what is their strategy for simplifying their language? Further to this, in terms of self-knowledge and pedagogic choices, could a sub-system be developed for analyzing TT and providing feedback to teachers in their conscious or unconscious choices?

Another worthwhile study would be a survey or quantitative study of what L2 learners perceive as simplification. This could include information on what they find beneficial or not beneficial in terms of TT.

Conclusion

In conclusion, this study has added a dimension to the study of vocabulary forms and vocabulary commonness with respect to FT which to date has remained largely unexplored. By examining vocabulary forms in a quantitative and comprehensive manner, we are better able to profile the nature the language adjustments teachers make when speaking to NNS learners. With this profile, we can begin a discussion as to the notion of simplification and its relevance to pedagogic decisions regarding the type of language input chosen with NNS students. Hopefully, this will ensure that simplification is not practised in a way that will result in cognitive simplification also.

Most importantly, this study has pointed to a number of more specific and valuable research avenues for further study in this area. The computer program PARSAID has been an invaluable tool in undertaking this type of comprehensive and quantitative study. With PARSAID, the possibilities for further detailed research are very attractive and immediate. Detailing and counting large corpora is no longer forbidding and it is this type of research which will result in a better understanding the nature of language in the second language acquisition process.

References

Anderson, W.L. and Stageberg, N.C. (1975). *Introductory Readings on Language*. New York: Holt, Rinehart and Winston.

Bailey, K.M., Long, M.H. and Peck, S. (Eds). (1983). *Second Language Acquisition Studies.* Rowley: Newbury House Publishers.

Beebe, L.M. (1985). Input: Choosing the Right Stuff. In S.M. Gass and G.G. Madden (Eds), *Input in Second Language Acquisition*. Rowley: Newbury House.

Blau, E.K. (1982). The Effect of Syntax, Speed, and Pauses on Listening Comprehension. *TESOL Quarterly, 16*(4), 517-528.

Brown, J.D. (1988). *Understanding Research in Second Language Learning*. Cambridge: Cambridge University Press.

Brulhart, M. (1986). Foreigner Talk in the ESL Classroom: International Adjustments to Adult Students at Two Language Proficiency Levels. *TESL Canada Journal Special Edition 1,* 29-42.

Carroll, J.B., Davies, P. and Richman, B. (1971). *Word Frequency Book.* Boston: Houghton Mifflin Co.

Carty, M. (N.D.). *Strategies Used By Native Speakers in Native Non-Native Conversations*. Unpublished dissertation.

Celca-Murcia, M. and Larsen-Freeman, D. (1983). *The Grammar Book an ESL/EFL Teachers Course.* Rowley: Newbury House.

Chaudron, C. (1983). Foreigner Talk in the Classroom - An Aid to Learning? In H. W. Selinger and M.H. Long (Eds.), *Classroom Oriented Research in Second Language Acquisition.* Rowley: Newbury House.

Chaudron, C. (1988). *Second Language Classrooms Research on Teaching and Learning.* Cambridge: Cambridge University Press.

Chaudron, C. (1983). Research Notes, Vocabulary Elaboration in Teachers' Speed to L2 Learners. *Studies in Second Language Acquisition. 4*(2)170-180.

Clyne, M.G.(Ed.) (1981). *International Journal of the Sociology of Language. 28.* Special Issue on Foreigner Talk.

Dahl, Hartvig. (1979). *Word Frequencies of Spoken American English.* Essex: Verbatim.

Derwing, T. (1991). The Role of NS Personality and Experience in NS-MNS Interaction. *TESL Canada Journal. 9*(1), 9-28.

Early, M. (1987). Linguistic Input and Interaction in the Content Classroom. *TESL Canada Journal, 4*(2), 41-58.

Ellis, R. (1985). Teacher-Pupil Interaction in Second Language Development. In S.M.Gass and G.G. Madden (Eds), *Input in Second Language Acquisition.* Rowley: Newbury House.

Ferguson, C. (1971). Absence of Copula and the Notion of Simplicity. In D.Hymes (Ed.) *Pidginization and Creolization of Languages.* New York: Camridge University Press.

Ferguson, C.A. (1981). Foreigner Talk as the Name of Simplified Register. *Int'l Journal of the Sociology of Language. 18.*9-18.

Freed, B.F. (1981). Foreigner Talk, Baby Talk, Native Talk. *Int'l Journal of Sociology of Language. 18*.19-39.

Freed, B.F. (1980). Talking to Foreigners versus Talking to Children: Similiarities and Differences. In R.C. Scarcella and S.D. Krashen (Eds.), *Research in Second Language Acquisition.* Rowley: Newbury House.

Fromkin, V. and Rodman, R. (1983). *An Introduction to Language (3rd ed.)* New York: Holt, Rinehart & Winston.

Fink, D.R. (1987). *A Descriptive Study of Teacher and Student Questioning Patterns in University ESL and Native Writing Classes.* Unpublished doctoral dissertation. University of Pennsylvania.

Gass, S.M. and Madden, C.G.(Eds.) (1985). *Input in Second Language Acquisition.* Rowley: Newbury House.

Hatch, E. (1978). Discourse Analysis and Second Language Acquisition. In E.M. Hatch(Ed.) *Second Language Acquisition A Book of Readings.* Rowley: Newbury House.

Hatch, E.M. (Ed.) (1978). *Second Language Acquisition A Book of Readings.* Rowley: Newbury House.

Hatch, E., Shapira, R. and Gough, J. (1975). Foreigner Talk Discourse. *ITL: Review of Applied Linguistics.* 39-60.

Henzl, V.M. (1974). Linguistic Register of Foreign Language Instruction. *Language Learning, 23*(2), 207-222.

Jones, L.V. and Wepman, J. (1966). *A Spoken Word Count.* Chicago: Language Research Assoc.

Kleifgen, J. (1985). Skilled Variation in a Teacher's Use of Foreigner Talk. In S.M. Gass and G.G. Madden(Eds.). *Input in Second Language Acquisition.* Rowley: Newbury House.

Krashen, S.D. (1981). *Second Language Acquisition and Second Language Learning.* Oxford: Permagon Press.

Krashen, S.D. (1982). *Principles and Practise in Second Language Acquisition.* Oxford: Permagon Press.

Krashen, S.D. and Terrell, T.D. (1983). *The Natural Approach.* Hayward: Alemany Press.

Kucera, H. and Francis, W.N. (1967). *Computational Analysis of Present Day American English.* Providence: Brown University Press.

Larsen-Freeman, D. (1983). Second Language Acquisition: Getting the Whole Picture. In K.M. Bailey, M.H. Long and S. Peck (Eds.), *Second Language Acquisition.* Rowley: Newbury House.

Larsen-Freeman, D. (1985). State of the Art on Input in Second Language Acquisition. In S. M. Gass and G.G. Madden (Eds.), *Input in Second Language Acquisition.* Rowley: Newbury House.

Lightbown, P.M. (1983). Exploring Relationships between Developmental and Instructional Sequences in L2 Acquisition. In H. Selinger and M. Long (Eds.). *Classroom Oriented Research in Second Language Acquisition.* Rowley: Newbury House.

Long, M.H. (1985). Classroom Foreigner Talk Forms and Functions of Teachers Questions. In S. M. Gass and G.G. Madden (Eds.). *Input in Second Language Acquisition A Book of Readings.* Rowley: Newbury House.

Long, M. (1981-82). Questions in Foreigner Talk Discourse. *Language Learning A Journal of Applied Linguistics, 31*(1), 135-157.

Long, M.H. and Sato, C.J. (1983). Classroom Foreigner Talk Discourse: Forms and Functions of Teachers' Questions. In H. W. Selinger and M.H. Long (Eds.). *Classroom Oriented Research in Second Language Acquisition.* Rowleey: Newbury House.

Lovejoy, E.P. (1975). *Statistics for Math Haters.* New York: Harper & Row.

MacPherson, E.D. *An Introduction to Educational Research for Practitioners* (5th Ed.) Unpublished Course Notes for 43:503, U. of Manitoba.

Mahnke, K.M. (1988). *Foreigner Talk Interaction in the Post-Secondary Content Classroom.* Unpublished doctoral dissertation. Indiana University.

Mason, E.J. and Bramble, W.J. (1989). *Understanding and Conducting Research. Applications in Education and the Behavioural Sciences.* New York: McGraw-Hill Book Co.

Meisel, J.M. (1976). Linguistic Simplification: A Study of Immigrant Workers' Speech and Foreigner Talk. In S. Corder and E. Roulet (Eds.) *Actes Du 5èmes Colloque De Linguistique Appliquée De Neuchatel.* Paris. AIMAV/Didier.

Pearson, C.R. (1983). *Foreigner Register: Characteristics and Implications.* Paper presented at CATESL Bay Area Regional Conference 1981 aand the Central California Conference on Teaching ESL 1982.

Pica, T. and Doughty, C. (1985). Input and Interaction in the Communicative Language Classroom: A Comparison of Teacher Fronted and Group Activities. In S. M. Gass and G.G. Madden (Eds.). *Second Language Acquisition: A Book of Readings.* Rowley: Newbury House.

Quirk, R. and Greenbaum, S. (1973). *A Concise Grammar of Contemporary English.* San Diego: Harcourt, Brace, Jovanovich Publishers.

Robinett, B.W. (1978). *Teaching English to Speakers of Other Languages.* Minneapolis: U of Minnesota Press.

Scarcella, R.C. and Krashen, S.D. (Eds.) (1980). *Research in Second Language Acquisition.* Rowley: Newbury House.

Shavelson. R.J. (1981). *Statistical Reasoning for the Behavioural Sciences.* Boston: Allyn and Bacon Inc.

Shinke-Llano (1981) *English Foreigner Talk in Content Classrooms.* Phd. dissertation. Northwestern University.

Shinke-Llano, L.A. (1983). Foreigner Talk in Content Classrooms. In H.W. Peck and M.H. Long (Eds.). *Classroom Oriented Research in Second Language Acquisition.* Rowley: Newbury House.

Snow, C.E. van Eeden, R. and and Muysken, P. (1981). The Interactional Origins of FT. *Int'l Journal of the Sociology of Language, 28,* 81-91.

Swan, M. (1980). *Practical English Usage.* New York: Oxford University Press.

Terrell, T.D. (1982). The Natural Approach to Language Teaching: An Update. *Modern Language Journal, 66,* 121-132.

Thorndike, E. and Lorge, I. (1944). *The Teacher`s Word Book of 30,000 Words.* New York: Columbia University Teachers` College.

Valdman, A. (1981). Sociolinguistic Aspects of Talk. *Int'l Jounrnal of the Sociology of Language. 28.* 41-52.

Welkowitz, J. Ewen, R.B. and Cohen, J. (1982). *Introductory Statistics for the Behavioral Sciences.* New York: Academic Press Inc.

Wesche, B.M. and Ready, D. (1985). Foreigner Talk in the University Classroom. In S. M. Gass and G.G. Madden(Eds.). *Input in Second Language Acquisition.* Rowley: Newbury House.

West, M. Ed. (1967). *A General Service List of English Words.* London: Longmans, Green and Co. Ltd.

Yuker, Harold E. (1958). *A Guide to Statistical Calculations.* New York: G.P. Putnam's Sons.

Zuengler, J. (1985). Phonological Aspects of Input in NS and NNS Interactions. In S. M. Gass and G. G. Madden(Eds.). *Input in Second Language Acquisition.* Rowley: Newbury House.

APPENDIX A: CODING

Basic Coding Procedure

Every word or connected group of words in the text is coded by a series of digits in square brackets. Each digit represents a particular sub-category within a main category. Where a main category is not applicable to the word a zero is entered. There are 8 main categories in the coding scheme, so the maximum length of a word code is 8 digits. However, most codes are shorter that this, since coding ceases with the last non-zero categorization in every case. For example, in a sentence beginning, "I went....", "went" is coded [22024].

> Column 1: General classification - "went" is a content word so it is coded "2".
> Column 2: Grammatical Form - "went" is a verb so it is coded "2".
> Column 3: Pronoun Form - a zero is coded here as it does not apply to this word.
> Column 4: Verb Form - "went" is a finite main verb so "2" is coded.
> Column 5: Inflectional Morpheme - "went" is marked for tense so it is coded with a "4" which is past tense.

In the description below, the 8 main categories are labelled by the letters A to H as a way of naming each column.

The coding has been described generally in Chapter 3. What follows is an expansion of this description.

The coding is not as precise as is possible, since the linguistic model used for classification is not as complex as is possible. The purpose was to impose a generalized classification system as an indicator of types of forms.

Specifics of Coding Procedure

A. *General Classification*

1. Function Words; often called "structure words" in the literature (Robinett, 1978).

The most frequently occurring are the following:

a. articles - e.g., *the, an.*
b. prepositions - e.g., *in, on.*
c. auxiliary verbs - e.g., *do, might.*
d. co-ordinating conjunctions - e.g., *and, or.*
e. subordinating conjunctions - e.g., *because, although.*
f. interrogative pronouns - e.g., *where, why.*
g. relative pronouns - e.g., *who, which.*
h. complementizer - e.g., *that.*

Function Words characteristically indicate the type of grammatical structure which is to follow in the form of a content word. They are the frame on which content words are hung. They have little referential meaning and belong to a closed class, i.e., a class to which new words are added only rarely in the historical development of the language. While actually few in number, they are frequently used.

2. Content Words (Robinett, 1978)

Content Words belong to an open class of words which can be endlessly created and freely substituted. These words usually have a denotative meaning, i.e., they refer to some existing entity, condition, action, manner, etc. They include:

a. nouns - common, e.g., *dog;* proper, e.g., *Elizabeth.*
b. verbs - e.g. *run, eat.*
c. adjectives - e.g., *heavy, pretty.*
d. adverbs - e.g., *very, quickly*

e. pronouns - e.g., *I, those*. Note: Pronouns are included here because they too are denotative, i.e., they refer to actual persons, things etc.

Derivational morphemes (prefixes and suffixes) can be added to Content Words to create words with new meanings, e.g., disavow. See main category H for details about coding.

3. Interjections

Interjections are often used by speakers to create a pause and are not usually intended to convey meaning. These types of words and sounds include okay, so, uh, and um. To code these words, it was first determined if they were an answer to a question or just interjected for pause. For example, "okay" can be an answer to a question or just a pausing device. If it was an answer to a question it was coded as an adverb (see Column B, number 5 below). If there was a series of interjections, each was coded as follows: uh[3] uh[3] okay[3].

B. *Grammatical Form*

1. Noun

This is a word used for a person place or thing. It is simply coded: dog[21].

2. Verb

This form includes lexical and auxiliary verbs. Each verb that appears is counted separately. For example, "I

must wash the car", must[1201] wash[2202], wherein the auxiliary is coded separately from the main verb.

Some verbs are considered to be phrasal verbs and their meaning cannot be separated from the particle which follows. For example, in "I gave in to their demands", gave in[2202] is coded as one. Further explanation of verbs and coding is contained below in "D". Auxiliaries are coded as structure words and the remainder of verbs as content words.

3. Pronoun

Pronouns, words which can substitute for nouns, are easily identified and coded. For example, "She took her book", she[231] her[233]. Definitions of the different types of pronouns are given in "C" below. Relative and interrogative pronouns are considered to be Function Words and the remainder of types are Content Words. The following pronouns can function as determiners and when occurring in this position are coded as Function Words and determiner:

every, each, some, either, any, neither, these, those, that, this.

4. Adjectives

These are words which modify nouns. For example, "subordinate clause", subordinate[24] clause[21].

5. Adverb/multi-word adverbial

A word or phrase of this nature further defines, qualifies or modifies a verb, adjective or another adverb. For example, in "He ran quickly", quickly [25] modifies the

verb. In "He ran very quickly", coding is very [25] quickly [25], with "very" further defining the adverb.

In "She is a very pretty girl", very[25] further defines the adjective "pretty".

Some adverbs appear as multi-word units and were coded as one. For example, "I showed you this last time", last time[25].

Conjuncts and disjuncts are coded as adverbs. For example, "Moreover, he left.." Moreover[25]

"Frankly, I think...." Frankly[25]

Also, responses to questions such as "yes, no, okay" were coded as adverbs. All adverbs were coded as content words.

6. Prepositions

These are structure words which help to show a relationship between a noun and some other word(s) in the sentence. They can appear as phrases and are coded as one. For example, "I'll take this instead of that", instead of[16]

The following are considered as prepositions:

about, above, across, after, against, along, among, around, at, before, behind, below, beneath, beside(s), between, beyond, by, despite, down, during, for, from, in, into, like, near, of, off, on, out, over, since, through, throughout, to, toward(s), under, until, up, upon, with, within, without, according to, along with, apart from, as far as, aside from, as to, because of, by way of, except for, in regard to, in addition to, in spite of, instead of, on account of, on top of, out of, together with.

7. Co-ordinating conjunctions

These are words which join two structures of equal grammatical type (nouns, verbs, phrases, clauses etc.). They include:

and, or, but, nor, for, so, yet, plus.

8. Subordinating Conjunctions

These are words which connect a dependent and independent clause wherein the dependent clause relies on the independent clause for its full meaning. Subordinating conjunctions include:

after, although, as, as if, as long as, as soon as, as though, because, before, since, so that, than, though, unless, until, when, whenever, where, wherever, whether, while.

9. Determiner (Swan, 1980)

These are structure words which appear before a noun, and add further definition or specification to the noun, but are not adjectives. The following are determiners:

a/an, the, this, that, these, those, some, any, no, each, every, either, neither, what, whatever, whichever, which, all, both

Swan would also include possessive pronouns, but for this study these are excluded as determiners and classified as pronouns (Column B, number 3). Determiners are coded as follows: "The house is..." the[19]

C. Pronouns (Quirk & Greenbaum, 1973)

1. Personal

These include both subject and object forms:

I, me, we, us, you, he, him, she, her, it, they, them.

2. Reflexive

These are:

myself, ourselves, yourself, yourselves, himself, herself, itself, themselves.

3. Possessive

Possessive pronouns include:
my, mine, our, ours, your, yours, his, her, hers, its, their, theirs.

Please note that words in this form are not coded as being possessive in "E" below.

4. Reciprocal

These are:

each other, one another.

5. Interrogative

These are:

who, whom, whose, what, which.

6. Demonstrative

Demonstrative pronouns include:

this, these, that, those.

7. Universal

These include:

everyone, everybody, each, everything, all, both.

8. Partitive

These are:

anyone, anything, anybody, either, any, no one, nobody, nothing, neither, none.

9. Relative

Relative pronouns include:

who, whom, whose, that, there.

"There" when used in an existential sentence as a "slot-filler" for an indefinite noun phrase (Quirk & Greenbaum, 1973), has been coded as a relative pronoun and Function Word [139]. However, existential "there" is reported as a separate count in the results and was done manually.

D. *Verbs*

1. Auxiliary in finite form

Auxiliaries include:

i. Modals - will, would, should, may, might, can, could, must.
ii. Periphrastic Modals - would rather, be able to, be going to, be about to, have to, have got to, be to, be supposed to, used to, be allowed to, had better.
iii. Have - in perfect verb forms- e.g., He *has* gone.
iv. Be - in progressive or passive verb forms - e.g., He *is* running., The cake *was* eaten.
v. Do - when used as auxiliary form - e.g., *Did* he go?
Note: "Have" , "be" and "do" can all function as main verbs in sentences, whereas modals cannot.

They are classed as Function Words and in the finite form would be associated with the main verb and subject and

could carry tense and aspect. For example, "I am going...", am[1201].

2. Main verb in finite form

This would be a verb associated with a subject which can carry tense, mood, aspect and voice. For example, "I went..." , went[22024].

3. Auxiliary in non-finite form

This is an auxiliary which is in a verb phrase, is not directly associated to a subject, and is not showing tense or aspect. For example, "Having been offended before, John left" , been[1203].

4. Main verb in non-finite form

This is a main verb not associated with a subject and not showing tense or mood. It can still show aspect and voice such as the present participle or the past participle and also includes the infinitive. For example, "I found him working", working[22046].

5. Infinitive with "to"

The infinitive with "to" is coded as one. For example, "I want to go to the store", to go[2205].

6. Bare Infinitive

This is the infinitive or root form of a verb without the "to" included. For example, "He made me go", go[2206].

7. Copula

When the verb "be" is used to describe a state of being and is not in an auxiliary position, then it is coded as the Copula. An example is: "He is 25 years old", is[2207].

8. Imperative

This is the imperative voice or command and the subject is an implied "you". For example, "Please get the police", (you) get[2208].

E. *Inflectional Morphemes (Fromkin & Rodman, 1983)*

A morpheme is generally defined as the basic element of meaning. Inflectional morphemes never change the grammatical form of the word to which they are attached and simply add further grammatical meaning to the word. The coding was straightforward: for example, "The cows went to greener pastures", cows[21001]. The final digit "1" represents the plural "s" morpheme.

F. and G. *Contractions (Jones & Wepman, 1966)*

Contractions are a characteristic of spoken language, particularly as it is speeded up and is less formal in nature. The coding was simply: "I'd like to go" I[231] 'd[120106]. Two columns were needed as only 9 sub-categories per column are permitted with PARSAID and there are ten contractions.

H. *Productive Derivational Morphemes (Fromkin & Rodman, 1983)*

Morphemes of this type can be added to Content Words to create new Content Words with new meaning. Derivational suffixes often change the grammatical class of the word. For example, sing as a verb plus "er" gives us singer as a noun form. These morphemes are called prefix and suffix. An example of the coding would be; "She was unhappy about that", unhappy[24000001].

APPENDIX B: SAMPLE CORPUS

here is you have to uh imitate me. Imitate the music of my voice. I don't care about the sounds. I don't care if you're mispronouncing sounds. I want you to imitate the music of my voice. So you do this. Do it in the privacy of your home somewhere you know where you can uh sound as silly as you think you sound. Okay? So you practise reading this and once you figure you sound close to what I sound like then I want you to record yourself. S Doing the poem by yourself. Okay? I've also recorded for you the first few paragraphs of this. Okay. This is just a newspaper article on medicare. And I want you to do the same thing. Practise reading along with me. Then what I've asked you to do is go ahead and read the next three paragraphs by yourself into the tape. Okay. And they're not paragraphs that I've recorded but that you'll record by yourself. Okay? All the instructions are on the tape too.S Okay? So you can go ahead and do that. First off, before I set you off doing that let's talk a little bit about um English and about stress and rhythm and intonation because you indicated to me that you felt that you were having some problem . That you sounded too choppy. Okay. Uh I have some material here for you. Okay. Have you done any lessons on stress? Word stress and and sentence stress before? S You have hey? So are you familiar with this concept that English is a stress timed rhythm like a piece of music? Kay it has regular beats and depending on the kinds of words that are in a sentence or a passage we try and cram them in between the regular beats. Okay. So its timed like this and however

APPENDIX C: SAMPLE CODED CORPUS

The following sample is printed from PARSAID files.

here[25] is[2207] you[231] have to[1201]imitate[2206] me[231] . Imitate[2208] the[19] music[21] of[16] my[231] voice[21]. I[231] do[1201] n't[15002] care[2202] about[16] the[19] sounds[21001]. I[231] do[1201] n't[15002] care[2202]if[16] you[231] 're[120109] mispronouncing[22026001]sounds[21001]. I[231] want[2202] you[231] to imitate[2205] the[19] music[21] of[16] my[231] voice[21]. So[15] you[231] do[2208] this[236]. Do[2208] it[231] in[16] the[19] privacy[21] of[16] your[233] home[21] somewhere[238] you[231] know[2202] where[28] you[231] can[1201] uh[3] sound[2202] as[18] silly[21] as[18] you[231] think[2202] you[231] sound[2202]. Okay[3] So[15] you[231] practise[2202] reading[22046]this[236] and[17] once[25] you[231] figure[2202] you[231] sound[2202] close to[16] what[139] I[231] sound[2202] like[25] then[18] I[231] want[2202] you[231] to record[2205] yourself[232] . Doing[22046]the[19] poem[21] by[16] yourself[232] . Okay[3] ? I[231] 've[120108] also[25] recorded[22025]for[16] you[231] the[19] first[24] few[24] paragraphs[21001] of[16] this[236]. Okay[3]. This[236] is[2207] just[25] a[19]

newspaper[24] article[21] on[16] medicare[21]. And[17]
I[231] want[2202] you[231] to do[2205] the[19]
same[24] thing[21]. Practise[2208] reading[22046]
along[25] with[16] me[231] . Then[18] what[139]
I[231] 've[120108]asked[22025]you[231]
to do[2205] is[2207] go[2206] and[17] read[2206]
the[19] next[25] three[24] paragraphs[21001] by[16]
yourself[232] into[16] the[19] tape[21].
Okay[3]. And[17] they[231] 're[220709]
not[25] paragraphs[21001] that[139] I[231]
've[120108]recorded[22025]but[17] that[18]
you[231] 'll[120103]record[2202] by[16]
yourself[232] . Okay[3] ? All[24]
the[19] instructions[21001002] are[2207] on[16]
the[19] tape[21] too[25]. Okay[3] ?
So[15] you[231] can[1201] go[2202] ahead[25]
and[17] do[2202] that[236] . First[25] off[25]
before[25] I[231] set[2202] you[231] off[25]
doing[22046]that[236] let[2208] 's[231007]
talk[2204] a[19] little[24] bit[21] about[16]
um[3] English[21] and[17] about[16] stress[21]
and[17] rhythm[21] and[17] intonation[21000002] because[18]
you[231] indicated[22024]to[16] me[231]
that[139]you[231] felt[22024]that[18] you[231]
were[12014]having[22026]some[29] problems[21001].

That[18] you[231] sounded[22024]too[25]
choppy[25000002]. Okay[3]. Uh[3] I[231]
have[2202] some[29] material[21] here[25] for[16]
you[231] . Okay[3]. Have[1201] you[231]
done[22025]any[24] lessons[21001] on[16] stress[21]?
Word[24] stress[21] and[17] and[17] sentence[24]
stress[21] before[25]? You[231] have[2201]
hey[3] ? So[15] are[2207] you[231]
familiar[24] with[16] this[239] concept[21]
that[18] English[21] is[2207] a[19] stress[24]
timed[22045]rhythm[21] like[16] a[19] piece[21]
of[16] music[21]? Okay[3] it[231]
has[2202] regular[24] beats[21001] and[17] depending[22046]
on[16] the[19] kinds[21001] of[16] words[21001] that[18]
are[2207] in[16] a[19] sentence[21] or[17] a[19] passage[21]
we[231] try[2202] and[17] cram[2202] them[231]
in between[16] the[19] regular[24] beats[21001].
Okay[3]. So[15] it[231] 's[220701] timed[22045]
like[16] this[236] and[17] however[18]

APPENDIX D

FREQUENCY OF VOCABULARY FORM

TABLE 1 - GROUP A

	ESL TEACHER											
VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\bar{X}$
A.General Classification												
1.Function Word	96	99	52	50	75	73	46	69	57	80	697	69.7
2.Content Word	225	197	137	64	109	155	117	173	152	220	1549	154.9
3.Interjection	7	14	5	7	4	4	16	5	8	7	77	7.7
B.Grammatical Form												
1.Noun	43	41	24	11	53	40	22	35	33	39	341	34.1
2.Verb	77	72	46	30	44	58	43	63	63	78	574	57.4
3.Pronoun	67	67	44	23	28	43	34	53	31	61	451	45.1
4.Adjective	23	14	19	11	14	24	14	19	15	33	186	18.6
5.Adverb	41	25	18	10	9	15	14	20	26	36	214	21.4

APPENDIX D TABLE 1 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\bar{X}$
6.Preposition	21	29	12	13	18	18	14	23	17	22	187	18.7
7.Co-ordinating Conjunction	15	15	2	4	7	7	4	5	10	15	85	8.5
8.Subordinating Conjunction	2	12	2	5	7	4	5	8	3	8	56	5.6
9.Determiner	32	21	21	7	24	19	13	16	11	8	172	17.2
C.Pronoun Form												
1.Personal	46	49	28	8	20	31	25	31	22	38	298	29.8
2.Reflexive	0	4	0	0	0	1	0	0	0	0	5	.5
3.Possesive	1	1	2	0	1	2	1	7	2	1	18	1.8
4.Reciprocal	0	0	0	0	0	0	0	0	0	0	0	0
5.Interrogative	1	0	3	6	0	0	0	0	1	0	11	1.1
6.Demonstrative	5	7	2	5	3	5	2	7	5	11	52	5.2
7.Universal	6	0	0	0	0	1	0	0	0	1	8	.8

APPENDIX D TABLE 1 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\bar{X}$
8.Partitive	3	1	6	3	0	1	5	6	0	4	29	2.9
9.Relative	6	5	3	1	3	3	1	2	1	6	31	3.1
D.Verb Form												
1.Auxiliary in Finite	15	13	4	6	9	18	11	16	12	16	120	12.0
2.Main Verb in Finite	32	30	18	12	17	25	18	27	23	29	231	23.1
3.Auxiliary in Non-Finite	0	0	0	5	3	0	0	0	0	0	8	.8
4.Main Verb in Non-Finite	5	8	4	3	5	0	5	4	11	3	48	4.8
5.Infinitve with "to"	4	4	4	1	1	5	1	6	9	10	45	4.5
6.Bare Infinitive	1	2	2	1	0	1	0	1	0	2	11	1.1
7.Copula	11	10	10	1	9	8	8	8	7	12	83	8.3
8.Imperative	8	4	4	1	0	1	0	1	1	5	26	2.6

APPENDIX D TABLE 1 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
E.Inflectional Morphemes												
1.Plural	5	4	4	2	0	8	9	6	5	27	78	7.8
2.Possessive	2	0	0	0	0	2	0	1	2	0	9	.9
3.3rd Person Singular	2	3	3	2	3	2	1	0	1	0	14	1.4
4.Past Tense	5	3	3	0	7	1	0	2	2	4	28	2.8
5.Past Participle	3	0	0	0	0	1	3	0	3	1	17	1.7
6."ing" Form	5	3	3	1	2	2	4	6	8	5	43	4.3
7.Comparative	0	0	0	0	1	0	0	0	0	0	1	.1
8.Superlative	0	0	0	0	0	0	0	0	0	0	0	0
F. & G. Contractions												
1.is - 's	4	6	6	1	3	3	4	1	2	5	30	3.0
2.not - n't	2	0	0	1	3	3	1	1	2	4	17	1.7

APPENDIX D TABLE 1 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
3.will - 'll	4	0	0	1	1	0	0	4	0	7	18	1.8
4.am - 'm	1	0	0	0	0	0	0	0	0	0	1	.1
5.has - 's	0	0	0	0	0	0	0	0	0	0	0	0
6.would - 'd	0	0	0	0	0	0	0	0	0	0	0	0
7.had - 'd	0	0	0	0	0	0	0	0	0	0	0	0
8.have - 've	0	0	0	0	0	0	0	0	0	1	1	.1
9.are - 're	0	1	1	1	1	0	2	4	3	3	17	1.7
10.us - 's	4	3	3	1	0	0	0	0	0	1	9	.9
H.Derivational Morphemes												
1.Prefix	1	3	3	0	2	0	0	0	0	0	7	.7
2.Suffix	8	0	0	2	0	2	4	4	2	4	29	2.9
3.Prefix and Suffix	0	0	0	0	0	0	0	0	0	0	0	0

APPENDIX D TABLE 1 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
Total Words	337	317	201	123	189	240	188	256	232	318	2401	240.1
Total Parsed Words	322	310	192	121	185	224	177	246	217	301	2296	229.6
Average Frequency	5909.22	8057.92	8715.0	7257.8	11747.72	7050.04	6803.33	5335.46	6707.11	5110.4	***	7269.4

APPENDIX D
TABLE 2 - GROUP B

FREQUENCY OF VOCABULARY FORM

	NON-ESL TEACHER											
VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\bar{X}$
A.General Classification												
1.Function Word	113	138	9	114	134	113	107	99	120	105	1052	105.2
2.Content Word	244	235	139	173	245	198	191	146	210	207	1988	198.8
3.Interjection	7	8	6	12	4	9	6	5	6	14	77	7.7
B.Grammatical Form												
1.Noun	44	58	31	56	73	57	52	41	60	45	517	51.7
2.Verb	85	78	50	58	86	81	78	66	97	69	748	74.8
3.Pronoun	85	64	33	41	51	50	44	37	54	60	519	51.9
4.Adjective	35	38	21	20	32	19	23	15	23	25	251	25.1
5.Adverb	26	34	23	24	31	19	17	25	18	30	247	24.7

APPENDIX D TABLE 2 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
6.Preposition	20	41	18	40	44	30	34	25	44	33	329	32.9
7.Co-ordinating Conjunction	20	16	5	7	19	15	12	6	13	14	127	12.7
8.Subordinating Conjunction	7	9	7	8	11	11	7	3	5	5	73	7.3
9.Determiner	29	35	24	33	32	28	30	27	35	31	304	30.4
C.Pronoun Form												
1.Personal	66	31	17	25	35	34	27	19	39	32	325	32.5
2.Reflexive	0	0	0	0	0	0	0	1	1	1	3	.3
3.Possesive	3	10	3	2	1	4	2	2	3	2	32	3.2
4.Reciprocal	1	0	0	0	0	0	0	0	0	0	1	.1
5.Interrogative	2	0	1	2	1	0	0	5	0	4	15	1.5
6.Demonstrative	3	8	3	1	4	3	5	2	4	10	43	4.3
7.Universal	1	0	2	1	0	0	1	1	0	1	7	.7

APPENDIX D TABLE 2 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\bar{X}$
8.Partitive	1	5	5	3	4	3	2	1	3	5	32	3.2
9.Relative	9	10	2	7	6	6	7	6	4	6	63	6.3
D.Verb Form												
1.Auxiliary in Finite	12	15	14	14	16	16	14	22	15	10	148	14.8
2.Main Verb in Finite	47	31	21	25	33	35	20	26	39	29	306	30.6
3.Auxiliary in Non-Finite	0	0	0	0	0	0	0	0	0	0	0	0
4.Main Verb in Non-Finite	4	5	4	4	6	7	15	4	6	8	63	6.3
5.Infinitve with "to"	8	5	4	4	15	8	9	3	3	5	64	6.4
6.Bare Infinitive	1	0	0	2	0	0	3	2	0	0	8	.8
7.Copula	10	22	3	9	13	15	16	7	13	16	114	11.4
8.Imperative	2	0	4	0	3	0	3	2	1	0	15	1.5

APPENDIX D TABLE 2 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
E.Inflectional Morphemes												
1.Plural	8	11	5	15	17	24	10	4	11	9	114	11.4
2.Possessive	0	1	0	0	0	0	4	0	0	0	5	.5
3.3rd Person Singular	0	3	2	6	3	1	3	5	4	7	34	3.4
4.Past Tense	18	13	0	1	6	1	6	3	5	2	55	5.5
5.Past Participle	4	2	5	5	0	5	11	2	3	4	41	4.1
6."ing" Form	9	7	4	4	9	5	11	16	8	7	80	8.0
7.Comparative	0	0	0	0	2	0	1	0	0	0	3	.3
8.Superlative	0	0	0	0	0	0	0	0	0	0	0	0
F. & G. Contractions												
1.is - 's	5	7	1	9	2	4	5	12	4	15	64	6.4
2.not - n't	0	4	0	2	0	6	2	2	0	2	18	1.8

APPENDIX D TABLE 2 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
3.will - 'll	0	1	0	0	0	3	0	0	0	4	8	.8
4.am - 'm	0	3	0	1	0	0	0	0	1	0	5	.5
5.has - 's	0	0	0	0	0	0	1	3	1	0	5	.5
6.would - 'd	0	0	0	0	2	0	0	0	0	0	2	.2
7.had - 'd	0	0	0	0	0	0	0	0	0	0	0	0
8.have - 've	0	0	0	1	0	0	1	0	0	0	2	.2
9.are - 're	3	0	0	1	3	1	3	0	1	0	12	1.2
10.us - 's	0	0	0	0	0	0	0	1	0	0	1	.1
H.Derivational Morphemes												
1.Prefix	1	0	0	0	0	0	0	0	0	0	1	.1
2.Suffix	2	9	3	12	5	7	12	1	6	6	63	6.3
3.Prefix and Suffix	0	1	0	0	0	0	1	0	0	0	2	.2

APPENDIX D TABLE 2 (cont)

VOCABULARY FORM	1	2	3	4	5	6	7	8	9	10	TOT	$\overline{X}$
Total Words	379	398	231	303	410	339	321	250	348	335	3314	331.4
Total Parsed Words	364	378	215	298	383	318	302	249	335	326	3169	316.9
Average Frequency	8771.91	9370.96	9301.56	8775.93	8590.7	7960.13	8183.63	9433.13	9544.74	8380.25	***	8831.29

APPENDIX E FREQUENCY OF OCCURRENCE PER MILLION

WORD TYPE	FREQUENCY/MILLION	RANK
the	73123	1
of	28462	2
and	26172	3
a	24442	4
to	23653	5
in	19366	6
is	11643	7
you	9573	8
that	9266	9
it	9179	10
he	8245	11
for	7687	12
was	7457	13
on	7136	14
are	6743	15
as	6296	16
with	5933	17
his	5316	18
they	5285	19
at	4678	20
be	4554	21
this	4517	22
from	4456	23
I	4405	24
have	4346	25
or	3992	26
by	3924	27
one	3908	28
had	3634	29
not	3631	30
but	3620	31
what	3350	32
all	3280	33
were	3200	34
when	3103	35
we	3067	36
there	2917	37
can	2878	38
an	2814	39

WORD TYPE	FREQUENCY PER MILLION	RANK
your	2764	40
which	2678	41
their	2530	42
said	2469	43
if	2461	44
do	2440	45
will	2438	46
each	2426	47
about	2423	48
how	2413	49
up	2409	50
out	2335	51
them	2307	52
then	2306	53
she	2275	54
many	2263	55
some	2234	56
so	2227	57
these	2149	58
would	2138	59
other	2073	60
into	2062	61
has	2003	62
more	1949	63
her	1877	64
two	1876	65
like	1868	66
him	1762	67
see	1634	68
time	1634	69
could	1607	70
no	1606	71
make	1538	72
than	1530	73
first	1478	74
been	1447	75
its	1411	76
who	1372	77
now	1371	78
people	1344	79
my	1328	80

WORD TYPE	FREQUENCY PER MILLION	RANK
made	1326	81
over	1314	82
did	1307	83
down	1302	84
only	1280	85
way	1279	86
find	1229	87
use	1219	88
may	1217	89
water	1207	90
long	1176	91
little	1150	92
very	1149	93
after	1143	94
words	1124	95
called	1110	96
just	1105	97
where	1082	98
most	1078	99
know	1072	100

From *Word Frequency Book* (Carroll, Davies & Richman, 1971)

APPENDIX F FREQUENCY OF OCCURRENCE PER MILLION

TABLE 1 - GROUP A, ESL TEACHERS

	RANGE				
TEACHER	3351 - 73,123	1604 - 3350	1072 - 1600	11 - 1071	0 - 10
1	112	63	13	104	30
2	130	35	7	110	28
3	66	25	8	73	20
4	36	29	6	41	9
5	84	37	4	54	6
6	85	30	9	94	6
7	68	18	7	53	31
8	76	50	6	86	28
9	59	38	18	89	16
10	87	60	12	121	21
Mean	80	39	9	83	20

APPENDIX F FREQUENCY OF OCCURRENCE PER MILLION

TABLE 2 - GROUP B, NON-ESL TEACHERS

	RANGE				
TEACHER	3351 - 73,123	1604 - 3350	1072 - 1600	11 - 1071	0 - 10
1	151	58	20	117	18
2	159	55	19	122	23
3	80	28	16	75	16
4	111	31	10	100	46
5	149	147	32	111	41
6	116	56	21	101	24
7	124	35	13	95	65
8	94	36	11	92	16
9	146	33	21	109	26
10	127	49	15	108	27
Mean	126	42	18	103	28