

The Language of Oral Presentations Given by PhD Researchers in an EAP class: Level of
Performance and Disciplinary Differences

by

Ricardo Nausa

A thesis submitted to the University of Birmingham for the degree of

DOCTOR IN PHILOSOPHY

Department of English Language and Applied Linguistics
School of English, Drama and American & Canadian Studies
University of Birmingham

August 2018

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis explores the language of 88 Colombian PhD researchers' oral presentations (OPs) in an EAP class, in five studies on their strategies to engage the audience and clarify content for them. In a parallel corpus of 88 pairs of OP transcriptions and essays (n=128228 tokens), corpus and statistical significance procedures identify features that discriminate among students' levels of oral achievement and disciplines. These features are: gestural-verbal deixis (chapter 4), audience (ch.5) and impersonal identity projection (ch.6), code glosses (ch.7), and transformation of written into oral content (ch.8). Features analyses include distribution across the levels and disciplines subcorpora, recurrent patterns, discourse functions, and pragmatic appropriacy and grammatical variety. The studies reveal that levels differ in the way that presenters mark stance authorship, anticipate the audience need for help, and vary their strategies grammatically. Disciplinary differences re-present the ways in which disciplines (re)produce knowledge. Hard-field OPs' focus on research methods and outcomes is observed in presenters' interaction with images, academic identity projection, and technical terms explanation. Soft-field OPs' focus on interpretations is observed in the opinions towards existing knowledge or the use of folk examples for the audience. Language choices also reveal the non-expert character of the audience. This thesis contributes to the study of oral academic genres by demonstrating the importance of multimodal, across modes (vs written), non-deficiency analyses; confirming disciplinary differences; and proposing ways of understanding levels of achievement not based solely on grammatical accuracy but on the pragmatic success of the features used.

Key words: *Oral presentations, disciplinary differences, levels of achievement, English for Academic Purposes, deixis, identity projection, audience, code glosses, written to spoken discourse transition*

This thesis is dedicated to the memory of José Felipe Pardo
Maestro, después de tantos años, sigo aprendiendo de usted.

Acknowledgements

This is supposedly the easiest part to write in a thesis, but it is not. Lots of names and moments come to my mind, but the most important thing is the feeling of gratitude in my heart. I want to start by thanking Susan Hunston, my supervisor. Susan, I really thank you for all your help and support in these four years. From the very first Skype supervision meeting, I felt that I was going through the most amazing learning experience I have ever had. Not only did I feel honoured to have the chance to learn from such an accomplished and influential scholar, but I also was so lucky to have gotten to know you as a superb human being who was always warm and supportive. The saddest thing about finishing the PhD is that our Skype meetings that I love so much will be over.

Besides my supervisor, I would like to thank Dr Nick Groom and Dr Hilary Nesi for their comments on my work in the first two modules and the final thesis. Nick, I am thankful for the very specific comments; as I said to you in the Viva, the comments in module 1 were crucial to have my research published as two articles, and I am sure that your feedback in module 2 and the thesis chapters will also be of great help for my publication plans.

I also want to thank Universidad de los Andes for providing me with the funding and resources to achieve the goals of the thesis. My gratitude goes out to Dr Hugo Ramírez, my boss and head of the Languages and Culture department, for his help in making it possible for me to complete my studies at the same time I was working full time in the department.

Also, thank you, Gerriet Janssen my friend and colleague, for pushing me and insisting on my following my PhD studies. Our work in the IPD program and your own PhD were two of my main motivations to start this incredible journey.

I am also grateful to my colleagues and friends Ana Milena and Nicole Bruskewitz for their help in the validation procedures in three chapters of this thesis and the second module.

I owe thanks to my former IPD2 students-PhD researchers who gave their consent to participate in this study. As I always said, the great thing about teaching in IPD is that we teachers also get to learn a lot from your research. Doing this thesis, I learned even more.

With a special mention to my former students and research assistants Alejandra Bustos Cabrera, Paola Herreño, and Luciano Mejía Peroni for their help in the video transcription process and corpus compilation. Without a doubt, you were of great help in the not so gratifying and probably least interesting part of the research process. The nicest part came when all the corpus was completed. 😊

Last but not least (a phrase that nobody has ever used to express gratitude), I want to share this happy moment with my two children, Juana and Felipe, and my girlfriend, Diana Mendoza. Thank you for being there and bearing with me in my best and worst moments.

List of Contents

Abstract.....	2
Acknowledgements.....	4
List of Contents.....	6
List of Tables	12
List of Figures	15
List of abbreviations	17
The Language of Oral Presentations Given by PhD Researchers in an EAP class: Level of Performance and Disciplinary Differences	20
1. CHAPTER 1	20
INTRODUCTION	20
1.1. Rationale.....	20
1.1.1. Background: The challenges of public speaking for PhD students and scholars	20
1.1.2. Creation of an EAP program to face private university internationalization challenges	21
1.1.3. IPD2 and its OPs challenges	22
1.2. Questions and general aim.....	24
1.3. Summary of the chapters	25
2. CHAPTER 2	27
GENERAL REVIEW OF THE LITERATURE.....	27
2.1. The study of (English) academic discourse.....	27
2.2. Academic discourse analysis as genre analysis	28
2.3. Oral academic discourse.....	30
2.4. Oral presentations (OPs).....	34
2.4.1. Multiple terms, one genre.....	34
2.4.2. The study of OPs as a student genre	35
2.5. Conclusion.....	39
3. CHAPTER 3	40
METHODOLOGY	40
3.1. Introduction	40
3.2. Participants	40
3.3. The corpus	43
3.3.1. The task	43
3.3.2. Transcription, anonymization, and indexing.....	44
3.3.3. Corpus design.....	44

3.3.4.	Subcorpora selection and modification	47
3.4.	Quantitative analyses	48
3.4.1.	Raw and normalised frequencies, and percentages	48
3.4.2.	Significance and effect size analysis: hypothesis testing	49
3.5.	Qualitative (discourse analysis) analyses	51
3.5.1.	Ad hoc, eclectic, theoretical perspective	51
3.5.2.	Alternative wordings in discourse analyses	52
3.5.3.	Validation procedures	53
3.6.	Software	54
PART I: ENGAGING THE AUDIENCE		55
4.	CHAPTER 4	56
<i>AS WE CAN SEE HERE...: SHOWING THINGS TO THE AUDIENCE</i>		56
4.1.	Pointing behaviours (deixis) in Colombian PhD researchers' OPs	56
4.2.	Verbal deixis (pointing with words)	57
4.2.1.	Reference to preceding or subsequent items in the discourse (<i>anaphora and cataphora</i>) ...	58
4.2.2.	Reference to an item in the physical context of the discourse (<i>deixis</i>)	59
4.2.3.	Reference to an item in the symbolic or virtual context of the discourse (<i>symbolic deixis</i>) .	59
4.3.	Gestural spatial deixis (pointing with gestures)	61
4.4.	Synchronization of verbal and gestural spatial deixis, and integration of images to talk (the deictic process)	62
4.5.	Multimodal analysis of oral academic discourse	64
4.5.1.	Analysis of spatial deixis	67
4.6.	Methods and statistical analyses	70
4.6.1.	Selection of a subcorpus and data clean-up	71
4.6.2.	Statistical significance analysis	74
4.7.	How Colombian PhD researchers perform deixis	75
4.7.1.	<i>This</i> (hard science)	76
4.7.2.	<i>This</i> (soft science)	79
4.7.3.	<i>These</i> (hard-science)	82
4.7.4.	<i>These</i> (soft science)	84
4.7.5.	<i>Here</i> (hard science)	87
4.7.6.	<i>Here</i> (soft science)	89
4.8.	Conclusion	92
5.	CHAPTER 5	95
<i>YOU: GIVING AN IDENTITY TO THE AUDIENCE</i>		95

5.1.	Introduction	95
5.2.	Self-mentions and the projection of identity roles in OPs.....	96
5.3.	The study of <i>you</i> in oral academic discourse: review of the literature	100
5.3.1.	<i>You</i> in the territory of information and dynamic-inferential view of communication models 101	
5.3.2.	<i>You</i> in the study of oral academic discourses.....	104
5.4.	This chapter: <i>you</i> and the projection of audience identity roles	111
5.4.1.	Model for the analysis of <i>you</i> to assign identity roles to the OP audience.....	113
5.5.	Method and quantitative analyses.....	121
5.5.1.	Selection of sentences for analysis and data clean-up.....	121
5.5.2.	Frequency and statistical significance of <i>you</i> in the level of achievement subcorpus	122
5.5.3.	Frequency and statistical significance of <i>you</i> in the disciplinary divide corpus	125
5.6.	Discourse analyses.....	127
5.6.1.	<i>You</i> -projections in the level of achievement divide	127
5.6.2.	<i>You</i> -projections across the disciplinary divide.....	132
5.7.	Conclusion.....	142
6.	CHAPTER 6	147
	<i>INFORMATION SHOULD BE ACCESSIBLE...: IMPERSONAL PROJECTION OF IDENTITIES</i> 147	
6.1.	Introduction: from personal to impersonal modalized projection of academic identities in OPs 147	
6.2.	Identity projection with impersonal modalized expressions: model of analysis	150
6.2.1.	Modality	150
6.2.2.	Identity projection	157
6.3.	The study of impersonal identity projection in OPs	164
6.4.	Methods.....	166
6.4.1.	Selection of sentences for analysis and data clean-up.....	166
6.4.2.	Quantitative and discourse analyses.....	167
6.5.	Quantitative analyses.....	168
6.5.1.	Modals, adverbs, and modal-like realisations in the impersonal expression of identity roles 168	
6.5.2.	Identity roles: significance of frequency differences and effect size in the level of achievement and disciplinary sub-corpora.....	170
6.6.	Discourse analyses: impersonal academic identity projection	175
6.6.1.	By level of achievement.....	175
6.6.2.	By discipline.....	183
6.7.	Conclusion.....	187

PART II: MAKING CONTENT EASY FOR THE AUDIENCE	192
7. CHAPTER 7	193
<i>IN THIS CASE...: INCLUSION OF CODE GLOSSES</i>	193
7.1. Introduction: code glosses as a mark of high oral achievement	193
7.2. Code glosses: an interactional and interactive type of metadiscourse.....	194
7.3. Types of code glosses	195
7.3.1. Reformulations.....	195
7.3.2. Exemplifications	197
7.4. Previous research on code glosses in academic oral discourse	199
7.5. Methods.....	201
7.5.1. Update of the list of code gloss markers	201
7.5.2. Selection of sentences for analysis and data clean-up.....	202
7.6. Quantitative analyses.....	206
7.6.1. Code glosses: reformulations and exemplifications.....	206
7.6.2. Reformulations: explanations, implications, paraphrases, and specifications	208
7.6.3. Exemplifications: category instantiations, similar cases, and rule instantiations.....	209
7.6.4. Analysis by CG markers: frequency and significance analysis.....	212
7.7. Discourse analysis of code glosses.....	214
7.7.1. Explanations.....	214
7.7.2. Specifications	216
7.7.3. Category instantiations.....	218
7.8. Conclusion.....	220
8. CHAPTER 8	223
<i>WE NEED EDUCATION...: TRANSFORMING WRITTEN INTO ORAL CONTENT</i>	223
8.1. Introduction: a follow-up study on the mechanisms to transition from written to oral discourse 223	
8.1.1. Questions in the follow-up study	224
8.2. Mechanisms of change to transition from written to oral discourse: main concepts and review of available research	225
8.2.1. Clause structure changes and information structure.....	225
8.2.2. Noun phrases: heavy modification, nominalisation, denominalisation, and grammatical metaphors.....	229
8.2.3. Modality	235
8.2.4. Code glosses.....	236
8.3. Methods.....	238

8.3.1. Selection of subcorpora for analysis	238
8.3.2. Identification and analysis of sample sentences.....	239
8.4. Quantitative analyses.....	240
8.4.1. By level of achievement.....	241
8.4.2. By discipline.....	244
8.5. Discourse analyses.....	247
8.5.1. By level of achievement.....	248
8.5.2. By discipline.....	256
8.6. Conclusion.....	268
9. CHAPTER 9	272
CONCLUSION.....	272
9.1. Summary of findings	272
9.1.1. Engaging the audience	273
9.1.2. Making content easy for the audience.....	275
9.2. Implications	276
9.2.1. Methodological	276
9.2.2. Theoretical.....	279
9.2.3. Pedagogical	281
9.3. Limitations and further research.....	283
9.4. Concluding remarks.....	286
References.....	287
Appendixes	328
Appendix A: RUBRIC TO EVALUATE ESSAYS.....	328
Appendix B. RUBRIC TO EVALUATE ORAL PRESENTATIONS	329
Appendix C. UNIVERSIDAD DE LOS ANDES ETHICS RESEARCH APPROVAL LETTER	330
Appendix D. GENERAL TRANSCRIPTION CONVENTIONS	331
Appendix E. LIST OF ESSAY/OPs PARALLEL TEXTS.....	332
Appendix F. MULTIMODAL ANALYSIS TRANSCRIPTION CONVENTIONS (taken and adapted from Rendle-Short, 2006)	335
Appendix G. LOG LIKE LIKELIHOOD AND BIC (Bayes Factor Approximation) TESTS EXCEL WORKSHEETS (Rayson, 2017)	337
10. Appendix H. TABLES FROM PERSONAL PRONOUNS STUDY (Nausa, 2016).....	338
Appendix I. KEYWORD LIST IN OPs SUBCORPUS / REFERENCE CORPUS: BROWN (Francis & Kučera, 1964).....	340
Appendix J. CONCORDANCE SEARCH FOR DEICTICS	341

Appendix K. CONCORDANCE SEARCH FOR <i>YOU</i> -IDENTITY CASES	342
Appendix L. CONCORDANCE SEARCH FOR ADJ+THAT PATTERNS	343
Appendix M. CONCORDANCE SEARCH FOR CODE GLOSSES (<i>IN THIS CASE</i>).....	344
Appendix N. MANUAL IDENTIFICATION OF PAIRS OF ESSAY-OPs SENTENCES	345
Appendix O. ANALYSIS AND QUANTIFICATION OF MECHANISMS IN WRITTEN-ORAL PAIRS OF SENTENCES	346

List of Tables

Table 3.1. Corpus of Colombian PhD researchers essays and OPs by level of achievement and disciplines	46
Table 3.2. Updates to Corpus of Colombian PhD researchers essays and OPs	47
Table 3.3. Subcorpora selected for each chapter.....	47
Table 3.4. Significance of LL values based on (Rayson, 2017).....	49
Table 3.5. Approximate Bayes Factor (BIC)	50
Table 3.6. Language and semiotic features analysed in each chapter.....	52
Table 3.7. Krippendorff 's alpha values for interrater reliability	53
Table 4.1. Use of spatial deictics by level of achievement: raw and normalized (N) frequencies	72
Table 4.2. Use of spatial deictics by discipline: raw and normalized (N) frequencies	72
Table 4.3. Spatial deictics expressed in raw and normalised (N) frequencies by level of achievement	73
Table 4.4. Spatial deictics expressed in raw and normalised (N) frequencies by discipline	73
Table 4.5. Significance analyses of spatial deictics use by levels of achievement.....	74
Table 4.6. Significance analyses of spatial deictics use by discipline	74
Table 4.7. Summary of findings deixis study	94
Table 5.1. Raw (R) and normalised (N) frequencies, and percentages (%) of you audience roles realisations by level of achievement.....	123
Table 5.2. Log likelihood and Bayes Factor (BIC) values by levels of achievement.....	124
Table 5.3. Raw (R) and normalised (N) frequencies, and percentages (%) of "you" audience roles realisations by discipline.....	125
Table 5.4. Log likelihood and Bayes Factor (BIC) values by Discipline	126

Table 5.5. Summary of findings you-audience study	146
Table 6.1. Modals and modal-like realisations, and identity roles expressed in raw frequencies and percentages	169
Table 6.2. Raw frequency and significance analysis of impersonal modalized academic identity role realisations by level of achievement.....	171
Table 6.3. Raw frequency and significance analysis of impersonal modalized academic identity role realisations by discipline.....	173
Table 6.4. Modal and modal-like raw and normalised (N) frequencies in the expression of the opinion holder role.....	176
Table 6.5. Modal and modal-like raw and normalised (N) frequencies in the expression of the representative role.....	180
Table 6.6. Modal and modal-like raw and normalised frequencies in the expression of the recounter/announcer role	183
Table 6.7. Modal and modal-like raw and normalised frequencies in the expression of the <i>opinion holder</i> role.....	186
Table 6.8. Summary of findings impersonal modality study	191
Table 7.1. Code glosses use expressed in raw frequencies and percentages by level of achievement	206
Table 7.2. Code glosses use expressed in raw frequencies and percentages by discipline.....	206
Table 7.3. Significance analyses of code glosses use by level of achievement	207
Table 7.4. Significance analyses of code glosses use by discipline.....	207
Table 7.5. Reformulations use expressed in raw frequencies and percentages by level of achievement	208

Table 7.6. Reformulations use expressed in raw frequencies and percentages by discipline	208
Table 7.7. Significance analyses of reformulations use by level of achievement.....	209
Table 7.8. Significance analyses of reformulations use by discipline	209
Table 7.9. Exemplifications use expressed in raw frequencies and percentages by level of achievement	210
Table 7.10. Exemplifications use expressed in raw frequencies and percentages by discipline.	210
Table 7.11. Significance analyses of exemplifications use by level of achievement.....	210
Table 7.12. Significance analyses of exemplifications use by discipline	211
Table 7.13. Code gloss markers expressed in raw frequencies and Log likelihood (LL) and Bayes factor (BIC) values by level of achievement and discipline	213
Table 7.14: Summary of findings code glosses study.....	222
Table 8.1. syntactic modifications raw (R) and normalised (N) frequencies, and percentages (%) by level of achievement	242
Table 8.2. Syntactic changes raw frequency and significance analysis by level of achievement	244
Table 8.3. Syntactic modifications raw (R) and normalised (N) frequencies, and percentages (%) by discipline	244
Table 8.4. Syntactic changes raw frequency and significance analysis by discipline	246
Table 8.5: Summary of findings in written to oral transition study	271
Table 10.1.Raw (R) and normalised (N) frequencies, and percentages (%) of self-mention realizations by pronoun number.....	338
Table 10.2. Log likelihood and Bayes Factor (BIC) values by levels of achievement.....	339
Table 10.3. Log likelihood and Bayes Factor (BIC) values by disciplines.....	339

List of Figures

Figure 1.1. Questions, OPs' aspects, and features for analysis in the thesis	25
Figure 2.1. Taxonomy of academic genres (Hyland, 2009).....	29
Figure 2.2. Giménez' (2000) taxonomy of oral academic genres	31
Figure 2.3. Fortanet's (2005, p. 32) classification of oral academic genres by purpose	32
Figure 2.4. Survey of literature on non-linguistic aspects of academic OPs	38
Figure 3.1. Distribution of students per disciplines and programs	41
Figure 3.2. Distribution of students by discipline and level	42
Figure 3.3. Information included in text codes	44
Figure 4.1. Coordination of verbal and gestural deixis in as we can see "here"	62
Figure 4.2. Use of "this picture" in a hard-field OP	77
Figure 4.3. Use of "this picture" in soft-field OP	80
Figure 4.4. Use of "these" in a hard-field OP	83
Figure 4.5. Use of "here" in a soft-field OP	85
Figure 4.6. Use of "here" in a hard-field OP	88
Figure 4.7. Use of "here" in a soft-field OP	90
Figure 5.1. Typology of identities behind the first person pronoun (Nausa, 2016)	99
Figure 5.2. Kamio's representation of territory of information (Kamio, 2001, p. 1113) adapted to OPs.....	102
Figure 5.3. Meanings of "you" in OPs adapted from Gast et al. (2015).....	103
Figure 5.4. Meanings of "you" based on the territory of information (Kamio, 2011) and dynamic view of communication (Gas et al., 2005) theories	104
Figure 5.5. Identity projection dyads	114

Figure 6.1. Modalization taxonomy based on Halliday and Matthiessen's (2014) and Quirk et al's., (1985) taxonomies	152
Figure 6.2. Tang and John's (1999, p.S29) typology of identities behind the first person pronoun in the light of proximity and positioning processes (Hyland, 2012).....	162
Figure 6.3. Reinterpretation of identity typology based on the expansion of the representative role	162
Figure 6.4. Nature of knowledge in the hard and soft disciplines (adapted from Becher & Trowler, 2001, p. 36).....	174
Figure 7.1. A taxonomy of code glosses and expressions to introduce them (adapted from Hyland, 2007)	199
Figure 7.2. Provisional list of code gloss markers in OPs	202
<i>Figure 7.3. Definite list of code gloss markers in OPs</i>	<i>205</i>
Figure 8.1. Taxonomy of changes in the transition from written to oral discourse	237
Figure 8.2. distribution of syntactic modifications in the e-op subcorpus.....	241
Figure 8.3. distribution of syntactic modifications in level of achievement subcorpora.....	242
Figure 8.4. Distribution of the four types of syntactic modifications by level of achievement..	243
Figure 8.5. Distribution of syntactic modifications in disciplines subcorpora	245
Figure 8.6. Distribution of the four types of syntactic modifications by discipline	246
Figure 8.7. Distribution of heavy NPs reduction mechanisms by level of achievement	248
Figure 8.8. Distribution of modality change mechanisms by level of achievement.....	253
Figure 8.9. Distribution of clause structure change mechanisms by discipline.....	257
Figure 8.10. Distribution of code glosses by discipline.....	261

List of abbreviations

ADMI: administración (management)

ANTR: antropología (anthropology)

AOP: academic oral presentations

BASE: The British Academic Spoken English

BIC: Bayesian Information Criterion

CA: Conversation Analysis

CBIO: Ciencias biológicas (biological sciences)

CEFR: Common European Framework of Reference

COCA: Corpus of Contemporary American English

CP: conference presentations

CPOL: ciencia política (political science)

CQUI: Ciencia química (chemical science)

DERE: derecho (law)

EAP: English for Academic Purposes

e-op: essay-oral presentation pair

ECON: economía (economics)

EDUC: educación (education)

EFL: English as a foreign language

ERPP: English for research publication purposes

FILO: filosofía (philosophy)

FISI: física (physics)

HIST: historia (history)

INGE: ingeniería (engineering)

IMRD: introduction, methods, results, and discussion

IPD: Inglés Para Doctorados (English for PhD students)

ITAs: international teaching assistants

L1: First language

L2: Second language

LCR: learner corpus research

LITE: literatura (literature)

LL: log likelihood

MATE: matemáticas (mathematics)

MDA: Multimodal Discourse Analysis

MICASE: Michigan Corpus of Academic Spoken English

NNS: Non-native speaker (of English)

NS: native speaker (of English)

OAP: oral academic presentations

OP: Oral presentation

OPC: oral presentation competence

p : probability value

PhD: Philosophiæ doctor

POS: parts of speech

PPP: prosody, paralinguistics, and paraphernalia

PSIC: psicología (psychology)

Q&A: questions and answers section

SFL: Systemic Functional Linguistics

TAs: Teacher assistants

TEFL: Teaching English as a Foreign Language

TESOL: Teaching English to Speakers of other Languages

The Language of Oral Presentations Given by PhD Researchers in an EAP class: Level of
Performance and Disciplinary Differences

CHAPTER 1

INTRODUCTION

This thesis examines oral presentations (OPs), a student academic oral genre common in EAP courses that helps PhD researchers to prepare for public speaking in academic conferences. The thesis contains five studies that focus on the language used in OPs by a group of Colombian PhD researchers enrolled in an EAP course. The purpose is to identify differences among this student population that correlate with their levels of oral achievement and the disciplines that they are enrolled in as observed in the English language features that they use to engage the audience and make content easy for them.

1.1. Rationale

1.1.1. Background: The challenges of public speaking for PhD students and scholars

Speaking in public is worse than death for many (Burgess, 2013). This phobia can be so terrifying that there is a plethora of self-help literature on how to overcome it (e.g. Esposito, 2000) and therapeutic approaches to deal with it like cognitive behavioural (Anderson, Zimand, Hodges, & Rothbaum, 2005), exposure (Cuncic, 2017) or virtual reality therapy (North, North, & Coble, 2015; Pertaub, Slater, & Barker, 2002; Slater, Pertaub, Barker, & Clark, 2002). Speaking in front of an audience can be even more challenging when it is done in a language different to one's own mother tongue. Apart from making the talk intelligible for the audience, presenters are expected to have a well-organized, clear, easy to follow, and interesting delivery. Also, they

should be able to attract and retain the audience attention, as well as interact, engage with, and deal with their questions and comments. All these things must be done in real time, in a short time frame, and with limited possibilities to change those things that go wrong.

Many PhD students and scholars around the world must face these OP challenges. In addition to the publications demands of their institutions, scholars are expected to attend conferences to share their academic production with members of the international communities they are (want to be) part of, which can be a frightening experience for some. Nonetheless, overcoming these challenges is worth the effort, for speaking at a conference can bring important strategic benefits such as gaining visibility, informing academic communities of research findings, interacting to receive feedback, establishing new contacts, collaborating with research groups, or getting funding for further research (Wallwork, 2010).

1.1.2. Creation of an EAP program to face private university internationalization challenges

These conference participation benefits are conceived in the development goals of a private university in Colombia (Universidad de los Andes, 2016). These goals include strategic internationalization for the university and the country (p.9) through

- high-quality, innovative, flexible, and interdisciplinary undergraduate and graduate programs (p. 14);
- a strengthened presence of English in the programs (p. 14),
- the incremental improvement in the quality and quantity of research production (p.15), and

- the establishment of programs and projects with high quality international organizations to create research and collaboration networks that incorporate faculty, alumni, and students (p.18).

In their plans to achieve these goals, several departments approached Departamento de Lenguas y Cultura to request a program that would help PhD students develop strategies for success in the writing of articles for publication and speaking for professional presentations (Janssen, Ángel, & Nausa, 2011). The result of this petition was the creation of an EAP program: programa IPD (inglés para doctorados). Among the procedures in IPD development were the administration of needs analysis surveys, interviews to PhD programs directors, and questionnaires for PhD candidates (Janssen, Nausa, & Rico, 2012). In the interviews and surveys, two findings were crucial in the definition of the program. First, the study of English for this population decreased as they advanced in their educational development (p. 54). Second, academic presentations in English speaking contexts were perceived by students and stakeholders as more important than other skills (e.g. reading) (p. 58), which is in stark contrast with the reported tendency for public speaking to be addressed empirically in academic settings (Mauranen, 1994, 2002). As a result, the program included four intensive EAP courses whose main goals are writing articles and public speaking in English.

1.1.3. IPD2 and its OPs challenges

I have taught the second course of this program (IPD 2) since 2011. In alignment with the IPD program goals, IPD2 teaches the basics of essay writing and short OPs delivery. In this course, essays are usually well written in terms of content, organization, and language use (see

Appendix A); however, some students experience difficulties in OPs such as the lack or misuse of linguistic resources, hesitation disfluencies, or constantly reading from slides or scripts.

Hence, writing a good essay is not necessarily an indication of a good OP. IPD instructors anecdotally observe that public speaking is an area in which discrepancies between writing and speaking and between high and low-achieving students are evident.

In 2014, when I started my PhD studies with the university of Birmingham, I conducted a study (Nausa, 2015, 2017, 2018) on these discrepancies and compared high and low-achieving students' mechanisms to modify the contents of their essays to present them as OPs to their classmates (see 8.1 for a summary of this study). As expected, high-achievers used more mechanisms and submechanisms in the transition and did so in a more consistent manner. That study also suggested that students' disciplinary background influenced their use of English. This observation motivated the second module (Nausa, 2016) on students' academic identity projection in their OPs through the use of first person pronouns (see 5.2 for a summary). In this second study, I included the *medium* level of achievement and the disciplinary divide *hard-soft*. Level comparisons showed that discrepancies were at the level of the pragmatic appropriacy and grammatical variety of the language choices made. Disciplinary comparisons showed that language choices correlated with the way disciplines accept and (re)produce knowledge.

A further conclusion from the two studies was that the traits that characterize these student's use of English in OPs can be classified into two broader categories: strategies to (1) engage and (2) make content easy for the audience. Engaging strategies include the use of pointing expressions and gestures and the projection of authorial identities; content facilitation strategies comprised the use of code glosses and mechanisms to translate originally written content into the oral mode, among others. These studies left other questions unanswered,

methodological gaps (e.g. need for a bigger corpus) and suggested other lines of research. The purpose of this thesis is to continue the analysis of the language of Colombian PhD researchers in OPs, including new areas of analysis and filling gaps identified in my previous studies or available research.

1.2. Questions and general aim

The general aim of this thesis is to demonstrate how English use variation in the oral presentations given by PhD researchers' is correlated with their level of oral achievement, knowledge of their disciplines rhetorical conventions, and the non-expert character of their audience. The research questions build on and are narrowed down to the same two variables of analysis (levels of oral achievement and disciplines) in my previous studies. The variation analyses proposed in the questions focus on the audience engagement and content facilitation strategies in OPs as observed in five language features (*Figure 1.1*).

QUESTIONS		STRATEGIES (PARTS)		FEATURES (CHAPTERS)
1. What are the characteristics of the language that Colombian PhD researchers use in their OPs to...		I. engage the audience	<i>as observed in their use of...</i>	1. spatial and gestural deixis, 2. <i>you</i> to assign the audience an identity, and 3. modalized impersonal constructions?
VARIABLES	<i>in the</i>			

2. What are the differences between high, medium and low-rated OPs...		II. make content easy for their audience		4. code glosses and 5. mechanisms to translate written content into the oral mode?
3. What are the differences between hard and soft-field OPs...				

Figure 1.1. Questions, OPs' aspects, and features for analysis in the thesis

1.3. Summary of the chapters

This thesis has three main sections: the first has three chapters: introduction, review of the literature, and methodology. The second section is divided into two parts: Part I contains three chapters on engagement strategies; Part II has two chapters on content facilitation strategies. The last section contains the conclusion chapter.

In the first section, chapter 1 (this chapter) has introduced the background to the research: the OP challenges of a group of Colombian PhD researchers taking an EAP class, reference to previous studies on their use of English, and the questions that continue to explore their oral academic English. These general questions are explored in more specific questions of the 5 different studies (chapters 4 to 8) reported in this thesis. Chapter 2 provides a general review of the literature. This review focuses on the study of oral academic discourse as genre analysis, proposes a definition of OPs as a PhD-level EAP student training genre, and explores studies on academic oral monologues: lectures, conference presentations, and classroom presentations. More specific reviews of the literature are provided in chapters 4—8. In chapter 3 (methodology), I describe the participants; the procedures for corpus construction, compilation, and update; and the general procedures to perform and validate quantitative and discourse analyses. Again, as with the review of the literature chapter, chapters 5—8 contain specific

methods sections which explain specific procedures to select subcorpora, analyse language features, and validate classifications.

The second section organizes the five studies into two main parts: engaging the audience (chapters 4—6) and making content easy (chapters 7—8). In part I, chapter 4 explores spatial gestural and verbal deixis with an emphasis on how students use these resources to interact with their audience and the images in their presentations. Chapters 5—6 revisit a model of analysis for academic identity projection with the use of personal pronouns (Nausa, 2016) and adapt it for the analysis of identities projected and assigned to the audience with the use of *you* (ch.5) and the authorial identities projected with impersonal modalized expressions (ch.6). In part II, chapter 7 focuses on the use of code glosses as anticipation of potential sources of confusion. Chapter 8 explores four mechanisms to adapt the way content is expressed in essays to present it orally.

Chapter 9 (third section) presents a summary of the five studies' findings, their implications, limitations, and perspectives for future research.

CHAPTER 2

GENERAL REVIEW OF THE LITERATURE

2.1. The study of (English) academic discourse

A widely accepted definition of academic discourse is that of the ways of thinking and using language in academia (Hyland, 2011). Universities, institutes, research centres, and other discourse communities (Swales, 1990) are *par excellence* places where knowledge is created and disseminated. A good deal of the academic dissemination genres is in English, recognized by many as the language of scientific communication (Drubin & Kellogg, 2012). The rapid advent of globalisation and the turn of the century have brought about an unprecedented number of international university programmes run in English in countries where it is not an L1 (Graddol, 2006; Hynninen, 2013; Wächter, 2008) and the need for these universities to be recognized in international academic communities. To be a recognized member, any person, native or non-native speaker of English, needs to have knowledge and mastery of the communities' ways of thinking, (re)producing knowledge, genres, and lexis; in other words, they need the English of academia. Academic English has been studied in applied linguistics in areas such as EAP - English for Academic Purposes- (Jones, 1972 as cited by Jordan, 2002), ERPP -English for research publication purposes- (Flowerdew, 2013), academic discourse analysis (Hyland, 2011), contrastive rhetoric (Connor, 1996), and scientific English (Halliday, 1989; Swales, 1971) to understand how NS and NNS scholars use this language.

2.2. Academic discourse analysis as genre analysis

The study of academic discourses has been predominantly a study of academic genres. Genre analysis focuses on recurrent elements of language use that are representative of academic communities' rhetorical practices and that can give insights of the language in the texts and the communities themselves (Hyland, 2011). The term *genre* has been defined differently in several disparate areas like stylistics, anthropology, and literature. In academic discourse analysis, genres are defined as discourse whose structure, content, style, and intended audience is defined by sets of communicative purposes shared by members of a given discourse community (Swales, 1990). Tardy (2011) defines *genres* as typified forms of discourse that emerge as regularized responses to specific demands. Biber and Conrad (2009), in distinguishing *genre* from related concepts *register* and *style*, refer to *genres* not as types of text, but as an approach to the study of complete text structuring linguistic mechanisms. *Register* and *style* approaches, on the other hand, focus on common linguistic traits of representative text extracts from a given variety. Similarly, Bhatia, (2002) uses *genre* to refer to the analysis of "...linguistic behaviour in institutionalized academic or professional settings" (p. 22). In this study, I adopt a notion common to all definitions of *genre* as institutional regularized texts that reflect the epistemological, rhetorical, and linguistic practices and expectations of a given discourse community.

The concept *genre* is more easily understood in one of its uses: the classification of types of texts produced by a given community. Hyland (2011) proposes an (incomplete)¹ taxonomy of genres (*Figure 2.1*) whose main classification criteria is the academic status of text producers

¹ This is not meant to be a criticism but an observation.

(researchers, instructors, students) or lay readers (general public). In the description of text varieties, Hyland uses the terms *genre* and *discourse* interchangeably.

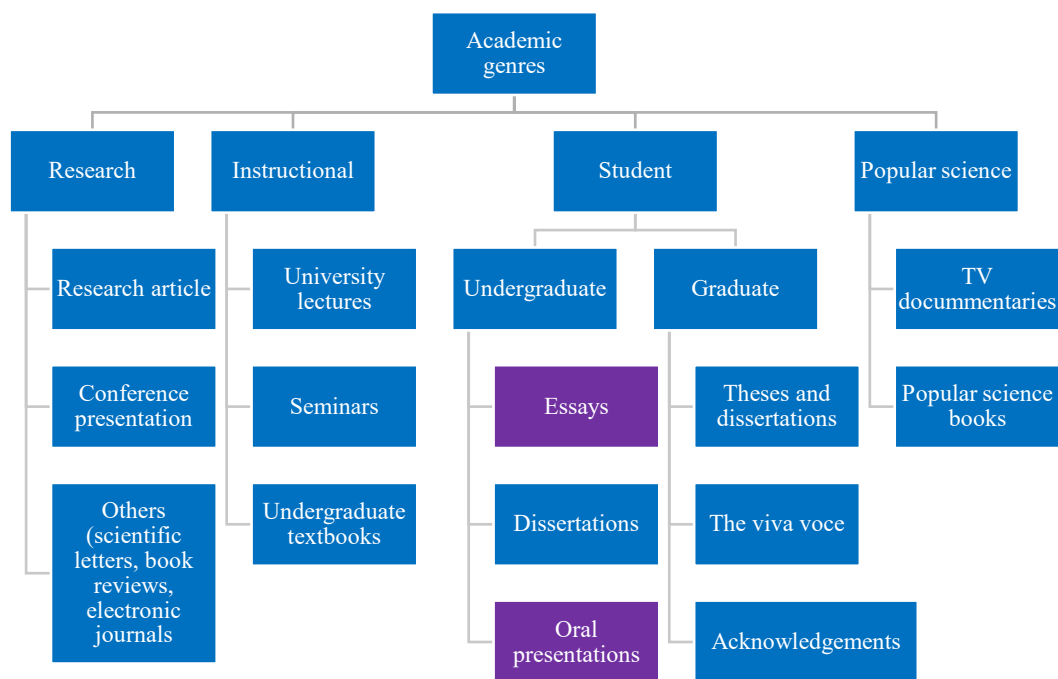


Figure 2.1. Taxonomy of academic genres (Hyland, 2009)

This taxonomy represents the roles of members in academia and their ways to (re)produce and disseminate knowledge. In this thesis, the genres that are analysed are OPs (all chapters) and essays (CHAPTER 8). Although Hyland classifies them as undergraduate student

genres, the way they will be analysed can take two dimensions: as PhD researcher language training genres or as EAP student genres.²

The study of English academic discourse, highly motivated by language student needs, has mostly seen studies on written genres, mainly research articles (e.g. Cao & Hu, 2014; Hyland, 1996b; Khedri, Heng, & Ebrahimi, 2013; Kuo, 1999a; Lafuente Millán, 2010; Murillo, 2012; Norton, 1997; Valero-Garcés, 1996; Yang, Zheng, & Ge, 2015) and student essays (e.g. Aull & Lancaster, 2014; Bruce, 2016; Crosthwaite & Jiang, 2017; Gardner & Nesi, 2013; Gilquin, Granger, & Paquot, 2007; S. Lee, 2008; Leedham & Fernandez-Parra, 2017; Nesi & Gardner, 2012; S. North, 2005). Studies on oral genres are by far fewer, among other things, given the lack of clarity in the definition of genre boundaries and the difficulties in the transcription process such as the number of hours it takes and the difficulty in understanding segments (McCarthy, 1998) especially when they are produced by language learners (Gilquin & Granger, 2015).

2.3. Oral academic discourse

The definition of genres is not simple because the differences among them are not minor (Swales, 1990). Genres differ from each other in several respects, among them complexity of rhetorical purpose, degree and stages of preparation (e.g. rough and final drafts), and mode of expression (written and oral) (p. 62). A research paper, for example, can go through several drafts before it is finally published; and then, it can be presented at a conference as an OP. In general,

² The inclusion of these genres in the EAP course has pedagogical motivations. They prepare PhD students for more complex research genres that are the main aim of the program: research articles and conference presentations in English (not their L1). Their level of English (see 3.2) requires the development of basic academic language knowledge and skills.

oral academic genres are not exclusively spoken but tend to connect to written genres or are based on some form of written expression. As pointed out by Giménez (2000 as cited by Fortanet, 2005), oral genres are often based on written texts that are read out loud.

Two authors have proposed classifications of oral genres. Giménez (2000 as cited by Fortanet, 2005) classifies genres according to the relationship between speaker and audience (*Figure 2.2*).

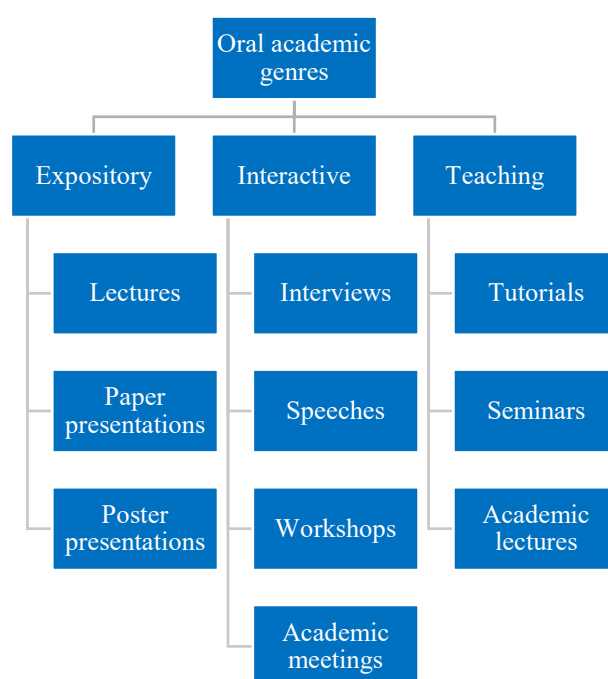


Figure 2.2. Giménez' (2000) taxonomy of oral academic genres

Fortanet (2005) finds that this classification criterion confuses the relationship between speaker and hearer (exposition and interaction) with purpose (teaching) pointing out that teaching genres can be expository and interactive too. Fortanet proposes a taxonomy based solely on purpose (*Figure 2.3*).

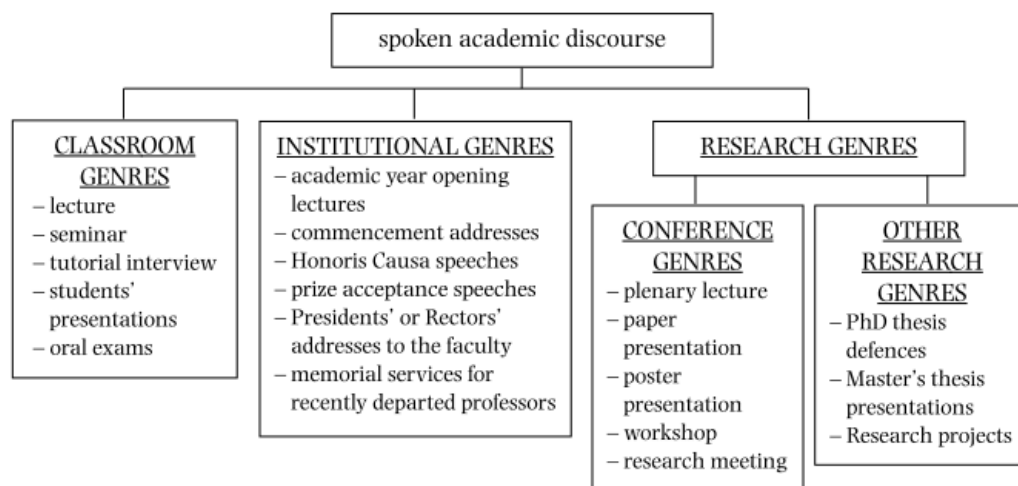


Figure 2.3. Fortanet's (2005, p. 32) classification of oral academic genres by purpose

At the epistemological level, this classification is more convenient than Gimenez' (for the reasons already discussed) or Hyland's. Hyland's classification by text producer status does not capture specific situations in which a genre typical of a person with a certain academic status is produced by individuals with a different status (e.g. those cases in which undergraduate students write research papers). Henceforth, I will refer to essays and OPs, not as undergraduate genres, but as EAP classroom genres. I will keep the term oral presentation, though.

Like written genres research, two oral genres have been more widely studied: lectures and conference presentations (CPs).

Lectures studies are mainly motivated by university students' comprehension difficulties. They aim at understanding lectures from the perspective of students' comprehension and instructors' use of English. Discourse studies that focus on students explore aspects like comprehension (S. Thompson, 1994; Zare & Keivanloo-Shahrestanaki, 2017), strategies to take notes (Siegel, 2018) and summarize information (Österholm, 2012), participation enhancement (Morell, 2007), and engagement (Yeo & Ting, 2014). Studies that focus on instructors discourse

explore their speech speed (Nesi, 2005), academic lexicon (Dang & Webb, 2014) and its density (Nesi, 2005); uses of language (Deroey, 2012; Deroey & Taverniers, 2011); moves and functions in introductions (Yaakob, 2013); metadiscourse (Aguilar & Arnó, 2002; Bu, 2014; J. Lee & Subtirelu, 2015) as reformulations (Murillo, 2008), elaboration of explanations (Crawford, 2015); discourse organization (J. Lee, 2016) and cohesion (Nesi & Baştürkmen, 2006); and the use of storytelling (Alsop, Moreton, & Nesi, 2013). Lecture multimodality is also explored in aspects such as gestural and symbolic deixis (Bamford, 2004), the uses of humour (Fortanet & Ruiz-Madrid, 2016) and laughter (Nesi, 2012), and interaction with slides (Knoblauch, 2008). Another area of interest is the analysis of ITAs (international teaching assistants) and the difficulties that their NNS status brings about for NS students. These studies focus on non-linguistic aspects such as how students perceive ITAs (Damron, 2000; Fitch & Morgan, 2003; Rubin, 1992, 1998) and their cultural assimilation (Gorsuch, 2003). Linguistic studies focus on ITA's pronunciation and accent (Hendriks, Van Meurs, & Reimer, 2018; Kang, 2010; Pickering, 2001, 2004), their use of discourse markers (Bellés & Fortanet, 2006), or how they modify written discourse to present it orally (Levis, Levis, & Slater, 2012).

Conference presentation (CP) studies seek to provide information to help novice researchers or NNS researchers. Like lecture studies, CP studies analyse aspects that generate difficulties to presenters like the expression of stance (Fernández-Polo, 2014; Hood & Forey, 2005; Querol-Julián & Fortanet, 2012; Rowley-Jolivet & Carter-Thomas, 2005b), interactivity (Webber, 2005), persuasion (Valeiras, 2015), talk organization (Fernández-Polo, 2014; Rowley-Jolivet & Carter-Thomas, 2005a), quantification (Rowley-Jolivet, 2015); visual aids use (Diani, 2015; Dubois, 1980, 1982, 1985, Rowley-Jolivet, 2002, 2012) compared to visuals use advice in self-help guides (Anthony, Orr, & Yamazaki, 2007); or the conversion of written into oral

language (Carter-Thomas & Rowley-Jolivet, 2001; Rowley-Jolivet, 2012). CP studies also analyse its parts: introduction (Hood & Forey, 2005; Rowley-Jolivet & Carter-Thomas, 2005c), conclusion (Kite, 2008), and discussion or Q&A session (Querol-Julián, 2010; Querol-Julián & Fortanet, 2012; Webber, 2002; Wulff, Swales, & Keller, 2009).

In this thesis, I include reference to lecture and CP³ studies in the literature review sections of specific chapters when the aspects discussed are of relevance to my OPs studies.

2.4. Oral presentations (OPs)

2.4.1. Multiple terms, one genre

One problematic aspect in the search for studies is the variety of terms and acronyms used to refer to OPs. Among them are OPs –oral presentations (Andeweg, de Jong, & Hoeken, 1998; X. Li, 2018), AOP –academic oral presentations– (Kaur & Ali, 2017), OAP⁴ –oral academic presentations– (Morita, 2000; Wu, 2010), student presentations (Fortanet, 2005; Yaakob, 2013), speeches (Giménez, 2000; Reinhart, 2005), academic speeches (Simpson-Vlach, 2006) academic presentations (Reinhart, 2005; Rendle-Short, 2006), academic student presentations (Ágnes, 2002), student academic presentations (Zareva, 2009). I find the use of these terms problematic at least for two reasons. The first is the delimitation of genres. The term *presentation* implies speaking in front of an audience; therefore, *presentation* could cover lectures, workshops, class presentations, and other monologues. The addition of the adjectives

³ Other intermediate genres like the graduate seminar (Weissberg, 1993)³ although not as frequent, will also be referred to.

⁴ I had originally adopted this acronym but given its other meanings (e.g. old age pensioner) and probable negative connotations, I adopted OPs.

oral and/or *academic*, and the alternation of their order does not really help to clarify the nuance in which they should be interpreted. This is evident in some reviews of the literature (e.g. Kaur & Ali, 2017) in which *oral presentations* is used in the public speaking and not the genre (*mode, tenor, field*) sense. When terms like *student, classroom, engineering, or conference* are added, the delimitation of the genre is clearer. The vague character of the term *presentation*, and this is the second reason, also makes the search for literature more difficult than it is when searching for clearly delimited genres like *research article*. This is probably a reflection of the difficulties in defining oral genres that McCarthy (1998) refers to.⁵

2.4.2. The study of OPs as a student genre

OPs are an under-researched genre if compared with written genres or oral genres like lectures and CPs. However, studies have grown in number, which is evidenced in recent OPs literature reviews (Barrett & Liu, 2016; Kaur & Ali, 2017; van Ginkel, Gulikers, Biemans, & Mulder, 2015).

Until recently, literature on oral academic discourse focused on aspects other than linguistic or discourse descriptions (Zareva, 2009), and OPs literature was not the exception. Non-linguistic studies on OPs can be located in four strands: pedagogical, vocational orientation, professional settings, and textbooks (*Figure 2.4*). Pedagogical based studies focus on students' needs in different disciplines, challenges they face, and individual factors affecting their performance. Studies also include teaching innovations to approach OPs, reflections on what

⁵ I do not seem to be consistent in my critique. I use oral presentation and OPs in this thesis although I claim that doing so is problematic. I came to the conclusion of the confusion of terms recently, after I had written two modules and two articles using the term *oral presentation* and its acronym *OP*. I decided to keep using them just to keep consistency across my studies.

classroom aspects improve with OPs, and performance assessment. Vocational orientation and professional settings literature explores aspects of projection, development, and professional growth related to OPs. Finally, available public speaking textbooks can be classified in two groups: general advice for different audiences and specific instruction, study, and practice of linguistic aspects (*Figure 2.4*).

Pedagogical								
Needs analysis		<ul style="list-style-type: none"> - EAP (Ferris, 1998; Ferris & Tagg, 1996), - Medicine (Haber & Lingard, 2001) - Geography (Hay, 1994), - Engineering (Bhattacharyya & Shaari, 2012) - Marketing (M. R. Young & Murphy, 2003) - Non-engineering fields (Kim, 2006); 						
	Challenges	<ul style="list-style-type: none"> - Language knowledge (Mahfoodh, 2014; Stapa, Murad, & Ahmad, 2014) or choices (Chanock, 2005) - L2 socialization across disciplines (Zappa, 2001) - Speech suppression caused by slide use (Wecker, 2012) - Lack of experience or practice opportunities (Evans, 2013; L. Yang, 2010) - Face-threatening events (Thompson & Collins, 1995) - Sources of anxiety (Chen, 2009; Mohd Radzuan & Kaur, 2011; Woodrow, 2006) 						
Individual factors	<i>Styles and strategies for OPs</i>	<ul style="list-style-type: none"> - Self-directed learning (Tsai, 2011) - Verbal guidance (T. Brown & Morrissey, 2004) - Connection to learning strategies (Chou, 2011) - Goal orientation and self-reflection (De Grez, Valcke, & Roozen, 2009b) 						
	<i>Perceptions of... in relation to OPs</i>	<ul style="list-style-type: none"> - Own competence and actual Performance (Alwi & Sidhu, 2013) - Peer (Girard, Pinar, & Trapp, 2011) and self (Miles, 2014) evaluations - Video-assisted self-reflection (VASR)(X. Li, 2018) - Factors predicting (Otoshi, J., & Heffernan, 2008) and leading to (Soureshjani & Ghanbari, 2012) OPs effectiveness in EFL 						
	<i>Expectations of...</i>	<ul style="list-style-type: none"> - Genre (Seliman, 1996) 						
	<i>Attitudes towards...</i>	<ul style="list-style-type: none"> - OPs (De Grez, Valcke, & Berings, 2010b; Gedamu, 2016) 						
Instructional innovations		<p>Instruction mode: multimedia-based (De Grez, Valcke, & Roozen, 2009a), online (Hill & Storey, 2003), task-based (Rahman, 2010), clinical reasoning (Wiese, Varosy, & Tierney, 2002), experiential learning (Qurban & Austria, 2009), service-Learning (Tucker & McCarthy, 2001), expert behaviour analysis (K. L. Taylor & Toews, 1999), model-based (Green et al., 2005; Swanson, Spooner, Reeder, Haight, & Senthilselvan, 1992; P. Taylor, 1992), learning by design (LBDM) (Devi, Amir, & Krish, 2014), OPs conditions modification (Bayless, 2004)</p> <p>Presentation types: goal-based embedded team and individual (Kerby & Romine, 2009), simultaneous (Shimo, 2011)</p> <p>Portfolios: public speaking (Jensen & Harris, 1999) video (Moore & Voth, 1997),</p> <p>Assessment: use of real-world standards (Pittenger, Miller, & Mott, 2004), audience-in-charge format (Shaw, 2001), student-taught review sessions (M. R. Nilsson, 2001)</p>						
	Improvement of classroom practices	<ul style="list-style-type: none"> - Embedding, teaching and assessing oral communication in university science subjects (Chan, 2011), - Optimum number (Calcich & Weilbaker, 1992) - Benefits (Al-Issa & Al-Qubtan, 2010; Munby, 2011), - L2 oral performance improvement with OPs (Miles, 2003; J. Wilson & Brooks, 2014) or teaching oral Academic discourse (Cheong, n.d.) 						
Assessment		<ul style="list-style-type: none"> - Peer (W. Cheng & Warren, 2005; De Grez, Valcke, & Berings, 2010a; Mitchell & Bakewell, 1995; Smith & King, 2004) and/or - Self (Campbell, Mothersbaugh, Brammer, & Taylor, 2001; Reitmeier & Vrchota, 2009) 						
<table> <tr> <th>Vocational orientation</th><th>Professional settings</th><th>Textbooks</th></tr> <tr> <td> <ul style="list-style-type: none"> • Professional projection (Fallows & Steven, 2000; Yusoff, 2010; Živković, 2015) • Research related skills and presentation of findings (Bankowski, 2010; Wu, 2010) </td><td> <ul style="list-style-type: none"> • Professional development (Boyd, 1989; Rowley, 2012) • Gaming as socialization (Castronova, 2013). </td><td> <ul style="list-style-type: none"> • Advice and information for Graduate student spoken genres (Huang, 2010) Public speaking (Beebe & Beebe, 2015; Osborn, Osborn, & Osborn, 2012) </td></tr> </table>			Vocational orientation	Professional settings	Textbooks	<ul style="list-style-type: none"> • Professional projection (Fallows & Steven, 2000; Yusoff, 2010; Živković, 2015) • Research related skills and presentation of findings (Bankowski, 2010; Wu, 2010) 	<ul style="list-style-type: none"> • Professional development (Boyd, 1989; Rowley, 2012) • Gaming as socialization (Castronova, 2013). 	<ul style="list-style-type: none"> • Advice and information for Graduate student spoken genres (Huang, 2010) Public speaking (Beebe & Beebe, 2015; Osborn, Osborn, & Osborn, 2012)
Vocational orientation	Professional settings	Textbooks						
<ul style="list-style-type: none"> • Professional projection (Fallows & Steven, 2000; Yusoff, 2010; Živković, 2015) • Research related skills and presentation of findings (Bankowski, 2010; Wu, 2010) 	<ul style="list-style-type: none"> • Professional development (Boyd, 1989; Rowley, 2012) • Gaming as socialization (Castronova, 2013). 	<ul style="list-style-type: none"> • Advice and information for Graduate student spoken genres (Huang, 2010) Public speaking (Beebe & Beebe, 2015; Osborn, Osborn, & Osborn, 2012) 						

		International conference presentations (Wallwork, 2010) <ul style="list-style-type: none"> • Information on language (Wallwork, 2010), models analysis, and practice exercises on academic presentations (Reinhart, 2005)
--	--	--

Figure 2.4. Survey of literature on non-linguistic aspects of academic OPs

These non-linguistic aspects are important to inform theory users in the design of plans to help students improve their OPC (oral presentation competence (De Grez, 2009)). However, discourse-based descriptions of (non)professional and (non)successful OP language behaviour are important for curricular planning, class instruction, and materials design for college and NNS students.

Studies on OPs as a student (NNS/novice researcher) genre are few and can be classified into 3 categories based on the area they focus on: general structure and moves, lexicogrammar features, and multimodality.

Studies on moves or parts include the analysis of attention-getting techniques in the introduction (Andeweg et al., 1998), moves and structures of the OPs body (Seliman, Salbiah; Affendi, Irwan; Pendidikan, 2010) disagreement, involvement, and intrusion in the Q&A section (Lin, 2017). A recent study (Hu & Liu, 2018) focuses on 3-minute thesis presentations moves.

Most of the studies can be placed in the lexicogrammar/functional strand. These studies comprise the analysis of multi-word verbs (Zareva, 2016), metadiscourse (Ágnes, 2002; Alessi, 2005; Rui & Xin, 2009), students perception of OPs' formality as reflected in circumstantial adverbials (Zareva, 2009), OPs' structural organization (Yeereem, 2013), pronouns and rhetorical questions (Vassileva, 2002), or from a multidimensional analysis stance (Iberri-Shea, 2011). Other studies analyse how presenters interact with the audience (Rendle-Short, 2006), express stance with pronouns (Morton, 2009, 2011; Nausa, 2016; Zareva, 2013) and adverbials or other

structures (Zareva, 2012). Finally, another area of interest is the lexicogrammar mechanisms to transition from written to oral discourse (Nausa, 2017, 2018).

The analysis of multimodality in OPs has mainly focused on slide transitions and includes spatial deictics use and image integration (Morton, 2006; Rendle-Short, 2006).

2.5. Conclusion

Research on oral academic discourse has increased in recent years, but specific research on student OPs is still scarce. Literature review articles or sections in research reports tend to mix studies on OPs given by students and professional speakers or other types of academic public speaking like lectures and seminars, which is evident in the types of studies reported and the array of terms to refer to this genre. Although not necessarily wrong, not setting clear genre limits could lead to language behaviour generalizations applicable to some individuals but not others. Most of the literature focuses on OPs' non-linguistic aspects and their benefits for L2 development. Studies on linguistic behaviour are still scarce and deemed necessary for the design of strategies and material to effectively help novice researchers and NNS of English. This study seeks to contribute to the understanding of the genre by focusing on PhD researchers in an EFL context, whose L1 is Spanish, a population not studied yet. Also, this thesis hopes to contribute to learner corpus research (LCR) studies at earlier learning stages. As pointed out by Gilquin and Granger (2015), "...most corpora to date represent the more advanced stages [and] ...the number of written corpora by far outnumbers that of spoken corpora" (p. 419).

CHAPTER 3

METHODOLOGY

3.1. Introduction

This chapter presents general methodological aspects of the studies in this thesis. It is organized as follows:

1. participants: selection and ethical procedures;
2. the corpus: description of the task it is based on, texts preparation, and design principles;
3. description of quantitative analyses; and
4. description of qualitative (discourse) analyses.

Each of chapters 4—8 has an additional methodology section, describing the subcorpora employed and specific analytical processes adopted.

3.2. Participants

The research took place in the second course of an EAP program: Programa IPD (see 1.1.2 and 1.1.3). IPD comprises 4 courses part of the English language requirement for all PhD programs at Universidad de los Andes. 88 students (ages 26-56; women:50, men:38) enrolled in the IPD2 courses taught between 2011 and 2016 provided consent to participate in this study. 33% (29) had passed the first course of the program (IPD1) while the other 67% (59) had been allocated directly to IPD2 through the in-house placement test. This test has not yet been aligned with standard tests like IELTS; nonetheless, our rough estimations place students in

the A2-B1 CEFR levels. For the purposes of the study, students' selection was based on the program they were enrolled in (disciplines) and the level of achievement in OPs.

The disciplinary selection and organization of students was based on Becher and Trowler's (2001) disciplines characterization ⁶. The 88 selected students fairly represent the disciplinary distribution of the PhD student population of the university. 45 (51.1%) students were enrolled in soft-field programs and 43 (48.9%) in hard-field ones (*Figure 3.1*).

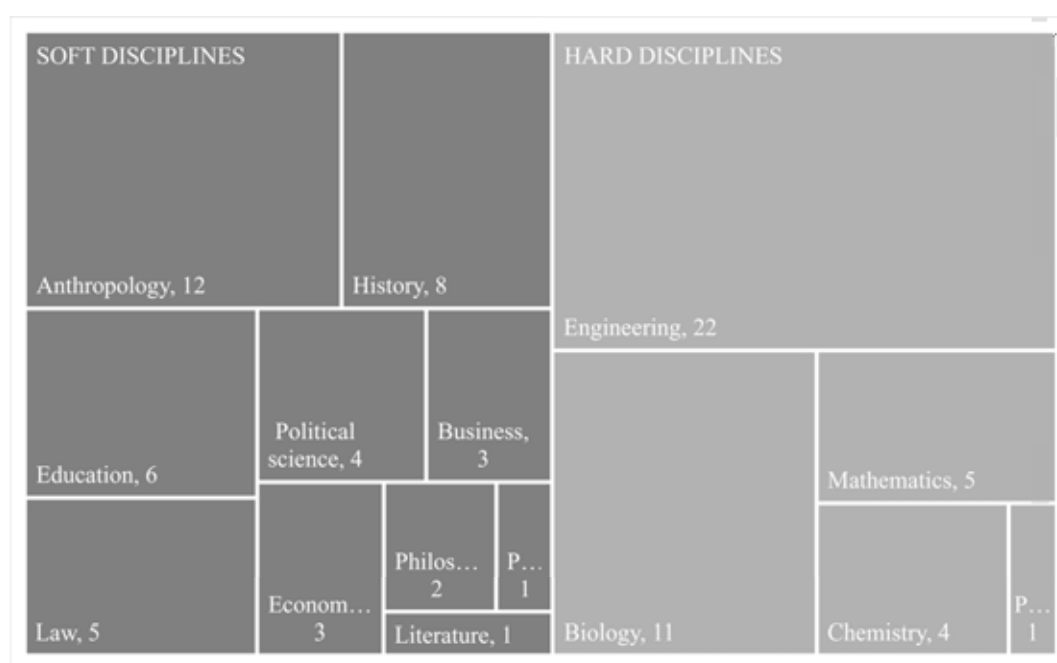


Figure 3.1. Distribution of students per disciplines and programs

It was not possible to have similar numbers of students per program because of the number of enrolments, among other factors beyond the scope of this study. Nonetheless,

⁶ However, only the soft-hard dichotomy was taken into consideration. Other variables like pure-applied or humanities vs. social sciences were not considered. The soft-hard distinction was useful to guarantee balance in the corpus comparisons.

disciplinary representation is balanced if we make a binary distinction between hard sciences and engineering, and the rest. With this classification, there is a roughly hard-soft disciplines opposition 50-50 split.

In terms of the distribution by level of achievement (*Figure 3.2*), students were classified into three categories: high, medium, and low. These levels were set by the score assigned to OPs (1 to 5) with 5 being the maximum passing grade (Appendix B). The average grade (3.5) was the point of reference to define the ranges: high (4.3 – 5.0), medium (3.5 – 4.2), low (2.7 – 3.4).

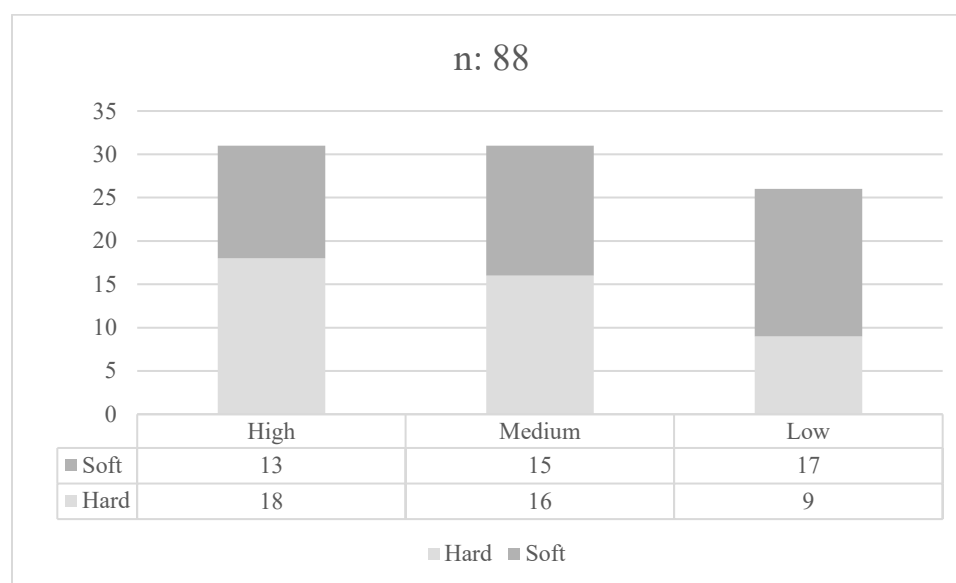


Figure 3.2. Distribution of students by discipline and level

Although the distribution by discipline is balanced, the inclusion of levels of achievement makes that some subgroups might be misrepresented. For example, there are twice as many high achievers (18) as low achievers (9) in the hard fields.

To guarantee the integrity of the studies, the 149 students who had taken the course between 2011 and 2016 were informed about the study via e-mail; 126 completed electronic

online consent forms, and 125 agreed to participate. Students had the chance to provide consent for different aspects of their essays and OPs, ranging from the mere use of what they said or wrote to the inclusion of pictures showing what they were doing while giving the OP. The project was submitted for approval to Universidad de los Andes and University of Birmingham research ethics committees at the beginning of my PhD studies in 2014. Authorization was granted for the pilot and subsequent studies (Appendix C). In the selection of texts, only essays and OPs from students who expressed consent were considered. I videotaped OPs, kept them in a hard drive, and modified the essays and OPs transcriptions to guarantee students' confidentiality and anonymity.

3.3. The corpus

3.3.1. The task

In the first unit of IPD2, students write an essay about their research and present it orally to their classmates, a multi-department audience. The essays (Oshima & Hogue, 2006) are written in three stages: outlining, drafting, and revising. Comments and support are provided in every step. To prepare for the OP, students study OP models (Reinhart, 2005) and use them as reference to adapt the essay to the oral mode.

Students have several opportunities to revise and rehearse for their OP. Two weeks before the OP, they can discuss how they will present the contents of the essay; classmates provide feedback on how to improve OPs (language, explanations, examples, visuals, etc). The week of the OP, students have several small-group rehearsals, in which further recommendations are given. OPs are delivered after the rough draft of the essay has been submitted.

On the day of the delivery, OPs must be 5-10 minutes, include visual aids (e.g. Power Point or Prezi presentation), and have a Q&A section. OP roles are assigned to other students in advance (chair, time keeper, and audience).

Students write the final version of the essay based on the comments made on the rough draft and OP.

3.3.2. Transcription, anonymization, and indexing

Universidad de los Andes sponsored the recruitment and training of three assistants. Training sessions were scheduled to teach assistants transcription procedures. Meetings were programmed to review transcriptions together and negotiate discrepancies.

To anonymize students' information and facilitate corpus searches, codes were assigned to pairs of texts (Reppen, 2010) (*Figure 3.3* and Appendix E).

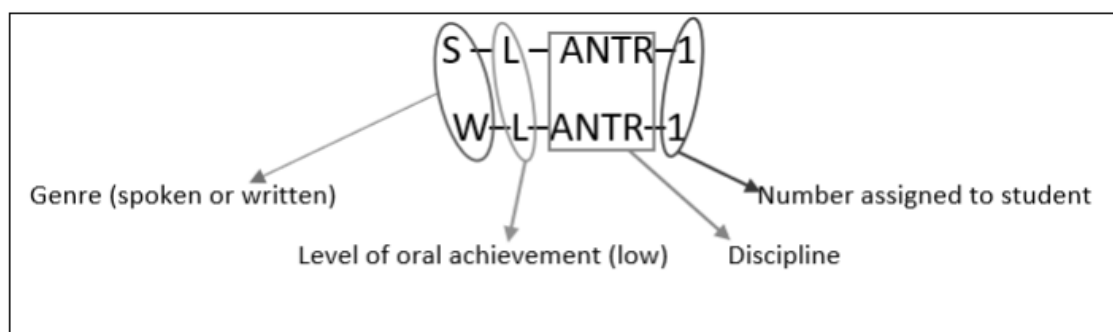


Figure 3.3. Information included in text codes

3.3.3. Corpus design

The resulting parallel corpus is composed of 88 pairs of essays and OP transcripts. Each pair is about the same content and created by the same author. Five aspects of corpus design in Hunston (2002) were considered: representativeness, size, content, balance, and permanence.

This corpus represents the English of Colombian PhD researchers, whose level is estimated to be within A1-B2 CEFR levels (see 3.2), in two different academic student genres (Hyland, 2009; McEnery et al, 2006): essays and OPs. To guarantee that texts include an accurate representation of students' written and spoken English, only rough drafts were considered, as these do not include edits based on the instructor's feedback; OPs transcripts include sequential repetitions and tags for hesitation disfluencies (Corley & Stewart, 2008) like false starts ([fs]), hesitation marks (*eh*), and reading moments ([reading]) (Appendix D).

Another aspect to consider was content (Hunston, 2002) or domain (McEnery et al., 2006), which is based on the corpus intended use: the analysis of Colombian PhD researcher-student English. Here I consider not only the types of genres included in the corpus (student genres), but also what the texts are about (PhD researcher contents).

The size of the corpus can be said to be appropriate. The terms *sampling* (McEnery et al., 2006) and *complete representation* (Reppen, 2010) are used to refer to this complementary aspect in corpus representativeness. Size appropriateness in my corpus is based on the available resources, the purpose of the research, and observed tendencies in research on spoken language. As explained in 3.2 and 3.3.1, I included texts by 88 (59%) of the 149 students who have taken the course⁷. The size is also appropriate for the identification of evidence for the phenomena studied, apart from a few cases with negative effect size values (3.4) that seem to indicate the need for a larger corpus. Finally, with the exception of large-scale studies on spoken varieties (e.g. Biber, 2006a), corpus-based studies on oral academic discourse tend to be based on small

⁷ This percentage probably makes the sample representative of Colombian PhD researchers studying EAP. However, there is not enough information to characterize this population. On average, there are about 1950 PhD students per year in Colombia doing their studies in 22 universities (Acreditación, 2008), but there is no available information on the type of language requirements in each program.

corpora, for example: 26559-token corpus (Rowley-Jolivet, 2012), 40985 (Zareva, 2013); 54717 (Fernández-Polo, 2018), 170000 (Rowley-Jolivet & Carter-Thomas, 2005b).

The balance of the corpus was achieved considering the three main variables: genres, level of achievement, and disciplinary divide. The number of texts was the same for each genre (88); however, the number of words was higher in the OPs subcorpus for several reasons: the presence of additional information that speakers use to introduce themselves; the inclusion of further explanations, humour, and examples; the appearance of repetitions, false starts, hesitation devices, and the like. Levels and disciplinary divides balance were also satisfactory in terms of texts and tokens, although some differences are observed. High-achievers' OPs and essays have more words. Also, soft-field essays have more words than hard-field ones. To remedy possible result drawbacks caused by size differences, normalization procedures (Evison, 2010) are used. Table 3.1 summarizes the distribution of the corpus and subcorpora.

Table 3.1. Corpus of Colombian PhD researchers essays and OPs by level of achievement and disciplines

	Oral (O)		Written (W)		Total (O+W)	
	Texts	Words	Texts	Words	Texts	Words
High	31	27038	31	20902	62	47940
Medium	31	26117	31	18984	62	45101
Low	26	19656	26	15531	52	35187
	88	72811	88	55417	176	128228
Hard	43	36579	43	25125	86	61704
Soft	45	36232	45	30292	90	66524
	88	72811	88	55417	176	128228

Finally, corpus permanence has been guaranteed by the additions made during the last three years. A corpus that is not updated on a regular basis might eventually become unrepresentative of the language variety it contains samples of (Hunston, 2002). Instructors in

the IPD2 course, record OP videos and keep both essays and OPs files in their PCs. During the last three years, the corpus has had three updates (Table 3.2).

Table 3.2. Updates to Corpus of Colombian PhD researchers essays and OPs

	Oral		Written		Total	
	Texts	Words	Texts	Words	Texts	Words
2014-2015	8	5809	8	5255	16	11064
2015-2016	58	47728	58	37027	116	84755
2016-2017	88	72811	88	55417	176	128228

3.3.4. Subcorpora selection and modification

The 128228-token corpus is not used in its entirety in each of the 5 studies presented in this thesis. Specific subcorpora were selected depending on the analyses needs or limitations. CHAPTER 4, for example, required gesture analysis, but not all videos showed the moments in which verbal and gestural deixis were combined and the transcription process was highly time-consuming. Table 3.3 presents the subcorpora that are used in each chapter⁸.

Table 3.3. Subcorpora selected for each chapter

Chapter	Subcorpora	Texts	Tokens	Additional Procedures
4: Deixis	OPs	30	24175	Gestures and deixis mark up, elimination of flawed sentences
5: You	OPs	88	72811	Elimination of flawed sentences
6: Impersonal modalization	OPs	88	72811	Elimination of flawed sentences, POS tagging
7: Code Glosses	OPs	88	72811	Elimination of flawed sentences
8: Written to oral transition	OPs/essays	60	45558	Elimination of flawed sentences, Manual selection of sentences

⁸ Other reasons for subcorpora selection and the procedures to prepare the texts for analyses will be explained in the methods section in each chapter.

3.4. Quantitative analyses

One of the advantages of using a corpus is that it provides quantitative information about language features (McEnery et al., 2006). This quantitative information is in several cases more reliable than native speaker intuition (Hunston, 2002).

The studies in this thesis are based on the following statistics: raw frequencies, percentages, normalised frequencies, significance and effect size tests.

3.4.1. Raw and normalised frequencies, and percentages

Raw frequencies are the arithmetic count of linguistic elements (tokens) (McEnery et al., 2006). They are used to provide the size of the corpus and subcorpora, and the number of occurrences of language features (e.g. *you*). These counts are complemented with percentages to give an intuitive proportion of features occurrence in the corpus.

When features frequencies need to be compared across subcorpora (e.g. *you* between hard and soft fields) normalised frequencies (per 10,000 words) are provided. Normalisation is used when compared corpora are of different sizes (Evison, 2010). When this is the case, comparisons are made by expressing the frequencies in each corpus by a common factor. As the subcorpora in this study are measured in ten-thousands of words, frequencies are normalised to a 10,000-word base. This is done by dividing the feature frequency by the corpus token-size; then, this result is multiplied by 10,000. Normalised frequencies indicate features' overuse or underuse in a subcorpus in relation to another.

3.4.2. Significance and effect size analysis: hypothesis testing

Although normalised frequencies are useful in subcorpora comparisons, they are not enough for hypothesis testing (McEnery et al., 2006). Each study in this thesis is based on hypotheses that students' levels of achievement and disciplines correlate with their choice of specific language features. To validate hypotheses like these, an established practice in corpus linguistics is the use of inferential statistics tests. In this thesis I use statistical significance (log likelihood) and effect size (Bayes Factor) tests.

Log likelihood (LL) tests "... do not assume that the data is normally distributed..." (McEnery et al., 2006. p.55) and provide probability values (p) to claim that observed frequency differences are not the result of random chance. LL p values close to 0 are of high statistical significance, not the result of chance. Also, p values close to 0 allow us to infer that null hypotheses "...are unlikely to be true..." (A. Wilson, 2013, p.4). p close to 1 indicates that frequency differences are random. For a hypothesis to be accepted, a p value of <0.05 is expected, meaning that we can be 95% confident that the observed differences are not random.

Table 3.4 provides the interpretation of LL values in probability terms.

Table 3.4. Significance of LL values based on (Rayson, 2017)

LL value ranges	Assigned significance values (p)	Degree of certainty
3.84 - 6.62	0.05	95%
6.63 - 10.82	0.01	99%
10.83 - 15.12	0.001	99.90%
> 15.53	0.0001	99.99%

LL p values are complemented with effect size calculations, Bayes Factor (BF). LL values only provide an indication of how much evidence there is for a difference between corpora while effect size calculations indicate how big that difference is (Hardie, 2014). The

effect size statistic that is used in this thesis is Bayes Factors. “Bayesian statistics focuses on the probability of hypotheses in the light of observed data, rather than on the probability of observed (and more extreme) data in the light of hypotheses” (A. Wilson, 2013, p.5.). They are expressed as BIC (Bayesian Information Criterion) approximate values. BIC values (Table 3.5) are used because true Bayes Factors are difficult to calculate as they use integrals, which are arguably difficult to understand or explain for a non-mathematician since they are used to assign numbers to functions to describe concepts such as area or volume. BIC value ranges and their interpretation provide a more intuitive way to explain effect size. Also, because BIC values test the likelihood of a hypothesis based on available data, they provide degrees of evidence against null hypotheses (H_0).

Table 3.5. Approximate Bayes Factor (BIC)

BIC ranges	Degree of evidence against (H_0)
0 - 2	not worth more than a bare mention
2 - 6	positive evidence against H_0
6 - 10	strong evidence against H_0
>10	very strong evidence against H_0
For negative scores, the scale is read as "in favour of" H_0 (Wilson in personal communication with Rayson, 2017)	

In this thesis, when a frequency difference is significant, but BIC values are negative, this is interpreted as there is a good amount of evidence for the hypothesis to be accepted, but the feature frequency differences between the corpora are not big enough to consider such feature to be representative of one group or another.

LL values or a combination of BIC and LL values are also used as filters to determine what frequency differences are worth analysing and reporting from a discourse perspective.

These values will be mentioned in the quantitative analysis subsection in each chapter.

3.5. Qualitative (discourse analysis) analyses

3.5.1. Ad hoc, eclectic, theoretical perspective

This research can be framed within discourse analysis research as genre analysis. It is discourse analysis as it implements a series of methods to look at language in action by analysing texts as they are related to their social context. More specifically, it can also be said to be genre analysis (Hyland, 2011) because it examines “...element[s] of recurrent language use, including grammar and lexis ... relevant to the analyst’s interests” (p. 174).

The purpose of the discourse (genre) analysis sections in this thesis is to describe features that significantly discriminate among subcorpora in the study. These analyses will focus on corpus linguistic phenomena (concordance lines, collocation, patterns) and discourse functions.

Feature descriptions will be complemented by analyses of grammar and its connection to pragmatic aims to provide a linguistic discourse-based explanation of how levels of achievement and disciplines correlate with language-semiotic choices in OPs.

The studies cannot be placed in one specific paradigm. Different paradigms were chosen depending on how appropriately they explain phenomena. They include Systemic Functional Linguistics (Halliday & Matthiessen, 2014), English Grammar (Quirk, Greenbaum, Leech, & Svartvik, 1985), Academic Discourse Analysis (Biber, 2002; Hyland, 2009; Tang & John, 1999), Corpus Linguistics (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Hunston, 2002; McEnery, Xiao, & Tono, 2008), and gesture, deixis, and multimodality (Gast, Deringer, Haas, & Rudolf, 2015; Kamio, 2001; McNeill, 2006; Rendle-Short, 2006).

Discourse analyses are based on the features in Table 3.6.

Table 3.6. Language and semiotic features analysed in each chapter

<i>Chapter</i>	<i>Aspect</i>	<i>Language feature</i>
4: Deixis	Audience Engagement	- Spatial deictics (<i>this, that, these, those, here, there</i>) and body language
5. <i>You</i>		- Second person pronoun <i>you</i>
6. Impersonal modalization		- Modals, <i>ly</i> modal adverbs, and modal-like expressions
7. Code Glosses	Content facilitation	- Code glosses markers for reformulations and examples
8. Written to oral transition		- SVO modifications, NP reduction mechanisms, modality aspects, and code glosses

The selection of most features exhibiting big frequency differences among subcorpora as expressed by LL values is in alignment with the closed-class keyword strategy (Groom, 2010). This strategy discards open-class items (nouns, verbs, adverbs, adjectives) in a keyword list and focuses on words like deictics (*this, that*) pronouns (*you*), modals (*can*), and conjunctions (*because*). Closed-class words are valid items of semantic analysis (Groom, 2010, p.61) and more fruitful and manageable for concordance analysis (p. 71). Most of the features selected in this thesis are closed-class key words (See Appendix I).

Procedures for selection and analysis of features are explained in each chapter.

3.5.2. Alternative wordings in discourse analyses

Sometimes in discourse analyses, I will add alternative wordings to compare them with what students said or wrote. This is not a critique to their English but another explanatory procedure. Authors like Rounds (1987) make this kind of comparison to understand the difference between what is expressed and what can be interpreted. Comparing what was said to what is expected is a way of spotting differences in meaning or use (Ervin-Tripp, 1976). This view is complemented by the observation that "... a process analysis should presumably take an

interest not only in the paths that are taken but in those that are not but could have been.”

(Widdowson, 1979, p.146).

3.5.3. Validation procedures

To guarantee the validity of discourse analyses, I employed interrater reliability procedures. As the analyses adopt a listener/reader perspective and my interpretations could have been biased by my role as instructor, two colleagues were invited to a two-hour training session in which I explained the concepts and illustrated them with examples from the corpus. Then, I programmed meetings in which we would code randomly selected data samples. I kept record of agreements and disagreements to calculate Interrater agreement as expressed by Krippendorff's alpha values. Like other reliability values (Kappa, Cronbach), Krippendorff's values are ranged from 0 (poor or no agreement) to 1 (perfect agreement) among raters. I chose Krippendorff's nominal alpha values as they are applicable to different rating situations: any number of observers, any number of categories, scales, values or measures, any level of measurement (nominal or categorical, ordinal, etc), large and small sample sizes. (Krippendorff, 2011). Values range interpretation (Table 3.7) is adapted from Landis and Koch's (1977, p. 165).

Table 3.7. Krippendorff's alpha values for interrater reliability

<i>Krippendorff's nominal alpha value</i>	<i>Degree of agreement</i>
< 0.00	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Near perfect

Finally, instances of disagreement were resolved after analysis and discussion. I was only able to employ this validation procedure for the studies in chapters 5, 6, and 7⁹. Specific details will be provided in each chapter.

3.6. Software

To collect, organize, search, and analyse OPs features, I used word processing, spreadsheets, and corpus software. Word was used to transcribe essays and OPs. Files were converted into txt-files with AntFileConverter (Anthony, 2018a) to facilitate searches for corpus phenomena (concordance lines, collocations, and patterns) with AntConc (Anthony, 2014) (Appendix J—M). Concordance lines were transferred to Excel spreadsheets to facilitate manual discourse and statistical analyses (3.4). Statistical and effect size calculations were performed in Excel spreadsheets¹⁰ available in the *Log-likelihood and effect size calculator* web site (Rayson, 2017). Finally, POS tagging was done with TagAnt (Anthony, 2018b).

⁹ Although my 2 colleagues were eager to help, they just had time for three rating sessions.

¹⁰ This spreadsheet offers calculations for two or six corpora, not for other numbers. I modified the spreadsheets to calculate comparisons between three corpora in the level of achievement (high, medium, and low) comparisons (see Appendix G)

PART I: ENGAGING THE AUDIENCE

CHAPTER 4

AS WE CAN SEE HERE...: SHOWING THINGS TO THE AUDIENCE

4.1. Pointing behaviours (deixis) in Colombian PhD researchers' OPs

An initial exploration of Colombian PhD researcher OPs (Nausa, 2015) revealed that those students with higher performances used more mechanisms to engage the audience. Similarly, in module 2 (Nausa, 2016), I found that high-achievers more frequently and consistently showed things to the audience as evidenced in their use of self-references in the projection of the authorial stance role *guide* (e.g. *let **me** show you*). This sensitivity to one's interlocutors' linguistic needs is part of what scholars in the Conversation Analysis (CA) tradition refer to as *recipient design* (Sacks, Schegloff, & Jefferson, 1978). Among the recipient design mechanisms that students used was the inclusion and explanation of visuals. This means that at specific moments of the OP, presenters interacted with and referred to images on the OPs slides by using pointing expressions and actions to direct the audience's attention towards what they deemed important. In other words, their use of images implied the use of verbal and gestural deixis. This motivates a more in-depth study of verbal and gestural deixis in their presentations. This chapter seeks to identify features of verbal and gestural deixis that differentiate OPs performance in this group of students. The identification of differences is based on the following questions:

1. **Level of achievement:** What are the differences between high, medium, and low achievers in terms of how they perform deixis while showing images in their OPs?

2. **Disciplines:** What are the differences between hard and soft-discipline students in terms of how they perform deixis in their OPs?

4.2. Verbal deixis (pointing with words)

It is well-known that some language features, such as pronouns or some adverbs, are commonly used to refer to (point at) an entity, place, or time elsewhere in the text or in the surrounding context. For example:

- (1) ***Chit Poe** (...) was born in the Bornho refugee camp (now defunct) in 1990, shortly after his parents' flight from Burma. He lived most of his life in the Mae La refugee camp, where he attended school until eighth grade.* (Gilhooly & Lee, 2014, p.2.)
- (2) ... *he'll build models and hang them from the, the top of the studio. uh, and as you can see here, this is part of a larger body of work...* ("Twentieth Century Arts Lecture: LEL320JU147," 2001)

In (1) the two occurrences of the pronoun *he* refer to the noun *Chit Poe*. The pronouns point at a noun in the same text. In (2), the adverb *here* does not refer to another word in the text. *Here* is used to refer to a picture that is being shown to an audience in the context of an arts lecture.

The terminology used to refer to these features uses the morphemes *phora/phoric* (*endophoric, exophoric, anaphoric, cataphoric*) or the term *deixis*. The reference made in (1) is endophoric (inside the text) while the reference made in (2) is exophoric (out of the text). Different approaches to linguistics use these terms differently. Halliday and Matthiessen (2014) use the *phora* terms to refer to the in-text or out-of-text references exemplified here. Fillmore (1997), on the other hand, prefers *deixis*, and distinguishes between *discourse deixis* (equivalent

to *endophoric reference*) and proper *deixis* (equivalent to *exophoric reference*), distinguishing different sub-types of deictic aspects: *personal*, *social*, *time*, and *spatial*. Others (Bühler, 1965; Levinson, 1983) propose *symbolic deixis* (a sub-set of *exophoric reference* where the referent is imagined rather than physically present).

In this thesis, the terms *deixis* and the *phora* terms will be used, and the following uses are distinguished:

4.2.1. Reference to preceding or subsequent items in the discourse (*anaphora and cataphora*)

Anaphoric references occur when a unit refers to another unit backwards in the text. For an example of anaphora see sentence (1) above. Cataphoric references occur when a unit refers to another unit ahead in the text.

- (3) A few weeks before he died, (my father) gave me an old cigar box filled with faded letters. (Nordquist, 2017 as cited by Nilsson, 2017)
- (4) This idea that (the Elder Brother stories contain the law of the people) comes from many discussions I have had with (...) Maria Campbell. (Innes, 2009)

In (3) the pronoun *he* refers forwards to the noun phrase *my father*. In (4) *this idea that* refers to the proposition *the Elder brother... Campbell*. *He* in (1) and (3) and *this idea that* in (4) point at something within a text, not within the context of communication. Although pronouns have traditionally been identified as the units that perform this pointing-referring function in texts, other units can also perform this function. In (4) the signalling noun (Flowerdew, 2003) *idea* points forwards towards the proposition in parentheses.

4.2.2. Reference to an item in the physical context of the discourse (*deixis*)

Deixis, in this sense, is in the intersection between meaning and use, for it is in the context of use that the meaning of utterances is fully understood; in other words, utterances as acts of communication are not understandable unless we know the referents. Take for example the utterance *she is there*. If only the basic meaning of the compounding words is considered, we understand that the expression refers to a person (most probably a woman) and her location (not at the point where the person writing or uttering the sentence is). However, if some contextual information is provided: ‘the utterance is said to a parent picking up their daughter at nursery school and asking her caretaker where their daughter (who they can’t see) is’, the meaning expression capacity of the utterance is increased. Understanding *deixis* this way implies that the words used to perform this type of references (*deictics*) are almost devoid of meaning; they are empty shells which are only filled up with elements of the context of use. However, the opposite can also be stated: *deictics* signal important information about various aspects of the context of language use (Schiffrin, 1990). *Deictics* in this sense include key information such as (1) the roles of communication participants: addresser (*I*), addressee (*you*), and referred ones (*he, she, it, they*); (2) their status (*sir, professor*); (3) the time of enunciation (*now*); (4) the place (*here, there, this, that*), among others. Fillmore (1997) refers to the realisation of these aspects as *personal, social, time, and place deixis*. Example (2) above includes a case of *spatial deixis*.

4.2.3. Reference to an item in the symbolic or virtual context of the discourse (*symbolic deixis*)

In addition to pointing at things in contexts and texts, *deictics* can also be conceived as occurring to point at things in symbolic spaces, or the *symbolic field* (Bühler, 1965) which is part

of the systemic space (Köller, 2004 as cited by Pleyer, 2008) which we construct with language. This type of pointing is referred to as *symbolic deixis*. In symbolic deixis, symbolic spaces, are the result of a mutual understanding (common ground) between addressee and addresser. In her analysis of the spatial deictic *here* in university lectures, Bamford (2004) found that the spatial deictic *here* can play a pointing at context function (*gestural*) when used to point at visuals in a lecture. Also, *here* can be used to refer to an intellectually shared space situationally limited in space and time as an event (*this lecture*), not as a space or object in the room where the lecture takes place (example 5).

(5) *So I won't go into good econometrics or bad econometrics out here but you have to have a trade to make that your regressions are good. or at least not obviously bad.*
(P.K Siena 1998, example included in Bamford, 2004, p. 123)

This chapter is concerned with the way presenters point at things in the context of the OP, more specifically with how they interact with and point at slides in their presentation software (Power Point, Prezi, Keynote and the like) and the images (pictures, graphs, diagrams, charts, tables, text, etc) or text contained within. As such, the chapter deals with the analysis of exophoric references as *spatial deixis*; however, related concepts will be invoked specially in those situations in which what is referred to is words or ideas previously mentioned (*anaphora*) or shared cognitive spaces (*symbolic deixis*), discussed in the context of my previous module on classroom identity projection (Nausa, 2016) and 5.4.1.2 in this thesis as *co-constructer* actions.

4.3. Gestural spatial deixis (pointing with gestures)

Reference to images on slides can be made with words and reinforced by or substituted with gestures. Gestures are spontaneous movements made with the hands, face, or other parts of the body that often accompany speech. Different taxonomies to classify or approaches to study gestures have been proposed (Efron, 1941; Eisenstein & Davis, 2004; Ekman & Friesen, 1969; Freedman & Hoffman, 1967; Karam, 2006; Kendon, 1972, 1980; McNeill, 2006; McNeill, 1992; Mittelberg, 2008; Quek et al., 2002; Wexelblat, 1998; Xiong, Quek, & McNeill, 2002). From a very general standpoint, it could be argued that these approaches mostly differ in regard to terminology or nuances in the criteria for classification like type of interaction (human-human, computer-human) among others. For example, most taxonomies agree in identifying iconic gestures, which resemble actions to which they refer (e.g. when a person moves their arm and shapes their hand as if they were throwing a ball). In McNeill's taxonomy, these gestures are called *iconics*; *Physiographics* and *kinetographics* in Efron's; *literal-reproductive* in Freedman and Hoffman's; and *kinetographs* and *pictographs* in Ekman and Friesen's.

In general, the different categories in these taxonomies focus on gestures that play a signifying role; this is to say, gestures re-present something concrete or abstract that is not present. However, *deictic gestures* (not in Freedman and Hoffman's taxonomy) also play an indexical role; they point at something that is present in context. Pointing is prototypically performed with fingers (usually the index) although other parts of the body or manipulated artefacts can be used (McNeill, 1992). However, deictic gestures have also been found to point at things that are not present in context but in text or in a common symbolic space; this is particularly the case in narratives (McNeill, 1992).

4.4. Synchronization of verbal and gestural spatial deixis, and integration of images to talk (the deictic process)

To explain the co-occurrence of gesture and talk in time, McNeill (1992, p.26) following Kendon (1980) proposes three “rules” (McNeill’s quotation marks) for gesture and talk synchrony: phonological, semantic, and pragmatic synchrony rules. These synchronization rules are part of the second phase (stroke) in gesture production; the first phase, preparation, is optional. These rules are used to explain an example for an iconic gesture realisation in a narration but can be transferred to the analysis of the interplay between verbal and gestural deixis. *Figure 4.1* exemplifies the three rules. In phonological synchronization, the gesture happens before or ends at a phonological peak in the utterance. In the use of *as we can see here*, *here* is stressed (phonological peak), and the pointing gesture occurs at the same time *here* is uttered.



Figure 4.1. Coordination of verbal and gestural deixis in as we can see “here”

In semantic synchronization, both words and gesture express the same meaning at the same time (*here* as *this place/point*). The rule would be violated if *here* was used to refer to something not present physically or symbolically. In pragmatic synchrony, words and speech perform the same function: *getting the hearer to look at the thing being pointed at*. McNeill explains that violations of the rule can happen in cases of multiple gestures happening (semantic synchrony rule violation); however, in the type of analysis that will be done in this chapter, the semantic and pragmatic synchronization rules are expected to happen in a straightforward way, for gestural and verbal deixis are expected to happen with literal referential-pointing meanings, which will make the semantic and pragmatic synchrony rules easily identifiable. Also, in McNeill's theory it is not clear whether the three synchrony rules have to happen for the general gesture synchronicity to happen. Given that sentence stress in Spanish is not the same as in English (a stress based language), the stress peaks produced by this student population are not always salient. For the purposes of the analyses, the occurrence of semantic and pragmatic synchronicity rules will be considered.

In addition to McNeill's deictic synchronization rules, gestures will be analysed from other types of kinesics such as body alignment, gazing, and position based on Rendle-Short's (2006) definition of the deictic process. The deictic process accounts for the integration of images to talk in OPs and comprises three phases: (1) expectation that the image is relevant to the talk, (2) combination of verbal and non-verbal actions to invite the audience to focus on the image, and (3) incorporation of image into the talk to focus on a specific part of the image. However, as will be shown, this last step tends to be affected by the nature of the image being used.

4.5. Multimodal analysis of oral academic discourse

The study of the interplay between gestural and verbal spatial deixis in OPs can be placed in the multimodal analysis of oral academic discourses. Different oral discourse analysts (e.g. Carter-Thomas & Rowley-Jolivet, 2003; Charles & Ventola, 2002; Norris, 2004; Poyatos, 2002; Ventola, 2002) have advocated the need for oral discourses to be approached from a perspective that includes the analysis of other forms of representation and conveyance of knowledge different from the verbal mode. Different terms have been coined to refer to the analysis of the linguistic and non-linguistic aspects of representation, communication, and interaction in oral discourses: *semiotic spanning* as *semiotic micro-focus* (Ventola, 1999, 2002); *multimodality* (Jewitt, 2016; Kress & Van Leeuwen, 2001); *multimodal analysis* (Morell, 2015); *Multimodal Discourse Analysis MDA* (Querol-Julián & Fortanet, 2012).

In general, nonverbal semiotic resources in oral academic genres have been approached from three general perspectives, which I have named (PPP): *prosody*, *paralinguistics*, and *paraphernalia*. *Prosody* includes the analysis of sound aspects like syllable duration, prolongation, and volume; sentence rhythm and stress among others. *Paralinguistics* (also referred to as kinesics) includes the analysis of gestures, head and arms movement, body alignment, positioning, etc. *Paraphernalia* refers to the use of objects (e.g., computer, screen, pointers), multimedia apps (e.g., Power Point, Prezi, Keynote, etc) and visuals (e.g., tables, figures, graphs, images, etc).

The analysis of multimodality in academic oral discourse implies the creation of models that rely on theories that account for language use (e.g., SFL- Systemic Functional Linguistics, CA-Conversation Analysis) and theories that account for nonverbal aspects (e.g. McNeill's model for gesture in communication analysis or Bertin's (1973) image taxonomy). In their

analysis of evaluation in Q&A sections in conference presentations, for example, Querol-Julián and Fortanet-Gómez (2012) propose a model composed of the verbal analysis of evaluation based on Appraisal Theory (Martin & White, 2005); analysis of kinesics as gestures (McNeill, 1992), and prosody and paralanguage (Poyatos, 2002) among others.

Although multimodality in oral academic discourses has been researched in their three main genre categories – classroom genres, institutional genres, and research genres (*Figure 2.3*) the conference presentation (a research genre) is probably the subgenre that has received the greatest deal of attention. The use of multimodality in oral academic discourse can be said to be roughly approached from two general functional perspectives: engaging with (interpersonal) and facilitating understanding to (ideational) the audience.

Engagement with the audience, understood as the actions to create empathy or mitigate face-threatening events, has been studied in various subgenres. Hood and Forey (2005) analyse the linguistic (appraisal model) and non-linguistic (visual, gesture) resources that lecturers in plenaries use to express attitudes and create a sense of solidarity. Querol-Julián and Fortanet (2012) study discussion sessions in conference presentations to determine how presenters act and react verbally (appraisal model again) and nonverbally towards comments from the audience. Zhang (2015) also analyses how two plenary lecturers verbally and gesturally express and mitigate disagreement to save face and be perceived as collegial members and likable individuals. Fortanet and Ruiz-Madrid (2016) contrast the multimodal use of asides (digressions) by an English-speaking and a Spanish-speaking lecturer also as mechanisms to build rapport with the audience. These studies in general highlight the importance of mode orchestration during the oral delivery to enhance engagement and avoid face-threatening events.

The other line of studies approaching multimodality reflects speakers' motivation to facilitate understanding to the audience by providing them with alternative explanation mechanisms. One area of interest in this function of multimodality has been the use of slides. Slides in presentations were first analysed in Dubois' (1980) pioneering study on biomedical speeches. Similarly, Charles and Ventola (2002) contrast slide use between hard (physical science) and soft (ethnology) disciplines. In these studies, these visual aids perform semiotic supporting roles: as evidence in hard sciences and as illustration in soft disciplines. Diani (2015) analyses the types of visuals and differences in the macrostructure of the Power Point Presentations based on Swales' (1990) IMRD move analysis, and Bertin (1973) and Rowley-Jolivet (2002) taxonomies of visuals. Her findings confirm that hard disciplines tend to rely on numerical visuals as evidence while soft disciplines tend to rely on figurative-polisemic visuals. The reported use of images in the studies clearly reflect that the epistemological differences among disciplines are not exclusively seen in language use but in the general expression of meanings.

Nonetheless, the use of multimodality to facilitate understanding in oral academic discourses has seen more studies in classroom genres: lectures and student oral presentations. Morell (2015), through a multimodal analysis of two technical-science and two social-science successful presentations given by ELF scholars in a public speech workshop, determines the effectiveness of their OPs. Palmer-Silveira (2015) analyses three presentations of business communication master's students to determine how previously identified flaws in space, hands eye contact, head movement, and stage have been overcome. Similarly, Valeiras (2015) analyses one conference presentation in the business communication field to determine how the variety of semiotic resources available (intonation, gestures, head movements, visuals) is used by two

presenters to achieve persuasion. Ruiz-Garrido (2015) analyses the use of adverbs of degree in coordination with nonverbal resources in the expression of intensification in the presentations of two English and two Spanish speakers. Crawford (2015) studies the elaboration of explanations as seen in the coordination of linguistic markers and nonverbal features to create meaning and reinforce understanding in five Open Courses video-lectures. The conclusions in these studies have pedagogical implications from which instructors and students can benefit; instructor-student or student-student communication can be improved by learning how to master and orchestrate different modes of expression.

Without a doubt, these studies have paved the way for multimodal oral academic discourse study not only because they have provided models of analysis based on the articulation of different paradigms but also because they have pointed out specific areas of inquiry and subsequent applications in different contexts (e.g. EFL and EAP learning). However, as most of their authors openly admit, their findings need to be taken as indicative, but not conclusive, especially because they are based on case studies (e.g. Querol-Julian & Fortanet, 2012) or small corpora. Other studies' claims can be questioned on methodological grounds. For example, Morell (2015) reports the use of different semiotic modes as a characteristic of high-rated professional EFL presentations, but her analysis is based on only four high-rated OPs; she does not compare high and low-rated OPs to conclude that the orchestration of modes is absent in low-rated and therefore a differentiating trait of high-rated OPs.

4.5.1. Analysis of spatial deixis

The multimodal studies in the previous section were classified according to their interpersonal or ideational focus. Deixis based studies in this section can be placed in the

ideational strand for the role that deixis has in making things clear for the audience through the orchestration of the visual and oral channels. Methodologies in the studies vary in that they have focused on verbal deixis, gestural deixis, or the coordination of modes. Studies unsurprisingly also vary in the types of genres (lectures, student presentations, etc) and types of comparisons (fields, successful vs non-successful performances).

One area of analysis of gestural and verbal deixis in oral academic discourses is in classroom genres, specifically in emergent talk and lectures. Roth (2000) in a study of high school and college student talk over and about visual representations of scientific phenomena found that deictic and iconic gestures precede associated utterances during the initial appearance of scientific discourse. Bamford (2004) in an analysis of the deictic *here* in professional NS and NNS economics lectures found that when *here* has a literal spatial meaning, it is easily interpretable from the context; however, when its meaning refers to a symbolic common space (*this lecture*) created between the lecturer and listeners, listeners would find it difficult to interpret it as such. This study is based on lectures extracted from several corpora (BNC, MICASE, Università di Siena) and some of its transcriptions include gestural deictic information, but it is not based on images or videos. Yaakob (2013) in a comparison of lecture introductions in different disciplines found that arts and humanities disciplines used the “handout sub-function” more frequently to guide students and help them become familiar with the disciplinary canon. This study uses lectures from the British Academic Spoken English Corpus (BASE). Although the author does not carry out a multimodal analysis, the presence of gestural and verbal deixis can be inferred from some of the examples in which lecturers make reference to handouts.

Deixis has also been analysed in academia, in *in* and *out-of-the-classroom genres* according to disciplinary variation. Simpson-Vlach (2006) analyses academic speech in MICASE office hour meetings, study groups, discussion sections, lectures, and seminars in hard and soft disciplines with a focus on pronouns, deictics, and hedges and fillers. She found deictics *this*, *these*, *those*, *here*, *there*, and *then* to be the third most common group of expressions after pronouns, and hedges and fillers to be more common in the hard sciences because oral discourse in these fields “...revolves more centrally around visual aids...” (P. 309). This finding confirms Dubois (1980), Charles and Ventola (2002) and Rowley-Jolivet (2002) findings that the types of visuals that are used in the disciplines determine how speakers interact with them.

And of course, multimodal deixis has also been analysed in oral open genres (Swales & Feak, 2012), with emphasis on successful professional oral presentations. Rendle-Short (2006), in her Conversation-Analysis based study of science seminar presentations, which includes talk, prosody, gaze, body position, images on screen and their orchestration with speech, found that for deixis to be successful, the coordination of verbal and nonverbal resources is key. Morton (2006) in a multimodal analysis of spatial deictics such as *this way*, *over here* in 24 oral presentations of successful and unsuccessful first-year architecture students found that successful students better interacted with their images linguistically and kinesthetically.

Again, like the other surveyed multimodal studies, most of these studies provide multi-paradigm models of analysis and highlight areas of study and subsequent application. However, their findings need to be taken cautiously, for some of them are based on a few cases (e.g. Rendle-Short, 2006) or small corpora. Deixis studies like Simpson-Vlach (2006) or Bamford (2004) are based on reasonably large corpora and rely on the existence of easily identifiable deictics like *this* or *here*. Their identification in transcriptions, for example, can make the analyst

expect that a gesture also occurred. However, their corpus analysis is not complemented with video or image analyses.

The multimodal analysis of deixis that I propose in this chapter seeks to fill some of the identified gaps in the analysis of oral academic discourses. The most obvious one is the lack of studies that focus on EFL students of EAP at the graduate level (PhD) of education. Second, it seeks to propose a methodology that includes (1) corpus and statistically significance analyses to identify differences of use by level of achievement and discipline and (2) multimodal analysis to identify how the coordination of verbal and non-verbal resources vary between the groups. As Adolphs (2012) states:

The impetus towards multimodal corpora recognizes that natural language is an embodied phenomenon and that a deeper understanding of the relationship between talk and bodily actions—particular gestures—is required if we are to develop a more coherent understanding of the collaborative organization of communication (p.1).

4.6. Methods and statistical analyses

In this section, I explain the methods for the selection of a subcorpus for multimodal analysis and present an initial quantitative analysis. The purpose of the quantitative analysis of spatial deictics is to determine whether spatial deixis is more common in high-rated OPs and to confirm (like Simpson-Vlach; 2006) that spatial deictics are more frequent in hard disciplines. Verbal spatial deixis is prototypically performed with demonstratives (*this*, *that*, *these*, and *those*) either used as determiners (e.g. *this image*) or pronouns (e.g. *If you look at this*), and the most frequent adverbs of location *here* and *there*. Other adverbs of location (e.g. *around*, *everywhere*)

are not considered here given their low frequency in the corpus. It is expected that the use of these deictics is accompanied by the use of other semiotic modes.

4.6.1. Selection of a subcorpus and data clean-up

To perform the analyses, it was necessary to select a subcorpus due to the difficulties that multimodal analysis implies. The following is a description of the difficulties and the actions to overcome them.

This thesis uses a 72811-token oral subcorpus created from 88 OPs (see 3.3.4). The performance of statistical and multimodal analyses of spatial deictics (*this*, *that*, *these*, *those*, *here* and *there*) that I propose in this chapter presented a few difficulties. First, not all incidences of these expressions are necessarily cases of spatial deixis. Spatial deictic expressions refer to the immediate spatial context. In the case of oral presentations, spatial deictics refer to aspects such as graphs and visuals on slides, the screen and its parts, parts of the room where the presentation takes place, among others. Expressions like *that* can perform other functions: subordinator of noun and adjective clauses, anaphoras/cataphoras, or determiners. Similarly, *there* can be used as grammatical subject in the existential constructions *there be* or *there exist*. A second difficulty is that not all cases of spatial deictics and their corresponding gestural actions can be seen on the OP videos. In several cases, the camera is focused on other aspects such as the slides or the audience. Third, the analysis of spatial deictics and their corresponding actions on video is highly time-consuming.

To overcome these difficulties in the analyses, the following actions were taken.

First, a sub-corpus of 30 OPs videos and their corresponding transcriptions was selected: 14 hard-field and 16 soft-field OP videos. Similarly, 30 video-transcripts pairs include 10 pairs

from each level of oral performance (high, medium, and low). These distributions closely represent those distributions in the general corpus. On all videos, the actions performed by the presenters doing deixis can be seen. Concordance lines of *this*, *that*, *these*, *those*, *here*, and *there* were obtained using AntConc (Anthony, 2014) and classified in spreadsheets by level and discipline. Raw and normalised frequencies were calculated too. (Table 4.1 and Table 4.2)

Table 4.1. Use of spatial deictics by level of achievement: raw and normalized (N) frequencies

	Tokens	<i>This</i>	N	<i>That</i>	N	<i>These</i>	N	<i>Those</i>	N	<i>Here</i>	N	<i>There</i>	N	Total	N
High	8717	175	200.8	160	183.5	34	39.0	2	2.3	9	10.3	15	17.2	395	453.1
Medium	8423	155	184.0	112	133.0	35	41.6	3	3.6	13	15.4	17	20.2	335	397.7
Low	7035	128	181.9	52	73.9	38	54.0	2	2.8	0	0.0	21	29.9	241	342.6
	24175	458	189.5	324	134.0	107	44.3	7	2.9	22	9.1	53	21.9	971	401.7

Table 4.2. Use of spatial deictics by discipline: raw and normalized (N) frequencies

	Tokens	<i>This</i>	N	<i>That</i>	N	<i>These</i>	N	<i>Those</i>	N	<i>Here</i>	N	<i>There</i>	N	Total	N
Hard	10752	245	227.9	135	125.6	78	72.5	2	1.9	19	17.7	23	21.4	502	466.9
Soft	13423	213	158.7	189	140.8	29	21.6	5	3.7	3	2.2	30	22.3	469	349.4
	24175	458	189.5	324	134.0	107	44.3	7	2.9	22	9.1	53	21.9	971	401.7

Second, to guarantee that only occurrences of spatial deictics were included, the identified sentences were manually analysed in their context using the file viewer option in AntConc along with their corresponding video. The following cases were eliminated: (1) sentences containing occurrences of the target expressions that were not expressing spatial or symbolic deixis (e.g. subordinating *that*, anaphoric *this*); symbolic deixis cases including pointing were kept (2) sequential repetitions and false starts (e.g. *this*, *this*, *this microscope was used...*; *this use of [fs] we use the microscope...*); (3) cases where spatial deictic expressions alignment with a deictic action was not seen on video.

Third, the deictic actions that were considered were gazing, body alignment, pointing, and position in regard to the screen (Rendle-Short, 2006). Change of any of these aspects was also considered in the selection of sentences, for there were cases in which such gestures (e.g. pointing actions) were almost static, and it was not clear whether the coordination of the deictic gesture and the deictic expression was intentional.

Fourth, after the identification of cases of multimodal spatial deictics, new frequency and significance analyses were performed by level of achievement and discipline (Table 4.3 and Table 4.4).

Table 4.3. Spatial deictics expressed in raw and normalised (N) frequencies by level of achievement

	Tokens	<i>This</i>	N	<i>That</i>	N	<i>These</i>	N	<i>Those</i>	N	<i>Here</i>	N	<i>There</i>	N	Total	N
High	8717	80	91.8	15	17.2	21	24.1	1	1.1	5	5.7	4	4.6	126	144.5
Medium	8423	67	79.5	1	1.2	24	28.5	0	0.0	12	14.2	7	8.3	111	131.8
Low	7035	59	83.9	1	1.4	13	18.5	0	0.0	0	0.0	2	2.8	75	106.6
	24175	206	85.2	17	7.0	58	24.0	1	0.4	17	7.0	13	5.4	312	129.1

Table 4.4. Spatial deictics expressed in raw and normalised (N) frequencies by discipline

	Tokens	<i>This</i>	N	<i>That</i>	N	<i>These</i>	N	<i>Those</i>	N	<i>Here</i>	N	<i>There</i>	N	Total	N
Hard	10752	136	126.5	9	8.4	47	43.7	0	0.0	16	14.9	6	5.6	214	199.0
Soft	13423	70	52.1	8	6.0	11	8.2	1	0.7	1	0.7	7	5.2	98	73.0
	24175	206	85.2	17	7.0	58	24.0	1	0.4	17	7.0	13	5.4	312	129.1

The general analyses of normalised frequencies show that there are bigger differences in the disciplinary divide than in the level of achievement. These bigger differences are evident in the use of *this* (more recurrent in the hard disciplines than in the soft ones). Deictic *that* frequency of use difference, however, seems to be more relevant in the level of achievement divide. To know what frequency differences are statistically significant, log likelihood and BIC calculations were performed based on the raw (observed) frequencies.

4.6.2. Statistical significance analysis

The purpose of the significance analyses (3.4.2) is twofold: (1) identify what frequency differences are statistically significant as expressed by LL values and (2) determine the amount of evidence against null hypotheses. In this chapter, I decided to analyse and report cases whose Log Likelihood (Table 3.4) and BIC (Table 3.5) values were above 10.83 and 6 correspondingly.

The null hypotheses for the two comparisons are

- **Level of achievement:** *there is no difference between high, medium, and low achievers in terms of spatial deictics use frequencies.*
- **Disciplines:** *there is no difference between hard and soft discipline students in terms of spatial deictics use frequencies.*

Table 4.5. Significance analyses of spatial deictics use by levels of achievement

	Observed frequencies			Totals	log likelihood	Bayes Factor BIC
	Hig h	Medium	Low			
<i>This</i>	80	67	59	206	0.77	-19.42
<i>That</i>	15	1	1	17	20.09	-0.10
<i>These</i>	21	24	13	58	1.64	-18.55
<i>Those</i>	1	0	0	1	2.04	-18.15
<i>Here</i>	5	12	0	17	14.91	-5.28
<i>There</i>	4	7	2	13	2.28	-17.91
Corpus sizes	8717	8423	7035	24175		

Table 4.6. Significance analyses of spatial deictics use by discipline

	Observed frequencies		Totals	log likelihood	Bayes Factor BIC
	Hard	Soft			
<i>This</i>	136	70	206	38.70	28.60
<i>That</i>	9	8	17	0.49	-9.60
<i>These</i>	47	11	58	32.76	22.67
<i>Those</i>	0	1	1	1.18	-8.92
<i>Here</i>	16	1	17	19.50	9.40
<i>There</i>	6	7	13	0.01	-10.08
Corpus sizes	10752	13423	24175		

As can be seen in Table 4.5 Table 4.6, only deictics *this*, *these* and *here* for the disciplinary divide passed my test, which is in alignment with Simpson-Vlach's (2006) finding that *this*, *these*, *those*, *here*, *there*, and *then* were more common in the hard sciences. In my study, this means that we can be sure that the frequency differences among the disciplinary groups for the gestural and verbal use of *this*, *these*, and *here* are not random; additionally, the evidence against the null hypothesis is very strong.

4.7. How Colombian PhD researchers perform deixis

The multimodal analysis in this chapter focuses on two main aspects: the synchronization of talk and gesture and the interaction with images or things being pointed at. To facilitate these analyses, I adopted and adapted Rendle-Short's (2006) CA transcription conventions (Appendix F) which in turn were based on the system developed by conversation analyst Gail Jefferson (Jefferson, 2004). I adopted this system as it allows a straightforward representation of paralinguistic aspects in synchrony with words. One of the differences in the system that I developed is the inclusion of several pictures to capture changes in gaze direction, hands position, or body alignment, and the context of realisation of deictics. For gesture and talk orchestration, I relied on (McNeill, 1992) two of three rules of synchronization and Rendle-Short's (2006) definition of the deictic process explained in 4.4.

Two examples of use for each of the most significantly frequent spatial deictics (*this*, *these*, and *here*) will be analysed, one from the hard disciplines and another from the soft ones.

4.7.1. *This (hard science)*

Example (1)¹¹ shows the use of *this* by a hard science (chemical engineering) student. As can be observed in the sequence of pictures (*Figure 4.2*), deixis is performed both linguistically and gesturally complying with two of the three rules for gesture and talk synchrony: semantic and pragmatic synchrony rules (McNeill, 1992).

(1) *Copper is good ion because eh copper is present in (many cells) and can bind with proteins and enzymes. Eh **this picture** show the mechanism of action of eh ionophore of copper two.* (S-M-CQUI-1)

¹¹ In these examples, I first present the sentence that will be analyzed in the multimodal analysis charts (1). The underlined segment of the sentence is the part that is analyzed multimodally. Each multimodal analysis chart is composed of 2 columns and 4 rows. The left column has the transcription of verbal and nonverbal actions. The right column has the pictures corresponding to the descriptions on the left. Rows show the sequence of actions that happened as the presenter spoke. *Italics* show the words that the presenter (**Pres**) said. Words in normal type describe the actions that occur while the presenter was speaking. **Bold type** marks the spatial deictics.

Gaze: □□□□¹²
Pres: *Copper [is good ion*
Hands: BH home position¹³
Place: LL of □
Body: /////////////// (steps back)



Gaze: _____
Pres: *because eh copper is [present in many cells*
Hands: BH hp [BH draw two circles
 outwards
Place: LL of □
Body: -----



Gaze: _____
Pres: *and can [bind [with proteins and enzymes.*
Hands: BH hp [LH moves up and down
Place: LL of □ [steps forward
Body: -----



Gaze: □□□□□□□□□□□□□□□□
Pres: *Eh [this picture show the mechanism of action*
Hands: [LH↗□ with a pointer
Place: LL of □
Body: //



Figure 4.2. Use of “this picture” in a hard-field OP

¹² For the conventions used in the transcription, see Appendix F.

¹³ Home position (HP) is the default or resting position. Usually, after a gesture is done, the hands tend to return to this position. HP tends to vary from speaker to speaker.

It could be argued that the semantic synchrony rule is not complied with since the speaker is far from the point in the screen at which she is pointing and does not get close to explain the image. However, the pointer that she is holding serves the same function as pointing on and getting close to the screen. *This*, in this sense, although referring to a picture that is far from both the speaker and the audience, is in the common space of focused attention. No differences between this type of far pointing using *this* and pointing uses of *that* have been observed.

The three steps of the deictic process (Rendle-Short, 2006) are also observed. The first step –signalling (expectation) that the image is relevant to the talk– is performed when the slide is changed with the pointer. Second, the combination of verbal and non-verbal actions can be observed in pictures 2-4 when hand movements are used at the moments in which *present in* (hands drawing circles) and *bind* (left hand moving up and down) are uttered. The first hand gesture can be said to perform a symbolic deictic function, in which the circling hands reinforce the meaning of *present in* and the space which is pointed at is an imagined cell. The second gesture does not seem to perform a representative or iconic role, but more what McNeill (1992) refers to as beats (rapid with no discernible meaning flicks of fingers or hands) that co-occur with speech. These gestures happen while she is facing the audience; then, she turns gaze, hands, and body towards the screen. Third, the incorporation of the image to talk occurs when the slide introduced in the expectation stage is referred to and includes the coordination of various semiotic resources. In addition to the spatial deictic *this picture*, which is accompanied by the sense verb *show*, typical of *guide* academic identity role (Nausa, 2016), other three kinaesthetic mechanisms are used: gaze and hand (pointer) are directed towards the screen, the body which was facing the audience is now partially turned towards the screen too.

4.7.2. *This* (soft science)

Example (2) shows the use of *this* by a soft science (education) student. The pictures (Figure 4.3) show that deixis is also performed linguistically and gesturally but the way that the deictic process happens varies in the focalization of pointing and the use of other kinaesthetic processes. The synchronization of gestural and verbal deixis complies with the semantic and pragmatic synchrony rules (McNeill, 1992).

- (2) *Eh that's probably eh that you [fs] that [fs] that you had teach [fs] many teachers eh in secondary or in certain areas, because of that problem is situated in pri [fs] in primary school and kindergarten. Like you can see in this picture eh it is a [fs] like an informal meeting of teachers, all of them are women.* (S-L-EDUC-2)

Gaze: ☐☐☐☐☐☐☐☐☐☐
Pres: *because of [that problem*
Hands: BH HP [BH move apart
Place: L of ☐
Body: // // // // // // // [_ _ _ _ _ _ _ _



Gaze: _____
Pres: *is situated*
Hands: BH move close
Place: L of ☐
Body: _ _ _ _ _ _ _



Gaze: ☐☐☐☐☐☐☐☐☐☐
Pres: *in pri [fs] in [primary school and [kindergarten.*
Hands: BH HP [LH rises a little
Place: L of ☐
Body: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _



Gaze: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐
Pres: *Like you can see in this picture*
Hands: LH ☐
Place: L of ☐
Body: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _



Figure 4.3. Use of “this picture” in soft-field OP

The three-step deictic process (Rendle-Short, 2006) is slightly different. First, the expectation step is created with the display of an image, accompanied by the title of the OP, to help set the context in the introduction slide. This use of the slide confirms the observation that images in the soft fields tend to play a more illustrative, not evidential role (Charles & Ventola, 2002; Diani, 2015; Rowley-Jolivet, 2002). Second, verbal and non-verbal action orchestration to

focus the audience's attention is observed in hand movements and body alignment change (pictures 1-3). These changes in motion seem to correspond more to a beat function (McNeill, 1992) than to a meaning support function. Third, the incorporation of the image to talk does not take the coordination of various semiotic resources. Gaze and hand-pointing are directed towards the image as a whole, not at a specific point. Body alignment remains almost unaltered. Alignment with the audience (interpersonal function) seems to be more important than interacting with the image (ideational function).

The use of *this* to show pictures in the previous two cases shows the synchronization of gestural and verbal deixis. However, the orchestration of non-verbal deixis shows at least two differences that can be accounted for by the reported tendency of images use (Charles & Ventola, 2002; Diani, 2015; Rowley-Jolivet, 2002) in the hard-soft epistemological divide (Becher & Trowler, 2001). In the hard sciences, images are reported to be used as evidence, which oftentimes comes in the form of numbers or images representing what is talked about while in the soft-science the tendency is to use them as ornaments. This nature of the integrated images makes that although the language that is used to refer to them is similar (*as you can see, this image, this picture*), the way that presenters integrate images to talk is different. As can be seen in these two examples, focalized pointing in the chemistry presentation was accompanied by the integration of all the analysed elements: gaze, direction, body alignment, and proximity to the screen (slightly altered by the use of the pointer). In the education presentation, although gaze and pointing were aligned towards the screen, the body was aligned towards the audience arguably giving prominence to the interactional function. Additionally, the pointing was panoramic and not focalized.

4.7.3. *These* (hard-science)

Example (3) illustrates the use of *these* in an engineering OP. The synchronization of gestural and verbal deixis (*Figure 4.4*) complies with the phonological, semantic, and pragmatic synchrony rules (McNeill, 1992). The integration of the image in the deictic process includes a multi-focal pointing and, like the case of *this* for the chemical engineering OP, the alignment of the four kinesics factors.

- (3) *So what we are going to do is eh first try to modelate using a large scale modeling, how the seismic wave is going to propagate, so, **these examples** that I'm going to show to you, are notable simulations that have been developed in the Unites States they are called Terashake, Shakeout and Chino Hills.* (S-H-INGE-6)

Gaze: □□□□□□□□□□
Pres: *using a large scale modeling*
Hands: LH ↗ □
Place: L of □
Body: //////////////////////////////////



Gaze: □□□□□□□□□□
Pres: *how the seismic wave*
Hands: LH moved to face
Place: steps back to LL of □
Body: //////////////////////////////////



Gaze: □□□□□□□□□□
Pres: *is going to propagate*
Hands: LH ↘ □
Place: steps forward to L of □
Body: //////////////////////////////////



Gaze: □□□□□□□□□□□□□□□□
Pres: *so, these examples that I'm going to show to you*
Hands: LH → □
Place: moves to R of □ and walks backwards to L of □
Body: //////////////////////////////////



Figure 4.4. Use of “these” in a hard-field OP

The three steps of the deictic process (Rendle-Short, 2006) are performed as follows. Firstly, the expectation step, like the hard-science example (1) above, starts with the change of the slide and is complemented with the student’s pointing at the second bullet point on her slide (*realistic modelling...*). Secondly, multimodal alignment to focus the audience’s attention towards the image is observed when she quickly steps back and forth to then point at the sequence of images (*these examples*) under the text on the slide. The purpose of the images is to

present the examples that the presenter will explain in the reminder of her talk. The images play an evidential role (Charles & Ventola, 2002; Rowley-Jolivet, 2002). Thirdly, the incorporation of the sequence of images to talk is performed making use of several semiotic resources as she says *these examples that I'm going to show to you*. Gaze and hand-pointing are directed towards the sequence of images as a cluster. This cluster pointing is emphasized by her touching the three images as she moves forward and the way she opens her hand (as if trying to touch them all). Body alignment with the audience is partial; interacting with the image seems to be more important (ideational function). The way the presenter interacts with images has been previously described as typical of the academic identity *guide* and *architect* roles (Nausa, 2016; Tang & John, 1999; Zareva, 2013) and are evident in the way that the student announces what is going to come in the 'academic tour' at the same time that she gives organization to her discourse.

4.7.4. *These* (soft science)

Example (4) shows the use of *these* by a soft science (law) student. In this example, deixis is performed both linguistically and pragmatically (*Figure 4.5*) with the three rules for gesture and talk synchrony being observed. The multimodal deictic process includes a not focused pointing—like the case of *this* for soft-sciences—and the alignment of the four kinesics factors.

(4) *and the country who re fs which receives the investment is called host con country. Could be any investment. Between **these important actors** eh of international law there are several several obligations...* (S-M-DERE-1)

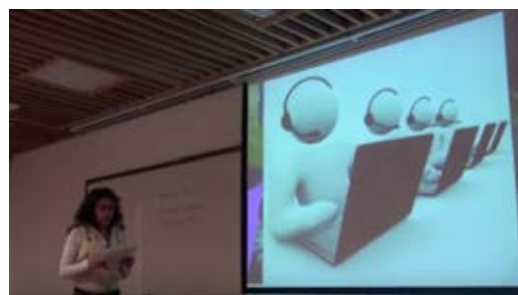
Gaze: _____
Slide: **FACTORY**
Pres: *the investment is called host con country.*
Hands: HP (RH slightly raised) script on LH
Place: LL of ☐
Body: //////////////////////////////////



Gaze: ☐☐☐☐☐☐☐☐☐
Slide: **FACTORY TO PEOPLE**
Pres: *Could be any investment.*
Hands: RH slightly raised holding script
Place: LL of ☐
Body: //////////////////////////////////



Gaze: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐
Slide: **PEOPLE**
Pres: (silently reading script)
Hands: BH holding script
Place: moves to L of ☐
Body: //////////////////////////////////



Gaze: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐
Slide: **PEOPLE**
Pres: *Between these important actors eh of inte...*
Hands: RH ☒☐
Place: L of ☐
Body: ////////////////////////////////// _ _ _ _ _



Figure 4.5. Use of “here” in a soft-field OP

In this presentation, we can observe the three-step deictic process with some variations related to the transition between slides and the type of images used. First, the expectation step, the same as other presentations, is created with the change of slides. The purpose of this change

of slides is to introduce a new topic: *actors of international law*. Like the images in the other soft field presentations, the iconic images in this presentation (a layout of a factory and a group of mannequins in front of computers) play a more illustrative, not evidential role (Charles & Ventola, 2002; Diani, 2015; Rowley-Jolivet, 2002). Second, the linguistic and paralinguistic modes coordination to indicate to the audience where to focus on is slightly evident in the way the presenter handles the script and directs her gaze towards the screen. Third, the incorporation of the iconic images to talk occurs in a not so highly multimodal way. Although gaze and hand-pointing are directed towards the image as a whole and the body is partially aligned towards the screen, the fact that the speaker does not get close to the screen and that focal pointing is not observed nor needed, given the abstract not specifying nature of the images, makes this coordination of modes a more incidental matter.

The use of *these* to show images in the last two examples confirms the observations in the analysis of *this*. The way that speakers coordinate modes is clearly influenced by the epistemological nature of the information conveyed by the images that their disciplines use. In the hard science example, images were used as evidence while in the soft science example, images played a more illustrative role: clearly, the presenter could have omitted them without compromising her content. In these cases, the expressions to refer to the images reflected the epistemological nature of information (*these examples*-evidence; *these actors*-abstractions); the examples are necessary to make the point, and the images serve the purpose of facilitating understanding for the audience while the actors in the law OP are the topic that is being discussed, and the images do not serve the purpose of identifying them. As a result, the way that presenters gesturally interact with images is different. This is confirmed by the focalized pointing (at the cluster of images-examples) in the engineering presentation with complete alignment of

gaze, pointing, body, movement, and proximity to the screen. In the law presentation, although gaze and pointing were aligned towards the screen, the absence of pointing focalization, proximity, and movement confirm the illustrative role of the images.

4.7.5. *Here* (hard science)

Example (5) illustrates the use of *here* in a hard field (biology) OP. The sequence of pictures (*Figure 4.6*) shows the synchronization of gestural and verbal deixis complying with the phonological, semantic, and pragmatic synchrony rules. The deictic process includes a focal pointing on a map and, like the other cases of hard-sciences analysed here, the alignment of the four nonverbal modes.

- (5) *In particular, the populations in Cueva de los guácharos. As we **can see here**, this is too close to the [fs] to the lowlands* (S-H-CBIO-6)

Gaze: _____

Pres: *in particular,*

Hands: □↗RH

Place: R of □

Body: -----

Home position (HP): BH clasped at waist holding a pencil



Gaze: □□□□□□□□

Pres: *the populations [in*

Hands: □↗RH [RH returns to HP]

Place: R of □

Body: //////////////



Gaze: _____ [□□□□□]

Pres: *Cueva de los [guacharos*

Hands: BH home position

Place: R of □

Body: -----



Gaze: □□□□□□□□

Pres: *As we can see here,*

Hands: □↗RH circling

Place: R of □

Body: //////////////



Figure 4.6. Use of “here” in a hard-field OP

The three-step deictic process is performed multimodally. The expectation step started with the introduction of the map slide. The audience attention focus step is observed in her

alternation of modes to point at the map and address the audience. The map clearly plays an evidential role: it shows the areas in which woollen monkeys can be found: blue dots (lowlands) and red dots (high lands). The incorporation of the maps and its areas to talk is performed in a highly multimodal way. Gaze and hand-pointing (aided by a pen) are directed towards specific areas and point in the map as she moves her hand from one dot to the other. Like the engineering OP, body alignment with the audience is partial; interacting with the image (evidence) is more important (ideational function).

4.7.6. *Here* (soft science)

This example of the use of *here* in a soft field (education) speech shows the synchronization of gestural and verbal deixis complying with the phonological, semantic, and pragmatic synchrony rules, but with a deictic process (*Figure 4.7*) that does not include pointing at the image and with a low alignment of the nonverbal modes.

- (6) *So the first challenge is [reading 3] globalization means a new social and economic and political order that necessarily presupposes a close relationship among different countries [reading 3]. So eh in this case, as you can notice **here for example in the university** there are some exchange programmes (S-H-EDUC-1)*

Gaze: ☐☐☐☐☐ _____
Pres: *presupposes a close relationship*
Hands: RH ☐ LH holds script
Place: L of ☐
Body: //////////////////////////////////



Gaze: _____ ☐☐☐ _____
Pres: *among different countries*
Hands: BH return to HP holding script
Place: L of ☐
Body: -----



Gaze: ____ ☐☐☐☐☐☐☐☐
Pres: *So eh in this case,*
Hands: BH home position
Place: L of ☐
Body: //////////////////////////////////



Gaze: _____
Pres: *as you can notice here for example in the*
Hands: LH moves horizontally and points at floor
Place: L of ☐
Body: -----



Figure 4.7. Use of “here” in a soft-field OP

In this speech, the three-step deictic process is very similar to the other two soft science examples in the low interactivity with the image, which also performs an illustrative role. The expectation step is created with the transition from one slide to the next. The purpose of this transition is to introduce one challenge (out of 3) why intercultural communication should be promoted in EFL classrooms. Like the images in the other soft-discipline presentations, the image on this slide (a World map) does not play an evidential role. Second, the linguistic and

paralinguistic modes coordination to get people's attention towards the slide is slightly evident in the way the presenter alternates her gaze between the computer screen, the video-beam screen, the script, and the audience. Third, the incorporation of the visual to talk occurs in a low multimodal way. The body is aligned towards the audience all the time and the pointing gesture is not directed towards the image but towards the ground to refer to the place where the presentation takes place (*here for example in the university*). The speaker does not get close to the screen nor interacts with the image on it.

It is interesting that although the last two examples share the use of maps and the use of *here*, the way deixis is performed highly varies in the amount and ways that multimodal resources are used. In the biology OP, the map and the areas pointed at were used as vital information to illustrate the point being made while in the education OP, the map was just used to represent the theme being discussed: interculturality and globalization. Like the other two soft-science presentations, the image could have been left out without serious effect on audience comprehension. Also, in this comparison, the way presenters use modalities to do deixis is determined by the images used. In the biology OP, complete alignment of gaze, pointing, body, movement, and proximity to the screen was observed while in the education OP, gaze and body were mostly aligned with the audience and pointing was performed more to emphasize than to specify. Lack of proximity to the screen and interaction with the image confirm the illustrative role of the image.

4.8. Conclusion

In this chapter, I have attempted to identify features of multimodal deixis that differentiate OPs performance in the group of Colombian PhD researchers. The identification of differences was based on the following hypotheses:

1. **Level of achievement:** *there are differences that discriminate between high, medium, and low achievers in terms of how they perform deixis in their OPs.*
2. **Disciplines:** *there are differences between hard and soft discipline students in terms of how they perform deixis in their OPs.*

To test these hypotheses, statistical significance and corpus analyses of deictics *this, that, these, those, here and there* were performed. Spatial deictic frequency differences that passed my statistics test were then considered for multimodal analysis. None of the frequency differences in the level of achievement divide passed the test. However, in the disciplinary divide, spatial deictics *this, these, and that* frequency differences were found to be statistically significant and very strong evidence against the null hypothesis. In the three cases, these deictics were much more frequent in the hard disciplines than in the soft disciplines, which is clearly in alignment with the findings in Simpson-Vlach (2006) study on oral academic discourse.

These deictics were then analysed from a multimodal discourse analysis perspective in 6 oral presentations (3 soft-discipline and 3 hard-discipline) to identify specific verbal and nonverbal differences between the speakers. The analyses confirmed that the way deixis is done in OPs is relevant in the disciplinary divide. Two frames of reference for the multimodal discourse analysis were used. McNeill's (1992) synchrony rules were useful in the selection of

cases; Rendle-short's (2006) description of the three-stage deictic process was useful especially in the third stage: integration of images to talk. The first (expectation) was similar in 5 of the six cases (slide change), so this was not found to be a specific difference in the way deixis is performed. The differences started to be noticeable in the second stage when speakers used multimodality to attract the audience's attention; the more important the image is in terms of reinforcing the content (ideational function) the more multimodal resources are geared towards it. Finally, in the third stage, integration of image to talk, the ideational function was highly evident in hard-discipline OPs as evidenced in the alignment of multimodal resources to point at images that were used as evidence. Soft discipline presentations tended to favor interaction as evidenced in the split use of verbal and nonverbal resources between interaction with the audience and integrating speech and images, which played an illustrative role.

Table 4.7. Summary of findings¹⁴ deixis study

Comparison	Spatial deictics (Log likelihood> 10.83; Bayes Factor>6)	Patterns	Overuse (↑) or underuse (↓)	Role of images	Gestural and verbal deixis orchestration (pragmatic and semantic)	The deictic process - Expectation image is important to talk - Combination of gaze, pointing, location, body alignment - Incorporation of image to talk
Level of oral achievement (high, medium, low)	<i>That</i> (LL: 20.09 / BF: -0.10)		High rated OPs (↑)			
	<i>Here</i> (LL: 14.91 / BF: -5.28)		Medium rated OPs (↑)			
Disciplinary divide (hard vs soft)	<i>This</i> (LL: 38.70 / BF: 28.60)	<i>This picture shows As you can see in this picture</i>	Hard-field OPs (↑)	HARD: evidence	HARD: yes	HARD Expectation (slide change) Combination (hands emphasize meanings expressed / other non-deictic functions (beats, iconic) / alignment with audience) Incorporation: multimodal resources to image (ideational) SOFT
	<i>These</i> (LL: 32.76 / BF: 22.67)	<i>These examples These NOUN</i>	Hard-field OPs (↑)	SOFT: Illustrative / background	SOFT: yes	Expectation (set at intro / slide change) Combination (hands emphasize meanings expressed / other non-deictic functions (beats, iconic) / alignment with audience) Incorporation: multimodal resources divided bw audience (interactional) and image (ideational)
	<i>Here</i> (LL: 19.50 / BF: 9.40)	<i>As we/you can see/notice here</i>	Hard-field OPs (↑)			

¹⁴ In addition to the summary of the main points in the thesis, this table includes patterns that were found in the analysed OPs. Similar summary tables are provided in the conclusion sections of the other 4 studies.

CHAPTER 5

***YOU*: GIVING AN IDENTITY TO THE AUDIENCE**

5.1. Introduction

This chapter focuses on the way that presenters engage the audience by assigning academic identities to them as evidenced in the presenters' use of *you*. The idea of doing this type of analysis emerged in module 2 from the observation that when presenters project a specific identity with first person pronouns, they almost automatically assign one to the audience. The purpose of the chapter is then two-fold: first, it seeks to demonstrate that whenever a speaker assigns an identity to themselves, a mirror identity is assigned to the audience; second in line with the main aims of the thesis, the chapter seeks to identify uses of *you* in OPs that discriminate students' oral performance. These aims are concretized in the following questions:

1. Do *I*-presenter identity roles imply *you*-audience roles in oral presentations?
2. What are the tendencies in audience identity role projections in OPs when students use *you*?
3. What *you*-audience identity roles are useful in discriminating among students' levels of oral performance and disciplines?

To answer these questions, this chapter first presents a summary of the research that I conducted in module 2 (section 5.2). Second, in the review of the literature (5.3 below), I present two theories that explain meanings and uses of *you* and the available discourse analysis research on the use of *you*. Third, I answer question 1 demonstrating that the identity roles identified in

the first study (Nausa, 2016) imply mirror image *you*-identities (section 5.4). The information in 5.4. provides the model of analysis for the remainder of the chapter: quantitative (5.5) and discourse (5.6) analyses, in which the answers to questions 2 and 3 are provided.

5.2. Self-mentions and the projection of identity roles in OPs

In module 2 (Nausa, 2016), I analysed first person pronouns and possessive adjectives (*I, my, me, we, us, our*) to identify how PhD researchers in this study interacted with their audience and positioned themselves in the oral presentations. Two variables were considered in the analysis: students' level of achievement in the OPs (high, medium, and low) and the disciplinary divide represented in the class (hard vs soft disciplines (Becher & Trowler, 2001)). Raw and normalised frequencies, and significance tests were calculated to identify what pronouns were more frequently used by each group and what academic identity roles (Tang & John, 1999) were invoked with each pronoun. Additionally, discourse analyses were performed to identify the discourse functions performed by each projected role and recurrent *I* patterns (Hunston & Francis, 2000), understood as "... a phraseology frequently associated with (...) a word, particularly in terms of the prepositions, groups, and clauses that follow the word" (p. 3).

To perform the identity role analyses, I expanded Tang and John's (1999) taxonomy of authorial stance identity roles to include two new role categories: knowledge contribution and language use roles. In this expanded taxonomy (*Figure 5.1*), academic identity roles show a continuum of appropriation of knowledge in the presenter's field; as such, individuals can present themselves with a very low authorial stance position as mere member of their communities (*representative*, see example 1) or with a very high authorial stance position as producers of knowledge (*originator*, see example 2).

- (1) *Hi, good afternoon, my name is (name) **I'm enrolled** in the law doctorate programme here in this university.* (S-H-DERE-2)
- (2) ***In my (unintelligible) research I study** all municipalities in Colombia and **I found**, un minute, **I found** that eh the democracy have different trajectories eh in all country.* (S-M-CPOL-2)

In the second category in the taxonomy, knowledge contribution roles¹⁵, students show themselves as contributors of novel information for the class (students from other PhD programs). In this set of identity roles, presenters use pronouns to project themselves as *receivers*, *co-constructors* or *providers* of the new shared knowledge. Sentence 3 shows an example of *co-constructer* projection.

- (3) *Yeah, finally, we release to the market this product and ideally the software I better, but **like we know** this doesn't happen, yeah? **I don't know if** you remember but some operative systems every time you have to change (them), yeah? **I don't know if** you remember Millennium or XP, yeah.* (S-H-INGE-3).

In (3) the presenter constructs her explanation by activating knowledge that is common to her and the audience. The use of *like we know* and *I don't know if you remember* can be interpreted to be a way to soften a potential- face threatening event: a PhD-level audience is expected to know about computer operating systems, so referring to them as something new could potentially be interpreted as insulting. The presenter construes the audience as her equal whilst at the same time giving the necessary information.

¹⁵ This category of taxonomy was originally proposed by Zareva (2013). However, there are two differences between her proposal and my review. First, she refers to this type of identity role as social, while I consider it a classroom role. Second, I include an intermediate role (co-constructer) between the extreme roles provider-learner.

The last category, language use roles, was inspired by how students use language or deal with communication breakdowns. The first two roles (*learner* and *independent user*) are similar in the sense that they are projected when students experience difficulties in language use. The difference lies in the fact that *learners* ask for help (*sorry, I don't remember the word for 'secuestro', what is it?*) and *independent users* make use of mechanisms such as paraphrase or circumlocution to deal with communication breakdowns. Finally, *providers* contribute new language knowledge usually in the form of specialized vocabulary. The following is an example of the projection of this role:

- (4) B: *Eh is [fs] is difficult because is eh is a, [fs] **we create** [fs] **we we use this word** [“entitlement”] because is not, is [fs] is is different than “right” but is kind of right and **we use “sujeto de derecho”** or [fs] in a positive way or “derechos extendidos” and **we try to** choose between two.* (S-H-ADMI-1)

These last two new category roles were inspired by Lave and Wenger (1991) concept of *communities of practice* and *peripheral participation* adapted to the EAP class.

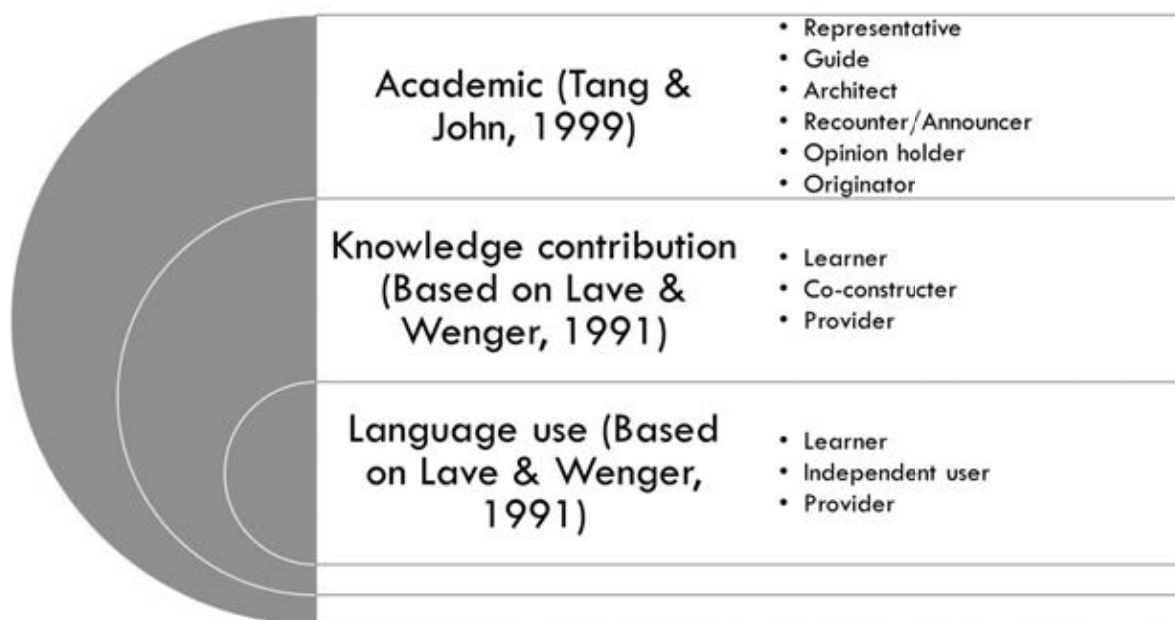


Figure 5.1. Typology of identities behind the first person pronoun (Nausa, 2016)

As expected, the roles that were projected the most were those related to how presenters position themselves in their academic communities, closely followed by knowledge contributions roles (mainly *co-constructor*). Language use roles were not projected as frequently. The analysis of the two variables: level of oral achievement and disciplinary divide show specific tendencies.

In the level of achievement analyses, *guide*, *originator* and *co-constructor* were identified as the roles that significantly discriminated among students' oral performance. In *guide* projections, presenters show and explain visuals to the audience (see deictics chapter). In *originator* projections, presenters clearly show what their contributions to their fields are. In the projection of the *co-constructor* identity, presenters anticipate moments in which their audience might need further content elaboration in a way that respects their audience's PhD researchers' status. These three roles imply a high degree of engagement between the presenter and the audience. High-rated OPs exhibit the projection of these roles on a more frequent basis. In other

words, high achievers more consistently use self-mentions to guide their audience through the presentation, claim ownership over findings or creations, and help hearers understand their contents in a way that does not underestimate their status as PhD students.

In the disciplinary analyses, *recounter/announcer* and *opinion-holder* were the identity roles whose projection frequency differences were identified as significant of soft and hard disciplines, correspondingly. Quantitative analyses demonstrated that hard discipline students more frequently project themselves as adopters of the methodological practices of their fields (example 5), while soft-discipline students more frequently invoke their positions regarding knowledge in their fields (example 6).

- (5) *For that, **we have to** know about decisions variables. First of all, **we have to** identify the different aspects around the decisions, but with the identification of the aspects is not eh [fs] is not enough.* (S-H-INGE-2)
- (6) *Eh, this approach, **I think that** is the solution to understand self-deception. Why? Because eh, **I think that** this approach eh try self-deception as another process [fs] psychological process and is eh [fs] is better no [fs] eh (certain) measure this mental state.* (S-H-PSIC-1)

5.3. The study of *you* in oral academic discourse: review of the literature

The purpose of this literature review is to provide the theoretical foundations that will be used to answer question 1 (*I*-roles imply *you*-roles) and to survey studies that have focused on the use of *you* as a mechanism of oral academic discourse producers to engage with their audience.

5.3.1. *You* in the territory of information and dynamic-inferential view of communication models

Although text producers can create a sense of engagement with the use of first person pronouns, it is with the use of second person pronoun that they directly address the audience. By directly addressing the audience, text producers establish what roles they are playing and automatically assign specific roles to their interlocutors (Wortham, 1996). Two models of analysis of *you* are useful in explaining second person pronoun use in semantic and pragmatic terms: territory of information (Kamio, 2001) and dynamic-inferential view of communication (Gast et al., 2015).

Kamio (2001) uses the concept of territory of information (*Figure 5.2*), which is based on two main assumptions. The first assumption is that, in conversation, there is a general perceived space (P2+D2) between hearer and speaker which can be divided into two spaces: the one where the conversation takes place (proximal: P2) and the area out of the conversational space (distal: D2). It has to be borne in mind, however, that Kamio warns us that these two spaces are conceived of psychologically rather than perceptually, although the two spaces (perceptual – psychological) might coincide. The second assumption is that the conversational space P2 is subdivided into two other subdomains: The speaker's space: P1 and the hearer's space: D1.

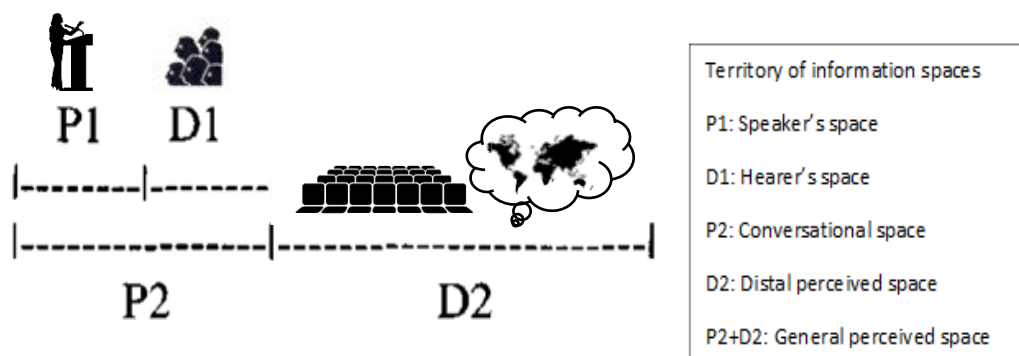


Figure 5.2. Kamio's representation of territory of information (Kamio, 2001, p. 1113) adapted to OPs

Kamio uses these four spaces to explain *I*, *we*, and *you* uses. *I* uses correspond to P1. Inclusive *we* uses correspond to P2, when both the audience and the presenter belong to the same group (e.g. the class) while exclusive *we* (Haas, 1969) uses correspond to P1, when the presenter belongs to a larger community which the audience is not part of (e.g. lab members). Personal uses of *you* correspond to D1; these uses are in contrast with inclusive *we*. Impersonal uses of *you* correspond to D1+D2; the audience is included in a larger group.

Gast et al (2015) in their dynamic view of communication specifically focus on *you* (Figure 5.3). They distinguish between *personal* (when the hearer(s) is(are) the only referent of the pronoun) and *impersonal* uses of the pronoun. Impersonal uses usually refer to larger groups and can be either *inclusive*, when the audience is construed as members of such groups, or *exclusive* when the members do not belong to the groups. Exclusive references sometimes invite the members of the audience to be part of two kinds of simulations. In *category simulations*, interlocutors are invited to imagine themselves as having properties of a kind of people (e.g. *doctors*) while *participant simulations* interlocutors are invited to imagine themselves as experiencing a situation (e.g. *you are on a blind date*).

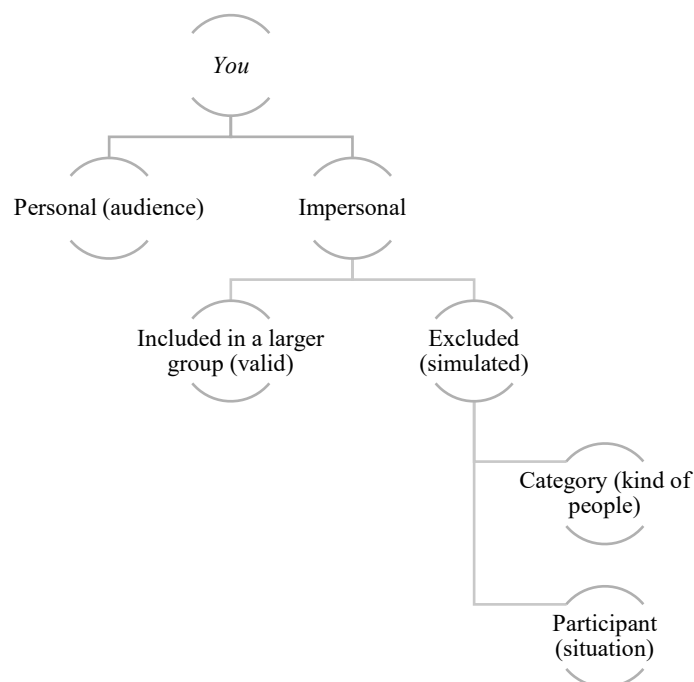


Figure 5.3. Meanings of "you" in OPs adapted from Gast et al. (2015)

Gas et al's (2015) types of *you* can be additionally described with Kamio's sub-territories (Figure 5.4). Personal uses of *you* can be described as belonging to D1: the hearer's space. Impersonal uses of *you* can be located in D1 and D2. Inclusive impersonal uses are between D1 and D2 since the audience is construed as belonging to larger groups not present in the conversation territory. Impersonal exclusive simulation uses of *you* (category and participant) can be said to be in D2: an area not in the perceived (but in the psychological) space in which are the members of the group to which the audience is invited to pretend they belong to. Although most of the studies that will be surveyed here use Kamio's model, I preferred Gas et al's, for simulated uses of *you* adequately explain three of the identity projection dyads that will be defined below, especially those in which the audience is invited to pretend they are *novice researchers* or *innovation users*.

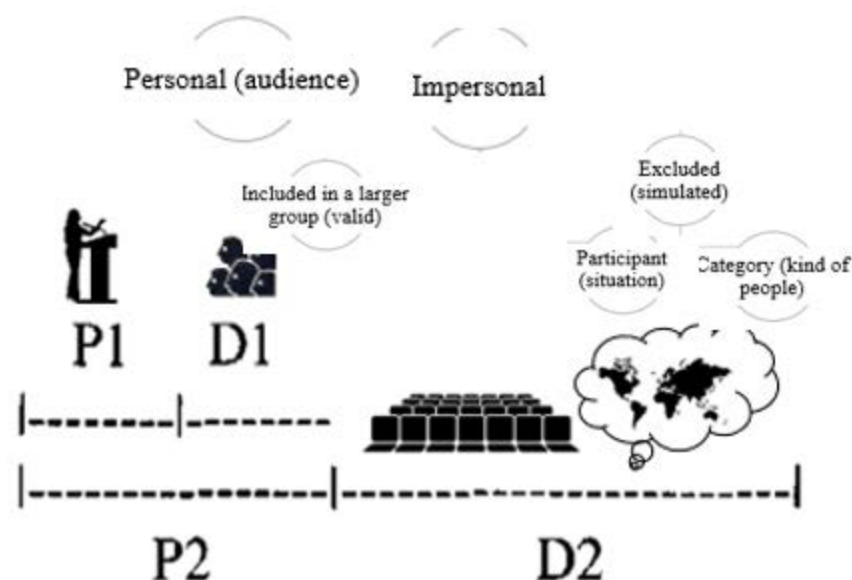


Figure 5.4. Meanings of “you” based on the territory of information (Kamio, 2011) and dynamic view of communication (Gas et al., 2005) theories

5.3.2. *You* in the study of oral academic discourses

The study of academic discourse has mainly focused on written genres. Even the study of personal pronouns use has received a greater deal of attention from the analysis of writing (e.g. Harwood, 2005; Hyland, 2001; Kuo, 1999; Lafuente, 2010; Luzón, 2009). Nonetheless, the study of *you* has been mainly approached in oral academic discourses, with a special focus on university lectures. Other genres like student conversations, student presentations, or interviews have seen studies of *you* use but on a less frequent basis. Oral uses of *you* in the academia have been approached in comparison to other pronouns frequency of use (*I*, *we*), from a functional approach to determine what it is used for (e.g. interaction with interlocutors), at different levels of instruction (e.g. undergraduate, postgraduate), at different moments of spoken events (introductions, middle, and end of lectures), in different types of language learning or use (EFL, ESL), from the perspective of language speaking nativeness (NS vs NNS), in success in oral

performance (effective vs less effective instructors), or in disciplinary variation (hard vs soft disciplines). The following is a brief description of found studies, their foci, and how they relate to each other and the present research.

Analysis of *you* in lectures has focused on instructors and students to determine the success of communication. Rounds (1987) analysed pronouns *I*, *we*, and *you* in calculus university classes taught by five English NSs and 1 NNS teacher assistants to determine the success of their performance. She found that *you* was more frequent in the monologic (as opposed to interactive) part of lectures and that it was mainly used in an impersonal sense. Non-successful teachers in this study were found to use *you* more frequently than their more successful counterparts. *You* was contrasted with *we* in this respect. Successful teachers, for example, tend to say *we mathematicians*, while non-successful ones say things like *you students*. The most frequent use of *you* in monologic sections of lectures in Rounds' study was later contradicted in a study (Morell, 2001) of two lectures (one expository and one participatory). Morell's study showed that *you* was more frequently used in the interactive lecture. One of the criticisms aimed at Morell's study is that its findings were limited given the small size of its corpus (Fortanet, 2006). Fortanet's (2006) own study of five soft science undergraduate and graduate soft-discipline lectures (a subcorpus taken from MICASE) analysing the use of *I* and *you* eventually confirmed the most frequent use of *you* in the monologic sections of lectures pointed out by Rounds' (1987) study. However, *you* in Fortanet's study was found to be more frequent than *we*; which is different to Rounds' observation that *we* was more frequent than *you*. Fortanet explains the higher frequency of *I* and *you* in her study as a change in the oral discursive practices in university classrooms that now tend to favour more interactivity than they did in the past. Other studies that followed upon Rounds' research analyse lectures' intros (J. Lee, 2009;

Yaakob, 2013) and closings (S. Cheng, 2012) also focusing on the frequencies and discourse functions of *I*, *we*, and *you* with a view on classrooms size (S. Cheng, 2012; J. Lee, 2009). Cheng and Lee's studies also use lectures from MICASE corpus while Yaakob's uses lectures from BASE. In his intros study, Lee (2009) found that the frequencies of *I* and *you* are higher in small classrooms lectures than in large classroom lectures. Based on Brown and Levinson's (1987) politeness theory, the author explains that the higher frequency of *I* and *you* in a small class might be due to the closer relationship between the lecturer and students. In Yaakob's study comparing disciplines, the higher use of *you* in physical science was found to refer to anyone in the field. The highest use of *we* in Life Sciences was found to be due to the presence of more than one lecturer in each session while in arts and humanities *we* highest frequency was found to be related to its use to refer to oneself. In the closings study, Cheng (2012) found that *I* and *you* are used on a more frequent basis than *we*, which suggests a higher number of lecturer-student interactions in closings. The author explains that a great number of *I* and *you* instances refer to the lecturer and the students, correspondingly. Pronoun use in lectures intros has also been studied in contexts where English is not an L1. In a study of Malaysian university English talk in lectures (Yeo & Ting, 2014), a comparison of *I*, *we*, and *you* showed that *you* is the most frequently used personal pronoun. *You* is used for prior knowledge activation, instructions, and announcements, which were found to be the main activities in lecture intros. Lectures use personal *you* (audience included) when referring to personal experiences or directing students' attention; they use impersonal *you* (audience excluded) in explanations of subject matter.

You and *I* in university classrooms have also been analysed from the perspective of NS and NNS student talk. O'Boyle (2014) analysed two corpora: UNITALK and ELLTTALK. Although both corpora include university talk, ELLTTALK focuses on upper intermediate to

advanced learners of English. Statistical significance analysis (log likelihood) shows that ELLTALK speakers use *you* more frequently than UNITALK students, which is interpreted as NNS students employing discourse management functions of *you* within their student-led discussions in a similar fashion to tutors in UNITALK. But the analysis of *you* in spoken registers has not been limited to the analysis of lectures. In his seminal work on university language (written and spoken), Biber (2006) found that pronouns *I*, *we* and *you* can be found in all spoken registers; however, *you* is more common in class management, office hours, and service encounters.

Another tendency in the study of *you* in oral academic registers is the study of phraseology and their associated functions. In an analysis of classroom teaching and textbooks, Biber, Conrad, and Cortes (2004) and (Biber, 2002) identified three kinds of *you* lexical bundles (patterns): stance expressions, discourse organizers, and referential bundles. One of the main findings in the study is that classroom teaching uses more stance and discourse organizing bundles than conversation. Among the *you*-bundles that the authors found are *if you want to... do you want to... you want to go* (attitudinal / modality stance bundles), *I want you to... you don't have to...* (discourse organizers) mainly present in classroom teaching. One *you*-lexical chunk that has sparked special interest given its frequency in academic talk has been *you know*, which is commonly analysed along with *I mean* in general talk (e.g. Fox Tree & Schrock, 2002; McCarthy, 2010; Stirling & Manderson, 2011). Csomay (2007) conducted a large scale study aimed at identifying how students and teachers talk vary in American university classrooms according to level of instruction (undergraduate and postgraduate) and disciplinary differences (hard and soft disciplines) in 196 transcribed class sessions in American universities. Three concepts were considered in the analyses: (1) contextual orientation, (2) conceptual focus, and

(3) personalized framing. Use of *you* was found to be a key element in contextual orientation along with other ones like first person use, modals, and non-passive constructions. Chunks like *you know* were key in the analysis of personalized framing, among others. Csomay found that teachers make use of more linguistic elements (including *you*) related to a contextual, directive orientation in Engineering, Business, and Education. The author also observed that students in five of the six analysed disciplines (with the exception of Engineering) mostly use linguistic features associated with personalized framing (*you know* included).

So far, the bulk of the surveyed studies relate to instructional (lectures) and undergraduate discourses. *You* has also been analysed, but at a lesser extent, in conference presentations. These studies are mainly comparative focusing on aspects like NS and NNS performance, comparisons with written genres, and discourse functions. In a study comparing NSs and Bulgarian (BE) speakers of English presenting their linguistics papers at conference presentations (CPs), Vassileva (2002) analysed the use of *I* and *we*, and direct address (*you*-perspective) among other interactive and identity projection features. The author found that seven out twelve BEs used the *you*-perspective to address the audience directly while nine out of ten NS did. Vassileva observes that those speakers who avoided the use of direct address are mainly young and inexperienced. Common uses of *you* in direct address to the audience include generic references (impersonal), reference to handouts, or reference to common experience or knowledge. Another study comparing NS and NNSs (French) CPs in the areas of physics and medicine (Rowley-Jolivet & Carter-Thomas, 2005b) sought to compare their organisational interactional strategies. Personal pronouns *I*, *we* and *you* in this study were analysed as mechanisms to avoid sentence subjects including nominalisations or heavily modified noun phrases, which can be difficult to process in real time for speaker and audience. This study confirms Vassileva's finding that inexperienced

and NNSs tend to underuse pronouns resulting in a diminished presenter-audience interaction. Similarly, this study also confirms the impersonal-generic references and the interactional character of *you*. In the disciplinary analysis, *you* is found to be more frequent in personal references (audience address) in medicine and in impersonal (generic) references in physics. In a follow-up study (Rowley-Jolivet & Carter-Thomas, 2005c) of conference presentation introductions in the areas of geology, medicine, and physics compared to their corresponding proceeding papers, the authors focused on how NS presenters facilitated information processing and created rapport with the audience. This study mainly focuses on the rhetorical move analysis of introductions, for which pronouns have a defining role. As expected, *I* and *you* were not found in the proceedings articles. Among the uses of *you* in the CP introductions are references to the scientific community (impersonal-generic), the audience (personal), the presentation of information on slides, reference to inanimate entities (category simulation), the purpose of which, in the authors' view, is to involve the audience in the research process. Quantitative information of *you* according to these functions is not provided for the disciplinary divide. Another study that compares CPs (in medicine) to written texts is Webber's (2005). This study confirms the rhetorical functions of *you* that most studies surveyed here have reported: *you* to address the audience, make general reference to the academic community. In addition to these functions, other uses of *you* include inviting the audience to share the claims made by the speaker, inviting the audience to pretend they are somebody else. Statistical information about *you* in this study includes the frequency of *you* occurrences in isolation or contractions (e.g. *you're*, *you've*, *you'll*) and personal or impersonal uses. Again, as expected and confirming the findings in Rowley-Jolivet and Carter-Thomas' (2005b) study, no instances of *you* were found in the written articles. The most recent study on the use of *you* in conference presentations

(Fernández-Polo, 2018) also analyses the pronoun discourse functions comparing NS and NNS. Unlike Vassileva (2002) and Rowley-Jolivet (2005a), this study finds *you* instances to be similarly distributed across the NS and NNS corpora. The author notices differences of *you* use when presenters use scripted versions of their talk, which he interprets as probably resembling written style. The referents (specific and general) and functions identified by Fernández are like those in previous studies and the ones I propose in my taxonomy of identity projection dyads below. These functions include the use of *you* to invite the audience to share data interpretation (Webber, 2005; *opinion evaluator*, this study), benefit from the action being performed (*architect's client*, *knowledge provider* and *co-constructer* in this study), thank the audience (*representative*), direct the audience attention to data (*guide-tourist*), emphasize shared knowledge (*co-constructer*). The author concludes that one of the most important differences between NSs and NNSs is the overuse of certain formulas like *as you can see* (probably the result of language instruction) that might lead to a negative politeness perception from the audience's part.

The survey of studies above allows me to draw the following conclusions. First, the polysemic character of *you* in terms of possible references and discourse functions makes the pronoun a versatile tool for the enhancement of interaction in academia but at the same time a source of trouble for inexperienced scholars or NNS of English. As some of the studies have shown, novice scholars and NNS might opt for other less interactive options or produce pragmatically inappropriate utterances. Similarly, *you* can be *double-edged*, for it can create interaction but can also be condescending (Rounds, 1987). Second, most of the research has focused on university lectures and CPs; other academic genres have not received such a great deal of attention. For my study, two genres are of special interest: student presentations and

conference presentations. Student presentations (an undergraduate or PhD researcher training genre) are important, for this is precisely what I analyse in my study, but no studies that analyse *you* were found. Conference presentations are of the utmost importance, for this is what students in my study are receiving instruction for and what grammatical accuracy and pragmatic performance in their OPs are evaluated against. These studies on CPs' use of *you* are highly relevant, for they are of a comparative nature focusing on the connection between written and spoken associated genres (aspects which I also analyse in this thesis), NS and NNS uses, and novice vs professional uses. Third, the NNSs in these studies are mainly of European descent (e.g. Bulgaria, France, Italy); no studies from Latin America or other third world countries were found. Finally, although these studies are based on corpus data, as some of their authors admit, their findings are based on small corpora and mainly based on raw or normalised frequency counts. It is expected that comparative studies also include statistical significance tests to have more sound conclusions. The purpose of this study is to contribute to the study of oral academic discourse by addressing these gaps.

5.4. This chapter: *you* and the projection of audience identity roles

This section provides the answer to the first question in this chapter: yes, it is possible to identify *you*-audience roles that are mirror images of the identity roles (5.2) in module 2 (Nausa, 2016). In the discourse analyses of OPs in that study, it was frequently observed that many of the sentences containing first person pronouns and projecting the abovementioned roles usually included a second person pronoun *you*. This can be interpreted as the presenter's not only projecting an identity but simultaneously assigning one to the audience as seen in (7).

(7) *To have great [fs] a better idea, **I**'m gonna show **you** this map. The blue dots are the populations in lowlands.* (S-H-CBIO-6).

Sentence (7) exemplifies the *guide* identity role, metaphorically explained as the presenter's taking the audience on a tour of specific knowledge in their areas. The metaphor can be extended by pointing out that if there is a *guide*, there must be a *tourist*¹⁶. By addressing the audience with second person pronoun *you*, the presenter ascribes the *tourist* identity to them. Therefore, it should be possible to identify mirror-image identities for the other *I*-roles in the cases in which presenters use *you*. As observed by (Wortham, 1996):

Speakers often use these forms [personal pronouns and other shifters] to establish what roles they are playing with respect to each other. Because of this, analysts can focus on personal pronoun use when they want to uncover participants' interactional positions. (p. 332)

To confirm the mirror-image *I-you* roles hypothesis, I created and expanded the identity role taxonomy (Figure 5.5) to analyse 690 sentences containing *you*.

The sentences containing *you* for audience identity roles analysis were retrieved from the corpus using the concordance function of AntConc (Anthony, 2014) (Appendix K). To define the new roles, I use a combination of Kamio's (2001) *territory of information* and Gast, Deringer, Haas, and Rudolf's (2015) *dynamic view of communication* theories adapted for the analysis of

¹⁶ The entailment effect observed in the description of these roles is analogous to the type of entailment observed in converses (relational antonyms) like *father – son*, *boss – employee*, or *teacher – student* (Griffiths, 2006).

the type of oral presentations in this study (section 5.3.1). The adaptation of these theories focuses on the generalization and empathy effects of *you* uses in the OPs.

5.4.1. Model for the analysis of *you* to assign identity roles to the OP audience

To define the identity role dyads, I took into consideration each role in the original presenters' roles taxonomy. I assigned to each *I*-presenter role its corresponding *you*-audience role taking into consideration converse (or relational antonym) words (Griffiths, 2006). In several cases, the definition of the role was straightforward (e.g. *guide – tourist*); in others, the same word was assigned implying that the presenter addresses the audience as equals (e.g. *co-constructor*). In other cases, in the absence of a converse, a noun phrase was coined (*architect – architect's client*).

Academic authorial stance roles	
<u>Presenter role</u>	<u>Audience role</u>
a. Representative	Representative
b. Guide	Tourist
c. Architect	Architect's client
d. Recounter / Announcer	Apprentice
e. Opinion holder	Opinion evaluator
f. Originator	Innovation user
Knowledge provision roles	
g. Learner	Provider
h. Co-constructor	Co-constructor
i. Provider	Learner
Language use roles	
j. Learner	Provider
k. Independent user	???
l. Provider	Learner

Figure 5.5. Identity projection dyads

The following are the descriptions of the roles and examples to illustrate them.

5.4.1.1. Authorial stance roles

In this category of identity roles, each role represents an authorial stance position ranging from the least powerful one (*representative-representative*) to the strongest (*originator-innovation user*)

*a. Representative - representative*¹⁷

In the projection of the *you-representative* role, the members of the audience are addressed as such, as witnesses of the communicative event taking place (the OP). *You* references are personal; the audience is directly addressed (included) making them the semantic referent of the pronoun *you*. This role projection is usually performed at the beginning of the OP, before the Q&A section, and at the end of the OP.

(8) *Good morning. Thank **you** for coming to my presentation. Eh my name is (name)*
(S-H-CBIO1)

(9) *I don't know if **you** want any question and I'm glad to answer it.* (S-H-INGE-5)

b. Guide – tourist

Like projections of the *you-representative* role, the audience is directly addressed in projections of the *you-tourist* role. Also, the semantic referent of the pronoun is the audience; that is, the audience is included in the referential scope of *you*. The projection of this role includes the use of multimodal resources (pointing, body alignment, gaze direction, etc) from the part of the *I-guide* to show visuals to the *you-tourist* audience (see [chapter 4](#)) and the audience corresponding gazing behaviours.

(10) *the property right has eh the related eh ah very bad effects. **You** can see here in this picture [fs] in this picture eh some...* (S-H-DERE-1)

¹⁷ For ease of reference, the first member of each pair is a presenter *I*-role, and the second, its complementary audience *you*-role. Descriptions will be provided for the audience roles only. When necessary, reference to the original presenter role will be made.

c. *Architect - architect's client*

In this third *you*-role, the *I-architect* metaphor implies the organization of disciplinary discourse for the *you*-audience (*the client*). *You* references are still personal, and the semantic referent of *you* is the audience, which is included in the semantic scope of the pronoun.

- (11) *I'm study a second semester. Today I'll explain **you** eh autonomous under vehicle, AUV. Eh I explain you some problems...* (S-L-INGE-2)

d. *Recounter-announcer - research apprentice*

In *recounter-announcer* self-mentions, presenters show themselves as followers of the methodological practices in their disciplines. However, when references to such practices are made with *you*, the audience is invited to be part of a category simulation (pretend they are researchers in the presenter's academic community) making these *you-research apprentice* references impersonal by excluding the audience from the referential scope of the pronoun. *You* refers to the researchers in the presenter's discipline; therefore, *you* occurrences could be replaced by exclusive *we* (Haas, 1969) pronouns.

- (12) *And they **you** take the species into the laboratory, and **you** pretend [try] to [fs] to have the same conditions in the laboratory. Conditions like temperature, humidity to maintain the [fs] the species okay. And after that **you** reintroduce the species in the eh natural environment.* (S-H-CBIO2)
- (13) *And if **you** need to obtain 100 mg of Taxol **you** need to extract of big amount of the tree.* (S-H-CQUI-2)
- (14) *First is the lineal synthesis, second is convergent synthesis and eh third is formal synthesis. First eh, the lineal synthesis is called to [fs] eh synthesis by steps. This means, eh that eh **you** can use one chemical reaction, A eh plus B eh produce C.* (S-H-CQUI-2)

According to several authors (Gast et al., 2015; Stirling & Manderson, 2011), this use of *you* has the effect of creating empathy on the audience¹⁸.

e. Opinion-holder - opinion-evaluator

In the projection of this audience identity role, the audience is directly addressed and invited to be a judge of the opinion/statement the presenter has expressed. References are personal again and the audience is included in the referential scope of the pronoun.

(15) *I am convinced that the bamboo fiber is a good alternative to replace the asbestos fiber; eh I hope that **you agree with me**. But now I need to develop or [Fs] develop a mechanism to put that fiber into cement in order to produce a appropriate composite fiber cement material.* (S-H-INGE-5)

f. Originator - innovator user

I-originator self-mentions convey the idea that the presenter is the creator of new knowledge or inventions. When presenters use *you* to construe their own intellectual production, the audience is invited to be part of a simulation: pretend they are end-users of the knowledge or technology that the presenter is contributing to their field. As in *you-research apprentice projections*, references are impersonal.

¹⁸ An interesting example of the empathy effect of *you* is that of prisoners' stabbing narratives (O'connor, 1994).

- (16) *Yeah. Okay, another eh application is domotics. Domotics is [fs] is called intelligent homes. **You** can open the doors or close the doors, or turn on or turn off the lights, but not in **your** home, far for **your** home. Maybe with a [fs] just a cellular call or local call, or using internet it's the same thing. Yes, also eh it's very very very (usive) at the air conditioner and heaters. On the air conditioner, for example if **you** are in a country that the summer is on July approximately, and if **you** are eh at **your** office **your** home is very very very very hot, **you** can turn on the air conditioner and when **you** eh come to **your** home, **your** home is (confort). (S-H-INGE-1)*
- (17) *You know this solution because **you** have opera [fs] Eh Microsoft operative system and all the time **you** have to install this patches, yeah? The filosi [fs] philosophy of this eh solution is [reading 1] if it works, don't fix it [reading 1], yeah? Then if no one client eh complain that the bug, **you** don't have to fix it, yeah? (S-H-INGE-3)*

5.4.1.2. Knowledge provision roles

In this category, presenters and audience relationship is based on who teaches what to whom. In this PhD student community of practice (Lave & Wenger, 1991), sharing one's research with the class in a way that everybody understands was key for legitimate participation.

g. Provider - Learner (of knowledge)

This was hypothesized as the acknowledgement of a classmate's previous contribution, but no cases of this role were found. In module 2, projections of this role were found with *us*.

h. Co-constructer - co-constructer

The use of and category-simulation *you* and inclusive and exclusive *we* in this dyad help to create a feeling of joint endeavour between presenter and audience in the unfolding of the OP. *You*-references can be either personal or impersonal: personal, when direct questions are made,

or instructions given to the audience; impersonal, when presenters refer to common shared knowledge that is needed to explain something or when reference to general knowledge is made by assuming that the audience is part of larger groups of people.

- (18) *To contextualize my presentation I'm going to remember a situation that happened in Bogotá a few years ago eh and that is possible that **you** have [fs] **you** have eh seen this on the news.* (S-H-DERE-1)
- (19) *As a result of this kind of reparation, Unión Patriótica was able to participate in 2014 political elections with its president Aida Avella, if **you** remember, she was the vice-president of the Polo Democrático alternative with Clara López the last year, and Unión Patriótica will be able to do the same in 2018 elections* (S-H-DERE-2)
- (20) *... blood and they rub their provosis along the eh stridulatory organ. Do **you** know what is a guacharaca, the musical instrument here in Colombia?* (S-H-CBIO-4)

i. *Learner- provider*

In the projection of this role, references are personal, and the audience is directly addressed (included). Presenters use *you* to refer to the audience as the receivers of new information.

- (21) *And I want **you** to keep some messages to **your** home. Eh [reading1] the use of vibrations in insects is more common than one would expect [reading1].* (S-H-CBIO-4)
- (22) *In most companies we're gonna find a [fs] a lot of related processes in this easy operation. Let me give **you** one example, ok, once the lack of a product is identified in the company, then a person has to fill a form requiring that product* (S-H-INGE-7)

5.4.1.3. *Language learning-use roles*

In this category, presenters and audience relationship is based on how language is used or how communication breakdowns are faced. Also, teaching new language (usually technical vocabulary) is key in the definition of the three roles. In this community of practice (Lave & Wenger, 1991), always using English was key for legitimate participation.

j. Provider / learner (of language)

In this role you-projections, references are personal and inclusive. The audience is addressed in cases in which the presenter needs help with language.

(23)W: *three times. Okay, first, you predict. Eh when you play, how do **you** say “ajedrez”?*

A: *chess*

W: *I’m sorry. When you play eh*

(S-L-INGE-2)

k. Independent user - ???? (of language)

In this dyad, the *I*-role was described as the presenter facing difficulties in the use of English but being able to solve them by themselves. In module 2, the identification of this role was straightforward, and it was based on self-initiated, self-repaired (Schegloff, Jefferson, & Sacks, 1977) stretches of discourse. However, in the present study, it was difficult to imagine what the mirror *you-audience* role would be. No cases of this hypothesized role were found.

l. Learner / provider (of language)

Although this dyad was easy to define, no cases of *you* being used to construe the audience as providers of language were found.

The definitions and examples in this section have demonstrated that the academic, knowledge provision, and language use identity roles can be redefined as dyads in which one presenter identity role implies an audience identity role. The following sections of the chapter focus on how the projection of *you*-audience identity roles can be used as marks to describe these PhD researchers OP language.

5.5. Method and quantitative analyses

5.5.1. Selection of sentences for analysis and data clean-up

To determine how presenters use *you* in their oral presentations, the following procedures were followed. First, all cases of *you* were retrieved from the 72811-token oral subcorpus. Anomalous sentences that contained pronouns occurring in a false start and then abandoned, sequential repetitions, and words transcribed as unintelligible were eliminated to avoid inflation (Zareva, 2009). Second, the resulting 690 sentences were manually analysed and classified in Excel spreadsheets according to the audience identity projection model described in 5.4.1 by level of achievement and the disciplinary divide. Third, to validate the classification of sentences, a colleague was trained and invited to recode 50 randomly selected instances of the data. Krippendorff's alpha was calculated for interrater agreement. The value (0.715) was found to be substantial (Table 3.7). After conjoined analysis and discussion, all instances of disagreement were resolved. Fourth, raw and normalised (per 10,000 words) frequencies, and percentages of *you* by projected role were calculated (3.4.1) as well as log likelihood and Bayes

Factor approximation (BIC) values (3.4.2). For the purposes of this chapter, I set a log likelihood critical value of >6.63 ($p < 0.01$).

5.5.2. Frequency and statistical significance of *you* in the level of achievement subcorpus

As can be seen in the raw frequency values in Table 5.1, the three most recurrent roles in the corpus are *research apprentice* (196), *co-constructer* (169), and *representative* (92). A comparison of normalised frequencies (per 10,000 words) by level of achievement shows that *apprentice* projections are slightly more frequent for low achievers (29.5) than for high (27.7) and medium (24.1) achievers. *Co-constructer* realisations are slightly higher for high achievers (25.5) than for medium (23.7) and low achievers (23.2). Finally, *representative* realisations of *you* show a similar situation. They are slightly more frequent for high achievers (14.1) than for medium (11.9) and low (11.7) achievers (see also figure 5.4).

Table 5.1. Raw (R) and normalised (N) frequencies, and percentages (%) of *you* audience roles realisations by level of achievement

	High			Medium			Low			Total		
	R	N	%	R	N	%	R	N	%	R	N	%
Academic Roles												
a. Representative	38	14.1	5.5	31	11.9	4.5	23	11.7	3.3	92	12.6	13.3
b. Tourist	41	15.2	5.9	33	12.6	4.8	11	5.6	1.6	85	11.7	12.3
c. Architect's client	24	8.9	3.5	18	6.9	2.6	8	4.1	1.2	50	6.9	7.2
d. R apprentice	75	27.7	10.9	63	24.1	9.1	58	29.5	8.4	196	26.9	28.4
e. Opinion-evaluator	1	0.4	0.1	4	1.5	0.6	0	0.0	0.0	5	0.7	0.7
f. Innovation user	54	20.0	7.8	16	6.1	2.3	10	5.1	1.4	80	11.0	11.6
subtotal	233	86.2	33.8	165	63.2	23.9	110	56.0	15.9	508	69.8	73.6
Knowledge Contribution Roles												
g. Provider	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
h. Co-constructer	69	25.5	10.0	62	23.7	9.0	38	19.3	5.5	169	23.2	24.5
i. Learner	3	1.1	0.4	4	1.5	0.6	2	1.0	0.3	9	1.2	1.3
subtotal	72	26.6	10.4	66	25.3	9.6	40	20.4	5.8	178	24.4	25.8
English Language Competence Roles												
j. Provider	0	0.0	0.0	2	0.8	0.3	2	1.0	0.3	4	0.5	0.6
k. Competence-in-progress user	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
l. Learner	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
subtotal	0	0.0	0.0	2	0.8	0.3	2	1.0	0.3	4	0.5	0.6
TOTAL	305	112.8	44.2	233	89.2	33.8	152	77.3	22.0	690	94.8	100.0

Given that the normalised frequencies of projected roles do not clearly show a distinction between the different achievement groups, significance (log likelihood) and effect size (Bayes Factor Approximation-BIC) tests are run for all *you* audience roles (Table 5.2).

Table 5.2. Log likelihood and Bayes Factor (BIC) values by levels of achievement

		Observed frequencies			log likelihood	BIC
		High	Medium	Low		
Academic roles						
a.	Representative	38	31	23	0.68	-21.71
b.	Tourist	41	33	11	10.49	-11.90
c.	Architect's client	24	18	8	4.08	-18.31
d.	R apprentice	75	63	58	1.33	-21.06
e.	Opinion-evaluator	1	4	0	5.18	-17.21
f.	Innovation user	54	16	10	30.45	8.06
Knowledge Contribution Roles						
g.	Provider	0	0	0	0.00	-22.39
h.	Co-constructer	69	62	38	1.98	-20.41
i.	Learner	3	4	2	0.29	-22.10
English Language Competence Roles						
j.	Provider	0	2	2	3.79	-18.60
k.	Competence-in-progress user	0	0	0	0.00	-22.39
l.	Learner	0	0	0	0.00	-22.39

In the significance and effect size analyses, the two most statistically significant frequency differences are observed in the *tourist* and *innovation user* roles. In module 2 (Nausa, 2016), their corresponding mirror image roles *guide* and *originator* also exhibited this trait but not the *co-constructer* role (see Table 10.2). The *innovation user* role log likelihood value of 30.45 is significant at $p < 0.0001$. Additionally, its effect size Bayes factor (BIC) value (8.06) indicates that there is strong evidence against the null hypothesis (there is no difference between the achievement groups in terms of *you* use to project the role). The *tourist* role log likelihood value of 10.49 is significant at $p < 0.01$. However, its BIC value is negative meaning that although the frequency differences are statistically significant, they are not big enough to constitute strong evidence against the null hypothesis.

5.5.3. Frequency and statistical significance of *you* in the disciplinary divide corpus

As evidenced in Table 5.3, a comparison of normalised frequencies (per 10,000 words) of the three most recurrent projections in the corpus shows that *research apprentice* projections are highly more frequent for hard sciences (47) than for soft sciences (6.6). *Co-constructor* realisations are higher for soft sciences (27.9) than for hard sciences (18.6). Finally, *representative* occurrences of *you* are slightly more frequent for soft sciences (13.8) than for hard sciences (11.5).

Table 5.3. Raw (R) and normalised (N) frequencies, and percentages (%) of "you" audience roles realisations by discipline

			Hard			Soft			Total		
Academic Roles			R	N	%	R	N	%	R	N	%
a.	Representative		42	11.5	6.1	50	13.8	7.2	92	12.6	13.3
b.	Tourist		49	13.4	7.1	36	9.9	5.2	85	11.7	12.3
c.	Architect's client		18	4.9	2.6	32	8.8	4.6	50	6.9	7.2
d.	R apprentice		172	47.0	24.9	24	6.6	3.5	196	26.9	28.4
e.	Opinion-evaluator		3	0.8	0.4	2	0.6	0.3	5	0.7	0.7
f.	Innovation user		59	16.1	8.6	21	5.8	3.0	80	11.0	11.6
subtotal			343	93.8	49.7	165	45.5	23.9	508	69.8	73.6
Classroom											
Knowledge Contribution Roles											
g.	Provider		0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
h.	Co-constructor		68	18.6	9.9	101	27.9	14.6	169	23.2	24.5
i.	Learner		4	1.1	0.6	5	1.4	0.7	9	1.2	1.3
subtotal			72	19.7	10.4	106	29.3	15.4	178	24.4	25.8
English Language Competence Roles											
j.	Provider		1	0.3	0.1	3	0.8	0.4	4	0.5	0.6
k.	Competence-in-progress user		0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
l.	Learner		0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
subtotal			1	0.3	0.1	3	0.8	0.4	4	0.5	0.6
TOTAL			416	113.7	60.3	274	75.6	39.7	690	94.8	100.0

Again, as the normalised frequencies of projected roles do not clearly show a distinction between the disciplinary groups (maybe only for the *research apprentice* role), log likelihood and BIC tests are run for all *you* audience roles to determine statistical significance and effect size.

Table 5.4. Log likelihood and Bayes Factor (BIC) values by Discipline

		Observed frequencies			
		Hard	Soft	log likelihood	BIC
Academic roles					
a.	Representative	42	50	0.77	-10.42
b.	Tourist	49	36	1.87	-9.32
c.	Architect's client	18	32	4.11	-7.09
d.	R apprentice	172	24	124.57	113.38
e.	Opinion-evaluator	3	2	0.19	-11.00
f.	Innovation user	59	21	18.44	7.24
Knowledge Contribution Roles					
g.	Provider	0	0	0.00	-11.20
h.	Co-constructer	68	101	6.80	-4.39
i.	Learner	4	5	0.12	-11.07
English Language Competence Roles					
j.	Provider	1	3	1.07	-10.13
k.	Competence-in-progress user	0	0	0.00	-11.20
l.	Learner	0	0	0.00	-11.20

These analyses show that the three most statistically significant frequency differences for the disciplinary divide are observed in the *research apprentice*, *innovation user*, and *co-constructer* roles. In module 2 (Nausa, 2016), only one of the corresponding mirror image roles: *recounter-announcer* exhibited this trait (Table 10.3). The other two mirror roles (*originator* and *co-constructer*) exhibited statistically significant frequency differences, but in the levels of achievement comparisons (Table 10.2).

The *research apprentice* and *innovation user* role log likelihood values are significant at $p < 0.0001$. Also, their BIC values indicate that there is strong evidence against the null hypothesis. The *co-constructer* role log likelihood value is significant at $p < 0.01$. However, its Bayes factor value is negative (see 3.4.2).

The statistical significance of the roles in the levels and disciplinary divides indicates that their frequency differences are not the result of random chance. This significance is probably made clearer with discourse analyses.

5.6. Discourse analyses

In this section, I will analyse the *you*-audience role projections that more significantly discriminate among the level of achievement and disciplinary groups. Each analysis will include common patterns and the discourse functions performed.

5.6.1. *You*-projections in the level of achievement divide

Significance tests and effect size scores indicate that the audience *you*-identity roles whose realisation can be considered as marks of oral performance among individuals in this study are *innovation user* and *tourist*.

5.6.1.1. *Innovation user (originator)*

The *innovation user* role is the audience mirror image of the presenter *originator* role. In Tang and John's (1999) taxonomy, *originator* is the most powerful authorial stance role because it implies claiming authority or ownership towards generated knowledge or innovations. In module 2 (Nausa, 2016), *originator* was also the one that exhibited the highest log likelihood value in the level of achievement comparisons, so it is not surprising that its mirror image role is also the one with the highest log likelihood value in this study.

As explained in 5.4.1.1, in the projection of this audience role, hearers are invited to be part of a category simulation (Gast et al., 2015) in which they pretend to be potential users of the

knowledge or innovation created by the *originator*. In the 80 cases of this role, all cases of you were in subject position in active voice clauses, which means that *innovation users* were construed as potential active users of the innovation.

(24) *So the idea of do an arrangement with a supplier that **you** trust is attractive because if **you** trust of him, then **you** don't have to do this operation, because the supplier, when he is making the product, he already have controls and then when **you** get the products **you** don't have to do the process again. So **you** save this part of the operation.* (S-H-INGE-7)

In the 80 *innovation user* projections, *you* collocates with activity (45%), existence (24%), mental (13%), and communication verbs (10%). Activity verbs include actions that can be done with innovations (e.g. *connect, use, save, apply, prevent, optimize*), or that could happen without the innovation (e.g. *lose, waste*), actions that imply the use of money (e.g. *pay, save, buy, get, invest*). Existence verbs (*be, have*) refer to states or things (mental or physical) that the *innovation user* might (or not) have. Mental verbs, the same as activity verbs, refer to actions that can be done with innovation or generated knowledge. Biber et al. (1999) propose various subtypes of mental verbs: cognitive state (e.g. *know, think*), cognitive dynamic activity (e.g. *calculate, consider, control*), emotional state (*love, want, hope, need*), perception (e.g. *see, hear*), receipt of communication (e.g. *read, hear*). In the 80 *innovation user* sentences, the most common verb was *need* (emotional state) used to convey necessity. Other not so recurrent mental verbs include *prepare, calculate, control* (cognitive state).

In *you-innovation user* projections, some of these verbs were used within specific patterns. The following are the two most recurrent ones.

- *You can* [activity/mental verb]

(25) *Because it's a time-consuming machine activity, so when you are eh operating these machines you are spending money. But eh the [fs] there is a [fs] a [fs] a like a [fs] an idea that **you can save money if you control the compaction process.*** (S-L-INGE-5)

- *You don't* [mental verb]

(26) *Now, this is another important part of the arrangement, storage costs, because when you get to the ultimate form of just-in-time philosophy you can cut off [fs] you don't need a warehouse, [unintelligible], **you don't need a big warehouse, you don't need a technology to control all those products** because you receive all your needs, all your goods, only when you need it.* (S-H-INGE-7)

The use of these patterns is not necessarily a mark of discrimination between the different levels of achievement. The *you can* pattern is almost similarly distributed among the three groups. However, the second pattern (*you don't* [mental verb]) is only used by high achievers.

Unexpectedly, this role is not mentioned in other studies of *you* in oral presentations neither directly nor in the form of the discourse functions it implies.

From the quantitative and qualitative analyses of the role, we can conclude that the difference between high, medium, and low achievers lies in the fact that high achievers more frequently use pronouns to address the audience as potential end users of innovations. Not surprisingly, more instances of their *you-innovation projections* include more types of verbs and *you-patterns*.

5.6.1.2. *Tourist (guide)*

The audience *tourist* role has as its mirror image the presenter *guide* role. In Tang and John's (1999) taxonomy, *guide* is one of the second least powerful authorial stance role because it just implies showing familiarity with field knowledge. In module 2 (Nausa, 2016), *guide* was the third role with the highest log likelihood value in the level of achievement comparisons, so it is expected that *tourist*, its audience mirror image role, is also one with the highest log likelihood values in this study.

In 5.4.1.1, this audience role was explained as one in which hearers are directly addressed; therefore, its use is personal (Gast et al., 2015). In the 85 cases of this role, 70 were in subject position and the remaining 15 in object position. *You-tourist* projections show *you* collocating with mental-perception (45%), activity (16%), and existence (1%) verbs. Mental-perception verbs (*see, observe*) cluster with *you* in subject position. Activity verbs (*present, show*) co-occur with *you* in object position.

Differences between levels of achievement are noticed at this point. First, as can be seen in the normalised frequencies in Table 5.1, low achievers (5.6) do not directly address the audience to explain their slides as much as medium (12.6) and high achievers (15.2) do. Second, although the sentences that they use are very similar in general structure, low-rated OPs tend to have more grammar and vocabulary errors.

- (27) *Eh I have a little question or suggestion for [fs] for you like a [fs] like an always. When you eh [fs] when you eh look this picture eh how you see? In your mind.* (S-L-HIST-5)
- (28) *...of photographs in anthropology from 1845 to 2006 [reading 1] Ok, eh you look some photo [fs] old photographs I suppose, in your house have...* (S-L-ANTR-1)
- (29) *So, how [fs] you can see at the picture it is a kindergarten eh she is...* (S-L-EDUC-2)

The following are the two most recurrent patterns in the projection of the *tourist* role.

- *You can see*

(30) *Eh human collecting of wild plants or animals at rates exceeding ability to those species [fs] species to recover. **You can see** in this slide the efficient industry, the tuna.* (S-H-CBIO2)

- *Show you*

(31) *Here I I **show you** the Google books in grand viewer is like a list for the publication of one topic in that period and again you can see the increase of that eh books reflects the importance of the [fs] of the matter, the subject.* (S-H-ECON-2)

These two patterns can be considered marks of level discrimination. Low achievers use of these is really low. The *you can see* pattern has 9 occurrences for high achievers, 4 for medium, and 1 for low. The *show you* pattern has 22 occurrences for high achievers, 2 for low, and 24 for medium. These raw frequencies have log likelihood values of 5.4 and 17.73, respectively. These patterns take as complements prepositional phrases (*in this picture*) and adverbs (*here*) that refer to the visuals in the OP. As described in the deixis chapter, the role dyad *guide-tourist* also implies the use of gestural deixis, which as it was demonstrated, is more consistently orchestrated with the verbal mode by high achievers.

Reference to visuals or the place in the oral presentation that this role implies is also reported in other studies of *you*. Vassileva (2002) reports the use of *you* by experienced linguistics researchers to refer to handouts given to the audience; Rowley-Jolivet and Carter-

Thomas (2005b), to refer to information on slides; and Fernández Polo (2018) to guide the audience attention to specific data.

From the quantitative and qualitative analyses of the role, it can be concluded that high and medium achievers outperform low achievers, not only in how often they address the audience, but also in how lexicographically correct sentences are produced. This confirms what other studies have reported and complements verbal and gestural deixis analysis in this thesis.

5.6.2. *You*-projections across the disciplinary divide

The audience *you*-identity roles that exhibited high log likelihood scores in the disciplinary divide were, *research apprentice*, *innovation user*, and *co-constructor*.

5.6.2.1. *Research apprentice*

The audience mirror image of the *research apprentice* role is *recounter/announcer*. In Tang and John's (1999) taxonomy, *recounter* is one of the less powerful authorial stance roles; it refers to the narration of the steps in an enquiry process. In module 2 (Nausa, 2016), I redefined the role for the purposes of the study and added the word *announcer*. This redefinition was also inspired by the concept *proximity* (Hyland, 2012), understood as the authors' adoption of the disciplinary practices for knowledge production in their areas, which can be either recounted (if they were done before the OP) or announced (if they are to be done in future research). In this chapter, when a presenter refers to these disciplinary practices and adds a second person pronoun, the rhetoric effect is that of the audience being invited to be part of a category simulation (Gast et al., 2015); they are invited to pretend they are novice researchers in the same

field as the presenter's. In my previous study, *recounter/announcer* was the second role with the highest log likelihood value in the hard-soft discipline comparisons (see Table 10.3). In this study, *research apprentice*, its audience mirror image role, is the one with the highest log likelihood (124.57) and Bayes Factor values (113.38) (Table 5.4).

In 5.4.1.1, this audience role was explained as impersonal, for the hearers are not directly addressed (Gast et al., 2015). The referent of the pronoun is a group to which the audience does not belong. All the 196 cases of *you* in this role are in subject position, and the pronoun collocates with activity (56%), mental (35%), existence (8%), and communication verbs.¹⁹

Disciplinary differences in the projection of this role are remarkable. Firstly, as seen in the normalised frequencies in Table 5.3, hard-discipline researchers (47) use approximately eight times as many *you*-references as their soft-discipline (6) counterparts. Secondly, a much higher number of occurrences in the hard-discipline corpus comes with a higher number of verbs and patterns of *you*-projections.

The following are the most recurrent patterns.

- *You* [verb (activity/mental)] (research instrument) (purpose clause)

¹⁹ In module 2 (Nausa, 2016), the identification and classification of verbs in the mirror image role (*recounter/announcer*), discriminating between mental and activity verbs as explained by Biber et al (1999) was not a straightforward task. Verbs like *classify* and *identify* are described as mental (dynamic cognitive process) in nature. However, they sometimes express a kind of physical manipulation of objects. Other verbs such as *work*, *excavate* apparently belonging to activity verbs imply the use of senses (sense verbs are part of mental verbs) or attention and memory (mental activities). Like I did in that study, I decided that when the action expressed by the verb expressed physical manipulation or a step-procedure to achieve a major research objective, the verb would be classified as *activity*. When the action expressed mental work and manipulation, if the verb referred to a bigger procedure or final product, the verb would be classified as *mental*.

- (32) *And **you** appli [fs] **apply** (some metrics) (to to take decision over your [fs] over this software), ok.* (S-H-INGE-3)
- (33) *Other part is diagnostic of the faults with the use of detectors, **you use** (this information) (for diagnostic).* (S-L-INGE-4)
- (34) *(With two eh tools) **you can analyze** (that items) and **you identify** (the relation clients, suppliers, products, external factors between others.)* (S-M-INGE-5)

I had already reported this pattern with exclusive *we* in module 2. As I explained there, the elements in parentheses are optional and can eventually be moved to theme position. Other authors (e.g. Kamio, 2001; Gast et al., 2015) have pointed out that these uses of *you* are near synonyms, and therefore, can be replaced with exclusive *we* or *one*.

Alternatively, this pattern is expressed with the modal verb *can* to convey methodological possibilities or alternatives.

- *You can [verb (activity/mental)] (research instrument) (purpose clause)*

- (35) *Because in one democracy [fs] one democracy eh [fs] one democracy can have one behavior, but inside her eh we can find many levels of democra [fs] democracy. What I want to explain. For example, **you you can think** in one national democracy, (in one democracy) **you can find** (many subnational democracy), and this subnational democracy are different between national democracy and between them.* (S-M-CPOL-1)

Other modal verbs can also be used in this pattern to express degrees of obligation or advisability.

- *You [obligation modal] [verb (activity/mental)] (research instrument) (purpose clause)*

- (36)... *but in pavement works, you have a lot of materials, so **you have to** define (the last question that it was very eh complex) (for a three or four materials), so that is impossible.* (S-L-INGE-5)
- (37)***you must** eh define (the complete eh elements) and for example here you **can** you **define** eh (four clients) (with specific information)...* (S-M-INGE-6)
- (38)*For example, this is an example of a failed road, because **if you** are in this part eh in this island eh then, (for getting out of the island) **you need to** cross (another bridge) two times and then this is a failed road, but eh this doesn't resolve eh the problem because it is possible that **you can** find (another good road) (that solves the problem).* (S-H-MATE-2)

Finally, these patterns are used in complex sentences, either in the main or the subordinate clause. The use of *if/when* clauses convey the idea of the need of a procedure, or the conditions under it should be followed. Example 38 contains one example of an *if* clause.

- *When/if you [verb (activity/mental)]*

- (39)*In response a epigenetic mechanics that is chemical change and I have here one one one example. **When you put in** (the DNA a methyl group), (the gene can turn off), but **when you retire** (this group), (the gene can turn on).* (S-M-CBIO-3)
- (40)*Eh, for example, eh Taxol, eh this is the structure eh of the Taxol, is extracted of the species taxus brevifolia. And **if you need** to obtain (100 mg of Taxol), you need to (extract of big amount of the tree).* (S-H-CQUI-2)

Reference to methods and procedures in this role projections is also discussed in other studies of *you* in conference presentations. Rowley-Jolivet and Carter-Thomas (2005b) explain that physics presenters use *you* to refer to "... a thinking entity or typical researcher" (p. 60). Webber (2005) says that impersonal *you* in medicine CPs can refer to a "... potential patient or

researcher...” (p. 163). The other oral academic discourse *you* studies surveyed here (e.g. Fernández Polo, 2018; Vassileva, 2002) do not report this function.

The use of these patterns is clearly a defining difference between the discipline groups. Hard-discipline presenters more frequently project this role than their soft-discipline counterparts. This difference is not only noticed in the raw and normalised frequencies of *you-research apprentice* occurrences, but in the specific patterns of use explained above. This confirms the tendency I reported in module 2 (Nausa, 2016), in which the *recountor/announcer* role exhibited the same quantitative and discourse characteristics, especially because exclusive-*we* can be replaced by simulation-*you* as near synonyms. This is also in line with the epistemological features of knowledge production in the hard sciences, which tend to favour standardization of and adherence to strict methodological procedures (Becher & Trowler, 2001). This can also be explained in terms of the nature of the assigned task: an OP on one’s research. Most hard-field research tends to be problem-based. As Becher and Trowler (2001) put it, “Another way of indicating the difference is to note that in hard, restricted fields, the available methods tend to determine the choice of problems; in soft, unrestricted ones, it seems rather that the problems determine the methods” (p. 185).

5.6.2.2. Innovation user

This role frequency difference was also significant in the level of achievement comparisons. (see 5.6.1.1).

The use of *you-innovation* user realisations and their corresponding patterns is also a distinctive characteristic of hard-science presenters. This is confirmed by raw and normalised frequency counts which show that hard discipline presenters use more than twice as many *you*

projections as soft discipline ones (Table 5.3). One possible explanation for this more frequent projection of the role in the hard sciences (e.g. engineering) is that they focus on the creation of innovations to solve every day problems that are usually sold to companies or specific clients. (see examples 16-17; 24-26)

5.6.2.3. *Co-constructer*

The *co-constructer* role is not a stance positioning role, but a knowledge contribution one. This is to say, it is related to one of the conditions for being a legitimate participant (Lave & Wenger, 1991) in class: students have to share their PhD research in a way that is easy for their classmates audience to understand. In module 2 (Nausa, 2016), the premise for the creation of the three roles was who teaches what to whom; as a result, a presenter can teach (*provider*), acknowledge that they were taught something in previous presentations (*learner*), or create the conditions to make (or pretend that) the experience in the OP is a construction with the audience (*co-constructer*). In the definition of dyad-roles for this study, two roles were simply reversed: if the presenter is projecting themselves as a *knowledge provider*, automatically, the audience is being construed as *learners*, and vice versa. In the case of the *co-constructer* role, however, the prefix *co* implies the other; therefore, the same term was used to name the audience role, to set the distinction, I will henceforth refer to the roles as *I-co-constructer* (presenter) and *you-co-constructer* (audience).

In the projection of *you-co-constructer*, hearers are either addressed directly or invited to be part of a category or situation simulations (Gast et al., 2015).

- (41) *I want to talk about eh [fs] about the Bogota's Panopticon, but eh first **I want to ask you** a question **if you [fs] if you know** which is this building today. Anyone knows? (S-M-HIST-2)*
- (42) *And what happen with this situation I [fs] imagine that **you are**, for example, in your house and **you have** this eh po power outlet, **you can plug in** to this power outlet devices and consume consume energy, but what happen **if you connect** and connect devices? (S-M-INGE-7)*

In the 169 *you-co-constructer* projections, *you* collocates with mental (56%), existence (17%), activity (13%), communication (11%), and occurrence (2%) verbs. Mental verbs include cognitive state verbs (e.g. *know, remember, understand, imagine*), dynamic verbs (e.g. *learn, read, study*), and perception verbs (*see, hear*).

Activity verbs usually refer to daily life actions (*go, walk, play*). Communication verbs refer to actions that will be performed by the presenter (*tell, ask, say*). Mental verbs are of special interest in the characterization of this role, for they can refer to different co-construction actions. Cognitive state verbs like *know* and *remember* have a clear past orientation and are useful in previous knowledge activation:

- (43) *Maybe **you remember** when you study philosophy eh the the example of the (unintelligible), great philosopher who who says that eh you can imagine a group of people in a in a cave, you know... (S-M-ANTR-1)*
- (44) *So, I know that **all of you know** the bottlenose dolphins because all of **you know** eh what eh flipper. (S-H-CBIO-5)*

Previous knowledge activation can also be performed with perception and activity verbs. Present perfect is common in this function.

(45) *Today I'm going to talk you [fs] to you the Programa de Familias en Acción that is the topic of my thesis. **Have you heard** som [fs] something about this program?* (S-M-ANTR-2)

Cognitive state verbs with a present-future orientation can also be useful in setting the context; with these, presenters can ask the audience, not to remember, but to imagine something.

(46) *Well, so, first that I want to do is that **you please imagine** that you are insects. And you are living in this landscape, yeah?* (S-H-CBIO-4)

In *you-co-constructer* projections, most of these verbs were used in rhetorical questions; others in regular clauses.

- *ask you* * (question)

(47) P: *Eh Unión Patriótica was a political party and I want to know [fs] **I want to ask you another question**, do you know who is she?*
A: *Aida.*

P: *Aida Avella, the candidate of the Unión Patriótica in two thousand eh fourteen*

(S-H-DERE-2.txt)

- *Yes-no know/heard* questions

(48)***Do you know** what is illegal damage in the context of the law?* (S-L-DERE-2)

(49)***Did you know** that in two two thousand eleven Colombia had thirty thousand twenty hundred thirty four million dollars for foreign investment?* (S-M-DERE-1)

(50)*eh my thesis topic that is about intercultural communicative competence, a new perspective in a global world. Eh **have you heard** that before or not? Not idea?* (S-H-EDUC-1)

Questions are at the core of knowledge contribution roles. The rhetorical questions in these patterns have the purpose of activating knowledge that is common for both presenter and audience. Clearly, the audience is addressed directly, and by doing this, presenters increase the possibility of engagement with them. By referring to knowledge that is common for both, the presenter creates a collegial atmosphere in which the audience is construed as someone who has access to the same type of knowledge; therefore, they are at the same level in the construction of arguments and explanations, for they are just not mere receivers of information, but contributors as well. The *ask you* * (question) pattern can take questions in either direct (47) or indirect (48) syntactic form. Direct yes-no questions (50) tend to include cognitive state and dynamic mental verbs.

Presenters can also anticipate moments in which the activation of common ground with the audience could be potentially face-threatening. In other words, asking questions about the things they (PhD researchers) know might be potentially insulting. When this is the case, presenters can make use of hedging mechanisms such as admitting ignorance about how much the audience knows with chunks like *I don't know if you*:

- *I don't know if you* [mental verb]

I had already reported this pattern (Nausa, 2016) as a way that presenters construe themselves as *co-constructers*. Like other mirror image dyads, this one explicitly includes both the presenter (*I*) and the audience (*you*).

(51) *for example in the [fs] I don't know if you know how is the the function of the incubators babies. It's [fs] for example the sensors eh take the temperature and eh control the device. (S-H-INGE-1)*

(52) *I don't know if you remember but some operative systems every time you have to change (them), yeah? I don't know if you remember Millennium or XP, yeah. Or Vista, yeah. (S-H-INGE-3)*

Another pattern performing this condescendence mitigating function is *you know* + (common knowledge noun or clause). This chunk is sometimes preceded by *as* or *but*. By using this expression, presenters construe the information as important for the presentation, but not as new information given to the audience.

- *You know* + common knowledge noun or clause

(53) *So eh about indigenous people, as you know in Colombia at this moment we have 102 groups, communities or indigenous people. (S-M-ANTR-4)*

(54) *For example, a (rate) interpret [fs] eh can interpreted eh the concept of eh an apple, for example, but you know that in the real life there are eh a lot of [fs] of [fs] or or a lot of apples that can be eh huge, that can be eh with colo [fs] with different colors but this is in the real life but the concept is the [fs] is the apple. (S-M-INGE-6)*

(55) *However, eh the efficacy of this of this right requi requires some conditions, for example, a specific legis legislation. In Colombia you know this with eh pe petition right and other eh second question is a fundamental principles to the access of information... (S-M-CPOL-2)*

Some of the discourse functions described here that are typical of the *you-co-constructer* role are also reported in other studies of *you* in conference presentations. Researchers in these studies are reported to use *you*-addresses to refer to common experience (Vassileva, 2002) or shared knowledge (Fernández-Polo, 2018; Vassileva, 2002) or ask the audience to participate in imagined (simulated) situations (Webber, 2005).

The use of *I-co-constructer* was identified as a mark of discrimination between the different levels of achievement in module 2 (Nausa, 2016); however, in this study, *you-co-constructer* is a mark of discrimination between the disciplinary groups. Normalised frequency values show the role to be more common in the soft disciplines. Disciplinary differences are not evident in the use of verbs and lexical chunks that have been described so far, though. Despite this, I decided to include this given the wealth of linguistic resources in its projection that, as has been shown, exhibits more patterns of use and more clearly reflects the interactive character of *you* in oral academic discourse.

5.7. Conclusion

This chapter has aimed at demonstrating that *I*-presenter identity roles in OPs imply the existence of *you*-identity roles and identifying uses of *you* in OPs that would differentiate the oral performances of PhD researchers in an EAP class. The following questions have guided this study:

1. Do *I*-presenter identity roles imply *you*-audience roles in OPs?
2. What are the tendencies in audience identity role projections in OPs when students use *you*?

3. What *you*-audience identity roles are useful in discriminating among students' levels of performance and disciplines?

The descriptions and sample sentences in 5.4.1 demonstrated that *I*-presenter identity roles have a corresponding *you*-audience identity role. Although some roles were not found in the corpus (e.g. *you*-provider of new language), they can be defined and examples can be imagined.

Similarly, the study demonstrated that the audience role projections are correlated to the PhD researchers level of oral achievement and disciplines.

The *you*-identity roles useful in discriminating among levels of oral achievement in the OPs are *innovation user* and *tourist*. The *innovation user* has as its mirror image the *I*-role *originator*, in which presenters claim authorship for findings or inventions. This dyad was proven to be typical of high achievers. When presenters use *you* to construe their audience as *innovation users*, they use the pronoun to invite them to participate in category simulations (Gast et al., 2015). Typical patterns in the projection of this role are *you can* + activity/mental verbs and *you don't* + mental verb. With these structures, presenters indicate to their imagined clients the possibilities and capabilities that their creations would bring. In the use of *you* to project the *tourist* identity role, presenters use patterns like *you can see* and *show you* to guide them through the information on their visuals. The audience is directly addressed, no simulation is performed or proposed. This role image mirror (*guide*) was also proven to be characteristic of high achievers in my previous study (Nausa, 2016). From the study, it can be concluded that high achievers more frequently and consistently engage the audience. This is not only evident in the use of *you* but also in the use of the required linguistic devices to perform the role-related

discourse functions. With *innovation user* projections, engagement is achieved by directly asking the audience to pretend they are somebody else. With *tourist* projections, engagement is achieved by directly telling the audience where to focus. Other resources (proxemics, gaze, etc) are used with this role as explained in the deixis chapter, in which high achievers are also described as more consistently using multimedia resources to guide the audience.

In the disciplinary division, the *you*-identity roles that discriminate between soft and hard-discipline students are *research apprentice*, *innovation user* and *co-constructer*. The *apprentice* role's mirror image is *recounter/announcer*, in which presenters show adherence to methodological practices for knowledge production in their fields. This dyad was also demonstrated to be typical of hard-discipline students. When *you* is used to construe the audience as *a research apprentice*, they are invited to participate in a category simulation (Gast et al., 2015). *Research apprentice* patterns include *You + verb (activity/mental) + (research instrument) + (purpose clause)*; *if you + verb (activity/mental)*, among others. These patterns resemble the ones I reported in module 2 and confirm the fact that impersonal uses of *you* can be paraphrased by exclusive *we*. The use of *you* to project the *innovation user role* (also useful in discriminating among levels of performance) was found to be more common in hard-science OPs, which is explained by the fact that hard-fields like engineering tend to be product (solution) oriented. Finally, the *co-constructer* role, whose role image had been found to be statistically significant for the level of achievement comparisons (Nausa, 2016), was more statistically significant for the disciplinary comparison in this study. However, two unexpected outcomes were found. One, effect size comparisons suggest that there is not enough evidence in favour of the hypothesis. Two, normalised frequencies show that this role is more frequent in soft disciplines, but a closer look at typical patterns like rhetorical questions (*have you heard, do you know*) or lexical chunks

(*I don't know if you remember* or *you know* + shared knowledge) do not clearly show differences other than the content/methodology expected ones.

A comparison with other studies (e.g. Fernández Polo, 2018; Rowley-Jolivet & Carter-Thomas, 2005a; Rowley-Jolivet & Carter-Thomas, 2005b; Vassileva, 2002; Webber, 2005) confirms several of the functions found in this study. However, this study can be said to go a few steps further by including statistical tests that account for statistical significance and effect size. Additionally, this study proposes a more systematic way of organizing information by means of a role-dyad taxonomy that not only complements well with *I*-projections (and their related research) but organizes discourse functions into general categories (authorial stance, knowledge contribution, and language use) and specific role subcategories (*guide-tourist*, *originator-innovation user*). Finally, *you* is often overlooked in studies of pronouns, presumably because it is (considered to be) relatively infrequent, or because it is assumed to be only a corollary of *I/we*. But several examples here show that *you* is often used without *I/we* and plays an important role by itself.

Table 5.5. Summary of findings you-audience study

Comparison	Identity role (Log likelihood and Bayes Factor)	Patterns	Over use (↑) or under use (↓)	Meanings of <i>you</i>	Discourse functions
of oral achievement (high, medium, low)	<i>Tourist (guide)</i> (LL: 10.49/BF: -11.90)	<i>You can see</i> <i>Show you</i>	High rated OPs (↑)	Audience as other PhD students	Guide audience through information on visuals Highlight important information
	<i>Innovation user (originator)</i> (LL: 30.45/BF: 8.06)	<i>you can</i> + activity/mental verbs <i>you don't</i> + mental verb	High rated OPs (↑)	Imagined client-user: category simulation	Expression of - authorship of own findings, Inventions, contributions - possibilities and capabilities that creations would bring
	<i>Research apprentice (Recounters)</i> (LL: 124.57/BF: 113.38)	You + verb (activity/mental) + (research instrument) + (purpose clause); if you + verb (activity/mental)	Hard-field OPs (↑)	Research apprentice: category simulation	Expression of - methodological practices for knowledge production in fields - outcomes (+ or -) if procedures are followed or not
Disciplinary divide (hard vs soft)	<i>Innovation user</i> (LL: 18.44/BF: 7.24)	See row 2	Hard-field OPs (↑)	See row 2	See row 2
	<i>Co-constructor</i> (LL: 6.80 / BF: -4.39)	Rhetorical questions (<i>have you heard, do you know</i>) <i>I don't know if you remember</i> <i>You know</i> + (shared knowledge)	Soft-field OPs (↑)	Audience as legitimate co-constructors of knowledge	- Activation of prior knowledge - Reference to common experience or knowledge - Mitigation of face-threatening knowledge activation

CHAPTER 6

INFORMATION SHOULD BE ACCESSIBLE...: IMPERSONAL PROJECTION OF IDENTITIES

6.1. Introduction: from personal to impersonal modalized projection of academic identities in OPs

In module 2 (Nausa, 2016), and in CHAPTER 5 in this thesis, I described the academic identity roles that Colombian PhD researchers assign to themselves and their audience in the oral presentations (OPs) given in an EAP class. The roles assigned to presenter and audience were identified through the discourse analysis of first person (*I, me, my, we, us* and *our*) and second person (*you*) pronouns occurrences in context. Pronoun study was the obvious choice since pronouns directly refer to the participants in the communication act. However, during the OP, speakers can also make use of other linguistic choices to assign identities to themselves and their interlocutors.

The construction and projection of academic identities with language choices can be understood in the context of Halliday's (1994) systemic functional model. This theory proposes three metafunctions (ideational, textual, and interpersonal), also understood as characteristics of language as a system of meaning choices or what we use language for. The first, the ideational function, is related to the ways in which we use language to make sense of the world. The textual function is concerned with how we manipulate language in the performance of the other two functions. Finally, the interpersonal function refers to the ways in which we use language to create and maintain social relationships.

In academic communities, in addition to the creation and maintenance of relationships with other scholars, identity construction depends on the adoption of the knowledge production and rhetorical practices of such communities. The purpose of adopting these practices is to be recognized as legitimate members of such organizations and to be equipped with the resources to evaluate or produce new knowledge. Hyland (2012) refers to these two processes of academic identity construction as *proximity* –showing oneself as being like the other members of the community– and *positioning* –showing oneself as contributor of something new or different to the community. For the discourse analysis of academic identity construction, Hyland recommends exploring interpersonal features of language, for these relate texts to a given context and express the author’s personality, credibility, sense of audience, and relationship to the message. Among interpersonal language choices are deictics, personal pronouns, and **modality**, among others.

It is therefore assumed in this chapter that the expression of modality is a mechanism of academic identity construction. More broadly speaking, this modality as identity projection can be explained with Du Bois' (2007) stance triangle. A stance-taking act implies three aspects: evaluation, positioning, and alignment. When a presenter takes a stance, they (1) evaluate something, (2) position themselves (or others) as legitimate judges of said knowledge, and (3) align with others, in this case the audience. As Hunston (2011) points out:

Evaluating something necessarily indicates ‘where you and I stand in relation to the object’ and also necessarily indicates ‘where you and I stand in relation to each other’.

These two aspects relate to the ideological aspect of evaluation (‘where you and I stand’) and the interactional aspect (‘how you and I relate to each other’) (p. 23).

In this thesis I understand modality as a type of evaluation or stance-taking act. Stance-taking is expressing an attitude towards something (e.g. person, situation, idea). Modality is taking a stance towards what is said or written. This chapter analyses the impersonal expression of modality as a language choice that allows Colombian PhD researchers to project specific identity roles in OPs, through the stances they take towards the material they present and the audience they address.

The purpose of this analysis is two-fold: first, it seeks to determine whether the distribution of impersonal devices mirrors the distribution of first person pronouns reported in Nausa (2016) for the same PhD population; second, it also seeks to identify whether personal or impersonal realisations correlate with the variables considered in this analysis: identity roles, level of achievement, and disciplines. As such, this chapter seeks to answer the following questions:

- **Tendencies:** what are the tendencies in impersonal expressions of modality in the oral presentations of Colombian PhD researchers as observed in their use of modals, adverbs, and some modal-like (Hunston, 2011) expressions?
- **Identity roles:** What identity roles are projected when presenters express modality in impersonal constructions? How does the impersonal projection of identity roles compare with personal projections?
- **Level of achievement:** What are the differences in the impersonal projection of identity among high, medium and low-rated presentations?
- **Disciplinary divide:** What are the differences in the impersonal projection of identity between hard-discipline and soft-discipline researchers?

- **Personal and impersonal realisations:** How do these tendencies (by level and disciplines) compare with the tendencies in identity projections with personal pronouns?

The analysis of modality as a mechanism to project identity in impersonal realisations focuses on the following language features: (1) features typically associated with the expression of modality: modal verbs and adverbs ending in *ly*, (2) features not typically associated with the expression of modality or *modal-like expressions* (Hunston, 2011): noun patterns (NOUN *that/to*), verb patterns (VERB *that/to*), adjective patterns (ADJ *that/to*).²⁰

6.2. Identity projection with impersonal modalized expressions: model of analysis

To identify projected identities in impersonal modalized constructions, I rely on (1) a combination of Halliday and Matthiessen's (2014) and Quirk, Greenbaum, Leech, and Svartvik's (1985) modality taxonomies, and (2) Tang and John's (1999) academic authorial stance identity roles taxonomy, which I refined and expanded in module 2 (Nausa, 2016) and section 5.4.1 (this thesis).

6.2.1. Modality

Modality has been defined as the judgement of what is being expressed (Halliday & Matthiessen, 2014) or the qualification of content in a way that it includes the speaker/writer's judgement (Quirk et al., 1985). Modality has been approached under different headings:

²⁰ Although other patterns could have been included, the ones surveyed here were chosen, for they can be automatically searched with corpus software.

propositional attitudes (Cresswell, 1985), *evaluation* (Hunston, 2011), *hedging* (Hyland, 1996a, 1996b), *stance* (Biber, 2006), *appraisal* (Martin & White, 2005), among others.

Several taxonomies of modality have been proposed; two seminal ones are considered here (*Figure 6.1*). First, Halliday and Matthiessen (2014) propose four types of modality: probability (*may be*), usuality (*sometimes*), obligation (*must*), and inclination (volition: *want to* or ability: *can*). Second, Quirk et al (1985) propose more types of modality: possibility, ability, permission, (logical) necessity, obligation, tentative inference, prediction, and volition. These two proposals have similar ways to cluster their different types of modality in subgroups. Halliday and Matthiessen group probability and usuality under the heading *modalization* –degree of certainty towards what is expressed– and obligation and inclination under *modulation* –degree of willingness to do what is proposed. Quirk et al (1985) classify modalities as *extrinsic* –no human control over what is expressed, just judgement of the likeliness of occurrence– and *intrinsic* –human control over what is expressed. The two subcategories in each proposal resemble the traditional logical semantics *epistemic-deontic* opposition in the analysis of modality. The types of modality that I analyse in this chapter can be grouped as follows. In the *modalization-extrinsic-epistemic* group, we can keep Halliday and Matthiessen’s *probability* and *usuality* while in the *modulation-intrinsic-deontic* group, we can include *ability*, *obligation*, *possibility*, and *necessity*. The following section defines the types of modality as understood in the research and provides examples from the corpus to explain the resulting taxonomy.

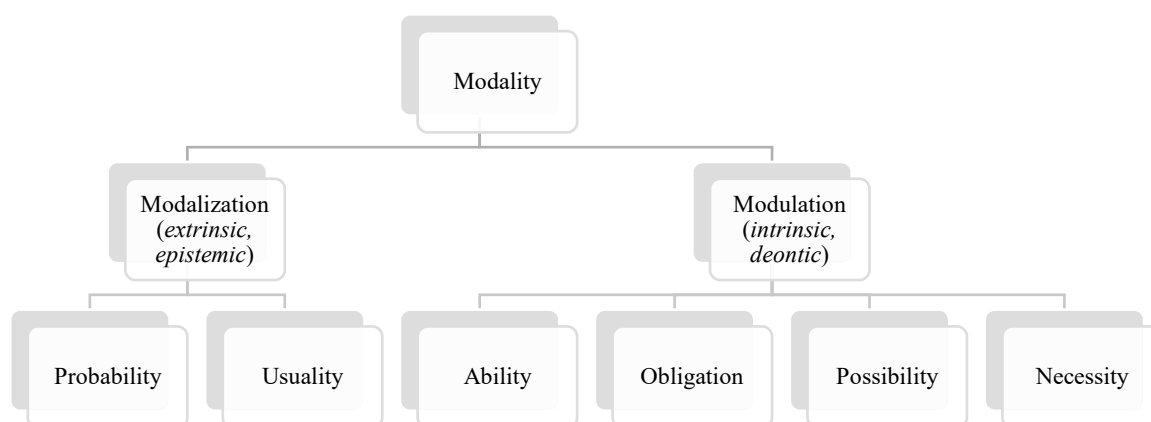


Figure 6.1. Modalization taxonomy based on Halliday and Matthiessen's (2014) and Quirk et al's., (1985) taxonomies

6.2.1.1. Modalization (extrinsic-epistemic)

In modalizations, the author refers to the degree of certainty towards the truth value of the proposition (probability) or the frequency of the state of affairs expressed by the proposition (usuality). The modalizing expression expresses judgement over the proposition, not human control. Two types of modalizations are considered here²¹.

²¹ Following Hunston (2011), I use underlines to mark the modalized propositions and **bold type** to mark the modalizing expressions.

Probability

- (1) *According to this hypothesis, the lack of nutrients **may** be due to two causes. First malnutrition and second unbalanced diet.* (S-H-ANTR-1)

Usuality

- (2) *In addition, most people **usually** spend more time, eh spend more money in order to get a eh college diploma.* (S-H-ECON-3)

6.2.1.2. Modulation (intrinsic-deontic)

In modulations, the author refers to the degree of willingness to do or inclination towards what is expressed in the proposition (Halliday and Matthiessen refer to these propositions as *proposals*). The modulating expression also expresses degree of human control over the proposal. Four subtypes of modulation are considered here.

Ability

- (3) *This is important because the people **can** interact with the nature and to learn to respect the life in all scales.* (S-H-CBIO-2)

Possibility

- (4) *The government maybe, eh **could** be eh eh invest more money in quality of education.* (S-H-ECON-3)

Ability refers to the actions humans can perform directly or aided by tools (e.g. apps, software, instruments and the like). Possibility refers to latent ability or the existence of alternatives for the performance of actions.

Similarly, in obligation and necessity, human control over proposals is implied. However, in these two modalities the idea of an external type of control variable is implied. In obligation, human action is construed as something to be done observing a moral or ethical scale. In necessity, actions are expressed as requiring the occurrence of other actions (e.g. laboratory procedures, methodological steps, etc). Without those required actions, related actions or states of affairs would not happen along expected lines.

Obligation

- (5) *And and that the news that the news the [fs] as the journalist say if there are another solution for that that kind of mortal fiber, the government should prohibit the the use of asbestos. (S-H-INGE-5)*

Necessity

- (6) *When I [fs] when an investment [fs] investor eh go to the ICSID to resolve the dispute, there are several steps that it must be follow. First the ICSID receives the claim and eh the [fs] accept jurisdiction. (S-M-DERE-1)*

6.2.1.3. Orientation (explicitness-subjectivity)

In addition to the types explained, Halliday and Matthiessen (2014) use the concept *orientation* to refer to the linguistic mechanisms to construe modalities. *Orientation* in turn includes two aspects: explicitness and subjectivity. The first *implicit-explicit* orientation refers to

the inclusion (or not) of the modalizing expression in the modalized proposition. The second *subjective-objective* orientation refers to the act of construing modalization as the speaker-writer's judgement, or somebody else's. In (5), for example, there is a case of *implicit-subjective* expression of obligation. It is implicit because the modalizing expression **should** is part of the modalized proposition (*the government ... prohibit the use of asbestos*). It is subjective, for the proposition is expressed as the speaker stating that the prohibition of the use of asbestos is something obligatory for the government to do. Objective realisations are expressed as the writer reporting the modalization as something out of their judgement, as something that others do. The combination of the different variables yields other three orientation alternatives. Examples 7 to 9 illustrate them.

- (7) ***I demand that the government prohibit the use of asbestos.*** (explicit-subjective)
- (8) ***The government is expected to prohibit the use of asbestos.*** (implicit-objective)
- (9) ***It is mandatory for the government to prohibit the use of asbestos.*** (explicit-objective)

6.2.1.4. *Value*

Another important aspect in the characterisation of the expression of modality is *value* (Halliday & Matthiessen, 2014). Modalizations and modulations are intermediate values between polarities. For example, in the expression of obligation, the imperative forms *do!* and *don't!* express polar positive and negative values. Intermediate values (modulations) range from high degree of obligation (*must*) to low degree of obligation (option – *allowed*). In fact, some of the modalities that Quirk et al (1985) propose can be rearranged as intermediate values in Halliday's taxonomy (e.g. logical necessity can be deemed as a modulation of probability).

6.2.1.5. *Phraseology: modals and modal-like expressions*

Modal verbs are the default, not marked, congruent—as opposed to metaphorical—realisations of modal meanings (Halliday & Matthiessen, 2014). However, other linguistic (marked, metaphorical) realisations reflect the interaction of the three variables (types, orientation, and value) in the expression of modality. For example, in the expression of probability in 10, the use of *must* indicates a subjective-implicit orientation and a high degree of certainty in relation to the modalized proposition.

(10) she **must** be tired

Changes in any of the aspects imply changes in the linguistic resources.

(11) *It is believed that* she is tired

In 11 the expression of probability undergoes three changes: the degree of certainty is reduced, the modalization is made explicit, and it is expressed as objective. The three changes were made by replacing *must* for *it is believed that*, which is not a modal verb.

Hunston (2011) coined the term *modal-like expressions* to refer to “...expressions other than modal auxiliaries which express modal meanings” (p. 68). In a survey of other studies on modality (e.g. Biber et al, 1999; Palmer, 1987; Stubbs, 1986), the author observes that the resources to express modality can range from simple words (modals), two-word expressions (*have to*, *ought to*), phrasal modals (*be (un)able to*, *be bound to*, *be going to*), to projecting clauses (*I think that*, *it is observed that*). As indicated above, this chapter analyses modals, adverbs, and modal-like expressions. The selection of modal-like expressions is informed by

Hunston's (2011) inventory of expressions that "attract modal meaning". Although the author provides specific words (e.g. *fear*, *speculate*, *essential*) and their corresponding patterns (Hunston & Francis, 2000) (*for fear of*, *speculate wh* clause, *it* v-link ADJ *that*), this chapter focuses on only three general patterns that attract modal meaning: VERB *that/to*, NOUN *that/to*, and ADJ *that/to*. The reasons for this will be explained in 6.4.1.

6.2.2. Identity projection

For the purposes of this chapter, I continue using and refining the taxonomy presented in 5.4.1. In this section, I include examples of impersonal modalized realisations from the corpus that exemplify each role projection, and when necessary, I provide further explanations.

6.2.2.1. *Authorial stance roles*

Authorial stance roles represent the stances that presenters take regarding disciplinary knowledge. Weak stance positions refer to disciplinary knowledge while strong positions either judge knowledge or claim authority for new knowledge or disciplinary contributions. For the purposes of this chapter, a reconceptualization of the *representative* academic identity role was necessary.

a. *Representative*

In the use of personal pronouns, this role is conceived "as a proxy for a larger group of people" (Tang & John, 1999, p.27). In module 2 (Nausa, 2016) *representative* was found to be a way of referring to membership to academic communities (programs, lab groups, research groups, and the like) as well as to obtained titles, current studies, or research group positions.

- (12) *I am* marine biologist but *I started* my master (intelligible) but *I transferred* to doctoral studies now (S-H-CBIO-5).

Nonetheless, it can be argued that it is not only by reference to groups or positions within them that a person can express membership to them. I argue that the mere fact of expressing ample disciplinary knowledge using the rhetorical conventions of a discipline is in itself an act of representativeness. There are two reasons from speech act theory for claiming this. The first relates to the idea that expressing ample knowledge as a way of projecting the *representative* identity partially resembles a characteristic of performative speech acts (Austin, 1962; Searle, 1969). These speech acts can only be appropriately performed by people who are invested with the authority to change a particular state of affairs. Classical examples of this type of speech act include priests saying *I pronounce you husband and wife* during marriage ceremonies or judges saying *as punishment for this crime, the court sentences you to...* in court trials. In the case of the oral presentations in this investigation, PhD students are invested with the disciplinary authority that comes from their experience as researchers.

The second argument comes from felicity conditions (Searle, 1969). They are the conditions that must be met for the speech act to be satisfactorily realised. In this research, these conditions are set by the task itself: OPs about students' research in their PhD programs; therefore, for assertions in OPs to be felicitous, they should meet at least the following conditions:

- Uttered sentences refer to students' research in their PhD program,
- **The presenter is enrolled in a PhD program (member of an academic community)**

- expressed contents are part of disciplinary knowledge or possible applications,
- the speaker believes that the audience will identify expressed contents as disciplinary.

However, one difficulty that arises from this line of reasoning is that almost everything that a presenter says in an OP should be considered as projecting the *representative* role. This begs the question of how to distinguish this role from others. For the purposes of the chapter, I decided that the presentation of disciplinary contents that were not heavily modalized as expressing an opinion or the generation of a new idea would be considered as projecting the representative role. More specifically, when such contents are expressed as attributed (Sinclair, 1988) to other authors, studies, theories, paradigms, etc, the expression of these contents would be considered a realisation of the *representative* role as connoisseur of disciplinary knowledge.

- (13) *In this area, there are fewer studies, however, **some studies indicates that or show that mycorrhizal specificity may help drive diversity indirectly indirectly by determining distribution patterns of orchids.*** (S-H-CBIO-7)
- (14) *Eh **consejo de Estado says that those requirements, the minimum number of citizens, for example 3,000 people, were impossible to accomplish... accomplish eh to Unión Patriótica*** (S-H-DERE-2)

To set the distinction between the nuances of the meaning of the role, where necessary, I will use the terms *representative-member* and *representative-connoisseur*.

b. *Guide*

A *guide* also exhibits disciplinary knowledge. The difference with *representative* is that presenters show the audience where to focus or what to see in the visuals they use in the OP.

- (15) *This is a man with a lot of eyes and that is the **the image that** Jeremy Bentham, a philosopher in the 18th century, eh they planned a new or a type of institutional building. (S-M-HIST-2)*

c. *Architect*

Architects organize and outline content for the audience.

- (16) ... *but the problem can can be divided in two sub problems: the first the first problem is with the coefficient eh and...* (S-L-MATE-2)

d. *Recounter/announcer*

Recounters/announcers refer to the implementation of actions to produce and validate knowledge in the disciplines. They might include failures or achievements in the application of methods or procedures, or reference to how easy or difficult their application is, among other things.

- (17) *In this case is so **easy to** compare this quantifier free sentence with the list again, the theorems T but eh where each of one is quantifier free, because it's possible. (S-H-MATE-1)*

e. *Opinion-holder*

Opinion-holders show their own stances regarding existing knowledge or procedures.

(18) *but the other is just **a philosophy, a crazy idea, that** you don't find mathematical models.* (S-H-INGE-7)

f. *Originator*

Originators claim ownership towards what is said. They come in the form of new knowledge, solutions to problems, novel procedures, innovations, contributions and the like.

(19) ... construction, diffusion and eh deliberation of these products **can be eh [fs] can contribute to the ethical and citizenship education in this country.** (S-M-EDUC-1)

In my 2016 study, I also reinterpreted the visual Tang and John (1999) created (*Figure 6.2*) to show how the continuum of roles represented the different stance positions. My reinterpretation includes two concepts that Hyland (2012) uses to explain the construction of academic identities to show how they are useful in explaining the roles: *proximity* (adhering to existing rhetorical and knowledge production and disclosure ways) and *positioning* (expressing different positions towards existing knowledge or claiming ownership over new contributions).

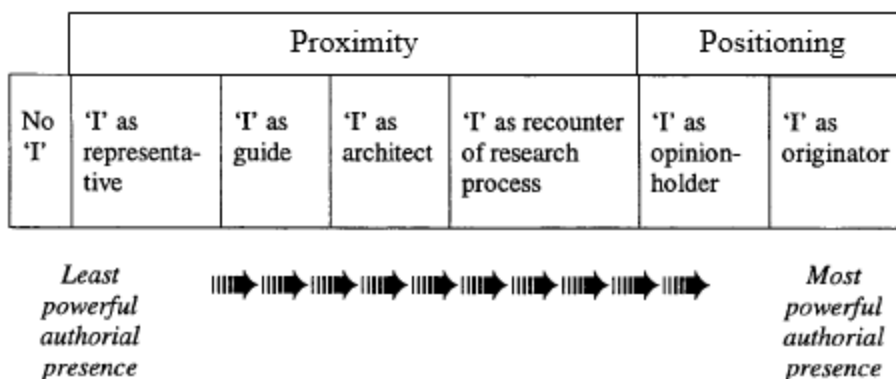


Figure 6.2. Tang and John's (1999, p.S29) typology of identities behind the first person pronoun in the light of proximity and positioning processes (Hyland, 2012).

My reinterpretation of *representative* requires a new visualization (Figure 6.3). Showing knowledge of disciplinary knowledge and its application is as strong an authorial stance as *recounter/announcer*.

PROXIMITY				POSITIONING	
'I' as representative (member of community)	'I' as guide	'I' as architect	'I' as recounter/announcer of the research process	'I' as opinion holder	'I' as originator
			'I' as representative (connoisseur)		

Figure 6.3. Reinterpretation of identity typology based on the expansion of the representative role

As previously argued, claiming membership to an academic community is not just a mere act of mentioning it; it is also an act of speaking as academics in those communities do. This way, listeners recognize that presenters talk like scholars (engineers, anthropologists, lawyers) do.

6.2.2.2. *Knowledge provision roles*

The identification of instances expressing knowledge contribution identities with impersonal modalized constructions was only possible for the *co-constructer* and *provider* roles. For the other (*learner*) (see 5.4.1.2) although it is possible to imagine instances for its expression, none were found in the corpus.

g. *Co-constructer*

Knowledge contributions are expressed as a presenter-audience conjoined creation. The audience is taken through a thinking process in which the presenter invites the audience to activate common knowledge or experience, usually general and not field-specific.

(20) *For example, eh we need eh have a school job [fs] a school activity. My machine could be could be a computer or a paper, eh and job is that activity that need perform a complete use of machines.* (S-M-INGE-4)

h. *Provider*

Knowledge contribution is construed as something given by the presenter, which probably the audience does not know.

(21) *Ok, but first what is a Panopticon? So, the name is also a refere [fs] a reference to panoptes from Greek mythology. He is a giant with a 100 eyes in [fs] in [fs] in his body [fs] in his body, and they can see everything.* (S-M-HIST-2)

6.2.2.3. *Language learning roles*

This subcategory of identity roles was conceived of as a continuum of language learning and use identities (see 5.4.1.3). No instances of this category were found being expressed with impersonal modalized constructions.

6.3. The study of impersonal identity projection in OPs

This chapter is based on the idea that impersonal expression of modality is a way of projecting academic identity roles. I have narrowed the search for studies that approach this theme to studies that focus on the use of modality in the projection of identities in student OPs.

Studies on the use of modality (or related concepts) in OPs have mainly focused on discourse socialization and the linguistic resources for its expression (Nausa, 2018). Discourse socialization is the study of the adaptation to a group's discourse practices. Morita (2000) studied the ways that TESOL graduate English NSs and NNSs expressed attitudes, judgements, and beliefs (epistemic stance) in preparation for OPs with tutors and classmates and during OPs. In OP preparation exercises, both NSs and NNSs use epistemic stance to acknowledge their (lack of) expertise in the delivery of OPs construing themselves as *novice* or *expert* presenters. In the OP delivery, students used epistemic stance to display (lack of) knowledge and engage with their audience. Discourse socialization in this study is closely related to the concept *proximity* (Hyland, 2012) in academic identity construction as the adoption and adaptation to a group's rhetorical practices that I also analyse under the headings *representative*, *guide*, *architect* and *recounter-announcer*. This study employs an ethnographic approach; the expression of epistemic stance is not approached with (corpus) linguistics methods. Kobayashi (2006) analysed the L2 socialization of Japanese undergraduate students in OPs of their experience as volunteers. Using

an SFL approach, the author identified that the use of relational and sensing verbs was a mechanism to describe thoughts and feelings. Also, L2 socialization was demonstrated in the use of linguistic resources to organize their talk and engage with the audience. Although not specifically stated, it is clear that the linguistic resources that student employed in L2 socialization allow them to construct identities of *opinion-holders*, *architect-guides* and *co-constructors*.

Other studies include a more linguistic approach to the expression of modality (as stance or identity projection) in OPs. Zareva (2012) analysed the expression of stance and persuasion in a corpus of OPs by TESOL and TEFL NS and NNS students. The analysis of the NSs and NNSs OP corpora shows that both groups considered it important to project themselves as experts as evidenced in the frequencies of use of 1st person stance structures, and stance adverbials. impersonal *it*-stance structures were found to be not very frequent in both groups. The NS group more frequently used these structures, but the author warns us that this should not be interpreted as this group setting a standard to be followed by the NNS group. In a similar study, (Zareva, 2013), the author used Tang and John's (1999) taxonomy of authorial stance roles to identify the academic identities projected by a group of TESOL students. She also included other types of identities (professional and institutional) in the analysis. The author found that the roles that were more frequently projected were the authorial stance ones.

Other studies that can be included in this survey are those that analyse modality in multimodal discourses. They are surveyed in section 4.5.

The study that I propose in this chapter seeks to fill some gaps in these studies. First, they were performed in ESL / EFL contexts with very similar populations: students in the second-foreign language teaching field; students in my study are all from at least 12 different fields.

Also, these students are either undergraduate or master's students; in my study, I analyse PhD-level students. Second, these studies make comparison of NSs and NNSs of English; I propose an analysis of three levels of oral achievement (high, medium, and low) within an NNS population. Finally, except for Zareva's (2012) stance adverbials, I propose an analysis of impersonal resources; studies on academic discourse tend to focus on personal realisations in the analysis of authorial stance identity projection.

6.4. Methods

This section explains the procedures to select impersonal modalized realisations and the types of analyses performed.

6.4.1. Selection of sentences for analysis and data clean-up

The identification of modalized sentences was performed in four phases. First, the 72811-token oral corpus was tagged for parts of speech (POS) with TagAnt (Anthony, 2018b). POS-tagging annotation allows the search of more complex grammar structures combining words and POS-tags (Groom, Charles, & John, 2015) like the modal-like patterns analysed in this chapter. To identify sentences containing modals, adverbs, and modal-like expressions, the concordance function of Antconc (Anthony, 2014) was used. Second, each sentence was analysed in context to identify the type of modality being expressed (6.2.1), the identity role being construed (6.2.2), associated discourse functions, and recurrent patterns (Hunston & Francis, 2000). Third, the following sentences were eliminated from the obtained concordance lines to avoid inflation of data: (1) sentences containing personal pronouns, as explained above, this chapter focuses on impersonal projection of identity through the expression of modality; (2) sentences with false

starts in which the expression of modality is abandoned; (3) sequential repetitions (e.g. *this law should be should be approved*); in cases like this, only one realisation was kept. To validate the classification of sentences into identity types, a colleague was trained and invited to recode 50 randomly selected sentences. I calculated Krippendorff's alpha value interrater agreement. The value (0.636) was found to be substantial (see Table 3.7). Fourth, raw frequencies and statistical significance tests were calculated to identify what identity roles discriminate among the different subcorpora.

6.4.2. Quantitative and discourse analyses

The answers to the questions that guide this study (6.1) include two types of analysis.

The first, quantitative analyses, focus on raw frequencies, percentages (3.4.1), and statistical significance tests and effect size scores (3.4.2). Raw frequencies and percentages are used to (1) determine the distribution of identity roles with impersonal modal and modal like realisations in the general corpus and (2) compare these identity role projection distributions with first person pronoun realisations in Nausa (2016). Statistical significance and effect size tests are performed among the subcorpora to (3) determine what identity roles frequency differences significantly discriminate among the disciplinary and level of achievement divides. I set a threshold log likelihood value of 10.83 ($p < 0.001$) (Table 3.4) to determine what identity role frequency differences are statistically significant in subcorpora comparisons. Those identity roles that exhibit values higher than the threshold I set are analysed in the second phase: discourse analyses.

Discourse analyses by role include two phases: one quantitative analysis and then a proper discourse analysis. The first phase starts with the normalised frequencies (per 10,000

words) of modal and modal-like impersonal realisations used in the projection of selected identity roles. Normalised frequency analyses seek to determine what modal and modal-like expressions more frequently occur in the level of achievement or disciplinary comparisons (overuse and underuse). The second phase seeks to describe and explain (1) the modalizations in the identity roles being projected, (2) recurrent patterns, and (3) discourse functions performed. For these qualitative analyses, I use the model explained in section 6.2.

6.5. Quantitative analyses

This section has two main purposes. The first is to determine the distribution of modal and modal-like realisations in the expression of different identity roles and how these distributions compare with personal pronoun realisations in Nausa (2016). The second is to identify what identity roles discriminate among levels of achievement and disciplinary subcorpora as expressed by statistical significance tests. This second part (6.5.2) also makes comparisons with the personal pronouns study and sets the stage for discourse analyses.

6.5.1. Modals, adverbs, and modal-like realisations in the impersonal expression of identity roles

Table 6.1 presents the linguistic realisations of modalities and their corresponding impersonal identity role projections in the 72811-token corpus of oral presentations. A total of 973 sentences were identified as containing modals, adverbs, and modal like expressions. In this 973-sentence subcorpus, 64% of modalized content is expressed with modal verbs (33.1%) and adverbs (30.9%). The remaining 36% is expressed with modal-like expressions: NOUN *that/to* (9.2%), VERB *that/to* (17.3%), and ADJ *that/to* (9.5%) constructions. This modals overuse

seems to confirm Aijmer's (2002) finding that learners of English use modal verbs more frequently than native speakers do.

In terms of the identity roles expressed, it can be observed that the most frequently expressed identity roles are of the academic authorial stance type: *representative-connoisseur* (42.3 %) *opinion holder* (27.5%), and *recountor/announcer* (14.8%). Knowledge provision roles only accounted for 6% of the realisations and comprised only *co-constructer* (3.5%) and *provider* (2.5%).

Table 6.1. Modals and modal-like realisations, and identity roles expressed in raw frequencies and percentages

	%		%		Noun + (that /to)	%	Verb s + (that /to)	%	Adjecti ves + (that / to)	%	Total		%
Modals	Adverb s												
Academic authorial stance roles													
Representative	100	10.3	180	18.5	42	4.3	85	8.7	5	0.5	412	42.3	
Guide	0	0.0	0	0.0	6	0.6	1	0.1	0	0.0	7	0.7	
Architect	2	0.2	7	0.7	0	0.0	23	2.4	8	0.8	40	4.1	
Recountor / Announcer	90	9.2	15	1.5	17	1.7	3	0.3	19	2.0	144	14.8	
Opinion holder	93	9.6	72	7.4	13	1.3	39	4.0	51	5.2	268	27.5	
Originator	23	2.4	4	0.4	11	1.1	2	0.2	4	0.4	44	4.5	
Knowledge provision roles													
Learner	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	
Co-constructer	9	0.9	21	2.2	1	0.1	0	0.0	3	0.3	34	3.5	
Provider	5	0.5	2	0.2	0	0.0	15	1.5	2	0.2	24	2.5	
Language use roles													
Learner	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	
User	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	
Provider	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	
Total	322	33.1	301	30.9	90	9.2	168	17.3	92	9.5	973	100.0	

The tendencies observed here partially differ from the ones in my previous study on first person realisations (*I, my, me, we, us, our*). In that study, the most frequently expressed roles were *recountor/announcer* (25.2%), *architect* (19.3%) and a knowledge contribution one, *co-*

constructor (15.6%) (see Table 10.1). From the comparison between the studies, it can be concluded that for the expression of the *recounter/announcer* role, students indistinctly make use of personal and impersonal realisations. However, impersonal constructions are preferred in the expression of strong authorial stances (*opinion-holder*) arguably as a safe facing strategy, or for the expression of knowledge attribution (*representative-connoisseur*). Personal pronoun realisations are the preferred choice in the organization of discourse (*architect*) and the mitigated facilitation of content (*co-creator*). In other words, when it comes to making the message clearer for the audience, more interactive features (pronouns) are used.

These distributions described in this section apply for the general corpus of PhD student OPs. The following section focuses on a statistical significance analysis of these distributions considering the two variables in the study: level or oral achievement and disciplinary divide. The purpose is to identify whether general distributions remain the same or change in the subcorpus analyses.

6.5.2. Identity roles: significance of frequency differences and effect size in the level of achievement and disciplinary sub-corpora

A closer analysis of the most frequent identity roles that modals and modal-like expressions serve to express reveals interesting tendencies.

6.5.2.1. by level of achievement

In the level of achievement frequency difference comparisons (Table 6.2), *opinion holder* (19.27), and *representative-connoisseur* (11.79) values are significant at $p < 0.001$. This indicates that the expression of these roles through impersonal modalized realisations can be seen as a

mark of oral achievement. For example, in the case of the *opinion holder* role, high achievers use more modalized realisations (133) than medium (85) and low (50) achievers. However, in the case of the two roles, the negative Bayes Factor effect size values seem to indicate that more evidence (a larger corpus) is needed to have positive evidence against the null hypothesis.

Table 6.2. Raw frequency and significance analysis of impersonal modalized academic identity role realisations by level of achievement

				Log	Bayes
Academic authorial stance roles	High	Medium	Low	Likelihood	Factor BIC
Representative	180	148	84	11.79	-10.6
Guide	2	2	3	0.81	-21.58
Architect	9	16	15	4.33	-18.06
Recounter / Announcer	69	46	29	7.52	-14.87
Opinion holder	133	85	50	19.27	-3.122
Originator	22	11	11	3.46	-18.93
Knowledge provision roles					
Learner	0	0	0	0.0	-22.4
Co-constructer	21	5	8	10.3	-12.1
Provider	7	12	5	2.0	-20.4
Language use roles					
Learner	0	0	0	0.0	-22.4
User	0	0	0	0.0	-22.4
Provider	0	0	0	0.0	-22.4
Total	443	325	205		
Corpora Sizes	27038	26117	19656		

Again, the tendencies observed in the impersonal modalized projection of identity roles differ from those of identity projection with first person pronouns. In Nausa (2016) two different roles *originator* and *co-constructer* were highly significant at $p < 0.0001$ and also with high positive effect size values (see Table 10.2). In both studies, nonetheless, identity projection language choices are more frequent in high achievers' OPs. Therefore, from both studies it can be concluded that in the projection of academic identity roles, high achievers more frequently

use personal pronouns to claim ownership over new generated knowledge (*originator*) and to facilitate the audience understanding (*co-constructer*). However, when it comes to expressing attributed disciplinary knowledge (*representative-connoisseur*) or taking specific stances (*opinion-holder*), the preferred choice is impersonal modalized realisations, probably, as argued above, as a face-saving strategy.

6.5.2.2.by discipline

In the disciplinary divide comparison (Table 6.3), *recountner/announcer* and *opinion-holder* not only exhibited high log likelihood values significant at $p < 0.0001$, but also positive effect size Bayes Factor values indicating that there is very strong evidence against the null hypothesis (impersonal identity projection with modalized expressions does not correlate with the discipline of the person expressing modality). This is evidenced in the raw frequencies. *Recountner/announcer* modalized projections were four times as frequent in the hard-disciplines OPs (116) as in the soft-discipline OPs (28). Opinion-holder realisations, on the other hand, were almost twice as frequent in the soft disciplines (177) as in the hard disciplines (91).

Table 6.3. Raw frequency and significance analysis of impersonal modalized academic identity role realisations by discipline

Academic authorial stance roles	Hard	Soft	Log Likelihood	Bayes Factor BIC
Representative	206	206	0.0	-11.2
Guide	1	6	4.0	-7.2
Architect	17	23	1.0	-10.2
Recountner / Announcer	116	28	56.9	45.7
Opinion holder	91	177	28.9	17.7
Originator	26	18	1.4	-9.8
Knowledge provision roles				
Learner	0	0	0.0	-11.2
Co-constructer	15	19	0.5	-10.7
Provider	15	9	1.5	-9.7
Language use roles				
Learner	0	0	0.0	-11.2
User	0	0	0.0	-11.2
Provider	0	0	0.0	-11.2
Total				
Corpora Sizes	36579	36232		

Interestingly, the same tendency is observed in my personal pronouns study of identity projection (Nausa, 2016). *Recountner/announcer* and *opinion holder* were also significant at $p < 0.0001$ as expressed by log likelihood values (Table 3.4) and also exhibited high positive effect size values. Also, in both studies, the distribution by discipline was the same. *Recountner/announcer* projections were more frequent in the hard disciplines while opinion holder projections were more frequent in the soft disciplines.

As explained in the pronouns study, these tendencies can be accounted for considering the epistemological features of knowledge in hard and soft fields (Becher & Trowler, 2001). In their characterization of disciplines based on hard-soft/pure-applied dyads, the authors provide a conceptualization of the nature of knowledge in the hard and soft disciplines (Figure 6.4).

Hard (pure and applied)	Soft (pure and applied)
<ul style="list-style-type: none"> • atomistic • concerned with universals, quantities, simplification • impersonal • value-free • clear criteria for knowledge verification and obsolescence • consensus over significant questions to address, now and in the future • results in discovery / explanation • results in products/techniques 	<ul style="list-style-type: none"> • holistic (organic/ river-like) • concerned with particulars, qualities, complication • personal • value-laden • dispute over criteria for knowledge verification and obsolescence • lack of consensus over significant questions to address • results in understanding/ interpretation • results in protocols/ procedures

Figure 6.4. Nature of knowledge in the hard and soft disciplines (adapted from Becher & Trowler, 2001, p. 36)

Opinion holder projections reflect several characteristics of knowledge in soft fields.

Opinions are concerned with particular qualities, personal, based on values, and interpretative.

Recounter/announcer realisations convey the adherence to clear criteria for knowledge verification techniques defining of the hard sciences.

In brief, the distribution of personal and impersonal realisations in the expression of academic identity roles vary in the level of achievement comparisons but remain the same in the disciplinary ones. As discussed, high achievers show a preference for the use of pronouns to project *originator* and *co-constructer* roles while they use impersonal modalized constructions to project *representative-connoisseur* and *opinion holder* roles. Use of pronouns seem to indicate the need of interactive features to express what their contributions are and to facilitate the audience understanding of contents. Impersonal modalizations seem to allow high achievers to save face when it comes to express their opinions or to attribute the origin of the contents they

use to others. In the disciplinary comparisons, the nature of the knowledge used in the discipline appeared to be the defining factor. Soft discipline students use personal and impersonal realisations to express opinions while hard science students use personal and impersonal realisations to express adherence to disciplinary knowledge production precepts.

The following section focuses on how impersonal modalized projections are used in the projection of identity roles.

6.6. Discourse analyses: impersonal academic identity projection

This section presents the typical patterns and the discourse functions of the identity roles that were determined to significantly discriminate among levels of achievement (*opinion-holder* and *representative*) and disciplines (*recounter/announcer* and *representative*). Although the purpose is to provide a qualitative discourse account of language choices with examples from the corpus, normalised frequencies will be used to make comparisons between the subcorpora and to determine what specific modals, adverbs, and modal-like expressions are preferred in the projection of identity roles.

6.6.1. By level of achievement

The academic identity roles that exhibit the highest log likelihood and Bayes Factor Approximation (BIC) values are *opinion holder* (19.20) and *representative* (11.76).

6.6.1.1. *Opinion holder*

Opinion holder is the first recurrent role in the oral corpus with 268 (out of 973) realisations and the one with greatest log likelihood value in the level of achievement

comparisons. The most frequently used linguistic devices in its expression are modals (12.8), adverbs (9.9), and adjectives + *that/to* (7.0) patterns as observed in the normalised per 10,000 words frequencies (Table 6.4). In the three cases, high achievers more frequently used these three resources than medium and low achievers.

Table 6.4. Modal and modal-like raw and normalised (N) frequencies in the expression of the opinion holder role

	Modals	N	Adverbs	N	Noun + (that / to)	N	Verb s + (that / to)	N	Adjectives + (that / to)	N	Total	N	Corpus size
High	51	18.9	34	12.6	10	3.7	16	5.9	22	8.1	133	49.2	27038
Medium	30	11.1	23	8.5	1	0.4	11	4.1	20	7.4	85	31.4	26117
Low	12	4.4	15	5.5	2	0.7	12	4.4	9	3.3	50	18.5	19656
Total	93	12.8	72	9.9	13	1.8	39	5.4	51	7.0	268	36.8	72811

When modal verbs were used in impersonal projections of *opinion holder*, the modality types that were frequent in its expression were probability (*could*, *would*), possibility (*can*, *could*), obligation (*should*, *must*), necessity (*need*, *must*), and ability (*can*).

- (21) *The models represent the counter reality, eh and it would be a suggestion it would be interesting development eh a solution that support [fs] the solution to support the interacting of the model and metamodel in this particular context because eh it is a problem that emerge.* (S-M-INGE-6)
- (22) *This is important because the people can interact with the nature and to learn to respect the life in all scales.* (S-H-CBIO-2)
- (23) *The information in any kind of of format, for example in paper, in digital information, all all all kind of of information should be accessible for the citizens.* (S-M-CPOL-2)
- (24) *The problem is that the current standard tools that eh does eh don't support the breaking eh of this conformity and in these eh tools eh the conformity always must be guaranteed and this is a a really problem in the in this eh kind of projects.* (S-M-INGE-6)
- (25) *Eh an additional problem is [fs] can be solved by molecular recognition is the human health.* (S-H-CQUI-1)

According to Halliday and Matthiessen's (2014) description of orientation, modal verbs are implicit subjective realisations of modalities. They are implicit, for they are within the proposition they modalize; they are subjective because the choice of a modal verb construes the proposition as modalized by the speaker/writer. It can then be argued that subjective orientations are at the core of authorial stance identity construction.

The use of adverbs in impersonal modalized projections of *opinion holder* include the expression of probability (*increasingly, only, simultaneously, obviously, necessarily*), obligation (*illegally, incorrectly*), usuality (*usually*), and possibility (*really, unfortunately, badly*).

- (26) *So the first challenge is [reading 3] globalization means a new social and economic and political order that **necessarily** presupposes a close relationship among different countries [reading 3]. (S-H-EDUC-1)*
- (27) *In developed and underdeveloped countries, the union between public policy and entrepreneurship is a powerful tool for promoting economic growth. Doing it **incorrectly** can have devastating consequences for modern capitalist society. (S-H-ADMI-2)*
- (28) *Eh and the conclusion eh is that eh **usually** the enterprise architecture projects analyze the reality of the enterprise through one model eh that is based in eh one metamodel and the model represent the abstract of the [fs] the abstraction of the reality. (S-M-INGE-6)*
- (29) *The problem is **unfortunately** compliance with family obligation and work obligation create conflict between both roles, situation known as work family conflict. (S-L-ADMIN-1)*

Adverbs are implicit objective realisations of modality (Halliday & Matthiessen, 2014). In their use, presenters construe stances not as their own, but as somebody else's or as guided by an external moral-ethical code (objective). However, this seems to apply to examples 26-88 but not to 29. In 26 – 28, modalizations seem to be out of the speaker's judgement or human control.

The expression of personal opinion in these sentences seem to be not only in the adverbs but also in other words (*challenge, means, presupposes, powerful, promoting, devastating, conclusion, etc*). In 29, the adverb *unfortunately* is hardly conceivable as something objective. Its use is arguably a subjective realisation of modality.

Finally, the third type of *opinion holder* realisation is ADJ *that/to* patterns. These include the expression of probability (*clear that, true that*), necessity (*necessary that, important that, necessary to*), possibility (*possible to, impossible to, difficult to*) and ability (*able to*). A number of sentences containing these constructions include *it is* projecting clauses. However, as it is common in this student population, the grammatical subject *it* was sometimes elided, probably because of Spanish not having this type of grammatical subject for projecting clauses.

- (30) *So eh this eh statement eh made by Fisher and Reuber in 2003 makes [fs] makes us **clear that** eh developed countries are already on the way of foster high growth firms. (S-H-ADMI-2)*
- (31) *So, what do we need to answering this question? So [fs] then, [reading 6] **is necessary to analyse plant consumption and dispersal patterns** [reading 6], and what kind of data do we need. (S-H-CBIO-6)*
- (32) *So, [reading 2] the good thing is that **it is possible** making wise choices among the huge range a [fs] that market offers [reading 2]. (S-H-ANTR-2)*
- (33) *As a result of this kind of reparation, Unión Patriótica **was able to participate in 2014 political elections with its president Aida Avella, if you remember she was the vice-president of the Polo Democrático alternative with Clara López the last year, and Unión Patriótica **will be able to do the same in 2018 elections with another candidate.** (S-H-DERE-2)***

All cases of ADJ *that/to* constructions construe modality as something objective. Some studies (e.g. Carter-Thomas & Rowley-Jolivet, 2001) have found *it* V-link ADJ *that* constructions to be more typical of written academic discourse. I do not interpret this to be

something to criticize these students' performances. On the contrary, what my study demonstrates is that they (high achievers) exhibit more semiotic choices to express their ideas.

In summary, modals, adverbs, and ADJ *that/to* constructions are useful in the expression of strong authorial role positions (*opinion-holder*). Modalities in the expression of this role include probability, possibility, obligation, necessity and usuality. Halliday and Matthiessen (2014) concept *orientation* is useful in explaining whether the expression of modality is implicit or explicit. However, when it comes to defining whether a realisation is subjective or objective, sometimes it is not clear whether subjectivity refers to the inclusion of the *opinion holder* in the modalizing expressions. If this was the case, none of the chosen sentences could be considered as subjective given that this chapter focuses on impersonal projections. However, as was discussed in the case of adverbs, described by the authors as objective realisations, cases like *unfortunately* are arguably a subjective way of judging content.

6.6.1.2. Representative-connoisseur

Normalised (x 10,000 words) frequencies (Table 6.5) show that adverbs (24.7), modal verbs (13.7), and VERB *that/to* constructions (11.7) are the most recurrent linguistic mechanisms in these identity projections. Adverbs and VERB *that/to* constructions are more frequently used in high-rated OPs than in medium and low rated OPs. Modal normalised frequency uses are notably lower for low-rated OPs (3.7).

Table 6.5. Modal and modal-like raw and normalised (N) frequencies in the expression of the representative role

	Modals	N	Adverbs	N	Noun + (that/to)	N	Verb + (that/to)	N	Adj + (that/to)	N	Total	N	Corpus sizes
High	37	13.7	82	30.3	16	5.9	42	15.5	3	1.1	180	66.6	27038
Medium	53	19.6	52	19.2	17	6.3	25	9.2	1	0.4	148	54.7	26117
Low	10	3.7	46	17.0	9	3.3	18	6.7	1	0.4	84	31.1	19656
Total	100	13.7	180	24.7	42	5.8	85	11.7	5	0.7	412	56.6	72811

Common modalities in *representative-connoisseur* projections include probability (*probably, basically*), usuality (*usually, constantly*), ability (*biologically, quickly, completely*) and possibility (*intentionally, voluntarily*). Adverbs in these modalities are placed in clause final position or before main verbs.

- (34) *It's the same thing in agriculture, for example, if the temperature of humidity is high or low the crops **maybe** or **probably** will be lost.* (S-H-INGE-1)
- (35) *in addition most people **usually** spend more time, a [fs] spend more money in order to get a eh college diploma.* (S-H-ECON-3)
- (36) *Ok, eh the construction was planned in 1849 but it took almost 30 years to get built **completely**, why?* (S-M-HIST-2)
- (37) *Eh, if we use this model to understand self-deception, self-deceivers **intentionally** get themselves to believe P while knowing or truly believing no P.* (S-H-PSIC-1)

Modal verbs are the second recurrent device in impersonal *representative-connoisseur* projections. Possibility (*can, could, may*), ability (*can*), and probability (*can, could, will*) are the most commonly expressed modalities. Modals appear in active and passive impersonal constructions.

- (38) *when you try to characterize characterize the the endomorphism for this extension is the multiplication, but the problem can can be divided in two sub problems.* (S-L-MATE-2)
- (39) *this is a a perfect example that both traffic or eh dolphins watch activities, local (intelligible) or or control could eh or have eh negative effects to long term and the population declined to the time.* (S-H-CBIO-5)
- (40) *Eh, for example [reading 10] if a person badly wants X but knows or feels that X cannot be true, it has evidence about that, then the person can activate a cognitive process in which he can avoid the true or evidence [reading 10].* (S-H-PSIC-1)

The third recurrent device in modalized *representative-connoisseur* projections is VERB *to/that* constructions. These are mainly used to express probability. Because the *representative-connoisseur* role implies the expression of others' modalizations, the most common patterns include activity verbs (*show that*), communication verbs (*say that*), and copula *BE (is that)*. Also, because they are not part of the content they modalize, the way they express modality is explicit. VERB *to* patterns are not frequent.

- (41) *Several rese research investigation **has shown that** eh host-guest complex can be eh recharged at least a hundred more cycles than a single metal.* (S-H-CQUI-1)
- (42) *Eh consejo de Estado **says that** those requirements, the minimum number of citizens, for example 3,000 people were impossible to accomplish accomplish eh to Unión Patriótica, why?* (S-H-DERE-2)
- (43) *The other the other **question is that** if you keep a lot of inventory, then if the products deteriorate with the time, then you're gonna be loss of ways, and then your costs are gonna be high, so you need a balance.* (S-H-INGE-7)
- (44) *This eh criteria is eh what companies use to try to win a contract with the Colombian state.* (S-H-INGE-2)

There are some differences among the patterns too. *Show* and *say that* are used to attribute (Sinclair, 1988) the modalization to somebody else (e.g. an author, an institution, a theory, previous studies). The *is that* pattern, however, tends to collocate with noun phrases that refer to what Hunston (2011) refers to as *status labels*. Hunston explains status as the "... identification, on named criteria, of categories of objects, and the labelling of each object according to the category it belongs to." (p. 30). Hunston proposes as examples of status labels FACT, HYPOTHESIS, SUPPOSITION, and the like. This concept is important in the expression of *representative-connoisseur*. Being able to appropriately use status words to mark propositional content as an epistemological object is something that is expected of a person who belongs to a given academic community. Some of the NPs that work as a status labels and tend to collocate with *is that* pattern include CONCEPT OF, TOPIC, QUESTION, PROBLEM, SOLUTION, IDEA. This pattern is also present in other academic identity roles; in the process of analysis it was sometimes difficult to define whether the authorial stance expressed was strong (as in *opinion holder*) or just indicative of knowledge of content as in *representative-connoisseur*.

To sum up, quantitative analyses show that the variety of devices for this role expression are indicative of the levels. Low achievers, for example, exhibit fewer sentences expressing these roles, and sentences in turn exhibit fewer modalization mechanisms. Discourse analyses show that an important difference between the impersonal expression of *representative-connoisseur* and *opinion-holder* is the modalization of content as disciplinary (objective-attributed) or as specific positions taken towards such content (subjective-averred).

6.6.2. By discipline

In the disciplinary divide, the academic identity roles with the highest log likelihood (LL) and Bayes Factor Approximation (BIC) values are *recountor announcer* (LL: 56.9 / BIC: 45.7) and *opinion holder* (LL: 28.9 / BIC: 17.7). Values in both cases indicate that frequency differences are highly significant ($p < 0.0001$) and there is very strong evidence against possible null hypotheses.

6.6.2.1. Recountor/announcer

In the general corpus, *recountor-announcer* is the third most recurrent role. There are 144 out of 973 modalized sentences that invoke the role whose main trait is the allegiance to methodological procedures in the disciplines. As expected, this role is more frequently projected in the hard-disciplines sub-corpus (31.7) than in the soft-disciplines one (7.7) confirming the tendency I reported in module 2 (Nausa, 2016) in which *recountor/announcer* projections with first person pronouns are more recurrent in the hard disciplines. The devices that are repeatedly used in these identity projections per ten thousand words (Table 6.6) are modals (12.4) and ADJ *that/to* (2.6).

Table 6.6. Modal and modal-like raw and normalised frequencies in the expression of the recountor/announcer role

	Modals	N	Adverbs	N	Noun + (that / to)	N	Verbs + (that / to)	N	Adje ctive s + (that / to)	N	Total	N	Corpu s size
Hard	81	22.1	10	2.7	10	2.7	2	0.5	13	3.6	116	31.7	36579
Soft	9	2.5	5	1.4	7	1.9	1	0.3	6	1.7	28	7.7	36232
Total	90	12.4	15	2.1	17	2.3	3	0.4	19	2.6	144	19.8	72811

Modals and ADJ *that/to* realisations refer to actions that people in academic fields are capable of doing, need to do, or can choose to do to generate knowledge that is accepted in such communities. Also, they may refer to the steps, stages, or conditions that have to be met before proceeding to other research related actions.

Firstly, use of modals in *recounter/announcer* realisations include deontic meanings such as ability (*can*), necessity (*must, should*), and possibility (*can*), expressed in active and passive constructions.

- (45) non-linear mathematical models can identify when the energy losses and improve this subsistent in the car. (S-L-INGE-1)
- (46) *The abstraction pro process of this eh of this metamodel it is not easy like the like the example in the real in the real life because there are a lot of elements, there are a lot of of aspects that could that eh **must be** eh **must be** analyse and the relation in this element is very complica complicated.* (S-M-INGE-6)
- (47) *Eh the in-depth interview eh don't have [fs] doesn't doesn't have eh questions eh eh but the interviewer eh **can** keep in mind eh different topics to interview, to question.* (S-M-EDUC-2)

Secondly, ADJ *that/to* patterns include the expression of almost the same modalities necessity (*necessary that/to, important that/to*), possibility (*possible that/to, easy to, difficult to*), but not ability. Like modal realisations, they refer to aspects of research procedures in the fields.

- (48) *because the populations living in highlands are particularly unknown, the [fs] the groups are [fs] are no habitual to the researchers, which makes **harder to follow them and get data from them**, and also because the floristic composition in mountains are quite different from [fs] to the lowlands, so is **important to know what are the relations up there**.* (S-H-CBIO-6)
- (49) *For the coordination occur need electron non-bonding pairs. Be eh [fs] in this case eh the pyrazole have two electron non-bonding pair in the site occur the combination or union with copper. **Is very important that these molecules have electron non-bonding pair.*** (S-M-CQUI-1)

These realisations are objective-explicit (Halliday & Matthiessen, 2014) and although they have been found to be more common in written discourse (Carter-Thomas & Rowley-Jolivet, 2001), they are effectively used in the expression of the role. They arguably prove that these EFL students bring to the classroom rhetorical knowledge that comes from their fields. Of course, as can be observed in some of the examples, they could benefit from feedback that points out common errors (e.g. dropping of *it*, inflection of past participle verbs in passive constructions and the like).

6.6.2.2. *Opinion holder*

As pointed out above, this is the most frequent role in the oral corpus comparisons and the second with the greatest log likelihood and effect size values in the disciplinary divide comparisons. Recurrent devices in its expression are modals (12.8), adverbs (9.9), and ADJ *that/to* patterns (7.0). In the level of achievement comparisons, high achievers more frequently use these language traits; here in the disciplinary comparisons, soft-discipline PhD researchers use almost twice as many realisations as hard-discipline researchers as observed in the normalised frequencies (Table 6.7). A look at raw and normalised frequencies of the three language traits shows the tendency of soft fields to use them more frequently. For specific examples see 21 – 33 in section 6.6.1.1.

Table 6.7. Modal and modal-like raw and normalised frequencies in the expression of the *opinion holder* role

	Modals	N	Adverbs	N	Nouns + (that / to)	N	Verbs + (that / to)	N	Adjectives + (that / to)	N	Total	N	Corpus size
Hard	31	8.5	28	7.7	5	1.4	9	2.5	18	4.9	91	24.9	36579
Soft	62	17.1	44	12.1	8	2.2	30	8.3	33	9.1	177	48.9	36232
Total	93	12.8	72	9.9	13	1.8	39	5.4	51	7.0	268	36.8	72811

As expected, *opinion-holder* projections are more frequent in the soft-field sub-corpus (48.9) than in the hard-field one (24.7), almost twice as much, again confirming the tendency reported in (Nausa, 2016), in which I found *opinion-holder* projections with first person pronouns to be more recurrent in the soft disciplines.

6.7. Conclusion

This chapter has analysed the projection of academic identity roles in the OPs of Colombian PhD researchers in an EAP class. More specifically, the chapter focused on how presenters projected their academic selves with the use of modals, adverbs, and modal-like expressions in impersonal realisations. As such, the chapter has attempted to answer the following questions:

- **Tendencies:** what are the tendencies in impersonal expressions of modality in the oral presentations of Colombian PhD researchers as observed in their use of modals and some modal-like (Hunston, 2011) expressions?
- **Identity roles:** What identity roles are projected when presenters express modality in impersonal constructions? How does the impersonal projection of identity roles compare with personal projections?
- **Level of achievement:** What are the differences in the impersonal projection of identity among high, medium and low-rated presentations?
- **Disciplinary divide:** What are the differences in the impersonal projection of identity between hard-discipline and soft-discipline researchers?
- **Personal and impersonal realisations:** How do these tendencies (by level and disciplines) compare with the tendencies in identity projections with personal pronouns?

The following is a summary of the answers to the questions (see Table 6.8):

Tendencies: Out of the 973 sentences containing impersonal modalizations, modals and adverbs accounted for 64% of the cases and modal-like constructions for 36%. The higher incidence of modals and adverbs can have two explanations that are not necessarily consistent. First, the preference for modals and adverbs could be due to the instruction students at this level (A2 and B1) have received. However, and this is the second explanation, the higher frequency of these devices could be explained by the nature of the task (oral presentation). Halliday and Matthiessen (2014) have found modals and adverbs to be more interpersonal and congruent (as opposed to metaphorical) expressions of modality. In general, congruent realisations are more typical of oral discourse.

Identity roles: 94% of the realisations convey academic authorial stance identities, while only a 6% conveys knowledge provision ones, and 0% convey language use identity roles. In personal projections (Nausa, 2016), 72.5% express academic identities, 25% express knowledge provision identities, and 2.47% express language use identities. The higher incidence of academic roles expressed by impersonal realisations might be because these roles are less interactional in nature than knowledge provision or language roles. In knowledge provision roles, students address the audience in a way that new knowledge is perceived as given or co-constructed. In language use roles, students ask for help, teach new expressions, or demonstrate that they can deal with language breakdowns.

Level of achievement: *opinion holder* and *representative* are the roles that more significantly discriminate among students. Modals and adverbs are frequent in both roles; *ADJ* that/to in *opinion holder* and *VERB* that/to in *representative*. These frequencies were expected since adjectives seem an appropriate choice in the expression of opinions and verbs (communication and mental) in the report of others' ideas. Also, as expected, these roles are

more frequent in high-rated OPs; these OPs also exhibit the use of more types of modalizations, resources to express them, and associated discourse functions. In comparison to Nausa (2016), it is observed that the roles change. The most significant roles in the pronouns study were *guide*, *originator* and *co-constructer*. Given that *guide* and *co-constructer* imply interaction with the audience, personal pronouns seem to be a congruent choice. In the *you-audience* study (5.6.1) the mirror image roles coincide with the ones in my second module study: *guide-tourist*, *originator-innovation user*. *Co-constructer* does not significantly discriminate among levels of oral achievement in the *you-audience* study.

Disciplinary divide: the roles that co-relate with the disciplinary divide are *recountner/announcer* and *opinion-holder*. Frequent linguistic resources in *co-constructer* are modals and ADJ *that/to* patterns and, in *opinion-holder*, modals, adverbs, and ADJ *that/to*. *Recountner/announcer* realisations are more frequent in hard-discipline OPs and express modalities referring to research actions that are necessary, obligatory, or optional to achieve research goals. *Opinion holder* realisations are more frequent in the soft-field OPs and include modalities that refer to how students position in regard to existing knowledge in their disciplines or the contexts related to them (probability, possibility, obligation, and the like). Interestingly, I also reported these roles as statistically significant in the frequency comparisons in my pronouns study. In the *you-audience* study (chapter 5), the *research apprentice-role* (the *recountner-announcer* role mirror image) was also found to be statistically significant in the disciplinary comparisons.

This study confirms several of the findings in similar studies on identity projection-modalization (Kobayashi, 2006; Morita, 2000; Zareva, 2013; Zareva, 2012). However, this study differs from those in its narrower focus on impersonal realisations and the inclusion of tests that

account for statistical significance and effect size. It also includes comparison variables not considered in the abovementioned studies: level of achievement, disciplines, PhD-level, and the analysis of modality from a discourse perspective. Additionally, this study continues to use and enrich Tang and John's (1999) taxonomy that has been successfully used in the authorial stance analysis of first person (Nausa, 2016; Zareva, 2013) and second person identity projections (CHAPTER 5).

Table 6.8. Summary of findings impersonal modality study

Comparison	Identity role (Log likelihood and Bayes Factor)	Modal and modal-like expressions (normalised frequencies *10,000)	Over use (↑) or under use (↓)	Modalities	Discourse functions
Level of oral achievement (high, medium, low)	<i>Opinion holder</i> (LL: 19.2/BF:-3.1)	Modals (12.8)	High rated OPs (↑)	<u>Probability</u> : <i>could, would, increasingly, only, simultaneously, obviously, necessarily, clear that, true that</i> ; <u>Possibility</u> : <i>can, could</i> ; <u>Obligation</u> : <i>should, must, illegally, incorrectly</i> ; <u>Necessity</u> : <i>need, must, necessary that, important that, necessary to</i> ; <u>Ability</u> : <i>can, able to</i> ; <u>Usuality</u> : <i>usually</i>	Own judgement of (averred): - Previous research - Theories - Common knowledge - Importance of own study - Future research venues
		Adverbs (9.9)			
		Adj + <i>that/to</i> (7.0)			
	<i>Representative</i> (LL: 11.7/BF:-10.6)	Adverbs (24.7)	High rated OPs (↑)	<u>Probability</u> : <i>probably, basically, show that, say that, STATUS NOUN + is that</i> ; <u>Usuality</u> : <i>usually, constantly</i> ; <u>Ability</u> : <i>biologically, quickly, completely</i> ; <u>Inclination</u> : <i>intentionally, voluntarily</i>	Expression of other authors, theories, paradigms, etc (attribution): - Concepts - Ideas - Theories - etc
		Modals (13.7),			
		Verb + <i>that/to</i> (11.7)			
Disciplinary divide (hard vs soft)	<i>Recountner/announcer</i> (LL: 56.9/BF: 45.7)	Modals (12.4)	Hard-field OPs (↑)	<u>Ability</u> : <i>can</i> ; <u>Necessity</u> : <i>must, should, necessary that/to, important that/to</i> ; and <u>Possibility</u> : <i>can, possible that/to, easy to, difficult to</i>	Expression of methodological procedures that... - other researchers in the field know - need to be followed - are possibilities ... to generate new knowledge or achieve research goals in disciplines
		Adjectives + <i>that/to</i> (2.6)			
	<i>Opinion holder</i> (LL: 28.9/BF: 17.7)	See row 2	Soft-field OPs (↑)	See row 2	See row 2

PART II: MAKING CONTENT EASY FOR THE AUDIENCE

CHAPTER 7

IN THIS CASE...: INCLUSION OF CODE GLOSSES

7.1. Introduction: code glosses as a mark of high oral achievement

In module 1 (Nausa, 2015) I conducted a pilot study whose purpose was to identify linguistic traits of the English written and spoken by the group of Colombian PhD researchers analysed in this thesis. The study focused on the strategies that students use to transition from written (essays) to spoken (oral presentations) discourse to express the same content. Level of achievement was the variable that was considered in the analysis of 8 pairs of essays and OPs (each pair by the same author) expressing the same content. Two groups were compared: 4 high and 4 low-achievers. Four mechanisms to modify written content were found: modifications to SVO clause structure, reduction of heavily modified NPs, changes to the expression of modality, and code glosses. The main finding in that study was that the four mechanisms and related sub-mechanisms were more consistently used (grammatically and pragmatically) by high achievers. Nonetheless, one of the limitations was that the findings were inconclusive given the small corpus and the lack of more consistent corpus and statistics procedures. Chapters 7 and 8 seek to remedy these limitations. This chapter focuses on code gloss use but only in the OPs as a mechanism that is used to make the content accessible to a non-expert audience.

The purpose of this chapter is to determine whether the distribution of code glosses correlates with the variables in the analysis of this population: level of achievement and disciplines. The chapter attempts to answer the following questions:

1. **Code glosses:** what are the tendencies in the use of code glosses in the OPs of Colombian PhD researchers?
2. **Level of achievement:** What are the differences in the use of code glosses among high, medium and low-rated presentations?
3. **Disciplinary divide:** What are the differences in the use of code glosses between hard-discipline and soft-discipline researchers?

7.2. Code glosses: an interactional and interactive type of metadiscourse

Code glosses are one type of *metadiscourse*. Metadiscourse has been defined as the commentary made by discourse producers to help their intended audience understand what they say or write (Hyland, 2017). Vande Kopple (1985) proposed seven types of metadiscourse which can be classified into SFL's textual and interpersonal metafunctions (Hyland, 2007). Textual metadiscourse (1) helps the audience understand the organization and connection of different types of a discourse, (2) the meaning of elements in text, (3) the authorship (or attribution) of ideas, and (4) the author's epistemological evaluation of what is said. With interpersonal metadiscourse, authors indicate to the audience (5) the type of action that they are performing as discourse unfolds, (6) their deontic attitude to what is said, and (7) the fact that they are directly addressing the audience. In a similar vein, Thompson (2001) proposes two types of resources that writers use to interact with their readers and that (Hyland, 2005) interpreted as metadiscourse: interactive and interactional. Interactive metadiscourse controls the flow of information and guides the reader through discourse. Interactional resources, on the other hand, have the purpose of engaging the audience in the reasoning process happening as the text unfolds. For Vande Kopple (1985) code glosses are a type of textual metadiscourse, for they are

mainly used to provide definitions of units in text. For Thompson (2001), code glosses perform an interactive role. However, Hyland (2005, 2007) argues that in addition to the textual metafunction suggested by Vande Kopple (1985), or the interactive role proposed by Thompson (2001), code glosses are also used in an interactional fashion in which text producers anticipate ways in which their message could create a processing burden for their audience and find ways to make it clear and easy to process.

This is the definition of code glosses that I adopt in this research: code glosses are units that refer to other portions of discourse used to help the audience in anticipation to potential comprehension pitfalls. The following is a description of the different types of code glosses; examples from the corpus are included to illustrate the code glosses.

7.3. Types of code glosses

The current study follows Hyland (2007) in proposing two general types of code glosses: reformulations and exemplifications.

7.3.1. Reformulations

Reformulations restate a previous unit from a different point of view, elaborate on it, or emphasize it. There are two subcategories: expansions and reductions.

7.3.1.1. Expansions

Expansions clarify a previously expressed idea. One type of expansion is *explanations*. Explanations can either introduce a technical term that refers to a previous idea or add a definition to a previous term. Sometimes what is introduced is a new nuance of meaning for a

known word. Expansions are introduced by markers such as *that is*, *known as*, *called* and *referred to as*.²²

- (1) (*Domotic system*) is frequently **called** [*Intelligent home*] where owners can control every device in their house only with a click from a computer or a phone call..
(W-H-INGE-1)

The second type of expansion is *implication*. Implications provide a conclusion derived from a previous unit and are frequently introduced by markers like *this implies that* and *this means that*.

7.3.1.2.Reductions

Reductions is the second kind of reformulation. They limit the scope of a previous idea; there are two subtypes: paraphrase and specification.

- (2) Secondly, (these reparations are a product of our own context of armed conflict.)
This implies that [political collective reparations are a Colombian product, and a unique experience around the world.] (W-H-DERE-2)

Paraphrases provide a summary of what was previously expressed with markers like *that is*, *in other words*, and *put another way*.

²² The following conventions are used in the sample sentences. Parenthesis are used to mark the unit that is elaborated; angular parentheses for the code glosses, and bold type for the expressions that introduce the gloss. It is important to set the distinction between the linguistic marker and the actual code gloss. It is common to find reports that confuse the markers of code glosses with the code glosses themselves.

- (3) *First, the construction of historical memory is a pedagogical strategy because it promotes a participative democracy. (Its construction is far from representative democracy because each individual provides to the construction of meaning from shared experience.) In other words, [historical memory values that each person contributes his voice to reconstructs the common experience.]* (W-M-EDUC-1)

The other type of reduction is *specification*. Specifications add details that restrict the interpretation of a previous unit. *Specifically* and *in particular* are expressions that introduce specifications.

- (4) *Despite this consensus, among the returns to education literature (there are different perspectives which try to explain the relationship between education and salary in terms of years and quality of education.) In particular, [there are two main theories which explain the positive effects of education on wages: human capital and signalling.]* (W-H-ECON-3)

7.3.2. Exemplifications

The second type of gloss that Hyland proposes is *exemplification*. Exemplifications provide ways to interpret ideas that are close to knowledge or prior experience. Hyland (2007) describes three kinds of examples: category instantiations, parallel or similar cases, and rule instantiations.

Category instantiations present members or subclasses of general classes of things and are commonly introduced by *such as* and *like*.

- (5) *And the community are suff are suffering of 22 kinds of (diseases) such as [asphyxia, eh asthma, eh bronchitis, skin rashes and fungi.]* (S-M-INGE-10)

Parallel cases can be presented in the form of similar cases, analogies, or metaphors and are introduced by markers like *like*.

- (6) *We can have a lot of other universes. And each of one [fs] each of these universes have a lot of different actions to talk about the decision. All of (these actions) are **like** [parallel universes] because if we choose one of this, all of the other universes cannot be possible for us.* (S-H-INGE-2)

The third type provides rules for a general law or precept and is introduced by expressions like *for example* or *for instance*.

- (7) *(There is a growing recognition of the complex interplay of mind and physical body and the contribution of environmental factors and emotion to the development of disease.) **For instance** [we can see how the links between economic status and social stress, smoking and low birth weight, health and social capital, have complex cultural constructs and interpretative frameworks.]* (W-M-ANTR-1)

The types of code glosses and some of their markers are summarized in *Figure 7.1*, which I used for my previous study (Nausa, 2015, 2018). This taxonomy is based on Hyland's (2007) lists of code glosses markers.

<u>Reformulation</u>				<u>Exemplification</u>		
Expansion		Reduction		Category Instantiation	Parallel / similar case	Rule instantiation
Explanation	Implication	Paraphrase	Specification			
<i>That is</i>	<i>In other</i>	<i>That is</i>	<i>More</i>	<i>Like</i>	<i>Like</i>	<i>Say</i>
<i>Known as</i>	<i>words</i>	<i>In other</i>	<i>specifically</i>	<i>e.g.</i>	<i>Much like</i>	<i>Such as</i>
<i>Called</i>		<i>words</i>	<i>In particular</i>	<i>For example</i>		<i>e.g.</i>
<i>Referred to as</i>	<i>This means that</i>	<i>Put another way</i>	<i>Accurately</i>			

Figure 7.1. A taxonomy of code glosses and expressions to introduce them (adapted from Hyland, 2007)

7.4. Previous research on code glosses in academic oral discourse

Research on code glosses in English academic discourse is commonly found as metadiscourse research and has mainly focused on written genres like argumentative responses in advanced EAP writing (Basturkmen & Randow, 2014) or PhD theses (Bunton, 1999). Most research is comparative in nature and focuses on variation across disciplines (Hyland, 1998), time and disciplines (Hyland & Jiang, 2018), genres and disciplines (Bondi, 2005); NS and NNS undergraduate (T. Li & Wharton, 2012); business management (Murillo, 2012), economics (Valero-Garcés, 1996) and sales promotion (Vergaro, 2004) writing, among others. Code glosses have also been studied in academic posters, a written genre that is used within OPs, in aspects such as the orchestrated use with visual resources by native speakers from different disciplines (D'Angelo, 2011), or comparing NS and NNS in the construction of arguments and the elaboration of interpretations in medicine (Talebinejad & Ghadyani, 2012).

In oral academic discourse, code glosses research has seen studies in instructors and students' discourses. Examples of code glosses research in instructors' use of English include the

analysis of lecture comprehension by learners (Aguilar & Arnó, 2002) or its role and use in the classroom (Bu, 2014; J. Lee & Subtirelu, 2015).

Like the reported tendency on other linguistic-discourse phenomena, only a few studies on the use of code glosses in OPs have surfaced. These studies have mainly focused on EFL tertiary education contexts and include the analysis of metadiscourse (code glosses included) in the light of the quality of Chinese majors' oral production (Rui & Xin, 2009). This study demonstrated a correlation between the score obtained and the amount of metadiscourse used. Another study on metadiscourse by Italian advanced learners of English in OPs (Alessi, 2005) found a low use of code glosses as reflected in the avoidance of traditional markers such as *which means*, which the author explains as due to the familiarity between the presenters and the small class audience. Finally, Nausa (2015, 2018) in an analysis of the mechanisms that PhD-level Colombian researchers used to transition from written to oral discourse, identified the inclusion of code glosses as one of the mechanisms that high achievers (compared to low achievers) more consistently used, which is in agreement with Rui and Xin's (2009) findings. Also, like Alessi (2005), Nausa found that traditional code gloss markers were avoided. In some cases, students introduced reformulations and examples without a marker or introduced an explanation with markers traditionally used for other functions (e.g. *because*) for the OP audience.

The study in this chapter seeks to fill some found gaps. The first is the obvious lack of studies in this area of oral academic discourse. Another is the absence of studies for Spanish speaking EFL populations, more specifically at PhD researcher level. Apart from the studies on these populations, this study will provide information about the use of code glosses with the two main variables in this thesis: levels of achievement and disciplines.

7.5. Methods

This section explains the procedures to identify code glosses, prepare the corpus, and remedy some of the limitations in the first approach to code glosses in the pilot study (Nausa, 2015, 2018).

7.5.1. Update of the list of code gloss markers

The update of the list of code markers used in this study included the identification of new ones and the reclassification of others. 1357 sentences containing code gloss markers were identified in the 72811-token oral corpus. The identification was performed in two stages: use of previously reported code gloss markers and identification and inclusion of new ones.

In the first stage, I used the code gloss markers list in *Figure 7.1*. Although originally used to analyse written academic discourse (Hyland, 2005, 2007), the list was useful in the retrieval and identification of sample sentences containing code glosses.

The original list underwent two main kinds of changes. One, some expressions (e.g. *indeed, put another way*) were not found in the corpus, so they were left out of the final taxonomy. Second, other expressions like *or, like, and for example* were found to perform other types of code gloss functions than the ones originally proposed by Hyland. *Or*, for example, was found to be used to introduce explanations and paraphrases (*Figure 7.2*).

<u>Reformulation</u>				<u>Exemplification</u>		
Expansion		Reduction		Category Instantiation	Parallel / similar case	Rule instanti ation
Explanation	Implication	Paraphrase	Specification			
<i>That is</i>	<i>In other</i>	<i>That is</i>	<i>More</i>	<i>Like</i>	<i>Like</i>	<i>Such as</i>
<i>Known as</i>	<i>words</i>	<i>That/this</i>	<i>specifically</i>	<i>For example</i>	<i>Much</i>	<i>For</i>
<i>Called</i>	<i>This means</i>	<i>is to say</i>	<i>In particular</i>	<i>For instance</i>	<i>like</i>	<i>exampl</i>
<i>Referred to</i>	<i>that</i>	<i>In other</i>	<i>As a matter</i>		<i>For</i>	<i>e</i>
<i>as</i>	<i>I mean</i>	<i>words</i>	<i>of fact</i>		<i>example</i>	
<i>Defined as</i>			<i>In fact</i>			
<i>Or</i>			<i>Namely</i>			

Figure 7.2. Provisional list of code gloss markers in OPs

7.5.2. Selection of sentences for analysis and data clean-up

A series of procedures in two stages were used to avoid over-estimation of frequencies and to identify further mechanisms to introduce code glosses.

In the clearing stage, the following cases were removed. First, sentences that contained code gloss markers occurring at a false start and then abandoned,

- (8) ... of all the world and can and permit establish different **eh [fs]** or the **eh [fs]** the five major population groups that **eh [fs]** that today we know. (S-M-CBIO-1)

Second, sequential repetitions in which the code gloss marker occurs more than once:

- (9) their strategies **eh** supported by **eh** armed actors or narcotrafic, **or** some **[fs]** or **s [fs]** **or** different undemocratic actors. And you can find in other (S-M-CPOL-1)

This case with three tokens of *or* was considered as one. The other generated concordance lines were eliminated and not considered in the calculations.

Finally, sentences containing code glosses with words transcribed as unintelligible:

(10)... *the chief has access to best food and the [fs] and **for example the (unintelligible)** or the corn or the others, but this case is a little criticized by...* (S-M-ANTR-3)

In the second stage, I did a cross-check by reading through 20 oral presentation transcripts to spot any code gloss markers or mechanisms missing. Six new markers/mechanisms were identified. These were added to the list presented in *Figure 7.2*; the new markers have an asterisk (*). Three are chunks like the ones in the original taxonomy: *kind of, in this case, not only... (but also)*. The others come as complete clauses: *what is/are* + (art) + NOUN?, NOUN + THAT/WHICH relative clauses, and other rhetorical questions. The following are some examples of sentences containing these markers.

- (11) *for example, museums for examples (truth commission, for example give again the legal status to Unión Patriótica.) As a result of **this kind of** [reparation], Unión Patriótica was able to participate in 2014 political elections with its president Aida Avella (S-H-DERE-2)*
- (12) *Then, eh the important feature of the problem is the next one: you can assign a number to the problem and (you can find some condition for that number), **in this case** [the number must be even, should be even], and this is the general procedure that uses algebraic topology for solve the problem. (S-H-MATE-2)*
- (13) *but that's the problem in the eh work division, the majority of eh principals are men and (the majority of teachers are women), and that is **not only** [in the teaching profession] eh it is **also** [in eh banks...] (S-L-EDUC-2)*
- (14) *we need to complete or define a process that can make an interaction between three processes that now are applied separately. These processes are: how the wave propagate in an heterogeneous media, an heterogeneous media is an (soil deposit) [**that is** compound by different kind of materials.] (S-H-INGE-6)*
- (15) *Eh, (however, although the domestic violence is eh a multidimensional with eh sociocultural, eh historical, clinical, (intelligible) factors, there aren't enough integral studies eh and there is a predominance of clinical or legal eh aspects.) **What are the reasons for this problem?** [There are many factors that difficult the study or the treatments of these problems.] (S-H-ANTR-3)*
- (16) *So following this example and the literature, we can define a (high growth firm) as [reading3] a firm which doubles or triples its size in terms of sales and employees or jobs created in a period ranging from 5 to 7 years [reading3]. But eh [reading4] **why are high grow firms so important?** [reading4] [Ah at least we can list three reasons.] (S-H-ADMI-2)*

<u>Reformulation</u>				<u>Exemplification</u>		
Expansion		Reduction				
1.explanation	2.implication	3.paraphrase	4.specification	5.Category Instantiation	6.Parallel / similar case	7.Rule instantiation
<i>That is</i>	<i>In other</i>	<i>That is</i>	<i>More</i>	<i>Like</i>	<i>Like</i>	<i>Say</i>
<i>Known as</i>	<i>words</i>	<i>That/this is</i>	<i>specifically</i>	<i>For</i>	<i>Similar</i>	<i>Such as</i>
<i>Called</i>	<i>This means</i>	<i>to say</i>	<i>In particular</i>	<i>example</i>	<i>to</i>	<i>For</i>
<i>Referred to as</i>	<i>that</i>	<i>In other</i>	<i>As a matter</i>	<i>For</i>	<i>Much</i>	<i>example</i>
<i>Defined as</i>	<i>I mean</i>	<i>words</i>	<i>of fact</i>	<i>instance</i>	<i>like</i>	
<i>or</i>		<i>Or *</i>	<i>In fact</i>			
<i>Rhetorical</i>			<i>Namely</i>			
<i>questions: what</i>			<i>In this case*</i>			
<i>is+(art)+noun?*</i>			<i>Not only, but</i>			
<i>(This) Kind of*</i>			<i>also*</i>			
<i>Noun+that</i>			<i>That/Which</i>			
<i>relative clauses*</i>			<i>relative</i>			
			<i>clauses*</i>			
			<i>Rhetorical</i>			
			<i>questions*</i>			

Figure 7.3. Definite list of code gloss markers in OPs

The new markers are used to introduce reformulations: explanations and specifications.

No new markers were found for implications, paraphrases or the three kinds of examples.

50 randomly selected sentences including the markers in the definite list (Figure 7.3) were given to a colleague to classify them into the taxonomy of seven types of code glosses.

Krippendorff's alpha value was calculated for interrater agreement. The value (0.816) was found to be near perfect (see Table 3.7).

After the corpus preparation and classification validation procedures were completed, two types of statistical analyses were performed: frequency and significance to determine if there

were categories, subcategories, or specific markers that serve to represent the difference between the subcorpora in this study.

7.6. Quantitative analyses

The quantitative analyses of code glosses use among the subcorpora is presented at three different levels: general types (reformulations and exemplifications), subtypes (explanations, paraphrase), and specific code gloss markers (*this kind of*, *in particular*).

7.6.1. Code glosses: reformulations and exemplifications

Tables 7.1 and 7.2 present the general frequencies of use in the divides.

Table 7.1. Code glosses use expressed in raw frequencies and percentages by level of achievement

	High	%	Medium	%	Low	%	Total	%
Reformulations	396	29.2	323	23.8	199	14.7	918	67.6
Exemplifications	180	13.3	162	11.9	97	7.1	439	32.4
	576	42.4	485	35.7	296	21.8	1357	100.0

Table 7.2. Code glosses use expressed in raw frequencies and percentages by discipline

	Hard	%	Soft	%	Total	%
Reformulations	451	33.2	467	34.4	918	67.6
Exemplifications	215	15.8	224	16.5	439	32.4
	666	49.1	691	50.9	1357	100.0

General analysis by percentage shows that reformulations occurrences are twice as frequent as exemplifications (67.6% vs 32.4%). This tendency changed with the inclusion of the new markers, specifically *that/which* relative clauses, *kind of*, and *in this case*. Before their inclusion, exemplifications were as frequent as reformulations.

Frequency analyses seem to show a different picture for the achievement and disciplinary divides. In the levels divide, code glosses use by high achievers (42.4%) is higher than that of medium (35.7%) and low (21.8%) achievers. In the disciplinary divide, the differences are not as marked. Hard discipline students use 49.1% of code glosses, and soft-discipline students use 50.9%. The tendencies seem to remain the same in the analysis by the two code gloss types: reformulations and exemplifications.

Significance analyses (log likelihood); however, indicate whether those general observed frequencies are significant or not.

Table 7.3. Significance analyses of code glosses use by level of achievement

	High	Medium	Low	log likelihood	Bayes Factor BIC
Reformulations	396	323	199	18.90	-3.49
Exemplifications	180	162	97	6.00	-16.39
CGs Total	576	485	296	24.32	1.93

Table 7.4. Significance analyses of code glosses use by discipline

	Hard	Soft	Log Likelihood	Bayes Factor BIC
Reformulations	451	467	0.45	-10.74
Exemplifications	215	224	0.28	-10.92
CGs Total	666	691	0.73	-10.47

In the levels of achievement divide, the frequency differences (log likelihood) among the corpora and the effect size (BIC) are significant for the use of code glosses in general. The effect size is reduced although the significance of the frequency difference is kept when reformulations and examples are analysed separately (see 3.4.2).

In the disciplinary divide, the frequency differences are not significant, nor is there positive evidence against the null hypotheses for code glosses in general and for their subcategories.

In order to identify whether these general tendencies in code gloss use remain the same at a closer look, I performed frequency and significance analyses at the following levels:

1. Types of reformulations and exemplifications
2. Specific markers of reformulations and exemplifications

7.6.2. Reformulations: explanations, implications, paraphrases, and specifications

Explanation (66.7%) and specification (25.2%) are the two most recurrent types of reformulation in the corpus.

Table 7.5. Reformulations use expressed in raw frequencies and percentages by level of achievement

	High	%	Medium	%	Low	%	Total	%
Explanation	258	28.1	218	23.7	136	14.8	612	66.7
Implication	19	2.1	13	1.4	8	0.9	40	4.4
Paraphrase	14	1.5	14	1.5	7	0.8	35	3.8
Specification	105	11.4	78	8.5	48	5.2	231	25.2
	396	43.1	323	35.2	199	21.7	918	100.0

Table 7.6. Reformulations use expressed in raw frequencies and percentages by discipline

	Hard	%	Soft	%	Total	%
Explanation	307	33.4	305	33.2	612	66.7
Implication	16	1.7	24	2.6	40	4.4
Paraphrase	16	1.7	19	2.1	35	3.8
Specification	112	12.2	119	13.0	231	25.2
	451	49.1	467	50.9	918	100.0

A closer look at the frequencies between divides shows the general tendency of code gloss use. High and medium achievers use more reformulations than low achievers. Soft-discipline student use of reformulations (50.9%) does not outnumber hard-discipline use (49.1%).

To identify whether this frequency differences are significant at this level, significance analyses were performed.

Table 7.7. Significance analyses of reformulations use by level of achievement

	High	Medium	Low	log likelihood	Bayes Factor BIC
Explanation	258	218	136	9.49	-12.90
Implication	19	13	8	1.99	-20.40
Paraphrase	14	14	7	0.93	-21.46
Specification	105	78	48	7.90	-14.49
Total	396	323	199	18.90	-3.49

Table 7.8. Significance analyses of reformulations use by discipline

	Hard	Soft	Log Likelihood	Bayes Factor BIC
Explanation	307	305	0.00	-11.19
Implication	16	24	1.69	-9.51
Paraphrase	16	19	0.29	-10.91
Specification	112	119	0.28	-10.91
	451	467	0.45	-10.74

In the levels of achievement divide (Table 7.7), the frequency differences are significant for explanations and specifications at $p < 0.01$ (Table 3.4), but Bayes Factor negative values indicate that the frequency differences are not big enough to represent positive evidence against the null hypotheses.

In the disciplinary divide (Table 7.8), frequency differences are not significant $p > 0.05$ (Table 3.4), nor is there positive evidence against null hypotheses (Table 3.5).

7.6.3. Exemplifications: category instantiations, similar cases, and rule instantiations

Tables 7.9 and 7.10 show that category (53.3%) and rule instantiation (33.9) are the two most recurrent types of exemplification in the corpus.

Table 7.9. Exemplifications use expressed in raw frequencies and percentages by level of achievement

	High	%	Medium	%	Low	%	Total	%
Category Ins	105	23.9	75	17.1	54	12.3	234	53.3
Similar Case	27	6.2	18	4.1	11	2.5	56	12.8
Rule Instantiation	48	10.9	69	15.7	32	7.3	149	33.9
	180	41.0	162	36.9	97	22.1	439	100.0

Table 7.10. Exemplifications use expressed in raw frequencies and percentages by discipline

	Hard	%	Soft	%	Total	%
Category Ins	96	21.9	138	31.4	234	53.3
Similar Case	34	7.7	22	5.0	56	12.8
Rule Instantiation	85	19.4	64	14.6	149	33.9
	215	49.0	224	51.0	439	100.0

In the levels divide, high achievers use more exemplifications than medium and low achievers. In the disciplines divide, soft-discipline student use of exemplifications (51%) is close to hard-discipline use of the same devices (49%). At a finer scale, soft discipline students seem to prefer category instantiations (31.4% vs 21.9%) while hard discipline students seem to prefer the use of rule instantiations (19.4% vs 14.6%).

To identify whether these frequency differences are significant at the exemplification level, significance analyses were performed.

Table 7.11. Significance analyses of exemplifications use by level of achievement

	High	Medium	Low	log likelihood	Bayes Factor BIC
Category	105	75	54	5,92	-16,47
Similar case	27	18	11	3,16	-19,24
Rule	48	69	32	6,97	-15,42
	180	162	97	6,00	-16,39

Table 7.12. Significance analyses of exemplifications use by discipline

	Hard	Soft	log likelihood	Bayes Factor BIC
Category Ins	96	138	7,99	-3,21
Similar Case	34	22	2,48	-8,72
Rule Instantiation	85	64	2,77	-8,42
Total	215	224	0,28	-10,92

In the levels of achievement divide (Table 7.11), the frequency differences are significant for category ($p < 0.05$) and rule instantiations ($p < 0.01$) (Table 3.4). The available evidence, on the other hand, is not positive against the null hypothesis as indicated by BIC values (Table 3.5).

In the disciplinary divide, the frequency differences of category instantiations are significant at $p < 0.01$, but the negative BIC values indicate that there is no positive evidence against the null hypothesis (i.e. frequency differences are significant but not big enough).

So far, the initial expectations have been confirmed: high achievers more frequently use code glosses while these devices use does not seem to be different among the disciplines. A closer look at the subcategories shows that reformulations in the form of explanations and specifications are the most recurrent types of code glosses used by this group of language learners, and the observed general tendency seems to remain the same: level of achievement, and not discipline, seems to be a more important factor in using code glosses. Significance analyses have shown that these differences can be significant (log likelihood values) but the available evidence (expressed as how big the frequency differences are) is probably not enough to reject the null hypotheses (BIC values).

7.6.4. Analysis by CG markers: frequency and significance analysis

This section focuses on the specific expressions that mark the presence of code glosses. This time the analysis focuses on raw frequencies and significance (Log Likelihood and Bayes Factor values), for they not only confirm the findings in the previous frequency analysis by percentages, but also serve as criteria for the selection of cases to report in the discourse analyses section. I have established a log likelihood value over 6.63 ($p < 0.01$). In none of the cases, there were positive BIC (effect size) values.

This analysis confirms that the frequency differences are statistically significant in the level of achievement opposition. This can be observed in Table 7.13 LL values for the explanation markers (*this*) *kind of* (8.61) and *that/which* relative clauses, the specification marker *in particular* (9.69), and the category instantiation marker *like* (17.52). An inspection of the raw values of these markers shows that they are more frequently used by high achievers, confirming the CG category analyses above.

In the disciplinary division, there is only one case that is above the established threshold: *like* (8.97) for specification. This use of *like is* more frequent in the soft disciplines subcorpus.

Table 7.13. Code gloss markers expressed in raw frequencies and Log likelihood (LL) and Bayes factor (BIC) values by level of achievement and discipline

	High	Medium	Low	LL	BIC	Hard	Soft	LL	BIC
Explanation									
<i>Known as</i>	0	0	2	5.24	-17.15	0	2	2.79	-8.40
<i>Called</i>	14	5	8	4.23	-18.17	17	10	1.77	-9.43
<i>Defined as</i>	1	0	0	1.98	-20.41	1	0	1.38	-9.82
<i>or</i>	56	42	41	1.98	-20.41	67	72	0.23	-10.96
<i>like</i>	1	0	0	1.98	-20.41	1	0	1.38	-9.82
<i>(This) kind of</i>	34	24	9	8.61	-13.78	30	37	0.80	-10.39
<i>Relative c</i>	124	117	59	8.77	-13.62	152	148	0.02	-11.17
<i>what is</i>	28	30	17	0.90	-21.50	39	36	0.09	-11.10
	258	218	136	9.49	-12.90	307	305	0.00	-11.19
Implication									
<i>In other words</i>	2	4	1	1.40	-20.99	2	5	1.36	-9.84
<i>This means</i>	17	9	7	2.82	-19.57	14	19	0.81	-10.39
	19	13	8	1.99	-20.40	16	24	1.69	-9.51
Paraphrase									
<i>That is to say</i>	0	0	2	5.24	-17.15	0	2	2.79	-8.40
<i>or</i>	9	8	2	3.18	-19.21	9	10	0.06	-11.13
<i>like</i>	5	6	3	0.36	-22.03	7	7	0.00	-11.20
	14	14	7	0.93	-21.46	16	19	0.29	-10.91
Specification									
<i>more specifically</i>	7	4	5	0.87	-21.52	11	5	2.25	-8.95
<i>in particular</i>	14	8	1	9.69	-12.70	15	8	2.10	-9.10
<i>or</i>	30	25	10	5.29	-17.10	35	30	0.34	-10.86
<i>in fact</i>	3	1	0	3.50	-18.90	2	2	0.00	-11.20
<i>specially</i>	6	4	1	2.55	-19.84	2	9	4.89	-6.31
<i>namely</i>	0	0	2	5.24	-17.15	0	2	2.79	-8.40
<i>like</i>	11	10	6	0.34	-22.05	6	21	8.97	-2.23
<i>in this case</i>	17	17	14	0.12	-22.27	28	20	1.26	-9.93
<i>not only</i>	11	6	3	2.97	-19.42	6	14	3.37	-7.83
<i>WH reth</i>	6	3	6	2.11	-20.29	7	8	0.08	-11.12
	105	78	48	7.90	-14.49	112	119	0.28	-10.91
Exemplification									
Category instantiation									
<i>like</i>	37	9	15	17.52	-4.87	22	39	4.96	-6.23
<i>for example</i>	38	27	22	1.64	-20.75	42	45	0.13	-11.06
<i>such as</i>	4	6	3	0.58	-21.81	8	5	0.67	-10.53
<i>for instance</i>	2	0	0	3.96	-18.43	0	2	2.79	-8.40
	105	75	54	5.92	-16.47	96	138	7.99	-3.21
Parallel similar case									
<i>like</i>	18	11	9	1.69	-20.70	24	14	2.57	-8.63
<i>similar to</i>	3	1	2	1.10	-21.30	4	2	0.66	-10.53
	27	18	11	3.16	-19.24	34	22	2.48	-8.72

Rule instantiation									
<i>for example</i>	48	64	30	5.49	-16.90	83	59	3.85	-7.35

The following section presents examples of code glosses including the markers found to be significant. These examples will include explanations about why code glosses are a more common differentiating feature in the level of achievement than in the disciplinary divide.

7.7. Discourse analysis of code glosses

7.7.1. Explanations

Two markers were found to be statistically significant in the level of achievement comparison: *(this)kind of* and *that/which* relative clauses.

7.7.1.1. *(This)kind of*

This was a new marker in the stage of new markers identification. Explanations provide new terms or their definition either because terms are perceived to be new for the audience or the presenter needs to provide a field-specific nuanced definition. One common procedure in definitions is the specification of the general class to which the referent of the word belongs (Barnbrook, 2002). This class specification is performed with mechanisms such as the use of hyperonyms for the general class, the copula BE or the use of terms like *kind of*, *type of*, *sort of* etc.

- (17) *for example, museums for examples (truth commission, for example give again the legal status to Unión Patriótica.) As a result of **this kind of** [reparation], Unión Patriótica was able to participate in 2014 political elections with its president Aida Avella* (S-H-DERE-2)
- (18) *Eh origin [Fs] originally the reinforcement [fs] reinforcement was asbestos due to low cost, fiber resistance, water lightness and light weight, and other useful properties. Eh the [fs] **this kind of** [product], the (fibrocement) today is consider a material physically suited for construction products and it have been demonstrate high [fs] high (fractural) and straight resistance.* (S-H-INGE-5)

In (17) the speaker provides the term *reparation* to refer to the previous unit, an example of *reparation*. In (18) the unit being elaborated on (*fibrocement*) does not appear before the marker *kind of*. However, the use of *kind of* presents fibrocement as a **product** with the characteristics suitable for the modern construction industry.

7.7.1.2. *That/which relative clauses*

In addition to the use of the copula or expressions like *this kind of* + hyperonyms to specify the class a referent belongs to, definitions also provide the specific attributes that make the element being described different to other similar elements in the same class. These defining attributes are expressed through relative clauses. This has been referred to as *definiendum* (the term that needs to be defined) and *definiens*: class (hyperonym) plus specifier (the information that the relative clause provides) (Barnbrook, 2002).

- (19) *The second factor is (climatic change). Eh this is a ([global phenomenon]) [**that is caused mainly by gases emission to atmosphere of cars, industries, eh deforestation.**]* (S-H-CBIO-2)

In (19), the speaker does not use a definition marker (e.g. *called*); however, she finds it necessary to define *climatic change* (sic). She does so by first specifying the class (global phenomenon) and a relative clause to provide the specific attribute that makes climate change different from other phenomena. In this sense, it could be argued that the use of adjective clauses to provide definitions could be considered both a case of explanation and specification.

As can be seen, the use of the two new markers of definitions makes sense in the context of the task on which this study is based: an OP for a non-expert audience. Given the instruction to share research with classmates, it is expected that presenters anticipate moments in which word meaning could create confusion. Probably, many of these definitions would not be provided if the talk was given to an expert audience. This tendency to align with the audience has previously been reported for the same population (Nausa, 2016) in which students invoked *knowledge-provider* or *co-constructer* identities.

7.7.2. Specifications

In specifications, the new unit presents characteristics to restrict previous units' interpretation. This allows speakers to include previous units within a more narrowed scope (Hyland, 2007). In the crosscheck for new markers, I identified four new markers (see *Figure 7.3*); however, only one (*in particular*) was above the established threshold.

7.7.2.1. *In particular*

Examples (20) and (21) illustrate the use of *in particular* in a high-rated and a middle-rated OPs.

- (20) *To have great [fs] a better idea, I'm gonna show you this map. The blue dots are the populations in lowlands, eh all of this area is part from the amazon. The red dot is one of the (populations at highlands), **in particular**, [the populations in Cueva de los guácharos.] (S-H-CBIO-6).*
- (21) *It means we [fs] the products [fs] the products of the memory eh can be analysed and discussed by different sectors and we can talk [fs] we could [fs] we can talk about [fs] about (the conflict) in [fs] **in particular** [about the causes, the origin, the actors, the impact, etc.] (S-M-EDUC-1)*

In both cases the presenters use the marker to introduce how the topic being discussed needs to be narrowed down to a more specific scope. Although *in particular* should not be difficult for learners to use since it is predictably easy to transfer from the Spanish *en particular*, low-rated OPs showed a preference for other expressions such as *or* and *in this case*.

7.7.2.2. *Like*

This is the only case of code gloss marker whose frequency difference was significant in the disciplinary divide.

- (22) *Six thousand of people: activists, human rights defenders, academics, say [fs] says [fs] say eh that it was a genocide in Colombia, may people died. Today, I want to talk about how (Unión Patriótica) **like** [a political party] died eh too. (S-H-DERE-2)*
- (23) *eh he or she receives eh an inde [fs] indemnization, compensation, satisfaction, but in the case of with indigenous people we have a (collective subject), you know? It's different, because they live **like** [a culture unity in the same territory, eh with the same group] (S-M-ANTR-4)*

Like is used to specify how *union patriótica* and *collective subject* should be interpreted. This usage was mainly found in the soft-disciplines subcorpus. Students in these examples were

attempting to use *like* as *as*. This is a common transfer error; in Spanish several uses of *as* and *like* are expressed with the word *como*.

7.7.3. Category instantiations

To elaborate a previous unit, the speaker presents a new unit as an instantiation that represents the type of entity, event, or phenomenon referred to (Hyland, 2007).

Only one instantiation marker was found to be statistically significant: *like* ($p < 0.0001$).

7.7.3.1. *Like*

Like was found to be highly frequent in the high-achievers OPs and less frequent in medium-achiever OPs. Examples 24-26 illustrate its use in the three levels of achievement.

- (24) *In (countries) **like** [Netherlands] poli [fs] like [the Netherlands], policy makers have chosen pol eh [fs] general policies* (S-H-ADMI-2)
- (25) *Basically, the research problem is this, eh the monitory this program says that contribute to resolve eh structural (problems of the society), something **like** [economy problems, social problems], and its eh it [fs] broke a problems worse eh* (S-M-CPOL-3)
- (26) *In conclusion from this viewpoint, research work developed by Agulhon and the other and the (other research) **like** [a Francoise Xavier Guerra] about the sociability eh category* (S-L-HIST-1)

These three cases are similar; they introduce members of a class of things (country-Netherlands; problems of society-economy problems, and researchers-Francoise Xavier Guerra). Apart from the frequency differences, the difference between the levels lie in the grammatical infelicities of the low-rated OP: the selection of the word form (*research* instead of *researcher*) and the lack of number agreement between the *other* and *research*. This is not to say that the

other OPs are not without errors or disfluencies, but the ones in (26) could be deemed as typical of a lower-level learner.

7.8. Conclusion

In this chapter, I have attempted to answer the following questions:

1. **Code glosses:** what are the tendencies in the use of code glosses in the OPs of Colombian PhD researchers?
2. **Level of achievement:** What are the differences in the use of code glosses among high, medium and low-rated presentations?
3. **Disciplinary divide:** What are the differences in the use of code glosses between hard-discipline and soft-discipline researchers?

Code glosses: in the preparation for analysis, new code gloss markers and mechanisms were identified and included in the inventory: the markers include *(this) kind of, in this case, not only... (but also)*. The mechanisms are *what is/are + art + NOUN?*, *NOUN + that/which* relative clauses, and other rhetorical questions. I set a distinction between markers and mechanisms, as code gloss markers have traditionally been conceived as chunks, not complete clauses. However, as it was demonstrated the new markers and mechanisms perform the rhetorical function of marking and introducing reformulations. This study also confirms the general finding in my pilot study (Nausa, 2015, 2018) that high achievers more consistently use code glosses as evidenced in their grammatical accuracy and pragmatic relevance.

Level of achievement over disciplinary divide: interestingly too, significance statistical analyses showed that the use of code glosses at any level (general taxonomy, categories, and specific, markers) was a discriminating factor of levels of achievement and not disciplines. In

other words, students with higher oral proficiency more consistently anticipated moments in which the audience could get confused and included explanations, clarifications and examples that guide them. The use of these resources reflects their sensitivity to context needs and the knowledge of a repertoire of options to mark the moments when reformulations or exemplifications are used.

Table 7.14: Summary of findings code glosses study

	Code glosses realisations (Log likelihood >6.63 <i>p</i> <0.01)	Overu se (↑) or under use (↓)	Patterns	Discourse functions
Level of oral achievem ent (high, medium, low)	Explanations (<i>This</i>) <i>kind of</i> (LL: 8.61) Relative clauses (LL: 8.77)	Low rated OPs (↓)	- Definiendum + definiens + specifier - This kind/sort/type of + HYPERONYM - HYPERONYM that clause	Provision of - Technical vocabulary not known by audience - field-specific nuanced definition of known terms
	Specification <i>In particular</i> (LL: 9.69)	High rated OPs (↑)	- NP, <i>in particular</i> NP/PP	Restriction of a discourse unit within a specific scope
	Exemplificati on (category instantiation) <i>Like</i> (LL: 17.52)	High rated OPs (↑)	- HYPERONYM <i>like</i> hyponym - NP <i>like</i> NP	Provision of instances to illustrate a previous unit
Disciplina ry divide (hard vs soft)	Specification <i>like</i> (LL: 8.97)	Soft- field OPs (↑)	- NP <i>like</i> NP	Restriction of a discourse unit within a specific scope Most cases should have used <i>as</i>

CHAPTER 8

WE NEED EDUCATION...: TRANSFORMING WRITTEN INTO ORAL CONTENT

8.1. Introduction: a follow-up study on the mechanisms to transition from written to oral discourse

As explained in 7.1, module 1 (Nausa, 2015) reports the first attempt to study the language of the oral presentations (OPs) of Colombian PhD researchers taking an EAP course. In that study, I analysed the strategies that students used during their OPs to modify and present contents that had originally been written in essays and how these strategies were useful marks in discriminating levels of oral performance in class: high vs low. I identified four types of mechanisms to transition from the written to the oral mode: (1) modifications to the SVO clause linear order, (2) reduction of heavily modified noun phrases (NPs), (3) changes to the expression of modality, and (4) inclusion of code glosses to elaborate on potentially problematic propositions. In all cases, high achievers were found to more consistently (grammatically and pragmatically) use the mechanisms. That study, however, had several limitations. One limitation was the small corpus, which made the findings indicative rather than conclusive. Another was the division of students between high and low achievers, which ignored intermediate levels of achievement, and therefore, other aspects in the use of the mechanisms. A third limitation was the non-inclusion of another relevant variable influencing student's performance: their disciplinary background. The purpose of this chapter is to follow up on the findings and remedy some of the limitations using a significantly larger corpus, including medium achievers and the disciplinary divide variable. This follow-up study seeks to determine whether the reported mechanisms to transition between modes discriminate three levels of achievement (high,

medium, and low) and if they are also relevant in the disciplinary divide (hard-field vs soft-field students). Also, the new study intends to identify new mechanisms and sub-mechanisms that had originally been predicted (e.g. denominalization) but were not found.

8.1.1. Questions in the follow-up study

To confirm the findings in module 1 (Nausa, 2015) and widen the scope of the original study, this new study seeks to answer the following questions:

1. What are the differences between the written and the oral versions of the same content produced by students in this class as observed in four mechanisms of change and their related submechanisms?
 - Change of clause structure
 - Reduction of heavily modified NPs
 - Changes of expression of modality
 - Inclusion of code glosses
2. What quantitative and qualitative differences are there between high, medium, and low rated OPs?
3. What quantitative and qualitative differences are there between hard-field and soft-field OPs?
4. How can the differences be explained in terms of the grammatical accuracy and pragmatic relevance of syntactic changes?

8.2. Mechanisms of change to transition from written to oral discourse: main concepts and review of available research

Language features (register) depend on physical and psychological aspects of context (Biber & Conrad, 2009). OPs imply the presentation of content in real time, which poses several demands on presenters: focusing on novelty, engaging with the audience, using visual information, and simplifying information (Carter-Thomas & Rowley-Jolivet, 2003). The simplification of information is related to the selection of language resources that make the conveyance of information an effective task (information structure) and to the actions to make ideas easy to understand (elimination of redundant information; attribution, mitigation or boosting of ideas; and inclusion of reformulations and examples). These actions are explained under the main constructs in this chapter: (1) clause/information structure, (2) nominalisation and NP modification, (3) expression of modality, and (4) code glosses.

8.2.1. Clause structure changes and information structure

The SVO linear order of clauses can be modified to make the conveyance of information an effective task. The concept *information structure* and its related concepts are useful in understanding these modifications. *Information structure* refers to one of the mechanisms to guarantee coherence in texts (Ward & Birner, 2004). Roughly speaking, texts are composed of clauses, and how clauses are connected might guarantee that a text is easily understood or not. Within a text, a clause is an information unit that comprises two elements: GIVEN and NEW (Halliday & Matthiessen, 2004). The GIVEN (or old) is the element that iterates a piece of information previously introduced. In pragmatic terms, the GIVEN is also defined as information

that is shared/known by the participants in the communication exchange. The NEW is the element that introduces a new unit, not known by the hearer/reader.

- (1) *Well I need eh explain **procedural justice** first. **Procedural justice** eh refers a justice in the process [fs] in the process of making a decision.* (S-H-ADMIN-1)

In example 1, *procedural justice* is the NEW element in the first clause. When it is iterated in the second clause, it becomes the GIVEN. There are different mechanisms for the iteration of a unit as GIVEN: repetition of the term (*procedural justice*), use of pronouns (*it*), use of deictics and hyperonyms (*this type of justice, this process*) among others.

In addition to the introduction of units as NEW or their iteration as GIVEN, their position in the clauses is also important to guarantee information flow. In the same example, it can be observed that *procedural justice* is placed at different positions in the clauses. In the first clause, it appears in final clause position (RHEME) while in the second clause, it is iterated in initial position (THEME)²³. The combination of NEW/GIVEN information and THEME/RHEME positions yields four possible structural arrangements²⁴.

²³ Theme and rheme have also been referred to with other terms such as *topic-comment* (Gundel, 1988), *topic-focus* (Quirk et al., 1985).

²⁴ The English grammar has a repertoire of syntactic choices to put NEW and GIVEN elements in THEME and RHEME positions (e.g. passive-active voice constructions, postpositions, cleft sentences, diversions of SVO clause structure, etc)

(1a) ***Procedural justice** will be explained first. **Procedural justice** refers to justice in the process of making a decision.* (THEME/NEW – THEME/GIVEN)

(1b) ***Procedural justice** will be explained first. Justice in the process of making a decision is referred to as **procedural justice**.* (THEME/NEW – RHEME/GIVEN)

(1c) *I need to explain **procedural justice** first. **Procedural justice** refers to justice in the process of making a decision.* (RHEME/NEW – THEME/GIVEN)

(1d) *I need to explain **procedural justice** first. Justice in the process of making a decision is referred to as **procedural justice**.* (RHEME/NEW – RHEME/GIVEN)

The placement of NEW elements in the RHEME (first clause in 1c/1d) is also referred to as *end-weight* (Quirk et al., 1985), *rhematization* (Lovejoy & Lance, 1989) or *old-before-new ordering*²⁵ (Ward & Birner, 2004). End-weight (rhematization) has also been defined as a mechanism to introduce information whose processing might be cumbersome for readers/listeners or to highlight information that the speaker/writer deems important. The placement of elements in clause initial position (THEME) is referred to as *thematization* (Lovejoy & Lance, 1989) or *topicalization*²⁶ (Gundel, 1988; Speyer, 2005) (second clause in 1a/1c). Arguably, the sentence structural arrangement that has better information flow is 1c, for it places NEW information in the RHEME of the first clause and iterates it in the THEME of the second; the fact that the two instances of *procedural justice* are close to each other improves coherence in the text and facilitates comprehension for the listener.

The role of syntactic modifications to clause structure in academic discourse has mainly been studied in the analysis of written texts (e.g. Keen, 2004; Wright, 2008). In oral academic

²⁵ Or new after old in this case.

²⁶ The term *topicalization* refers to the placement of elements different to the grammatical subject in clause initial position, before the subject. For the purposes of this study, I will use *thematization* to refer to placements of elements in initial position regardless of the fact that they are topicalizations or grammatical subjects.

discourse, the study of English clause syntactic modifications has been limited to the study of the effectiveness of discourse used by NS and NNS university teachers (Tyler, Jefferies, & Davies, 1988) and teacher assistants (Tyler, 1994). Only a few studies have focused on the transition from written to oral discourse (or vice versa). In a comparison of American, Indian, and Chinese teacher assistants (TAs) use of English, Levis, Levis, and Slater (2012) analysed the way TAs modified a written prompt to present it orally as if presented to undergraduate students. Although the authors identified lexicogrammatical modifications to the prompts, they were mainly related to the use of pronouns, modals choice, and the extension of nominal groups. In their SFL-informed analysis, they found that NS TAs more effectively used these traits. The opposite transition, from spoken (conference presentations) to written (proceeding papers), has also been approached in two investigations comparing presentation-paper pairs by NSs only (Carter-Thomas & Rowley-Jolivet, 2001) and oral presentation-paper pairs by NSs and NNSs (Rowley-Jolivet & Carter-Thomas, 2005c). These two studies analysed structures that allow information structure manipulation (*there* existential clauses, extraposition, passive voice). Two important conclusions in these studies in terms of information structure principles were that (1) *there* existential clauses were more common in the oral presentations than in the papers (2) NNS more frequently used passive voice than NS, who used more interpersonal choices like active clauses with personal pronouns. These two studies were the original source of inspiration for the comparison of parallel written-oral texts by the same authors in this study. However, the methodology that I propose differs from the one in those studies in that I select written-oral pairs of sentences by the same author, expressing the same content, to analyse the syntactic mechanisms in the written into spoken transition, identify and count the mechanisms used, and perform discourse analyses that explain the transition. Carter-Thomas and Rowley-Jolivet do not

select pairs of sentences; they count the occurrence of the information structure syntactic mechanisms in their oral and written corpora and then perform the discourse analyses to identify information structure principles use.

8.2.2. Noun phrases: heavy modification, nominalisation, denominalisation, and grammatical metaphors

In a task like the one described in this thesis (3.3.1), students are asked to move from written (the essay) to oral (the OP) discourse. This transition, other things being equal, could be described as a transition from *attic* to *doric* modes²⁷ of expression (Halliday & Martin, 2003). For these authors, the *attic* style is associated to the language of science and is considered simpler in terms of the grammar used to structure clauses (fewer subordinate or coordinated clauses, more clauses with SVO or SVC structure). The grammar in noun phrases is more complex, though. The *doric* style, on the other hand, is related to everyday discourse and considered more grammatically intricate in relation to clauses, but simpler for NPs. The two modes refer to the ways in which language represents reality. The *attic* mode represents the world as a world of things while the *doric* mode represents the world as processes and transformations. The focus on the world as a world of things makes the *attic* style more nominal whereas the focus on processes makes the *doric* style more clausal. These differences between the *attic* and the *doric* modes are confirmed by Biber and Gray's (2010) study on the stereotypes around academic writing and conversation stereotypes. In their corpus multidimensional analysis

²⁷ These two concepts are used to refer to the qualities that languages have to interpret and represent the world.

of these two registers, the authors found that speaking is more clausal and explicit while academic writing is more nominal and compressed.

The nominal character of the attic style is reflected in two aspects of written academic discourse: heavily modified noun phrases and frequent nominalisation.

8.2.2.1. Heavy noun modification

Noun phrases (or nominal groups) are syntactic units that comprise a head (NOUN) and determiners and modifiers (Biber et al., 1999). Determiners provide deictic information such as distance from the speaker (*this, that*) and definiteness (*a, the*) or quantification information (*many, much, seven, this, these*). Modifiers classify or describe the head and can be placed before (pre) or after (post) the noun. A noun can take as premodifiers adjectives, participial modifiers, and other nouns. Noun postmodification can be performed by relative clauses; *-ing*, *-ed*, and *to* infinitive clauses, prepositional phrases, and other noun phrases in apposition. Heavy noun modification can occur under two circumstances: when a noun has several pre and postmodifiers²⁸ (Fang, Schleppegrell, & Cox, 2006),

(2) [*Three-Dimensional*] [*Quantitative Structure–Property*] [*Relationship*] (**Models**) [*for Prediction of Thermodynamic Properties of Polychlorinated Biphenyls*] (Swati Puri, James S. Chickos, & Welsh, 2002; Trevor, 2006)

²⁸ Parentheses are used around the NP heads. The main head noun is **bolded**. Square brackets are placed around the noun head pre and postmodifiers.

or when modifiers are long and contain multiple levels of embedding (Biber & Gray, 2010):

- (3) [*a (**bifurcation**) [in the metaredundancy pattern,] [leading to the (duality) [of styles]] [that Rulon Wells spoke about at the (conference) [whose (aftermath) [we are celebrating here]]]*]. (Halliday & Martin, 2003, p.129)

Heavy NP modification is characteristic of written academic discourse (Biber & Gray, 2010; Biber, Grieve, & Iberri-Shea, 2009; Biber et al., 1999). Heavy NP modification allows writers to compress their meanings into NPs and therefore be more economical, which provides their expert readers with a “faster, more efficient reading” (Biber & Gray, 2010, p. 11). However, for novice readers, heavy NP modification eliminates the meaning relationships that are traditionally expressed with clausal links (connectors), which might make the processing of these units burdensome. Similar considerations apply to the real-time processing of heavy NPs in the oral mode.

In oral academic discourse, speakers avoid heavily modified NPs by using other structures like *there* existential clauses or passive voice (Carter-Thomas & Rowley-Jolivet, 2001). These authors do not report specific mechanisms to reduce heavily modified NPs; they just explain that NPs are lighter in conference presentations given the processing burden they might place on the audience. Studies contrasting the language used by NS and NNS engineering teaching assistants (Levis et al., 2012) have found that in those cases in which the spoken content (tutoring session) depends on originally written content (textbooks, notes) NNS international teaching assistants (ITAs) use shorter but more grammatically flawed NPs than their NS

counterparts. Although the study focuses on the efficiency of oral communication by ITAs, it does not explain heavy NPs reduction as a mechanism to facilitate content understanding.

In this study, I focus on the reduction of NPs through mechanisms such as the elimination or the change of position/function of modifiers, which I found as a mechanism to transition from written to oral discourse that high achievers consistently used (Nausa, 2015, 2017).

8.2.2.2.Nominalisation and grammatical metaphor

The nominal character of academic writing (attic style) is also reflected in the tendency to frequently use nominalisations. The term *nominalisation* refers to two different syntactic processes: (1) when a word from any given category is transformed into a noun or (2) when any word or group of words is placed in a slot typically occupied by a noun (Halliday & Matthiessen, 2004):

(4) *Dinosaurs became **extinct** after the Cretaceous meteor **struck**.*

(5) *The **extinction** of dinosaurs happened after the Cretaceous meteor **strike***

The transformation of 4 into 5 illustrates the two cases of nominalisation. The adjective and verb (*extinct* / *struck*) are transformed into noun form (*extinction* / *strike*); additionally, to express the same content, *extinction* occupies the subject slot in the main clause in 5 and *strike* becomes the head of the NP *the Cretaceous meteor strike*.

This type of nominalisation also exemplifies a semiotic phenomenon that Halliday and Matthiessen (2004) have termed *grammatical metaphor*. This term is opposed to the *congruent* expression of meaning. In congruent expression, there is a default one on one correspondence

between semantic categories (entity, process, attribute) and lexical choices (nouns, verbs, adjectives).

(6) *People in Colombia consume a lot of coffee*

In 6 there is a congruent expression of meaning: the nouns *people*, *Colombia*, and *coffee* denote entities; the verb *consume* denotes a process. The same idea can be made grammatically metaphorical by nominalising the noun (*consume-consumption*):

(7) *The high consumption of coffee in Colombia*

(8) *The high consumption of coffee in Colombia has reactivated the internal market.*

The resulting NP (7) can occupy the subject slot in (8). A process (*to consume*) is now expressed as an abstract entity-thing (the attic style represents the world as a world of things). This abstract entity in turn is construed as something that has agency; *consumption* is in the capacity to reactivate a market. The action that would be normally attributed to a human agent is now attributed to a non-human abstraction.

Like most academic discourse linguistic traits, nominalisation has been mainly studied in written discourse, for example, in aspects like the acquisition of nominalisation after explicit instruction (Crosthwaite, 2016), or as a mechanism for students to be socialized in their academic communities (Meunier & Gentil, 2014), among others. Nominalisation has also been studied in comparisons between written and spoken academic discourses. In an SFL comparison of how a physics textbooks and interactive teacher talk present contents (R. F. Young & Nguyen, 2002),

teacher talk was found to unpack the grammatical metaphors presented in the textbooks. Another study comparing written and spoken feedback provided to university students on their written production (Gardner, 2004) found that the unpacking of nominalisations (de-nominalisation) was also a common characteristic of the oral mode, along with others that make content more accessible to students. Nominalisation use was found to be low in a study in students' notes taken from mathematical lectures; their written discourse was also found to be more process oriented (doric) (Österholm, 2012). More specifically, in academic OPs, apart from the mechanisms to avoid heavy NPs or nominalisations (Carter-Thomas & Rowley-Jolivet, 2001), found research focuses on advice provided to postgraduate students (Chanock, 2002, 2005) on what lexicogrammatical aspects of oral academic discourse to use in OPs and what aspects to avoid from written discourse, nominalisations among them. Only one study comparing parallel pairs of paper-oral presentation that focuses on nominalisation was found (Umesaki, 1991). In this research, the author found nominalisations to be more frequent in the written than in the oral version; this research does not study the mechanisms used to unpack nominalised content that I propose in this chapter.

In summary, grammatical metaphor through nominalization increases the expression capacity of language by adding new nuances of meaning (Halliday & Matthiessen, 2004); however, when these nominalizations are accompanied by heavy pre and postmodification, aspects like the exact type of connection between ideas or the agency of actions can be hidden or obscured for the inexperienced reader/hearer (Fang et al., 2006; Biber & Gray, 2010). In OPs for non-expert audiences, a genre that can be considered as intermediate between the research paper and conversation, the need for elaboration, explicitness, and simplification of message makes it necessary for presenters to learn to reduce heavily modified NPs, make logical relationships

between ideas overt, and clarify or transform potentially difficult to process metaphors into congruent messages. As such, in OPs (doric style) that are based on originally written content (attic style), it is expected of presenters that they use heavy NP reduction mechanisms like elimination or repositioning of modifiers (Nausa, 2015, 2017) or the denominalization (unpackaging) of grammatically metaphorical nominalisations.

8.2.3. Modality

The expression of modality as the judgement of content is another key aspect in the engagement between presenter and audience in OPs. As explained above, the modification of type and orientation in the expression of modality was one of the aspects that was observed in the first study on this students' population (Nausa, 2015, 2018).

For a description of modality and related aspects (polarities, types, orientation, value), see section 6.2.1.

Studies of modality, or related concepts like *stance* (Biber, 2006; Hyland & Guinda, 2012), *evaluation* (Hunston, 2011), *appraisal* (Martin & White, 2005), in academic discourse are high in number for written discourse (e.g. Aull & Lancaster, 2014; Barton, 1993; Bruce, 2016; Crosthwaite, 2016; Crosthwaite & Jiang, 2017; Hyland, 1996a, 1996b; Lancaster, 2016; Lee, 2008). In oral academic presentations, the concept has been analysed in non-linguistic aspects like L2 learners' discourse socialization (Kobayashi, 2006; Morita, 2000). Modality has also been studied in linguistic comparisons of NS and NNS use of linguistics traits like *it*-stance structures, adverbials, and stance structures containing 1st person pronouns (Zareva, 2012) or the analysis of academic identity projection with first-person pronoun structures expressing stance (Zareva, 2013). Another line of studies focuses on the judgement of content with verbal and non-

verbal modes (Fortanet & Ruiz-Madrid, 2016; Hood & Forey, 2005; Querol-Julián & Fortanet, 2012; Zhang, 2015), which I also approach in this thesis in CHAPTER 4.

Nonetheless, no studies focus on mechanisms to transform written into oral content the way it is proposed in this study (written and spoken sentences by the same author expressing the same content). The study comparing NS and NNS TAs (teacher assistants) transforming a written prompt into the oral mode (Levis et al., 2012) reports Chinese TAs as using more modals than their American and Indian counterparts; however, this study does not analyse whether the way modalization is expressed changes from the written to the oral mode.

8.2.4. Code glosses

Code glosses have been defined as elaborations of content that the writer/speaker does when they anticipate that their audience might find their content difficult to understand. This thesis has already approached code glosses as used in the OPs (see sections 7.2 and 7.3.1 for the theory of code glosses and related research). However, the way code glosses are approached in this chapter is different in that the glosses that are analysed are those not present in the original written version. This is to say, this chapter analyses those cases in which (1) presenters recycle a proposition from their essay and (2) add a reformulation or example to help their OP audience understand such content.

Again, from the studies on the transition between modes surveyed in this chapter, only the one comparing Chinese, Indian, and American TAs (Levis et al., 2012) reports the inclusion of content not stated in the original written prompt. These additions came in the form of examples and explanations. Indian and Chinese examples were found to be obscure while American TAs' examples were found to be potentially closer to students' personal experiences.

In conclusion, speakers in OPs need to be aware of the needs of their audience, which they can address by organizing and presenting content in a way that (1) can be easily followed (information structure principles), (2) reduces heavy-to-process information, (3) clearly shows the presenter's stance, and (4) includes reformulations and examples to clarify potentially difficult to understand information. The processing of spoken information is different in several ways to the processing of written information (e.g. OPs are delivered in real time, presenters hardly ever repeat what they say, listeners can't go back to what they heard, etc), therefore, it is expected that the grammar choices that presenters make address the audience needs. The mechanisms explained here and summarized in *Figure 8.1.* were identified in Nausa (2015, 2017, 2018) and have also been studied in other oral genres. However, as far as the review of literature in this chapter is concerned, no studies approach them in the analysis of parallel corpora with pairs of written-spoken sentences by the same authors.

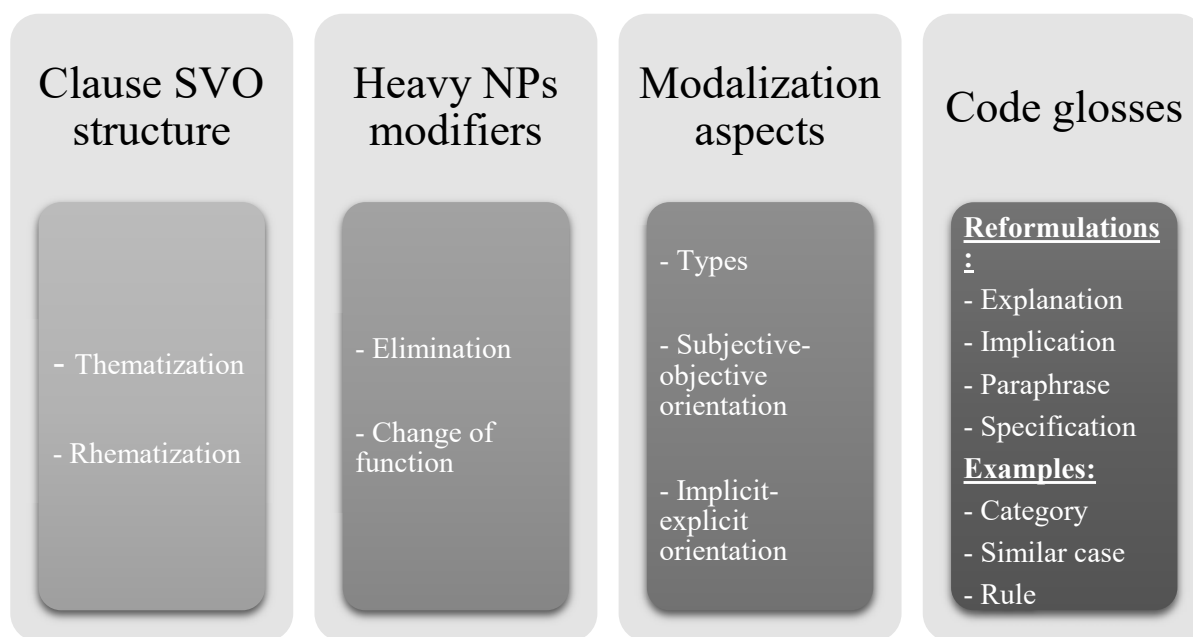


Figure 8.1. Taxonomy of changes in the transition from written to oral discourse

8.3. Methods

In this section, I explain the methodology that I implemented to achieve the goals of this follow-up study: confirm the mechanisms of change that I found in my first study, include other variables in the study (medium achiever, hard and soft disciplines, other mechanisms), and remedy the original study limitations (corpus size, lack of statistical information). The methodology is presented in two sections: the selection of two parallel subcorpora and the identification and analysis of sentences.

8.3.1. Selection of subcorpora for analysis

This research is based on a 128228-token corpus composed of 88 pairs of essays and oral presentations transcripts by the same authors. However, given that the automatic identification of sentences expressing the same content in an essay and an OP is not something that I was able to successfully do using corpus software, I opted for performing a manual search for those sentences.

The following are the steps that I followed. First, I selected a random sample of 30 pairs of essays and OP transcripts. Each e-op (essay-oral presentation) pair was created by the same author; therefore, the contents in each pair should be approximately the same. Second, to answer the questions in this chapter, the 30 pairs of texts were chosen as follows: from each level (10 high, 10 medium, and 10 low) and 15 from each disciplinary division (15 hard and 15 soft). (see 3.3.3 for the criteria for classifying texts).

8.3.2. Identification and analysis of sample sentences

The resulting parallel corpora were composed of 45558 words. Each e-op pair was colour coded and manually analysed to identify sentences expressing the same content (Appendix N). Those sentences were extracted and organized in Excel worksheets (Appendix O). Four pairs of sentences were extracted from each e-op pair to obtain 120 pairs of e-op sentences expressing the same propositions. This resulting corpus of sentences was composed of 6936 words. In the worksheets, every pair of sentences was analysed to identify the mechanisms for reworking content based on the taxonomy of mechanisms in 8.2.

The number of changes in each e-op pair was counted. Cases in which sentences were not modified because students recited or read from a slide or script were also counted. Once raw frequencies were obtained, the following quantitative analyses were performed. First, the percentage and normalised (per 1,000 words) frequencies of the distribution of changes in the e-op sentences parallel corpus and the specific sub-corpora (levels and disciplines) were calculated to make comparisons across corpora.

Significance (Log Likelihood) tests and effect size (BIC) scores were also calculated to determine what frequency differences between the corpora were statistically significant and large enough to be considered positive evidence against null hypotheses. Given the small size of the selected parallel corpus, I had to establish a not so high log likelihood value threshold >3.84 , which is significant at $p < 0.05$ (McEnery et al., 2006; Rayson, 2017), and ignore the effect size values, which as expected, were negative.

Finally, it is important to clarify that although the methodology of parallel written-oral corpora analysis is based on the one used by Carter-Thomas and Rowley-Jolivet (2001) for the analysis of parallel texts (conference presentations and proceeding papers) by the same author,

the methodology that I propose here differs in two important aspects. First, these authors did searches for specific syntactic mechanisms (Extraposition, existential *there*, Inversion, it-clefts, and wh-clefts) and how they distributed in the written and spoken corpora. My study selects e-op pairs of sentences to analyse how written content is reworked in the written to oral mode transition. The selection of syntactic changes is based on the mechanisms identified – thematization-rhematization, reduction of heavy NPs, changes in the expression of modality, and inclusion of code glosses-- and hypothesized (denominalization) in my first module (Nausa, 2015, 2017, 2018). Second, the authors do not calculate normalised frequencies nor significance tests to make comparisons. This is so probably because they only compare written and oral expression. Other variables like discipline or level of oral achievement are not considered.

8.4. Quantitative analyses

The types of modification made to originally written content had a similar distribution in the corpus. A total of 166 changes were identified in the 120 pairs of e-op sentences. 91 sentences underwent from one to six changes; 29 sentences were not modified. This means that each modified sentence underwent 1.8 changes on average. The most common type of modification was the inclusion of code glosses (33%), followed by changes in clause SVO structure (24%), reduction of heavily modified NPs (24%), and changes to the expression of modality (19%).

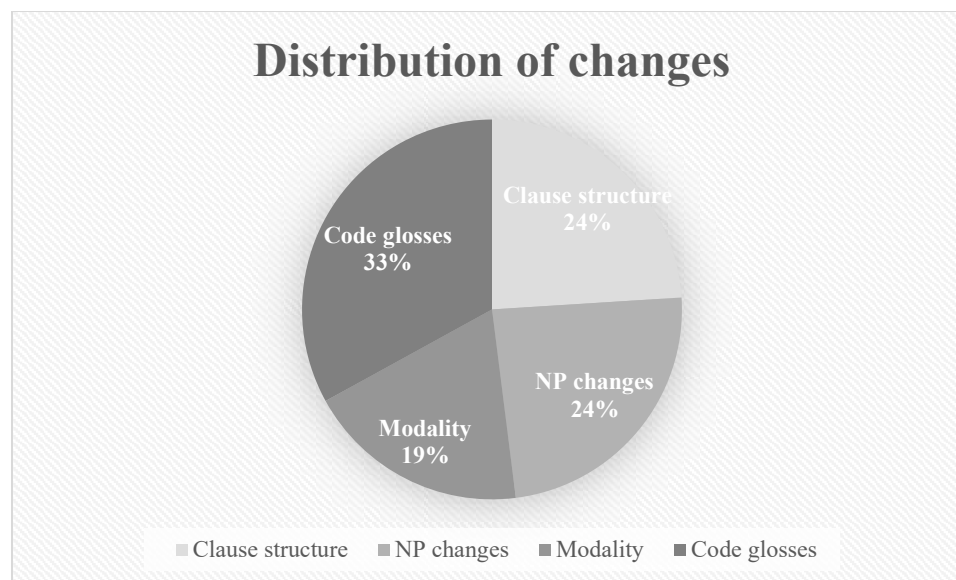


Figure 8.2. distribution of syntactic modifications in the e-op subcorpus

Table 8.1 presents the syntactic modifications in the e-op subcorpus distributed by level of achievement.

8.4.1. By level of achievement

As can be seen in *Figure 8.3*, high achievers made almost half of the changes (48%) almost twice as much as their medium (27%) and low achiever (25%) counterparts confirming my findings in the pilot study (Nausa, 2015, 2017, 2018), in which high achievers appeared to more consistently use the mechanisms.

Table 8.1. syntactic modifications raw (R) and normalised (N) frequencies, and percentages (%) by level of achievement

	Clause structure			NP reduction			Modality			Code glosses			Total Modifications			Corpus sizes
	R	%	N	R	%	N	R	%	N	R	%	N	R	%	N	
High	15	9	5.5	23	14	8.4	20	12	7.3	22	13	8.0	80	48	29.2	2740
Medium	13	8	6.1	8	5	3.7	10	6.02	4.7	14	8	6.5	45	27	21.0	2145
Low	12	7	5.9	9	5	4.4	2	1.2	1.0	18	11	8.8	41	25	20.0	2051
Totals	40	24	5.8	40	24	5.8	32	19.3	4.6	54	33	7.8	166	100	23.9	6936

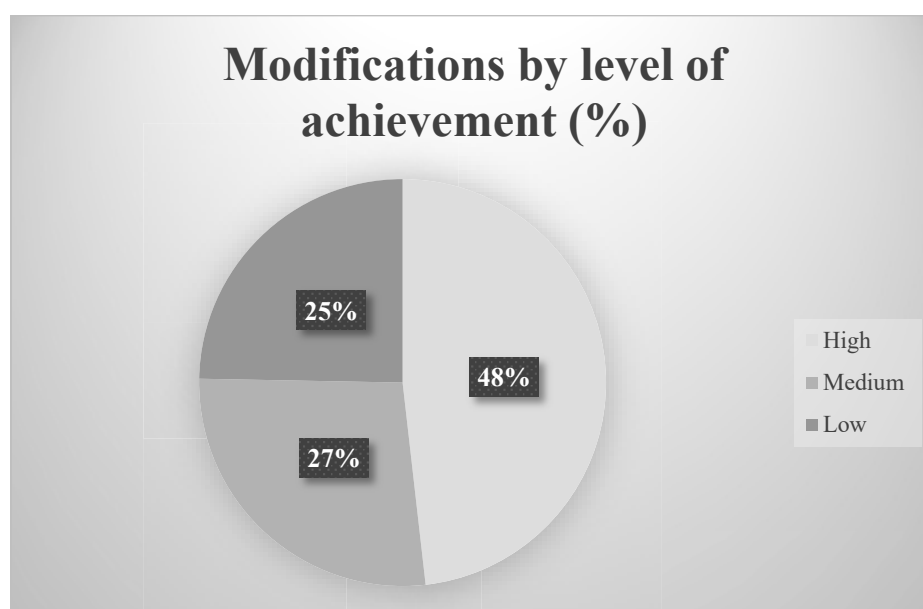


Figure 8.3. distribution of syntactic modifications in level of achievement subcorpora

The analysis of normalised frequencies (*Figure 8.4*) by specific changes shows a different distribution.

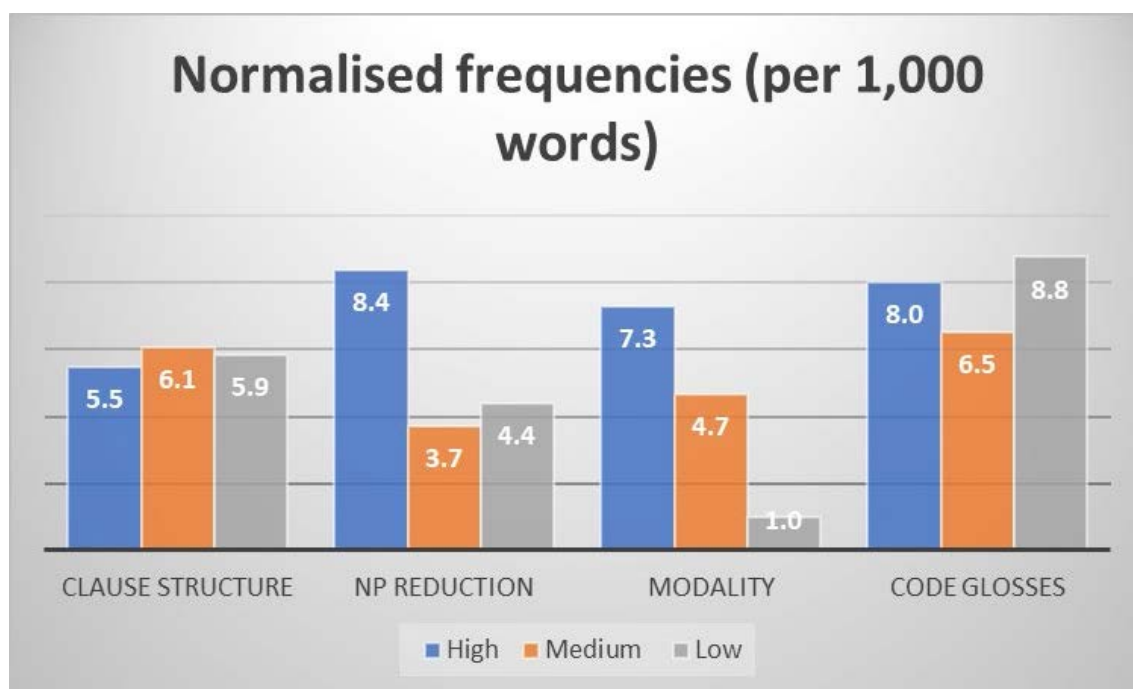


Figure 8.4. Distribution of the four types of syntactic modifications by level of achievement

Code glosses and clause structure modifications, two of the most common types of changes, do not exhibit sharp distribution differences among the achievement subcorpora. Code glosses frequency differences were expected to be significant given the results in chapter 7 (see Table 7.5) and module 1 (Nausa, 2015, 2018). NP reductions and modality changes, on the other hand, not only exhibit greater differences, but are also more frequently used by high-achievers. These frequency differences are statistically significant at $p < 0.001$ (modality) and $p < 0.05$ (NP reduction) (Table 3.4) as observed in the Log Likelihood values in Table 8.2.

Table 8.2. Syntactic changes raw frequency and significance analysis by level of achievement

	Observed frequencies			log likelihood	Bayes Factor BIC
	High	Medium	Low		
Clause structure	15	13	12	0.08	-17.61
NP changes	23	8	9	5.38	-12.31
Modality	20	10	2	12.34	-5.35
Code glosses	22	14	18	0.73	-16.96
Total	80	45	41	5.19	-12.50
Corpus size	2740	2145	2051		

8.4.2. By discipline

The analysis of distribution of the 166 syntactic modifications by discipline (*Figure 8.5*) shows a similar distribution for hard (56%) and soft (44%) disciplines. Therefore, the distribution of changes is a more significant discriminating mark in the level of achievement comparisons.

Table 8.3. Syntactic modifications raw (R) and normalised (N) frequencies, and percentages (%) by discipline

	Clause structure			NP reduction			Modality			Code glosses			Total Modifications			Corpus sizes
	R	%	N	R	%	N	R	%	N	R	%	N	R	%	N	
Hard	29	17	8.3	15	9	4.3	14	8	4.0	35	21	10.0	93	56	26.6	3501
Soft	11	7	3.2	25	15	7.3	18	11	5.2	19	11	5.5	73	44	21.3	3435
	40	24	5.8	40	24	5.8	32	19	4.6	54	33	7.8	166	100	23.9	6936

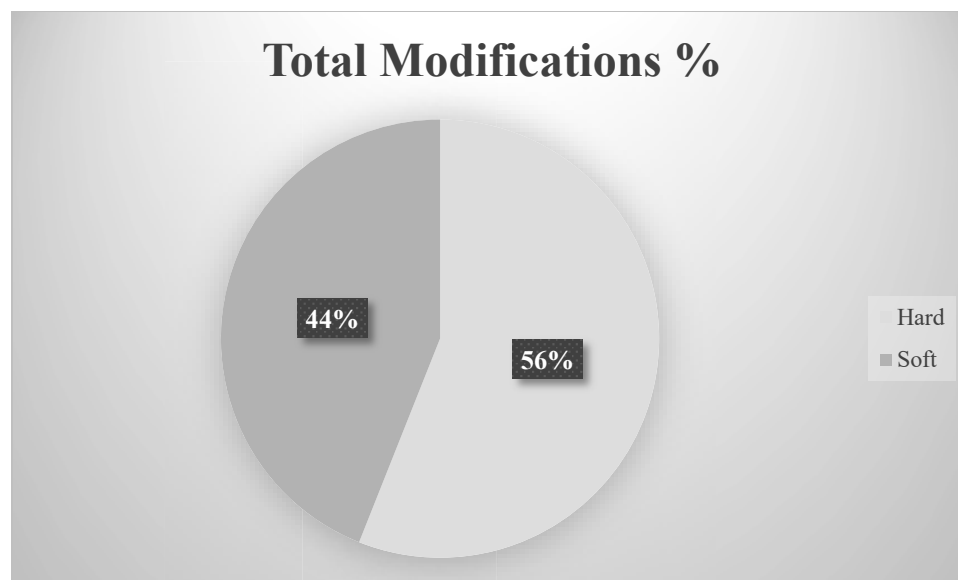


Figure 8.5. Distribution of syntactic modifications in disciplines subcorpora

However, a closer look at the different types of syntactic modifications as expressed by normalized frequencies (*Figure 8.6*) shows a different situation. Clause structure modifications and code glosses are more frequent in the hard science corpus while NP reduction and modality are more frequent in the soft fields.

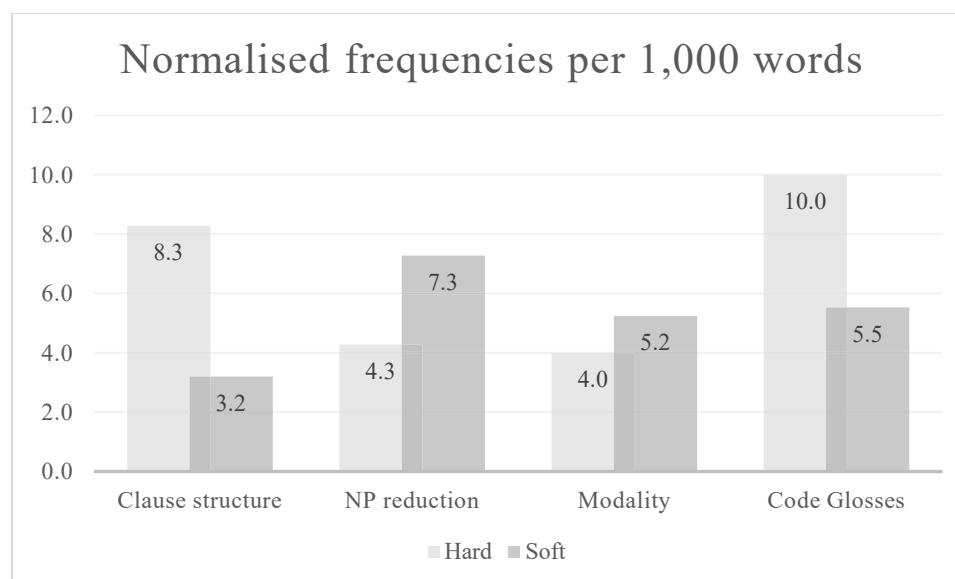


Figure 8.6. Distribution of the four types of syntactic modifications by discipline

Apparently, the tendency in the soft fields resembles that reported for the level of achievement divide; however, statistical significance analyses reveal a different tendency.

Table 8.4. Syntactic changes raw frequency and significance analysis by discipline

	observed frequencies		log likelihood	Bayes Factor BIC
	Hard	Soft		
Clause structure	29	11	8.06	-0.79
NP changes	15	25	2.72	-6.12
Modality	14	18	0.58	-8.26
Code glosses	35	19	4.51	-4.33
Total	93	73	2.05	-6.79
Corpus size	3501	3435		

In the disciplinary divide, clause structure modifications and the inclusion of code glosses had greater log likelihood values: 8.06 and 4.51, respectively. These frequency differences are statistically significant at $p < 0.01$ (clause structure) and $p < 0.05$ (code glosses).

Two interesting conclusions can be drawn from statistical analyses in this chapter. The first is that the general use of the mechanisms to transition from oral to written discourse seems to be a defining characteristic of high achievers given that they more frequently make use of them, almost twice as many as their medium and low achiever counterparts. A closer analysis of normalized frequencies and significance tests, and this is the second, shows that the tendencies of use of specific mechanisms can be affected by the two variables in the study: levels and disciplines. In the level of achievement divide, NP reductions and modification to the expression of modality are the mechanisms that more significantly discriminate among the three levels, with high achievers exhibiting higher frequencies of use. In the disciplinary divide, the other two studied mechanisms (clause structure changes and code glosses inclusion) more significantly discriminate between hard and soft field students. Also, interestingly, these two mechanisms were more frequently used by hard discipline students.

In the following section, I will report on discourse analyses of these mechanisms.

8.5. Discourse analyses

This section analyses the types of modifications and their corresponding sub-mechanisms that were identified to significantly discriminate among levels of achievement (*NP changes* and *modality*) and disciplines (*clause structure changes* and *code glosses inclusion*) as expressed by statistical significance log likelihood values. The purpose is to provide a qualitative discourse account of the mechanisms to rework originally written content to be expressed in the oral mode with examples from the corpus. Additional frequency counts are provided to explain the submechanisms.

8.5.1. By level of achievement

8.5.1.1. Heavy NP reduction

Three types of noun phrase reduction submechanisms were identified: elimination of modifiers, change of function (or movement) of head or modifiers, and denominalization (*Figure 8.3*). The most frequently used mechanism was the elimination of modifiers (32 instances). Unexpectedly, the other two, function change (2 instances) and denominalization (6 instances), were infrequent. Function change was identified twice in the smaller corpus in module 1 (Nausa, 2015), so it was expected to more frequently appear in this study. As explained above, denominalization has been reported as a mechanism to transition from written to spoken discourse (Gardner, 2004; Young & Nguyen, 2002). *Figure 8.7* shows the distribution of the three mechanisms in the levels subcorpora.

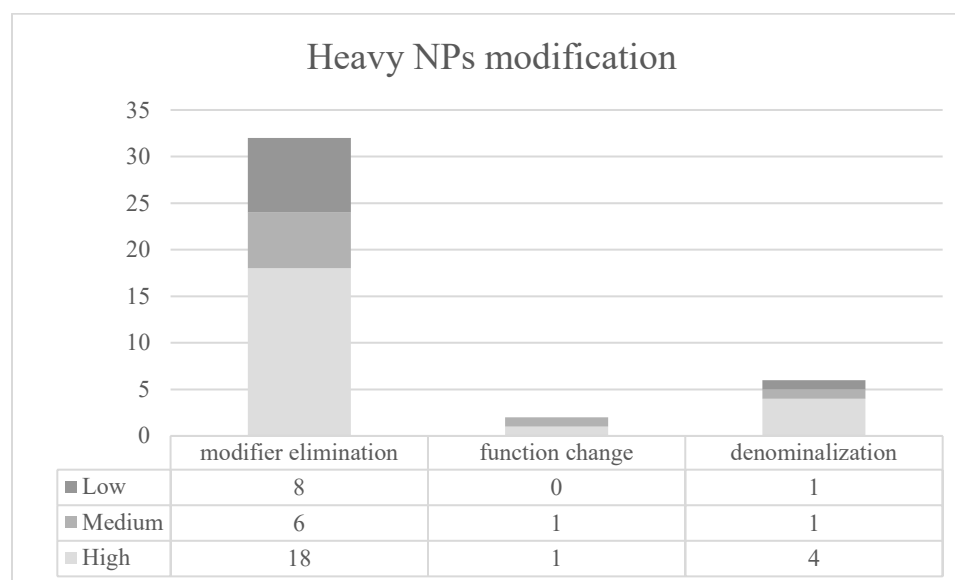


Figure 8.7. Distribution of heavy NPs reduction mechanisms by level of achievement

High-rated OPs clearly outnumber the others in the general use of the mechanisms.

Examples 1w and 1s²⁹ illustrate two of the heavy NP reduction submechanisms: modifier elimination and function change. In 1w there is a 21-word NP whose head is *increase*, which is reduced to a 16-word NP.

(1w) *In the last decade, the statistics have reported [a serious] (**increase**) [in the number of aggressions, homicides, and suicides] [within the Colombian domestic sphere], [especially related to loving matters].*

(1s) *In the last decades, eh the statistics have report eh a report an (**increase**) [on domestic violence in Colombia], eh [specially homicides, suicides, and aggressions related to loving subjects].*

(H-ANTR-3)

In the transition to 1s, the premodifier *serious* is removed. The first postmodifier undergoes three changes. First, it is reduced to *homicides, suicides and aggressions*. Second, the resulting NP is made the head of a new NP, which is in turn postmodified with the third postmodifier in 1w. Third, this new NP is moved to RHEME position in the clause.

An analysis of 1w and 1s co-texts, shows that the movement to RHEME position in 1s is motivated by information structure and information unpackaging (denominalization) principles.

²⁹ e-op sentence pairs are marked (1w) for the written version and (1s) for the spoken version. (1w) and (1s) are the essay and OP version of the same sentence. In most cases, sentences are presented in isolation and then within their co-text. To avoid confusion, sentences in isolation are marked (1w) or (1s), and within their larger written or oral co-text (1w-c) or (1s-c). In other cases, rewrites might be shown to demonstrate a particular point; those are marked as (1w-i) or (1s-i) meaning that they are an idealized version, not what presenters wrote or said. The reasons for using rewrites are explained in 3.5.2.

(1s-c) *I the last decades, eh the statistics have report eh a report an increase on domestic violence in Colombia, eh specially homicides, suicides, and aggressions related to loving subjects. Most of [fs] of victims are women a chil [fs] and children.* (H-ANTR-3)

The following clause in 1s-c iterates *homicides, suicides, and aggressions* as GIVEN by referring to *victims* (there is semantic link) and introduces *women and children* as NEW. Instead of iterating the abstract nouns in RHEME position, a human noun (*victim*) fills the subject slot, making the sentence more grammatically congruent.

The flow of information in 1w-c is different. The link to the previous clause is made with *these events* (a shell noun), making the following sentence grammatically metaphorical.

(1w-c) *In the last decade, the statistics have reported a serious increase in the number of aggressions, homicides, and suicides within the Colombian domestic sphere, especially related to loving matters. Since 2004, these events were the main preoccupation of the National Institute of Legal Medicine (2010) and some no governmental institutions (like Profamilia).* (H-ANTR-3)

The third submechanism for heavy NP reduction, denominalization, is illustrated in 2w and 2s.

(2w) *Besides the entitlement is based on the belief or (**perception**) [by individuals] [who are (**deserving**) [of a right to do or have something to aspire to something, or be someone in particular]], [without necessarily being linked to a legal right (can be informal)]. Without that is related to a real effort or input; right as its basis is the dignity of persons.*

(2s) *besides is based on [fs] in the belief on the belief that the people **perceive** they **deserve** from the others eh and the basis of this kind of entitlement is the dignity.* (H-ADMIN-1)

In the transition to 2s, apart from the evident eliminations and movements, two nouns expressing a process (*perception*) and an attribute (*deserving*) are denominalised and changed to a congruent verb form (*perceive* and *deserve*). The heavy NPs that contain *perception* and *deserving* are unpacked and turned into clauses with human subjects (*people-they*). With the denominalizations and the other movements and eliminations, this student turned a heavily modified grammatically metaphoric 62-word unit into a more congruent, and arguably, easier to process 33-word unit.

Although these mechanisms to reduce heavy or metaphorical NPs were found in medium and low achiever OPs, they were not as consistently used as they were used by high achievers.

(3w) *That is, that a single watchman could observe the prisoners from the (center) [of the jail], but they couldn't see him.*

(3s) *It means that one men can see everything, every eh [fs] all the prisoners from the (center), but the prisoners can't see the watchman.*

(M-HIST-2)

(4w) *In every part of the world the people need to take (drugs) [to control their diseases] and this treatment required the (control) [of both quantity and duration of drug in the human body.]*

(4s) *Eh in every part eh [fs] parts of the body eh the people needs eh to take drugs, and eh this treatment required the (control) [of the [fs] of the quantity and the duration the drug in the human body.]*

(L-CQUI-1)

In 3w-->3s, although there is a good amount of modification, the reduction of NPs can only be observed in one case: the elimination of *of the jail*. In 4w-->4s, the NPs only undergo a few changes: elimination of *to control their diseases* and *both*. Additionally, the substitution of

body for *world* distorts the meaning of 4w. Analysis of 4w and 4s co-texts doesn't seem to indicate that changes are motivated by information structure or NP reduction principles.

(4w-c) *In every part of the world the people need to take drugs to control their diseases and this treatment required the control of both quantity and duration of drug in the human body. The quantity is related with dose and the duration with the time that the drugs are in the organism.*

(4s-c) *Eh in every part eh [fs] parts of the body eh the people needs eh to take drugs, and eh this treatment required the control of the [fs] of the quantity and the duration the drug in the human body. Eh the quantity is related with the [fs] with a dose, and the duration is the time the [fs] the drug is in the human body.*

8.5.1.2. Changes to the expression of modality

Students made three types of changes to the expression of modality to transition from written to spoken discourse: change of types, implicit-explicit orientation, and objective-subjective orientation (*Figure 8.8*). Change to subjective-objective orientation was the most frequent type (15 instances), followed by implicit-explicit orientation change (10) and change of type (7). In all cases, most of the changes were found in the high achievers subcorpora. Low achievers had the fewest cases and they did not use the objective-subjective orientation change mechanism.

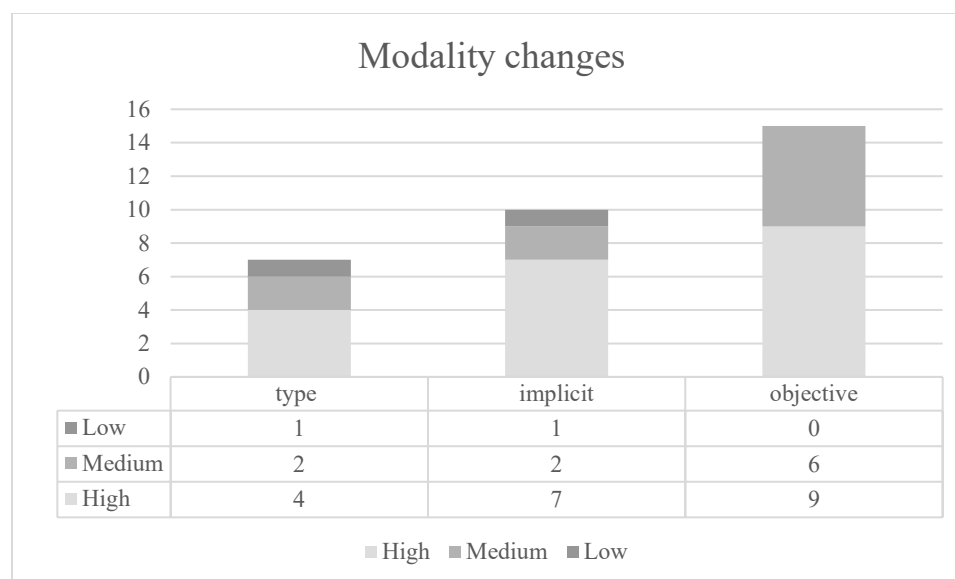


Figure 8.8. Distribution of modality change mechanisms by level of achievement

5w->5s³⁰ exemplifies a case of change of type. In 5w, the modalized content (underlined) is expressed as probability (*concluded that*) and necessity (*are needed*). In 5s, necessity is now expressed as possibility (*have chosen*) as something that depends on human control (Quirk et al., 1985).

(5w) *Therefore, Davidsson & Henrekson (2002) **concluded that** some institutional reforms **are needed** in order to foster de creation and development of HGF.*

(5s) *General policies eh have problems. [reading7] In countries like Netherlands poli [fs] like the Netherlands, policy makers **have chosen** pol eh [fs] general policies in order to foster the creation of high growth firms [reading7].*
(H-ADMIN-2)

³⁰ Modalizing expressions are marked in bold type. Modalized content is marked with underlines.

The change of modality type does not necessarily imply that the message has been distorted. The change makes sense given that the meaning in 5w is expressed as a theoretical recommendation from scholars, while in 5s, it is construed as a decision made by policy makers. The modalities are attributed to different actors; therefore, a change of modality type is the appropriate choice.

6w->6s includes the two orientation change mechanisms: subjective-objective / implicit-explicit. In 6w the modalized content (education about species) is expressed in an objective-explicit way (*it is necessary*). In the transition to 6s, the orientation is made subjective and implicit (*need*).

(6w) *The first mechanism is education about the species that we have in our regions. **It is necessary** first to know and then care. The education in the schools and universities **is important** because if our children and young people have a good knowledge about the species, they can convince their parents and friends to protect the species to local level.*

(6s) ***The first step** eh is education. We **need** education [fs] we **need** educate to our child because **is necessary that** the child know what is a specie and what is the role of the species in the ecosystem.*
(H-CBIO-2)

Like the cases of reduction of heavily modified NPs, medium and low achiever sentences not only exhibited fewer cases, but also grammar errors or pragmatic infelicities.

(7w) *In last years of twenty century Democracy had been adopted as system government for majority of countries of word. This movement was known how Third Wave and this described the fall of many dictatorial regimes in Latin America and East Europe and their transformation in democracies.*

(7s) *The context in my research are the last eh decades of the twenty century when many dictatorial governments, eh for example the military regimens in Argentina, Chile, eh Uruguay, or Bra Brazil, or the eh communist regimes in eh East Europe, (far) and they **can** to transform in democracies. Eh this [fs] this success was called terce eh [Fs] third wave of democratization and many countries [Fs] countries **can** democracy in this [fs] in this time.*
(M-CPOL-1)

(8w) *In addition, **this establishes** that the Colombian State **must** compensate to the persons when two elements are configured: imputation and damage*

(8s) ***the obligation** de raparate [fs] repair for the damage illegal occasioned of the citizens in relationship in the [unintelligible] public and citizens.*
(L-DERE-1)

In 7w->7s, although the modalization of content is appropriate, its grammar form is not; the modal verb takes up a verb in infinitive form and then a noun. In 8w->8s, although the student successfully changes from an implicit-subjective orientation to an explicit-objective-one, the expression of content includes the use of Spanish and several fluency disturbances that make the content difficult to understand. Additionally, the students does the opposite to what is expected – the spoken version is less personal / congruent than the written one as observed in the use of nominalization of *must* as *the obligation* in 8s.

The analysis of heavy NP reduction mechanisms and changes to the expression of modality in the levels of achievement subcorpus confirms the findings in module 1 (Nausa, 2015, 2017, 2018). It is clear that the higher the level of oral achievement the more frequent, the more varied, and more elaborated (grammatically and pragmatically) the use of transition mechanisms. The ability to eliminate or move modifiers reveals the grammar resources that high achievers

have at their disposal and their awareness of the needs of the audience. As pointed out in module 1, it could be argued that the mastery over grammar resources frees cognitive resources to analyse what information can be kept or moved so that the message is better understood. The ability to change the way modalization is expressed, apart from being evidence of students' grammar knowledge, shows their ability to engage the audience by selecting forms that allow them to take specific positions in relation to presented knowledge (what others say / what their contributions are). As argued in chapter 6, the expression of modalization assigns identities to both the presenter and the audience (e.g. *opinion holder-opinion judge*).

8.5.2. By discipline

8.5.2.1. Clause structure changes

In module 1 (Nausa, 2015, 2017) I identified two types of changes to clause SVO structure: thematization (topicalization or movement of adverbials to initial clause position) and the movement of adverbials between clauses. Like the other changes reported in that study, they were more frequently and consistently used by high achievers. In that module, I also identified the rhematization (movement to clause final position) of NP modifiers as something not effectively used, at least with clear pragmatic aims. In this study, based on a larger corpus, I found two clause structure change strategies to be more statistically significant in disciplinary comparisons: thematization and rhematization. Cases of movement of adverbials between clauses, although predicted, were not found. Both thematization and rhematization were found to be more frequent in the hard-discipline OPs than in the soft-discipline ones (see figure 5.8).

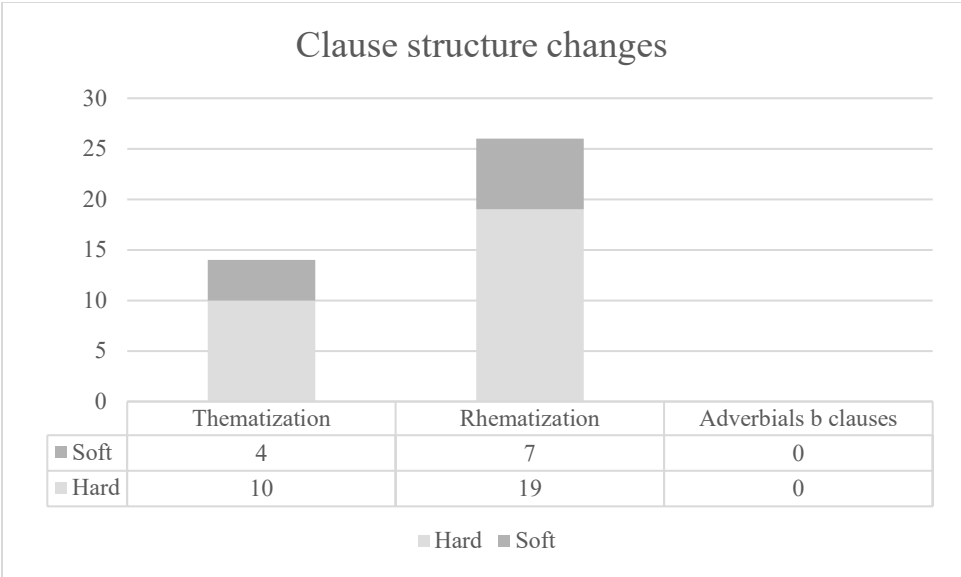


Figure 8.9. Distribution of clause structure change mechanisms by discipline

9w->9s illustrate a case of thematization. *In model theory* is topicalized as *in the theories* in 9s.

(9w) *The effective enumeration of theorems and non-theorems is the most natural way **in Model Theory** to establish if a first order theory *T*, in a countable language *L*, is decidable.*

(9s) **In the theories**, we say that a theory *T* is effectively numerable if we can list the theorems of *T* and we can list the non-theorems of *T*, and then eh we have the algorithm eh given by compare the sentence *p* with the missing list of these two.
(H-MATE-1)

An analysis of 9s in context shows that the change of expression and its thematization is motivated by information structuring principles (end-weight).

(9s-c) *Now, the problem is easy to state but it's so complex and we don't have a general solution. We have a partial solution and depends of the nature of the theory. For example, there are three techniques eh classic to show that a theory is eh decidable. The first, the effective numerability, the second, completeness, and the third the quantifier elimination.*

In the theories, we say that a theory *T* is effectively numerable if...

In the OP, the student introduces three theories/techniques to show that a theory is decidable in the RHEME. Then, in the following clause, he refers to them as *in the theories* in THEME position. The use of these techniques to guarantee information flow is a way to facilitate understanding to the hearer/speaker and therefore a sign of pragmatic competence.

However, thematization in the transition from the essay to the OP does not always seem to have a clear information structuring motivation. In the following example, the author had written *drought* in the RHEME of a clause, and then she iterated it as OLD in the THEME. However, in the OP, *drought* was put in the THEME of the first clause and iterated as *it* in the second.

(10w) *This phenomenon causes extreme drought. Drought is also presented in subtropical countries that presents seasons.*

(10s) *Ok, drought is one of these events. Eh it affects both topical countries and seasonal countries.*

(M-INGE-1)

An analysis of 10s in context shows that 10s was part of a cause-effect description in which there is a thematic progression. In this thematic progression, effects become causes.

(10s-c) *This eh global eh [fs] this eh was [fs] wastes eh causes extreme weather events eh, eh this eh extreme weather events eh causes [reading 2] changes in hydrol [fs] in hydrological system produced alterations in precipitations patterns, eh melting of snow and ice, increasing atmospheric models vapor, eh increasing evaporation and changes in soil moisture and runoff. These characteristics are related with the geographical location and geophysical formations [reading 2]. Ok, drought is one of these events*

It makes sense that the word *events* is used in 10s; however, to maintain a better flow of information, the positions of *drought* and *events* could have been reversed as in 10s-i. Also, probably *events* should have been iterated in the position in which *characteristics* is.

(10s-i) *These events are related with the geographical location and geophysical formations. Ok, one of these events is droughts. Droughts affect both...*

11w->11s illustrate rhematization, the second and most frequent mechanism to change the SVO clause structure in the transition to spoken discourse.

(11w) *The compounds that can transport ions are named ionophores.*

(11s) *Ionophores is a good solution, but what is ionophores? Basically a ionophore is a molecule that can transport ions.*

(M-CQUI-1)

In 11s there are two cases of rhematization. The first is the location of *ionophores* in the RHEME of the rhetorical question. This type of rhematization mechanism was the most commonly found and was mainly used to provide definitions or explanations (see 7.7.1 in code glosses chapter). The second is like the case described in 10w->10s; *ionophores* and *can*

transport ions are in RHEME and THEME positions in the essay. In the OP, their positions are switched; therefore, there are a thematization and a rhematization. Finally, an analysis in context of 11s shows that the rhematization of *can transport ions* is also motivated by the speaker's organizing information flow.

(11s-c) *Basically a ionophore is a molecule that can transport ions. In this in this picture, we can see the ionophore that transport ions from the extracellular cell to intracellular cell and the ionophore can eh travel through the cell member.*

In 11s co-text, the following clauses focus on the movement (*transport, travel*) of ions through cells.

8.5.2.2. *Inclusion of code glosses*

Code glosses are the second mechanism that significantly discriminates between disciplines in this study. Like changes to clause structure, code glosses³¹ are more frequent in hard-field OPs (*Figure 8.10*). For practical purposes, I only focus on the three most common cases: paraphrase, explanation, and specification.

³¹ It must be borne in mind that the way that I approach code glosses in this chapter is different from that in the general code glosses chapter 7. First, in this chapter, I focus on the reformulations or examples that are present in the OP and not in the essay. As such, it is possible that one example that was provided in an essay is repeated in an OP; that type of example is not the subject-matter of this chapter. Second, in chapter 7, I make great emphasis on the code glosses and their markers (*that is, for example*). In this chapter, following my conclusion in module 1 (Nausa, 2015, 2018), I focus on code glosses even though they are not marked linguistically, or another type of connector is used. In module 1, I found cases in which *because* was used for providing a cause for something, and the same expression of the cause was an explanation that was not in the essay and that was found to be aimed at facilitating understanding to the audience.

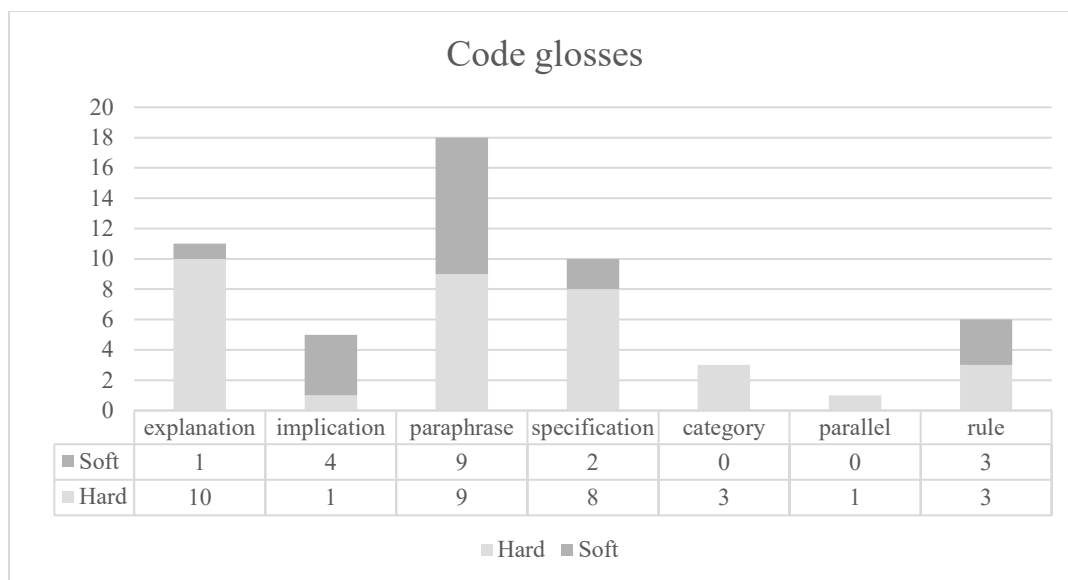


Figure 8.10. Distribution of code glosses by discipline

(12w) Furthermore Zitek et al. (2010) found that (the entitlement serves as a mediator of selfish behaviour), [ie, the victim of injustice acquires a sense of entitlement that leads to selfish behavior.]

(12s) Eh also Zitek at al eh found eh some moderator [fs] (some effect eh that is a mediator the entitlement in selfish behaviour) [when the people receive help, they become selfish] and this is eh the part that I want to explore.]
(H-ADMIN-1)

In 12w, the unit in parentheses is paraphrased as evidenced by the [*i.e.*,...] mark. In the transition to 12s³², the presenter paraphrases that paraphrase. In fact, the resulting paraphrase differs from the original in the focus on the person getting help (the original focuses on the victim). Focusing on entitlement as *getting of help* seems to make more sense to help the audience understand selfish behaviour.

³² In this section, examples are marked with parentheses for the unit that needs elaboration and square brackets for their code glosses elaboration (reformulations or examples).

In 13w, the author provides a definition for *habitat fragmentation* as *area reduction due to deforestation*. This definition is also provided in 13s, but the author paraphrases *area reduction* as *fragmentation of habitats* and provides extra information: *too small to support population*.

(13w) (*The habitat fragmentation*) is the **area reduction** due to *deforestation*. This new landscape is used for the human for agriculture, changing the soil nutrients.

(13s) (*Habitat fragmentation*) is the **fragmentation of habitats into patches that are too small to support population** and one of the [fs] of the main causes of *habitat fragmentation is deforestation for agriculture*.

(H-CBIO-2)

The second most common type of code gloss in the disciplines subcorpus is explanations. Explanations, as defined by Hyland (2007) “... elaborate the meaning of a preceding unit to make a concept more accessible by providing a gloss or a **definition**” (p.274). Definitions comprise two main elements: the *definiendum*, term that needs to be defined, and the *definiens*, the definition (Barnbrook, 2002). One common structure for definitions is one that links the *definiendum* to the *definiens* by providing (1) a superordinate term (the class of elements to which the referent of the term belongs to) and (2) specific information that makes the defined element different to other elements in the same class (specifier, discriminator). The introduction of the specifier can be done with relative pronouns (*that/which*), prepositions (*in, with*), past participle verbs, etc. 14w->14s introduces three instances of explanations (definitions).

(14w) *Emerging infection disease by fungal and oomycete species has high impact on several plant species. These species are common in different worldwide regions, but the last years it has emerged as a seriously problem with expansion crops in U.S.A and Latin America. Some examples included fungal such as Magnaporthe oryzae, Puccinia graminis and Ustilago maydis.*

(14s) *First, I will talk about emerging infe [fs] infection diseases. This is some pictures about some species. This is (magoaporthes oryzae), [that infect rice]; this is (puccinia graminis), [that infect wheat], and this is eh the most important or one the most important that is (ustilago maydis), [that infect eh maize].*

H-CBIO-1

In 14w, the student introduces three types of species (*Magnaporthe oryzae*, *Puccinia graminis* and *Ustilago maydis*). She does the same in 14s, but as can be seen in the example, this is done aided by pictures of the species. Each term that refers to each species (definienda) is defined. Although the provided definiens are appropriate for the context of the presentation, some grammar infelicities are evident: verbs are not inflected in third person form; *which* would have been a better word choice. The structure of the definitions requires the use of the default option for not restrictive relative clauses (*which/who/whose*), not *that*.

(15w) *For instance, an enzyme recognizes only one substrate, or a DNA section codifies a particular protein.*

(15s) *Eh one example is enzymes. One enzyme can react only in [a specific molecule called] (a substrate). Eh or another example is eh DNA, a section of a DNA chain eh can codify only one specific protein.*

(H-CQUI-1)

In 15w->15s, another linguistic realisation for definitions is observed. The definiens is presented first in the form of the specifier (*one enzyme can react only in*) and a superordinate

term (*molecule*), and then the term that is defined (*substrate*). The expected order of the definition would be:

(15s-i) *A substrate is the only molecule in which an enzyme can react.*

However, it is apparent that the student is following information structure principles (*rhematization-thematization*) as observed in the placement of *enzyme*.

(16w) *First of all, the pyrazoles should have high lipophilic groups for its use as transporter of copper ions (II). The lipophilic groups in this compounds stimulate the easy interactions with membrane lipids.*

(16s) *The first characteristic is high lipophilic groups, but what is (lipophilic integration or **high lipophilic integration**?) [is the same name]. the high lipo [fs] (lipophilic integration) is [the bind of two molecules that prefer repare eh [fs] repeal or [fs] or avoid water.]*

(M-CQUI-1)

16w->16s shows another case of explanation in which we not only observe the use of the definiendum-definiens structure with relative pronoun subordination, but also three other definition and information structure submechanisms. The first is the inclusion of another term for the term to be defined as observed in the marker *or*. The second is the use of rhetorical questions to introduce the term that needs to be defined in RHEME position. Rhetorical questions have also been found to be an interactive metadiscourse mechanism that undergraduate effective presenters use in their OPs (Ágnes, 2002). Third, the explanation that the two terms mean the same (*is the same name*). Of course, a more standard form like *both terms mean the same* would have been expected.

The third common type of code gloss in the disciplines subcorpus is specification.

Specifications present characteristics to restrict how the previous unit should be interpreted.

(17w) *For example, whether a person should invest in the purchase of goods what career he should study, where the company should locate a store or, even who should marry him.*

(17s) *For example, (personal decision), [which vehicle to buy], [how money to invest], [old or new, auto] eh [fs] it's a [fs] [automobile or truck], it's (a personal decision).*

(L-INGE-3)

In 17w->17s, there is a case of change of specifications. The specifications provided in 17w are in a different clause. In 17s the term that needs glossing and the specifications are in the same clause. These specifications are more of the “personal decision” kind than the ones in 17w; the one related to the location of a store relates to a company’s decision.

(18w) *In the branch of mathematics called (model theory) we study theories and their properties (from logic point view.)*

(18s) *The (model theory) is a large nice branch of mathematics where [fs] where we study the theories but specifically a theory is eh the following: in model theory we work with a language [that is a set of samples like connectors, logical symbols, operation symbols, relation symbols, eh and so on.]*

(H-MATE-1)

Both 18w and 18s have the same specification: *how model theory studies theories; from logic point of view* (sic). However, the way the specification is made in 18s is arguably closer to the audience’s knowledge and trying to avoid other nuances for the meaning of logic. The specification in 18s refers to the *language* of the field of knowledge we know as *logic*; this language is composed of symbols, connectors, and the like.

The analysis of clause structure changes and inclusion of code glosses in the disciplinary divide allows me to draw two conclusions. First, it is not surprising that these two mechanisms are more frequent in the hard disciplines. As I observed in the code glosses chapter in this thesis, although explanations in the form of definitions can be seen in the hard and soft disciplines, many definitions in the soft fields are related to how a known term must be interpreted in a different way. In the hard disciplines, most definitions are definitions of technical terms that people don't know or are not familiar with (e.g. *Puccinia graminis*). Similarly, the higher the technical knowledge implied in the OPs, it is expected that the use of other code glosses (paraphrases, specifications) is required. Of course, this is not something that is exclusive to the hard disciplines; soft disciplines can also have this type of high elaboration, but it can be argued that a good deal of terms and contents in social sciences and the humanities are more accessible to general audiences. Second, in relation to clause structuring principles and their higher frequency in the hard fields, the need to simplify information for the non-expert audience plays an important role. As it was shown in several of the examples for code glosses, their inclusion was usually accompanied by the use of rhematization and thematization. This was more clearly observed in the modification of the definiendum-definiens structure to guarantee that a term in the definition was close to the same term in the previous clause. In other words, the high frequency of code glosses and clause structure changes in the OPs is due to a combination of disciplinary concerns (sophistication of terms and knowledge) and the non-expert character of the audience. Based on these findings, I interpreted the use of information principles to be due to epistemological differences between the hard (supposedly more rigorous in organizational terms) and soft disciplines. Also, this interpretation was based on anecdotal accounts that students from the hard-disciplines usually learn and apply the basics of English academic writing more than

their soft-discipline counterparts. However, a study (S. North, 2005) comparing the use of these principles in essay writing contrasting hard and soft discipline students arrives to the opposite conclusion.

8.6. Conclusion

This chapter has presented a follow-up study on the mechanisms that Colombian PhD researchers use to transform originally written content into the oral mode. The following were the questions that guided the new study.

1. What are the differences between the written and the oral versions of the same content produced by students in this class as observed in four mechanisms of change and their related submechanisms?
 - Change of clause structure
 - Reduction of heavily modified NPs
 - Changes of expression of modality
 - Inclusion of code glosses
2. What quantitative and qualitative differences are there between high, medium, and low rated OPs?
3. What quantitative and qualitative differences are there between hard-field and soft-field OPs?

The following is a summary of the answers.

General distribution: frequency analyses showed that the four mechanisms were similarly distributed in the corpora. However, in the level of achievement divide, high achievers used almost twice as much modifications as medium and low achievers, partially confirming the

findings in the first study. In the disciplinary divide, the distribution between hard and soft-discipline subcorpora was similar.

Frequency differences in the subcorpora: statistical significance analyses showed that in the level of achievement subcorpora heavy NP reduction and modality change mechanisms exhibited the highest frequency difference significance values while the other two mechanisms, information structure changes and inclusion of code glosses, exhibited higher values in the disciplinary divide.

Differences in the level of achievement subcorpus: again, high achievers demonstrated that they more consistently and frequently used heavy NP reduction and modality expression change mechanisms. In NP reductions, three submechanisms were salient: elimination or change of modifiers, and denominalisation. In the expression of modality change, the most frequent submechanisms were change of type and orientation (subjective, explicit or both). In the NP reduction and modalization cases, it was common to find that successful transition to the oral mode implied the use of several mechanisms simultaneously without (or just a few) grammar mistakes. Pragmatically speaking, the changes were explained based on simplification of information principles and engagement. Heavy NP reduction and the denominalisation of grammatical metaphors are understood as strategies for removing information that could hinder comprehension by making information units too heavy to process. The change of modality expression was interpreted as a way to engage the audience. When presenters mark contents as belonging to themselves or somebody else, they project a specific authorial stance identity and assign one to the audience.

Differences in the disciplinary subcorpus: This aspect was not considered in the first study. The mechanisms of change that exhibited the highest frequency difference values were

SVO clause structure modification and inclusion of code glosses. Interestingly, these two mechanisms were found to be more frequent in the hard disciplines subcorpus. The most frequent SVO change mechanisms were thematization and rhematization. The most recurrent code glosses were explanations (definitions), paraphrase, and specifications. Interestingly, several of the examples of code gloss inclusion were accompanied by cases of rhematization or denominalization. The inclusion of these mechanisms was explained to be due to information structuring and clarification concerns. As discussed, terminology and some aspects of knowledge are arguably more sophisticated in the hard sciences. Therefore, the presentation of highly sophisticated knowledge to a non-expert audience implies the definition of terms (explanation), specification of referents, and probably the iteration of introduced terms in adjacent clauses.

Table 8.5: Summary of findings in written to oral transition study

Comparison	Mechanisms (Log likelihood >3.84 $p < 0.05$)	Sub- mechanisms	Overuse (↑) or underuse (↓)	Discourse functions and infelicities*
Level of oral achievement (high, medium, low)	NP changes (LL: 5.38)	Modifier elimination Denominalization	High rated OPs (↑)	<ul style="list-style-type: none"> - Elimination of potentially difficult to understand info - Simplification of info - Grammar errors / few mechanisms not motivated by information structure or simplification principles*
	Modality (LL: 12.34)	Objective to subjective Implicit to explicit Types	Low rated OPs (↓)	<ul style="list-style-type: none"> - Aversion or attribution - Simplification of info - Grammar errors / few mechanisms not motivated by information structure or simplification principles* - Not expected transition (subjective to objective)*
Disciplinary divide (hard vs soft)	Clause structure changes (LL: 8.06)	Rhematization Thematization	Hard- field OPs (↑)	<ul style="list-style-type: none"> - Organizing discourse GIVEN NEW fashion - Guiding the audience - Combined with code glosses
	Code glosses not in essay (LL: 4.51)	Paraphrase Explanation Specification	Hard- field OPs (↑)	<ul style="list-style-type: none"> - Facilitate understanding - Combined with rhematization and denominalization

CHAPTER 9

CONCLUSION

9.1. Summary of findings

This thesis has demonstrated the importance of the (1) level of oral performance, (2) knowledge of disciplinary rhetorical practices, and (3) type of audience in the diversity and variation of the language strategies that a group of Colombian PhD researchers selected when giving oral presentations in an EAP class. The following are the questions that have guided this thesis:

QUESTIONS		ASPECTS (PARTS)		FEATURES (CHAPTERS)
1. What are the characteristics of the language that Colombian PhD researchers use in their OPs to		I. engage the audience	<i>as observed in their use of</i>	1. spatial and gestural deixis, 2. <i>you</i> to assign the audience an identity, and 3. modalized impersonal constructions?
VARIABLES	<i>in the way that they</i>			
2. What are the differences between high, medium and low-rated OPs 3. What are the differences between hard and soft-field OPs		II. make content easy for their audience		4. code glosses and 5. mechanisms to translate written content into the oral mode?

To present a summary of the findings³³, I will describe what this PhD researcher population does to engage and make content easy for the audience (the two main parts of the thesis). These descriptions will be organized by level of achievement and disciplines (the variables used in the analyses).

9.1.1. How presenters engaged the audience

9.1.1.1. By level of achievement

When interaction with the audience is analysed from the perspective of students' level of oral achievement, it is evident that the higher the level of oral achievement, the more pragmatically relevant and grammatically versatile performances are. This is observed in how students assign academic identities to the audience and themselves. Students use *you* to address the audience, construing them as *tourists* and *innovation users*. With the first, they use gestural and verbal deixis to explain images as if the audience were on a tour of the presenter's academic production; with the second, they claim ownership over academic production by construing the audience as potential users of their findings. When it comes to expressing their opinions or when they have to attribute knowledge to others, high achievers prefer impersonal modalized constructions and by default construe the audience as *opinion-evaluators*. Being able to clearly mark ownership over generated knowledge makes presenters look like true *representative-connoisseurs* of their fields. In performing these functions, the difference among the levels is a

³³ More specific summaries (tables included) have been provided in the conclusion section of each chapter.

matter of degree marked by the level itself. For example, high achievers not only exhibit more instances of stance positioning and interaction with the audience, but grammatically speaking their utterances are more varied, use more linguistic resources, and exhibit fewer errors. Low achievers, on the contrary, do not mark stance as frequently, barely interact with the audience and their sentences tend to be not as varied and exhibit more lexicogrammatical errors.

No significant quantitative or qualitative differences were found among achievers in the way they show and explain images to their audience.

9.1.1.2. By discipline

In the disciplinary divide, interaction with the audience reflects the ways that knowledge is (re)produced in the disciplines.

In hard disciplines, methods to generate knowledge determine the problems to be researched and research outcomes are more important than interpretations (Becher & Trowler, 2001). In hard-field OPs, images are used as evidence of claims; therefore, it is common that presenters devote more deictic resources to interacting with them than with the audience. The importance of methods over interpretations is reflected in the identities that hard-field researchers assign to the audience and themselves. When they use *you* in their OPs, they construe the audience as *research-apprentices* and focus on the procedures and tools needed to generate knowledge. With the use of impersonal modalized constructions, they construe themselves as *recouters-announcers* of methods, emphasising the need/urge/advisability to follow certain procedures or the level of difficulty they imply.

Soft-field presenters' images, on the other hand, tend to play a more illustrative than evidential role (Diani, 2015; Rowley-Jolivet, 2002); as a result, semiotic resources are focused

more onto interacting with the audience. The importance of interpretations over methods in soft-disciplines is observed in soft-field presenters' projection of academic selves, in which they prefer the expression of opinions and the co-construction of knowledge with the audience. In the projection of *co-constructer* identity, soft-fielders use *you* to address the audience directly as equals respecting their PhD researcher status, and not underestimating their knowledge of information that can be either general or specialized. They might use impersonal modalized constructions to express opinions about disciplinary knowledge.

9.1.2. How presenters make content easy for the audience

Like the differences in audience engagement, the strategies to facilitate content that presenters use are also correlated with their level of achievement and disciplines.

9.1.2.1. By level of achievement

When giving information that might confuse the audience, high achievers use more reformulations and examples. When they recycle content from their writings they reduce heavily modified NPs and express modality in more congruent, less metaphorical, ways. Again, this reflects not only that they have more language resources at their disposal but also that their OPs are made with the audience in mind. Mastery over content and language resources seems to free cognitive space to accommodate language resources to the audience's needs. The lower the level of achievement, the fewer changes to content there are, the fewer types of resources are used, and the more likely grammar and performance errors are to occur.

9.1.2.2. By discipline

When hard-discipline students anticipate moments of confusion regarding knowledge in their fields, they use explanations to illustrate technical vocabulary. Explanations (understood as definitions or provision of new terms) are accompanied by information structuring mechanisms (rhematization, thematization) in which new words are iterated as GIVEN or new nuances of meaning are put in RHEME position.

Soft-field students also anticipate moments of confusion regarding knowledge in their fields. However, unlike their hard-field counterparts, they do not seem to have to explain as much technical vocabulary to the audience. They focus more on clearly marking stance and creating an atmosphere of collegiality as observed in their mechanisms to express modality.

9.2. Implications

These findings have methodological, theoretical, and pedagogical implications.

9.2.1. Methodological

This thesis contributes to the study of oral academic discourse by demonstrating the importance of (1) multimodal corpus analyses, (2) register/mode comparisons, (3) a non-deficiency approach, and by refining and expanding (4) academic identity projection analysis.

First, the multimodal corpus analysis described in the deictics chapter integrates inferential statistics, corpus analyses of closed-class keywords (Groom, 2010) (spatial deictics), and the gestural analysis of those words. A genre that relies on the use of several semiotic modes should also be described in terms of the orchestration of resources to express meanings. One difficulty that is often reported in multimodality studies is how time consuming the analysis of

non-verbal aspects can be. In this thesis, the previous identification of deictics and the implementation of statistical tests to determine what cases to analyse reduced the time that could have been spent in the analyses of hours of video and allowed me to select representative cases from the corpus. The subsequent analysis of concordance lines and specific video segments was useful in the identification and description of how researchers from different disciplines talk about and interact with images. An additional contribution in the way I approached multimodality was the inclusion of sequences of pictures that were key in the description of the deictic process (Rendle-Short, 2006). This procedure was useful to identify the three moments in the deictic process and to demonstrate that the way presenters explain images to their audience is more ideational (more focused on explaining the information on the image) in hard-field OPs and more interactional (more focused on interacting with the audience while showing the image) in soft-field OPs.

Second, this thesis also demonstrates the importance of comparing written and oral discourse by the same speakers (chapter 8) in genre analysis and in the description of oral performance differences. The contrast of a given genre (OPs) to other related ones (essays) can be useful in the identification of traits that make that genre a type of its own. This identification could arguably be more solid when the comparisons are made with a focus on the differences and guaranteeing that the similarities of various aspects are controlled (e.g. purpose, producer, content). Studies like Carter-Thomas & Rowley-Jolivet's (2001), compare oral and written texts by the same speakers/writers and with similar intents, but they do not select sentences expressing the same propositions to identify mechanisms to transition between modes of discourse. Although time consuming and not generating of evidence in high quantities, the identification of transition mechanisms by the same speaker was useful in the identification of mechanisms not

reported in similar studies and in the definition of marks to discriminate levels of oral achievement.

Third, this thesis also demonstrates that NNS academic discourse analyses do not necessarily need to be based on comparisons to NS performance as the desired target; NNS academic discourse can be analysed as ELF (English as a Lingua Franca). Firstly, comparisons of native and not native speakers' performance with a focus on the gaps distinguishing their production (deficiency model) do not really capture essential aspects in the study of academic discourse like the influence that disciplinary epistemologies or institutional practices can have over language use as it has been demonstrated in the different studies in this thesis. Secondly, given the more pervasive presence of NNSs in the global academic world and the pervasive role of English as the language of academia, descriptions of NNS written and oral discourses are key in gaining a deeper understanding of their uses of English for the dissemination of knowledge. Other authors go further and add that ELF, or the language of international communication, should be a variety in itself, belonging to its speakers, and a legitimate learning target (Mauranen, 2003, p. 517). Thirdly, when NNS EAP students like the ones in this research are presenting their work once they graduate, the role of ELF will be evident, for they will most probably be giving OPs to other NNS of English, making the model of the native speaker become increasingly irrelevant. These are some of the reasons this thesis treats deficiencies in comparisons of levels of oral achievement as aspects that could have affected engagement or clarity in the OP and not necessarily as language deficiencies or deviations from NS usages.

Finally, in relation to academic identity studies, the studies in chapters 5 and 6, consolidate, refine, and expand Tang and John's (1999) seminal academic roles taxonomy. This taxonomy's continuum of stance taking roles is useful in the description of academic language

behaviour in a functional and lexicogrammatical fashion. New contributed aspects like the clear definition of differences between the *guide* and *architect* roles; the new conceptualization of *representative* as *connoisseur* role; the specification of lexicogrammar personal (1st and 2nd person pronouns) and impersonal language choices, along with their associated patterns and discourse functions for each identity role, can be useful for those interested in the study of authorial stance or academic identity projection in discourse analysis.

9.2.2. Theoretical

At the theoretical level, the thesis also contributes to the analysis of oral academic discourse as genre analysis.

First, the way I have approached OPs considers aspects of variation according to use: mode, tenor, and field (Halliday & Hasan, 1989) in the definition of OPs as a PhD researcher public speaking training genre. Apart from the obvious mode (oral vs written, content) considerations, I have investigated the PhD OP as a genre that has a very particular kind of audience and I have demonstrated the audience's effect on what might be called the *tenor* (Halliday & Hasan, 1989) of the OP. The relationship among the participants in the PhD researcher OP is two-fold. On the one hand, the relationship is of equals, as both presenter and audience are PhD students. On the other, it is hierarchical in that the presenters have more expert knowledge than the multi-disciplinary audience does. The interplay between these two facets of the relationship is reflected in the *field* as well because of the extent to which the speakers change the content of their OPs (e.g. inclusion of simple examples or reference to common knowledge) as well as the identity roles assigned to themselves and their audience. In this sense, my approach to OPs resembles lecture studies as they tend to consider more the role of the

audience in terms of how difficult or easy it is for students to understand content based on the ways that lectures use English.

In addition to the aspects of register variation in my approach to this oral genre, I have also taken very seriously the notion of the OP as a multimodal genre as other oral discourse analysts urge (Adolphs, 2012; Carter-Thomas & Rowley-Jolivet, 2003; Charles & Ventola, 2002; Poyatos, 2002). I have investigated how deixis is performed gesturally in terms of how different nonverbal resources are orchestrated to show images to the audience. This analysis also considers the type of images being used, which in turn also reflects epistemological differences across disciplines. Deixis is also approached linguistically in terms of frequencies of use of spatial deictics and phraseology.

A third line of contribution of this thesis is how it helps (1) to confirm the findings of other discourse analysis studies adopting the hard-soft disciplinary divide (Becher & Trowler, 2001) variable as reference and (2) to add further arguments to this way of analysing academic discourse. Among the aspects that this thesis confirms are the use of images and how it affects the performance of deixis (Charles & Ventola, 2002; Dubois, 1980; Rowley-Jolivet, 2002) and the disciplinary distribution of deictics (Simpson-Vlach, 2006). The identity studies in chapters 5 and 6, as well as the findings in module 2 (Nausa, 2016), show how the epistemological practices described by Becher & Trowler (2001) (*Figure 6.3.*) influence these students' language choices to attribute or aver knowledge, or to explain technical content, regardless of their NNS status or level of oral achievement. These findings are not only in line with disciplinary discourse studies but also with pedagogical constructivist approaches to the study of language learning that value the knowledge that students bring to the classroom and do not assume that they are empty vessels to be filled with knowledge. The use of binary variables like hard-soft and pure-applied can be

criticized on the ground of their apparent absolutist character, but the uses of this group of NNS shows that these are not necessarily mistaken.

This thesis has also demonstrated that level difference ideas about what makes one level of achievement different from another go beyond the reductionist concept of *getting the grammar right*. In 4 out of the 5 studies (not the deixis study), I have found differences between the levels related to the way students use language resources to interact with and clarify content for the audience. When I look at the differences between levels, my studies raise the question to what extent the differences in levels are determined by grammatical accuracy and to what extent that is irrelevant. I have not studied grammatical accuracy directly; I have, nonetheless, demonstrated that even by not studying that directly, I can still find the differences between the levels. On the other hand, I have also found throughout that differences in things like pragmatic effectiveness actually go along with accuracy (see for example the use of *in particular* in examples 20 and 21; *like* examples 24-26 in chapter 7) or the ability to orchestrate and use different grammatical mechanisms accurately in the written and oral mode (see 7w->7s and 8w->8s in chapter 8). In these examples, grammatical errors made by medium and low achievers are analysed in terms of how their occurrence might have affected comprehension by the audience and not in terms of deviations from NS standards.

9.2.3. Pedagogical

The findings in the studies also have pedagogical applications. A teacher could take any of the findings and use them as the basis for advice to students or to design class activities. Here I will suggest three ways in which findings could be employed in the EAP classroom public speaking instruction.

A first piece of advice is related to the importance of the audience. Teachers can recommend and teach students to think of the audience and plan the things they need (e.g. images, definitions, explanations) based on whether their audience is composed of people with higher, similar, or lower levels of disciplinary expertise. Non-expert audiences, for example, could require the definition of terms, real life examples, and the continuous iteration of key information as GIVEN in new clauses; more expert audiences could require the careful planning of resources to mitigate discourse, or attribute or aver knowledge; any audience, no matter their level of expertise, could require the presenter's selection of multimodal resources to explain images on slides.

Another line of recommendations comes in the form of the specific things that high and low achievers do. The things that high achievers do could be directly taught to students and include: engaging the audience by using examples or references they can relate to; directly addressing them with the second person pronoun assigning roles to them like *research apprentice* or *innovation user*, claiming or attributing knowledge to others with impersonal constructions, among others. Additionally, samples of the things that are unsuccessfully done can be given to students for them to identify what went wrong and to propose alternative wordings or ways of structuring information.

Finally, students could also benefit of doing practice of rewriting texts for speech. Teachers can teach students to modify sentences they have created for written exercises and translate them to be presented orally. In addition to changes to vocabulary, students can include the changes to the SVO structure to guarantee cohesive flow among clauses, elimination of NP modifiers to reduce NP size, change of parts of speech to make sentences congruent (not grammatically metaphorical), inclusion of human subjects (e.g. pronouns) in sentences, and the

like. This type of instruction could develop the competence of grammars for speaking and writing and eventually build the foundations for different academic genres/registers competence.

9.3. Limitations and further research

Of course, the studies reported in this thesis are not without limitations and difficulties that should be overcome in future academic OPs studies. The following is a brief account of these limitations and suggestions for how to overcome them. Future research is proposed as quantitative and qualitative studies.

At the methodological level, limitations comprise the interrater reliability procedures in corpus design and sentence selection, the size of the corpus, and the inclusion of other relevant variables in statistical analyses.

The studies in the thesis could have implemented stricter interrater reliability procedures in corpus creation and revision and sentence classification procedures. Transcriptions for the oral corpus were revised, discussed, and edited but interrater agreement was not calculated in the editing process. In future updates of the corpus, the old transcripts could be subject of conjoined transcription accuracy analysis and the new ones could be done including the interrater tests and procedures proposed. Similarly, two of the studies (chapters 4 and 8) lacked interrater reliability procedures to guarantee the validity of classification of corpus sentences in the description of multimodal deixis or the written oral transition strategies; therefore, there is a risk that my interpretations could have been biased by my roles as instructor and researcher. The experience gained in the training of colleagues for rating sessions and in the rating sessions themselves can be used as the point of departure for the optimization of future rating procedures to guarantee that future studies fully comply with this validity requirement.

Another methodological limitation is the relatively small size of the corpus at least for two reasons. First, some effect size analyses suggested the need for a bigger corpus, which was concluded when significance tests values (LL) were positive and high, but effect size values (BIC) were negative (see Table 7.13 for an example). Second, the studies used three variables in the analyses of students' oral discourse (levels, disciplines, and modes), but others like previous experience giving oral presentations or studying English, level of English as measured by standardized tests, genre, socioeconomic status, attitudes towards speaking (English) in public, and others were not included. These language learning-use and demographic variables have been considered in other oral academic discourse studies with similar populations and have been marginally and informally observed in analyses procedures not reported in this study. It is my intuition that these factors might have had an impact on students' performance in OPs and could have been useful in explaining variation. However, given the space and scope limitations of my studies, the decision was made to exclude them.

Overcoming these corpus size limitations could open new possibilities in terms of quantitative future research. A bigger corpus and the use of the mentioned language learning-use and demographic variables would allow for the implementation of more sophisticated statistical analysis like MANOVA or regression analyses that would in turn allow for analyses with the control of variables or with the examination of the interaction of variables in the explanation of language behaviour.

Possibilities for future research in qualitative terms are also possible based on some of the conclusions and implication of the thesis.

One of the main conclusions is the effect of the (multi-department non-expert) audience on the OPs' *tenor* and *field* as observed in the presenters' language choices. Testing the findings

in this thesis with different expert audiences (colleagues or supervisors) would be useful to test some of the conclusions that I have drawn. This would help to clarify the extent to which the aspects emerging in the analysis are a consequence of the type of audience or if the presence of any audience would have the same types of consequences.

Another important conclusion is the correlation between level of oral achievement and the pragmatic relevance and grammar variety of language choices. However, it is not clear how much OP achievement depends on how well a student speaks English or the extent to which OP achievement depends on other things (rehearsing, use of images and carefully planned slides, previous experience giving OPs, confidence, etc). Therefore, future studies should aim at testing the findings with the same population (IPD2 former students) at more advanced levels (IPD3, IPD4) or with students who are placed in those levels with our in-house placement test. Similarly, the findings could be tested against individuals who exhibit the traits specified in the language learning-use and demographic variables mentioned above.

The thesis has also demonstrated the correlation between disciplinary knowledge and rhetorical practices, and presenters' language choices. However, specific discipline differences are not explored. The hard-soft distinction does not capture specific situations in which disciplines can exhibit traits of both sides of the split. For example, students in management (classified as a soft discipline) use statistical procedures, also typical in most hard-field disciplines. It would be interesting to test the conclusions in this thesis with studies focusing on specific disciplines (e.g. anthropology, mathematics).

Finally and based on the pedagogical implications of the thesis, It would also be interesting to know whether the interventions proposed (9.2.3) would make a measurable difference to students' oral presentation competence.

9.4. Concluding remarks

This investigation of the language of oral presentations given by Colombian PhD researchers using diverse approaches to genre analysis and corpus linguistics is a pioneer study in the way that it approaches PhD OPs as a genre that is defined by the public speaking needs of presenters and the multi-department character of the audience. This thesis demonstrates that although the language used by presenters in their OPs may seem influenced by their general speaking ability, the knowledge and rhetorical conventions of their disciplines and the non-expert character of their audience also play an important role. It is my hope that the studies reported in this thesis contribute to the existing literature on spoken academic genres and towards studies of NNS academic discourses and disciplinary differences, which provide EAP instructors with ideas to reflect on their practice, and a point of reference from which to help their students develop their oral presentation competence.

References

- Acreditación, C. N. de. (2008). *Situación actual de los doctorados en Colombia: Análisis de indicadores que tipifican características importantes*. Bogotá D.C. Retrieved from https://www.cna.gov.co/1741/articles-186363_SitDoc_Ind.pdf
- Adolphs, S. (2012). Corpora: Multimodal. *The Encyclopedia of Applied Linguistics*, 2006–2009. <https://doi.org/10.1002/9781405198431.wbeal0233>
- Ágnes, M. G. (2002). Are you with me? A metadiscursive analysis of interactive strategies in college students' course presentations. *Journal of English Studies*, 2(1), 55–78. <https://doi.org/10.6018/ijes.12.1.118281>
- Aguilar, M., & Arnó, E. (2002). Metadiscourse in lecture comprehension: Does it really help foreign language learners? *Atlantis*, 24(2), 3–21.
- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 55–76.
- Al-Issa, A. S., & Al-Qubtan, R. (2010). Taking the floor: Oral presentations in EFL classrooms. *TESOL Journal*, 1(2), 227–246. <https://doi.org/10.5054/tj.2010.220425>
- Alessi, G. (2005). The use of metadiscourse in EAP presentations by native Italian speakers. *Dialogue within Discourse Communities: Metadiscursive Perspectives on Academic Genres*, 28, 179.
- Alsop, S., Moreton, E., & Nesi, H. (2013). The uses of storytelling in university Engineering lectures. *ESP across Cultures*, 10, 7–20.
- Alwi, N. F. B., & Sidhu, G. K. (2013). Oral presentation: Self-perceived competence and actual performance among UiTM business faculty students. *Procedia - Social and Behavioral*

- Sciences*, 90(InCULT 2012), 98–106. <https://doi.org/10.1016/j.sbspro.2013.07.070>
- Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3), 156–158. <https://doi.org/10.1002/da.20090>
- Andeweg, B. A., de Jong, J. C., & Hoeken, H. (1998). “May I have your attention?”: Exordial techniques in informative oral presentations. *Technical Communication Quarterly*, 7(3), 271–284. <https://doi.org/10.1080/10572259809364631>
- Anthony, L. (2014). AntConc (Version 3.4.4)[windows]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Anthony, L. (2018a). AntFileConverter (Version 1.2.1)[Windows]. Tokyo, Japan, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Anthony, L. (2018b). TagAnt (Version 1.2.0)[Windows]. Tokyo, Japan, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Anthony, L., Orr, T., & Yamazaki, A. K. (2007). Signaling transitions in oral presentations: Language and strategies. *IEEE International Professional Communication Conference*, (November). <https://doi.org/10.1109/IPCC.2007.4464052>
- Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication*, 31(2), 151–183. <https://doi.org/10.1177/0741088314527055>
- Austin, J. L. (1962). How to do things with words: The William James lectures. *Cambridge, MA*.
- Bamford, J. (2004). Gestural and symbolic uses of the deictic "here" in academic lectures. In K. Aijmer & A.-B. Stenström (Eds.), *Discourse Patterns in Spoken and Written Corpora* (pp. 113–138). Amsterdam / Philadelphia: John Benjamins Publishing.

- Bankowski, E. (2010). Developing skills for effective academic presentations in EAP. *International Journal of Teaching and Learning in Higher Education*, 22(2), 187–196.
Retrieved from <http://www.isetl.org/ijtlhe/>
- Barnbrook, G. (2002). *Defining language : A local grammar of definition sentences*. Amsterdam / Philadelphia: John Benjamins Publishing.
- Barrett, N. E., & Liu, G.-Z. (2016). Global trends and research aims for English academic oral presentations: Changes, challenges, and opportunities for learning technology. *Review of Educational Research*, 86(4), 1227–1271. <https://doi.org/10.3102/0034654316628296>
- Barton, E. L. (1993). Evidentials, argumentation, and epistemological stance. *College English*, 55(7), 745. <https://doi.org/10.2307/378428>
- Basturkmen, H., & Randow, J. Von. (2014). Guiding the reader (or not) to re-create coherence : Observations on postgraduate student writing in an academic argumentative writing task. *Journal of English for Academic Purposes*, 16, 14–22.
<https://doi.org/10.1016/j.jeap.2014.07.005>
- Bayless, M. L. (2004). Change the placement, the pace, and the preparation for the oral presentation. *Business Communication Quarterly*, 67(2), 222–225.
<https://doi.org/10.1177/1080569904672011>
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: The Society for Research into Higher Education & Open University Press.
- Beebe, S. A., & Beebe, S. J. (2015). *A concise public speaking handbook* (4th ed.). London: Pearson.
- Bellés, B., & Fortanet, I. (2006). Discourse markers within the university lecture genre: A

- contrastive study between Spanish and North American lectures. *English Studies Department, PhD*, 333.
- Bertin, J. (1973). *Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes* (2nd ed.). Paris/La Haye: Gauthier-Villars.
- Bhatia, V. (2002). *A generic view of academic discourse. Academic Discourse*. Harlow: Longman.
- Bhattacharyya, E., & Shaari, Z. H. (2012). Communicative competence in technical oral presentation: That “Magic” perceived by ESL educators versus content experts. *World Academy of Science, Engineering and Technology International Journal of Cognitive and Language Sciences*, 6(12), 3540–3543. Retrieved from <http://waset.org/publications/6526>
- Biber, D. (2002). *University language: A corpus-based study of spoken and written registers*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116. <https://doi.org/10.1016/j.jeap.2006.05.001>
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press. <https://doi.org/10.1075/jhp.13.1.07leh>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. <https://doi.org/10.1016/j.jeap.2010.01.001>
- Biber, D., Grieve, J., & Iberri-Shea, G. (2009). Noun phrase modification. In Günter Rohdenburg

- & J. Schlüter (Eds.), *One Language, Two Grammars?: Differences between British and American English* (pp. 182–193). Cambridge: Cambridge University Press.
- <https://doi.org/10.1017/CBO9780511551970.010>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: London: Longman.
- Bondi, M. (2005). Metadiscursive practices in academic discourse: Variation across genres and disciplines. In J. Bamford & M. Bondi (Eds.), *Dialogue within discourse communities: Metadiscursive perspectives on academic genres* (pp. 3–30). Kempten: De Gruyter.
- Boyd, F. A. (1989). Developing presentation skills: A perspective derived from professional education. *English for Specific Purposes*, 8(2), 195–203. [https://doi.org/10.1016/0889-4906\(89\)90030-6](https://doi.org/10.1016/0889-4906(89)90030-6)
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.
- Brown, T., & Morrissey, L. (2004). The effectiveness of verbal self-guidance as a transfer of training intervention: Its impact on presentation performance, self efficacy and anxiety. *Innovations in Education and Teaching International*, 41(3), 255–271.
- <https://doi.org/10.1080/14703290410001733302>
- Bruce, I. (2016). Constructing critical stance in university essays in English literature and sociology. *English for Specific Purposes*, 42, 13–25.
- <https://doi.org/10.1016/j.esp.2015.10.005>
- Bu, J. (2014). Towards a pragmatic analysis of metadiscourse in academic lectures: From relevance to adaptation. *Discourse Studies*, 16(4), 449–472.
- <https://doi.org/10.1177/1461445613519019>

- Bühler, K. (1965). *Sprachtheorie: die Darstellungsfunktion der Sprache* (2nd ed.). Stuttgart: Gustav Fischer.
- Bunton, D. (1999). The use of higher level metatext in Ph.D theses. *English for Specific Purposes*, 18, S41–S46.
- Burgess, K. (2013). Speaking in public is worse than death for most. *The Times*. Retrieved from <https://www.thetimes.co.uk/article/speaking-in-public-is-worse-than-death-for-most-5l2bvqlmbnt>
- Calcich, S. E., & Weilbaker, D. C. (1992). Selecting the optimum number of in-class sales presentations. *Marketing Education Review*, 2(1), 31–33.
<https://doi.org/10.1080/10528008.1992.11488349>
- Campbell, K. S., Mothersbaugh, D. L., Brammer, C., & Taylor, T. (2001). Peer versus self assessment of oral business presentation performance. *Business Communication Quarterly*, 64(3), 23–40. <https://doi.org/10.1177/108056990106400303>
- Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, 66, 15–31.
<https://doi.org/10.1016/j.pragma.2014.02.007>
- Carter-Thomas, S., & Rowley-Jolivet, E. (2001). Syntactic differences in oral and written scientific discourse: The role of information structure. *ASp*, 33(31–33), 19–37.
<https://doi.org/10.4000/asp.1752>
- Carter-Thomas, S., & Rowley-Jolivet, E. (2003). Analysing the scientific conference presentation (CP): A methodological overview of a multimodal genre. *ASp: La Revue Du GERAS*, 40(39–40), 59–72. <https://doi.org/10.4000/asp.1295>
- Castronova, E. (2013). Down with dullness: Gaming the academic conference. *The Information*

- Society*, 29(2), 66–70. <https://doi.org/10.1080/01972243.2012.757262>
- Chan, V. (2011). Teaching oral communication in undergraduate science: Are we doing enough and doing it right? *Journal of Learning Design*, 4(3), 71–79. Retrieved from <https://files.eric.ed.gov/fulltext/EJ940652.pdf>
- Chanock, K. (2002). Using systemic functional linguistics in university teaching: A postgraduate class on making writing more 'readable'. *Monash University Linguistics Papers*, 2(1), 61.
- Chanock, K. (2005). Scholarship of teaching and learning: Investigating patterns and possibilities in an academic oral genre. *Communication Education*, 54(1), 92–99. <https://doi.org/10.1080/03634520500077115>
- Charles, C., & Ventola, E. (2002). A multi-semiotic genre: The conference slide show. *The Language of Conferencing*, 169–209.
- Chen, L. Y. (2009). *A study of EFL graduate students' oral presentation anxiety*. National Chung Cheng University.
- Cheng, S. (2012). "That's it for today": Academic lecture closings and the impact of class size. *English for Specific Purposes*, 31(4), 234–248. <https://doi.org/10.1016/j.esp.2012.05.004>
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93–121. <https://doi.org/10.1191/0265532205lt298oa>
- Cheong, E. (n.d.). *Effect of Teaching Academic Spoken English Discourse Features on L2 Oral Performances*. Retrieved from http://nau.edu/uploadedFiles/Academic/CAL/PIE/_Forms/Teaching_academic_discourse.pdf
- Chou, M. (2011). The influence of learner strategies on oral presentations: A comparison between group and individual performance. *English for Specific Purposes*, 30(4), 272–285.

<https://doi.org/10.1016/j.esp.2011.04.003>

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*.

Cambridge: Cambridge University Press.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of “um.” *Language and Linguistics Compass*, 2(4), 589–602.

<https://doi.org/10.1111/j.1749-818x.2008.00068.x>

Crawford, B. (2015). Elaborating explanations during OpenCourseWare humanities lectures: The interplay of verbal and nonverbal strategies. In B. Crawford Camiciottol & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings* (pp. 144–170). New York, NY, NY: Routledge.

Cresswell, M. J. (1985). *Structured meanings: The semantics of propositional attitudes*. MIT Press.

Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. <https://doi.org/10.1016/j.jeap.2016.04.005>

Crosthwaite, P., & Jiang, K. (2017). Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System*, 69, 92–107. <https://doi.org/10.1016/J.SYSTEM.2017.06.010>

Csomas, E. (2007). A corpus-based look at linguistic variation in classroom interaction: Teacher talk versus student talk in American University classes. *Journal of English for Academic Purposes*, 6(4), 336–355. <https://doi.org/10.1016/j.jeap.2007.09.004>

Cuncic, A. (2017). How do I get over my fear of public speaking? Retrieved June 15, 2018, from <https://www.verywellmind.com/how-do-i-get-over-my-fear-of-public-speaking-3024827>

D’Angelo, L. D. (2011). Academic posters across disciplines : A preliminary study. *Language Studies Working Papers*, 3, 15–28. Retrieved from

https://www.reading.ac.uk/web/files/english-language-and-literature/ell_language_3_D_Angelo_vol_3.pdf

- Damron, J. A. (2000). *Chinese 101, a prerequisite to Math 100? A look at undergraduate students' beliefs about their role in communication with international teaching assistants*. Purdue University, West Lafayette, IN. Retrieved from <https://docs.lib.purdue.edu/dissertations/AAI3018186>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33(1), 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>
- De Grez, L. (2009). *Optimizing the instructional environment to learn presentation skills*. Universiteit Gent. Retrieved from <https://biblio.ugent.be/publication/7051516/file/7051518>
- De Grez, L., Valcke, & Berings. (2010a). Peer assessment of oral presentation skills. *Procedia - Social and Behavioral Sciences*, 2(2), 1776–1780. <https://doi.org/10.1016/J.SBSPRO.2010.03.983>
- De Grez, L., Valcke, L., & Berings. (2010b). Student response system and learning oral presentation skills. *Procedia - Social and Behavioral Sciences*, 2(2), 1786–1789. <https://doi.org/10.1016/J.SBSPRO.2010.03.985>
- De Grez, L., Valcke, M., & Roozen, I. (2009a). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education*, 53(1), 112–120. <https://doi.org/10.1016/J.COMPEDU.2009.01.005>
- De Grez, L., Valcke, M., & Roozen, I. (2009b). The impact of goal orientation, self-reflection and personal characteristics on the acquisition of oral presentation skills. *European Journal of Psychology of Education*, 24(3), 293–306. <https://doi.org/10.1007/BF03174762>
- Deroey, K. L. B. (2012). What they highlight is...: The discourse functions of basic wh-clefts in

- lectures. *Journal of English for Academic Purposes*, 11(2), 112–124.
<https://doi.org/10.1016/j.jeap.2011.10.002>
- Deroey, K. L. B., & Taverniers, M. (2011). A corpus-based study of lecture functions. *Moderna Sprak*, 105(2), 1–22.
- Devi, I. S., Amir, Z., & Krish, P. (2014). Engaging undergraduate engineers in oral presentations: A multimodal approach. *Mediterranean Journal of Social Sciences*, 5(27), 1510–1520.
<https://doi.org/10.5901/mjss.2014.v5n27p1510>
- Diani, G. (2015). Visual communication in applied linguistics conference presentations. In B. Crawford & I. Fortanet (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 83–107). New York: Routledge.
- Drubin, D. G., & Kellogg, D. R. (2012). English as the universal language of science: Opportunities and challenges. *Molecular Biology of the Cell*, 23(8), 1399.
<https://doi.org/10.1091/mbc.E12-02-0108>
- Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Dubois, B. Lou. (1980). The use of slides in biomedical speeches. *The ESP Journal*, 1(1), 45–50.
[https://doi.org/10.1016/0272-2380\(80\)90009-8](https://doi.org/10.1016/0272-2380(80)90009-8)
- Dubois, B. Lou. (1982). “And the last slide, please”: Regulatory language function at biomedical meetings. *World Language English*, 1(4), 263–268.
- Dubois, B. Lou. (1985). Poster sessions at biomedical meetings: Design and presentation. *The ESP Journal*, 4, 37–48.
- Efron, D. (1941). *Gesture and environment*. Oxford, England: King’s Crown Press.

- Eisenstein, J., & Davis, R. (2004). Visual and linguistic information in gesture classification. *Proceedings of the 6th International Conference on Multimodal Interfaces - ICMI '04*, 113. <https://doi.org/10.1145/1027933.1027954>
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1), 49–98. <https://doi.org/10.1515/semi.1969.1.1.49>
- Ervin-Tripp, S. (1976). Is Sybil there? The structure of some American English directives. *Language in Society*, 5(1), 25–66.
- Esposito, J. (2000). *Overcome your fear of public speaking and performance* (5th ed.). Bridgewater, CT: In The SpotLight, LLC.
- Evans, S. (2013). “Just wanna give you guys a bit of an update”: Insider perspectives on business presentations in hong kong. *English for Specific Purposes*, 32(4), 195–207. <https://doi.org/10.1016/j.esp.2013.05.003>
- Evison, J. (2010). What are the basics of analysing a corpus? In A. O’Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122–135). London: Routledge.
- Fallows, S., & Steven, C. (2000). Building employability skills into the higher education curriculum: A university-wide initiative. *Education + Training*, 42(2), 75–83. <https://doi.org/10.1108/00400910010331620>
- Fang, Z., Schleppegrell, M., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, 38(3), 247–273.
- Fernández-Polo, F. J. (2014). The role of “I mean” in conference presentations by ELF speakers. *English for Specific Purposes*, 34(1), 58–67. <https://doi.org/10.1016/j.esp.2013.09.006>
- Fernández-Polo, F. J. (2018). Functions of “you” in conference presentations. *English for Specific Purposes*, 49, 14–25. <https://doi.org/10.1016/j.esp.2017.10.001>

- Ferris, D. (1998). Students' views of academic aural/oral skills: A comparative needs analysis. *TESOL Quarterly*, 32(2), 289. <https://doi.org/10.2307/3587585>
- Ferris, D., & Tagg, T. (1996). Academic oral communication needs of EAP learners: What subject-matter instructors actually require. *TESOL Quarterly*, 30(1), 31. <https://doi.org/10.2307/3587606>
- Fillmore, C. J. (1997). *Lectures on deixis*. Stanford, Calif: CSLI Publications.
- Fitch, F., & Morgan, S. (2003). "Not a lick of English": Constructing the ITA identity through student narratives. *Communication Education*, 52(3), 297–310. <https://doi.org/10.1080/0363452032000156262>
- Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes*, 22(4), 329–346. [https://doi.org/10.1016/S0889-4906\(02\)00017-0](https://doi.org/10.1016/S0889-4906(02)00017-0)
- Flowerdew, J. (2013). English for research publication purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 301–321). Malden, MA: Wiley-Blackwell.
- Fortanet, I. (2005). Honoris Causa speeches: An approach to structure. *Discourse Studies*, 7(1), 31–51. <https://doi.org/10.1177/1461445605048766>
- Fortanet, I. (2006). Interaction in academic spoken English: the use of 'I' and 'you' in the MICASE. *Information Technology in Languages for Specific Purposes*, 35–51.
- Fortanet, I., & Ruiz-Madrid, N. (2016). Multimodal humor in plenary lectures in English and in Spanish. *Multimodal Communication*, 5(1), 55–69.
- Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of "you know" and "I mean." *Journal of Pragmatics*, 34(6), 727–747. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0378216602000279>

- Freedman, N., & Hoffman, S. P. (1967). Kinetic behavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Perceptual and Motor Skills*, 24(2), 527–539.
- Gardner, S. (2004). Knock-on effects of mode change on academic discourse. *Journal of English for Academic Purposes*, 3, 23–38. [https://doi.org/10.1016/S1475-1585\(03\)00052-3](https://doi.org/10.1016/S1475-1585(03)00052-3)
- Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics*, 34(1), 25–52. <https://doi.org/10.1093/applin/ams024>
- Gast, V., Deringer, L., Haas, F., & Rudolf, O. (2015). Impersonal uses of the second person singular: A pragmatic analysis of generalization and empathy effects. *Journal of Pragmatics*, 88, 148–162. <https://doi.org/10.1016/j.pragma.2014.12.009>
- Gedamu, A. D. (2016). In-service EFL trainees' attitude to oral presentation and assessment at Arba Minch University. *International Journal of Current Research*, 8(08), 37437–37442. Retrieved from <https://www.researchgate.net/publication/322245637>
- Gilhooly, D., & Lee, E. (2014). The role of digital literacy practices on refugee resettlement. *Journal of Adolescent & Adult Literacy*, 57(5), 387–396. <https://doi.org/10.1002/jaal.254>
- Gilquin, G., & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 418–435). Cambridge: Cambridge University Press.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335. <https://doi.org/10.1016/j.jeap.2007.09.007>
- Giménez, R. (2000). *La repetición lingüística en el género de la clase magistral: El inglés académico oral en el ámbito de las Ciencias Sociales*. Universitat de València.

- Girard, T., Pinar, M., & Trapp, P. (2011). An exploratory study of class presentations and peer evaluations: Do students perceive the benefits? *Academy of Educational Leadership Journal*, 15(1), 77–94.
- Gorsuch, G. J. (2003). The educational cultures of international teaching assistants and US universities. *Tesl-Ej*, 7(3), 1–17.
- Graddol, D. (2006). *English Next: Why global English may mean the end of “English as a foreign language.”* British Council. <https://doi.org/10.1075/lplp.32.1.08amm>
- Green, E. H., Hershman, W., DeCherrie, L., Greenwald, J., Torres-Finnerty, N., & Wahi-Gururaj, S. (2005). Developing and implementing universal guidelines for oral patient presentation skills. *Teaching and Learning in Medicine*, 17(3), 263–267.
https://doi.org/10.1207/s15328015t1m1703_11
- Griffiths, P. (2006). *Introduction to English semantics and pragmatics*. Edinburgh: Edinburgh University Press.
- Groom, N. (2010). Closed-class keywords and corpus-driven discourse analysis. In M. Bondi & M. Scott (Eds.), *Keyness in Texts* (pp. 59–78). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Groom, N., Charles, M., & John, S. (2015). Corpora, grammar, and discourse analysis: Recent trends, current challenges. In N. Groom, M. Charles, & S. John (Eds.), *Corpora, Grammar and Discourse: In honour of Susan Hunston* (pp. 1–19). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in Syntactic Typology*, 17, 209–239.
- Haas, M. R. (1969). “Exclusive” and “Inclusive”: A look at early usage. *International Journal of*

American Linguistics, 35(1), 1–6.

Haber, R. J., & Lingard, L. A. (2001). Learning oral presentation skills. *Journal of General Internal Medicine*, 16(5), 308–314. <https://doi.org/10.1046/j.1525-1497.2001.00233.x>

Halliday, M. A. K. (1989). Some grammatical problems in scientific English. *Australian Review of Applied Linguistics. Supplement Series*, 6(1), 13–37.

Halliday, M. A. K., & Hasan, R. (1989). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective* (2nd ed.). Oxford: Oxford University Press.

Halliday, M. A. K., & Martin, J. R. (2003). *Writing science: Literacy and discursive power*. London: Taylor & Francis.

Halliday, M. A. K., & Matthiessen, C. (2004). *An Introduction to functional grammar. Language* (3rd ed.). New York, NY: Arnold, Hodder Headline Group. <https://doi.org/10.2307/414947>

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar*. London: Routledge.

Hardie, A. (2014). Log Ratio – an informal introduction | ESRC Centre for Corpus Approaches to Social Science (CASS). Retrieved May 30, 2018, from <http://cass.lancs.ac.uk/?p=1133>

Harwood, N. (2005). “Nowhere has anyone attempted ... In this article I aim to do just that” A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, 37(8 SPEC. ISS.), 1207–1231. <https://doi.org/10.1016/j.pragma.2005.01.012>

Hay, I. (1994). Justifying and applying oral presentations in geographical education. *Journal of Geography in Higher Education*, 18(1), 43–55. <https://doi.org/10.1080/03098269408709236>

Hendriks, B., Van Meurs, F., & Reimer, A.-K. (2018). The evaluation of lecturers’ nonnative-

- accented English: Dutch and German students' evaluations of different degrees of Dutch-accented and German-accented English of lecturers in higher education. *Journal of English for Academic Purposes*, 34, 28–45. <https://doi.org/10.1016/j.jeap.2018.03.001>
- Hill, M., & Storey, A. (2003). SpeakEasy: Online support for oral presentation skills. *ELT Journal*, 57(4), 370–376. <https://doi.org/10.1093/elt/57.4.370>
- Hood, S., & Forey, G. (2005). Introducing a conference paper: Getting interpersonal with your audience. *Journal of English for Academic Purposes*, 4(4), 291–306. <https://doi.org/10.1016/j.jeap.2005.07.003>
- Hu, G., & Liu, Y. (2018). Three minute thesis presentations as an academic genre: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes*, 35, 16–30. <https://doi.org/10.1016/j.jeap.2018.06.004>
- Huang, L. S. (2010). *Academic communication skills: Conversation strategies for international graduate students*. New York: University Press of America.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2011). *Corpus approaches to evaluation*. New York, NY: Routledge.
- Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Hyland, K. (1996a). Talking to the academy: Forms of hedging in science research articles. *Written Communication*, 13(2), 251–281.
- Hyland, K. (1996b). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17(4), 433–454.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30, 437–455.

- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), 207–226. [https://doi.org/10.1016/s0889-4906\(00\)00012-0](https://doi.org/10.1016/s0889-4906(00)00012-0)
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.
- Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2), 266–285. <https://doi.org/10.1093/applin/amm011>
- Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum International Publishing Group.
- Hyland, K. (2011). Academic discourse. In K. Hyland & B. Paltridge (Eds.), *Continuum companion to discourse analysis* (pp. 171–184). London: Continuum International Publishing Group.
- Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge: Cambridge University Press.
- Hyland, K. (2017). Metadiscourse: What is it and where is it going? *Journal of Pragmatics*, 113, 16–29. <https://doi.org/10.1016/j.pragma.2017.03.007>
- Hyland, K., & Guinda, C. S. (2012). *Stance and voice in written academic genres*. Palgrave Macmillan. <https://doi.org/10.1057/9781137030825>
- Hyland, K., & Jiang, F. (Kevin). (2018). “In this paper we suggest”: Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, 51, 18–30. <https://doi.org/10.1016/J.ESP.2018.02.001>
- Hynninen, N. (2013). *Language regulation in English as a lingua franca*. Boston / Berlin: De Gruyter Mouton.
- Iberri-Shea, G. (2011). Speaking in front of the class: A multi-dimensional comparison of university student public speech and university language. *Classroom Discourse*, 2(2), 251–

267. <https://doi.org/10.1080/19463014.2011.614061>
- Innes, R. A. (2009). “Wait a second. Who are you Anyways?”: The insider/outsider debate and American Indian studies. *The American Indian Quarterly*, 33(4), 440–461.
<https://doi.org/10.1353/aiq.0.0069>
- Janssen, G., Ángel, C., & Nausa, R. (2011). *Informe de la investigación: El desarrollo de un currículo para la escritura de inglés nivel posgrado, según las necesidades y habilidades de los estudiantes (Proyecto IPD)*. Bogotá.
- Janssen, G., Nausa, R., & Rico, C. (2012). Shaping the ESP curriculum of an English for PhD students program: A Colombian case study of questionnaire research. *Colombian Applied Linguistics Journal*, 14(2), 123–4641. Retrieved from
<http://www.scielo.org.co/pdf/calj/v14n2/v14n2a05.pdf>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp. 13–31). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Jensen, K. K., & Harris, V. (1999). The public speaking portfolio. *Communication Education*, 48(3), 211–227. <https://doi.org/10.1080/03634529909379170>
- Jewitt, C. (2016). *The Routledge handbook of multimodal analysis*. London: Routledge.
- Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, (1), 69–78. Retrieved from http://www.uefap.com/baleap/quotes/jordan_jeap.pdf
- Kamio, A. (2001). English generic “we”, “you”, and “they”: An analysis in terms of territory of information. *Journal of Pragmatics*, 33(7), 1111–1124. [https://doi.org/10.1016/S0378-2166\(00\)00052-7](https://doi.org/10.1016/S0378-2166(00)00052-7)
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2

comprehensibility and accentedness. *System*, 38(2), 301–315.

<https://doi.org/10.1016/J.SYSTEM.2010.01.005>

Karam, M. (2006). *A framework for research and design of gesture-based human computer interactions*. University of Southampton. Retrieved from <http://eprints.soton.ac.uk>

Kaur, K., & Ali, A. M. (2017). Exploring the genre of academic oral presentations : A critical review. *International Journal of Applied Linguistics & English Literature*, 7(1), 152–162.

Keen, J. (2004). Sentence-combining and redrafting processes in the writing of secondary school students in the UK. *Linguistics and Education*, 15(1–2), 81–97.

<https://doi.org/10.1016/j.linged.2004.10.005>

Kendon, A. (1972). Some relationships between body motion and speech. *Studies in Dyadic Communication*, 7(177), 90.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Kay (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton Publishers.

Kerby, D., & Romine, J. (2009). Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business*, 85(3), 172–179. <https://doi.org/10.1080/08832320903252389>

Khedri, M., Heng, C. S., & Ebrahimi, S. F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies*, 15(3), 319–331. <https://doi.org/10.1177/1461445613480588>

Kim, S. (2006). Academic oral communication needs of East Asian international graduate students in non-science and non-engineering fields. *English for Specific Purposes*, 25(4), 479–489. <https://doi.org/10.1016/j.esp.2005.10.001>

- Kite, Y. (2008). Language use by engineers at an academic conference. *Kansai University for Human Activity Theory Research Center CHAT Technical Reports*, (7), 75–89.
- Knoblauch, H. (2008). The performance of knowledge: Pointing and knowledge in powerpoint presentations. *Cultural Sociology*, 2(1), 75–97. <https://doi.org/10.1177/1749975507086275>
- Kobayashi, M. (2006). Second language socialization through an oral project presentation: Japanese university students' experience. In G. H. Beckett & P. C. Miller (Eds.), *Project-based second and foreign language education: Past, present, and future* (pp. 71–93). Greenwich, Connecticut: Information Age Publishing.
- Köller, W. (2004). *Perspektivität und sprache: zur struktur von objektivierungsformen in bildern, im denken und in der sprache*. Berlin: Walter de Gruyter.
- Kress, G., & Van Leeuwen, T. V. (2001). *Multimodal discourse: The modes and media of contemporary communication* (1st ed.). London / New York: Arnold / Oxford University Press.
- Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. *Departmental Papers (Annenberg School for Communication)*, 1–10. Retrieved from http://repository.upenn.edu/asc_papers
- Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121–138. [https://doi.org/10.1016/s0889-4906\(97\)00058-6](https://doi.org/10.1016/s0889-4906(97)00058-6)
- Lafuente, E. (2010). ' Extending this claim , we propose ... ' The writer's presence in research articles from different disciplines. *Ibérica*, 20, 35–56.
- Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline- specific and general qualities. *Journal of English for Academic Purposes*, 23, 16–30.

<https://doi.org/10.1016/j.jeap.2016.05.006>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lee, J. (2009). Size matters: an exploratory comparison of small- and large-class university lecture introductions. *English for Specific Purposes*, 28(1), 42–57.

<https://doi.org/10.1016/j.esp.2008.11.001>

Lee, J. (2016). “There’s intentionality behind it ...”: A genre analysis of EAP classroom lessons. *Journal of English for Academic Purposes*, 23, 99–112.

<https://doi.org/10.1016/j.jeap.2015.12.007>

Lee, J., & Subtirelu, N. C. (2015). Metadiscourse in the classroom: A comparative analysis of EAP lessons and university lectures. *English for Specific Purposes*, 37(1), 52–62.

<https://doi.org/10.1016/j.esp.2014.06.005>

Lee, S. (2008). Attitude in undergraduate persuasive essays. *Prospect*, (23), 43–58. Retrieved from

<https://www.researchonline.mq.edu.au/vital/access/services/Download/mq:36521/DS01>

Leedham, M., & Fernandez-Parra, M. (2017). Recounting and reflecting: The use of first person pronouns in Chinese, Greek and British students’ assignments in engineering. *Journal of*

English for Academic Purposes, 26, 66–77. <https://doi.org/10.1016/j.jeap.2017.02.001>

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levis, J., Levis, G. ., & Slater, T. (2012). Written English into spoken: A functional discourse analysis of American, Indian, and Chinese TA presentations. In *Working theories for*

- teaching assistant development* (pp. 529–574). Ames: New Forums Press.
- Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English : A cross-contextual , cross-disciplinary study. *Journal of English for Academic Purposes*, 11(4), 345–356. <https://doi.org/10.1016/j.jeap.2012.07.004>
- Li, X. (2018). Teaching English oral presentations as a situated task in an EFL classroom: A quasi- experimental study of the effect of video- assisted self-reflection. *Revista Signos: Estudios de Lingüística*, 51(98), 359–381. <https://doi.org/10.4067/S0718-09342018000300359>
- Lin, C.-Y. (2017). “I see absolutely nothing wrong with that in fact I think ...”: Functions of modifiers in shaping dynamic relationships in dissertation defenses.” *Journal of English for Academic Purposes*, 28, 14–24. <https://doi.org/10.1016/j.jeap.2017.05.001>
- Lovejoy, K. B., & Lance, D. (1989). Discourte analysis (Part I): Information management and cohesion. *ERIC Database. Report ED310383*, 1–59. Retrieved from <https://eric.ed.gov/?id=ED310383>
- Luzón, M. J. (2009). The use of “we” in a learner corpus of reports written by EFL engineering students. *Journal of English for Academic Purposes*, 8(3), 192–206. <https://doi.org/10.1016/j.jeap.2009.04.001>
- Mahfoodh, O. H. A. (2014). Oral academic discourse socialisation: Challenges faced by international undergraduate students in a Malaysian public university. *International Education Studies*, 7(2), 10–17. <https://doi.org/10.5539/ies.v7n2p10>
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230511910>
- Mauranen, A. (1994). Two Discourse Worlds: Study Genres in Britain and Finland. In A.

- Mauranen & R. Markkanen (Eds.), *Students Abroad. Aspects of exchange students' language* (Special is, pp. 1–40).
- Mauranen, A. (2002). "A Good Question". Expressing Evaluation in Academic Speech. *Domain-Specific English. Textual Practices across Communities and Classrooms*, 115–140.
- Retrieved from
- https://www.academia.edu/1858125/A_good_question_Expressing_evaluation_in_academic_speech
- Mauranen, A. (2003). The corpus of English as lingua franca. *TESOL Quarterly*, 37(3), 513–527.
- <https://doi.org/10.2307/3588402>
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, 1(01), e4.
- <https://doi.org/10.1017/S2041536210000012>
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London / New York: Routledge.
- McEnery, T., Xiao, R., & Tono, Y. (2008). *Corpus-based language studies: An advanced resource book*. London and New York: Routledge, 2006. Pp. xix, 386. Pb \$33.95. (Vol. 37). Cambridge University Press (CUP). <https://doi.org/10.1017/s0047404508080615>
- McNeill, D. (2006). Gesture: a psycholinguistic approach. In *Cambridge Encyclopedia of the Language Sciences* (pp. 344–346). Cambridge University Press.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago / London: The University of Chicago Press.
- Meunier, F., & Gentil, G. (2014). Tracking learners' progress in nominalization use: A

- quantitative and qualitative longitudinal corpus analysis. In *GURT 2014-Georgetown Roundtable in Linguistics: Usage-based approaches to language, language learning and multilingualism*.
- Miles, R. (2003). Oral presentations for English proficiency purposes. *Reflections on English Language Teaching*, 8(2), 103–110. Retrieved from <http://nus.edu.sg/celc/research/books/relt/vol8/no2/103-110miles.pdf>
- Miles, R. (2014). The learner's perspective on assessing and evaluating their oral presentations. In *Proceedings of CLaSIC 2014* (pp. 337–352).
- Mitchell, V.-W., & Bakewell, C. (1995). Learning without doing. *Management Learning*, 26(3), 353–366. <https://doi.org/10.1177/1350507695263005>
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 115–154). Amsterdam / Philadelphia: John Benjamins.
- Mohd Radzuan, N. R., & Kaur, S. (2011). Technical oral presentations in English: Qualitative analysis of Malaysian engineering undergraduates' sources of anxiety. *Procedia - Social and Behavioral Sciences*, 29, 1436–1445. <https://doi.org/10.1016/j.sbspro.2011.11.383>
- Moore, M. M., & Voth, B. (1997). The pedagogical values of video portfolios in basic public speaking courses. *Journal on Excellence in College Teaching*, 8(2), 87–105.
- Morell, T. (2001). The role of discourse markers and personal pronouns in lecture discourse. In A. I. Moreno & V. Cowell (Eds.), *Perspectivas recientes sobre el discurso* (pp. 202–206). León: Universidad de León.
- Morell, T. (2007). What enhances EFL students' participation in lecture discourse? Student, lecturer and discourse perspectives. *Journal of English for Academic Purposes*, 6(3), 222–

237. <https://doi.org/10.1016/j.jeap.2007.07.002>

Morell, T. (2015). International conference paper presentations: A multimodal analysis to determine effectiveness. *English for Specific Purposes*, 37, 137–150.

<https://doi.org/10.1016/j.esp.2014.10.002>

Morita, N. (2000). Discourse socialization through oral classroom activities in a TESL graduate program. *TESOL Quarterly*, 34(2), 279–310. <https://doi.org/10.2307/3587953>

Morton, J. (2006). The integration of images into architecture presentations: A semiotic analysis. *Art, Design & Communication in Higher Education*, 5(1), 21–37.

https://doi.org/10.1386/adch.5.1.21_1

Morton, J. (2009). Genre and disciplinary competence: A case study of contextualisation in an academic speech genre. *English for Specific Purposes*, 28(4), 217–229.

<https://doi.org/10.1016/j.esp.2009.04.005>

Morton, J. (2011). Persuasive presentations in engineering spoken discourse. *Australasian Journal of Engineering Education*, 17(2), 55–66.

Munby, I. (2011). The oral presentation: An EFL teacher's toolkit. *Studies in Culture*, 49, 143–168.

Murillo, S. (2008). The Role of Reformulation Markers in Academic Lectures. In A. M. Hornero, M. J. Luzón, & S. Murillo (Eds.), *Corpus linguistics: Applications for the study of English* (Vol. 25, pp. 353–364). Bern: Peter Lang.

Murillo, S. (2012). The use of reformulation markers in Business Management research articles: An intercultural analysis. *International Journal of Corpus Linguistics*, 17(1), 64–90.

<https://doi.org/10.1075/ijcl.17.1.03mur>

Nausa, R. (2015). *Syntactic mechanisms in the transition from academic written to oral*

- discourses: Performance differences in an EAP course*. University of Birmingham.
- Nausa, R. (2016). *Identity projection in the oral presentations of PhD-level EAP students: level of Performance and disciplinary differences*. University of Birmingham.
- Nausa, R. (2017). Syntactic mechanisms in the transition from academic written to oral discourses. *Colombian Applied Linguistics Journal*, 19(2), 234–249.
<https://doi.org/10.14483/22487085.11765>
- Nausa, R. (2018). Modality and code glosses to transition from academic written to oral discourses: An exploratory study. *Íkala, Revista de Lenguaje y Cultura*, 23(4).
<https://doi.org/10.17533/udea.ikala.v23n04aXX>
- Nesi, H. (2005). A corpus-based analysis of academic lectures across disciplines. In J. Cotterill & A. Ife (Eds.), *Language Across Boundaries: Selected Papers from the Annual Meeting of the British Association for Applied Linguistics* (pp. 201–218). London: BAAL in association with Continuum Press.
- Nesi, H. (2012). Laughter in university lectures. *Journal of English for Academic Purposes*, 11(2), 79–89. <https://doi.org/10.1016/j.jeap.2011.12.003>
- Nesi, H., & Baştürkmen, H. (2006). Lexical bundles and discourse signaling in academic lecturers. *International Journal of Corpus Linguistics*, 11(3), 283–304.
<https://doi.org/10.1075/ijcl.11.3.04nes>
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Nilsson, M. R. (2001). Student-taught review sessions: Fostering communication skills and reinforcing concepts. *Journal of Chemical Education*, 78(5), 628.
<https://doi.org/10.1021/ed078p628>

- Nilsson, N. (2017). *Pragmatics, newspapers and context: A study of how British tabloid newspapers construct context*. Halmstad University. Retrieved from <http://www.diva-portal.org/smash/get/diva2:1071143/FULLTEXT02>
- Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. New York / London: Routledge.
- North, M. M., North, S. M., & Coble, J. R. (2015). Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality (IJVR)*, 03(3), 1–6. Retrieved from <https://hal.archives-ouvertes.fr/hal-01530637/>
- North, S. (2005). Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education*, 30(5), 517–533. <https://doi.org/10.1080/03075070500249153>
- Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly*, 31(3), 409. <https://doi.org/10.2307/3587831>
- O’Boyle, A. (2014). “You” and “I” in university seminars and spoken learner discourse, 16, 40–56. <https://doi.org/10.1016/j.jeap.2014.08.003>
- O’connor, P. E. (1994). ‘You could feel it through the skin’: Agency and positioning in prisoners’ stabbing stories. *Text*, 14(1), 45–75. <https://doi.org/10.1515/text.1.1994.14.1.45>
- Osborn, M., Osborn, S., & Osborn, R. (2012). *Public speaking: finding your voice* (9th ed.). Boston: Pearson.
- Oshima, A., & Hogue, A. (2006). *Writing academic English* (4th ed.). White Plains, NY: Pearson/Longman.
- Österholm, M. (2012). Students’ summaries of mathematical lectures: Comparing the discourse of students with the discourse of lectures. In L. P. C. Dindyal & S. F. Ng (Eds.),

- Mathematics education: Expanding horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia)* (pp. 578–585). Singapore: Mathematics Education Research Group of Australasia Inc. Retrieved from <https://files.eric.ed.gov/fulltext/ED573338.pdf>
- Otoshi, J., & Heffernen, N. (2008). Factors predicting effective oral presentations in EFL classrooms. *The Asian EFL Journal Quarterly*, 10(1), 65–78.
- Palmer-Silveira, J. C. (2015). Multimodality in Business Communication. In B. Crawford Camiciottol & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 171–192). New York, NY, NY: Routledge.
- Palmer, F. (1987). *The English Verb* (2nd ed.). London: Longman.
- Pertaub, D.-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1), 68–78. <https://doi.org/10.1162/105474602317343668>
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233–255. <https://doi.org/10.2307/3587647>
- Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, 23(1), 19–43. [https://doi.org/10.1016/S0889-4906\(03\)00020-6](https://doi.org/10.1016/S0889-4906(03)00020-6)
- Pittenger, K. K. S., Miller, M. C., & Mott, J. (2004). Using real-world standards to enhance students' presentation skills. *Business Communication Quarterly*, 67(3), 327–336. <https://doi.org/10.1177/1080569904268084>
- Pleyer, M. (2008). Shared symbolic storage: The cognitive foundations of perspective. Retrieved May 8, 2018, from <http://sharedsymbolicstorage.blogspot.com.co/2008/01/cognitive->

foundations-of-perspective.html

Poyatos, F. (2002). *Nonverbal communication across disciplines. Volume I: Culture, sensory interaction, speech, conversation*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., ... Ansari, R. (2002).

Multimodal human discourse: Gesture and speech. *ACM Transactions on Computer-Human Interaction*, 9(3), 171–193. <https://doi.org/10.1145/568513.568514>

Querol-Julián, M. (2010). Multimodality in discussion sessions: Corpus compilation and pedagogical use. *Language Value Journal*, 2(1), 1–26.

Querol-Julián, M., & Fortanet, I. (2012). Multimodal evaluation in academic discussion sessions: How do presenters act and react? *English for Specific Purposes*, 31(4), 271–283.

<https://doi.org/10.1016/j.esp.2012.06.001>

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language* (1st ed.). London / New York: Longman.

Qurban, M. H., & Austria, R. D. (2009). Improving the communication skills of IS developers during requirements elicitation using experiential learning. *Journal of Information Systems Education*, 20(3), 301–311. Retrieved from

<https://search.proquest.com/openview/d0ec9b18a308b9402327085556b372ef/1?pq-origsite=gscholar&cbl=25848>

Rahman, M. M. (2010). Teaching oral communication skills : A task-based approach. *ESP World*, 9(1), 1–11. <https://doi.org/10.1080/10572259809364631>

Rayson, P. (2017). Log-likelihood and effect size calculator. Retrieved December 17, 2017, from <http://ucrel.lancs.ac.uk/llwizard.html>

- Reinhart, S. (2005). *Giving academic presentations*. Ann Arbor: University of Michigan Press.
- Reitmeier, C. A., & Vrchota, D. A. (2009). Self-assessment of oral communication presentations in food science and nutrition. *Journal of Food Science Education*, 8(4), 88–92.
<https://doi.org/10.1111/j.1541-4329.2009.00080.x>
- Rendle-Short, J. (2006). *The academic presentation: Situated talk in action*. Hampshire / Burlington: Ashgate Publishing Limited.
- Reppen, R. (2010). Building a corpus. In A. O’Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 31–37). New York: Routledge.
- Roth, W. (2000). From gesture to scientific language. *Journal of Pragmatics*, 32, 1683–1714.
[https://doi.org/10.1016/S0378-2166\(99\)00115-0](https://doi.org/10.1016/S0378-2166(99)00115-0)
- Rounds, P. L. (1987). Multifunctional personal pronoun use in an educational setting. *English for Specific Purposes*, 6(1), 13–29. [https://doi.org/10.1016/0889-4906\(87\)90072-X](https://doi.org/10.1016/0889-4906(87)90072-X)
- Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers: A genre-based study. *English for Specific Purposes*, 21(1), 19–40. [https://doi.org/10.1016/S0889-4906\(00\)00024-7](https://doi.org/10.1016/S0889-4906(00)00024-7)
- Rowley-Jolivet, E. (2012). Oralising text slides in scientific conference presentations: A multimodal corpus analysis. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 137–165). Amsterdam / Philadelphia: John Benjamins Publishing Company.
<https://doi.org/10.1075/sll.15.1.07mor>
- Rowley-Jolivet, E. (2015). Quantification in conference talks and proceedings articles in engineering. *English for Specific Purposes*, 38, 11–22.
<https://doi.org/10.1016/j.esp.2014.10.003>

- Rowley-Jolivet, E., & Carter-Thomas, S. (2005a). Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting. *English for Specific Purposes*, 24(1), 41–64.
<https://doi.org/10.1016/j.esp.2003.09.003>
- Rowley-Jolivet, E., & Carter-Thomas, S. (2005b). Scientific conference Englishes. In G. Cortese & A. Duszak (Eds.), *Identity, community, discourse: English in intercultural settings* (pp. 295–320). Bern: Peter Lang.
- Rowley-Jolivet, E., & Carter-Thomas, S. (2005c). The rhetoric of conference presentation introductions: context, argument and interaction. *International Journal of Applied Linguistics*, 15(1), 45–70. <https://doi.org/10.1111/j.1473-4192.2005.00080.x>
- Rowley, J. (2012). Six steps to successful academic conference presentation. *The Marketing Review*, 12(4), 435–448. <https://doi.org/10.1362/146934712X13546196910891>
- Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531.
<https://doi.org/10.1007/BF00973770>
- Rubin, D. (1998). Help! My teacher (or physician or boss) can't talk English. *Readings in Cultural Contexts*, 149–159.
- Rui, K., & Xin, X. (2009). Empirical study on metadiscourse in Chinese EFL learners oral Communication. *CELEA Journal*, 32(1), 52–64.
- Ruiz-Garrido, M. F. (2015). Intensifying adverbs in academic spoken discourse. In B. Crawford Camiciottol & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 61–82). New York: Routledge.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization

- of turn taking for conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 7–55). New York / London: Academic Press Inc.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 361–382.
- Schiffrin, D. (1990). Between text and context: Deixis, anaphora, and the meaning of then. *Text*, 10(3), 245–270. <https://doi.org/10.1515/text.1.1990.10.3.245>
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Seliman, Salbiah; Affendi, Irwan; Pendidikan, F. (2010). The genre of the body of oral presentations delivered by English for workplace communication students. Retrieved from http://eprints.utm.my/id/eprint/10452/1/The_Genre_Of_The_Body_Of_Oral_Presentations_Delivered_By_English_For_Workplace_Communication_Students.pdf
- Seliman, S. (1996). *The genre and the genre expectations of engineering oral presentations related to academic and professional contexts*. University of Stirling. Retrieved from https://dspace.stir.ac.uk/bitstream/1893/1778/1/The_Genre_and_Genre_Expectations_of_Engineering_Oral_Presentations_by_Salibah_Seliman_%281996%29.pdf
- Shaw, V. N. (2001). Training in presentation skills: An innovative method for college instruction. *Education*, 122(1), 140–144.
- Shimo, E. (2011). Implications for effective ways of conducting and assessing presentations in EFL classes. *Language Education in Asia*, 2(2), 227–236. <https://doi.org/10.5746/LEiA/11/V2/I2/A05/Shimo>
- Siegel, J. (2018). Did you take “good” notes?: On methods for evaluating student notetaking performance. *Journal of English for Academic Purposes*, 35, 85–92.

<https://doi.org/10.1016/J.JEAP.2018.07.001>

- Simpson-Vlach, R. (2006). Academic speech across disciplines: Lexical and phraseological distinctions. In M. Bondi & K. Hyland (Eds.), *Academic Discourse Across Disciplines* (pp. 295–316). Bern: Peter Lang.
- Sinclair, J. (1988). Mirror for a text. *Journal of English and Foreign Languages*, 1, 15–44.
- Slater, M., Pertaub, D.-P., Barker, C., & Clark, D. (2002). An experimental study on fear of public speaking using a virtual environment. *Presence: Teleoperators and Virtual Environments*, 11(1), 68–78. Retrieved from http://www.communicationcache.com/uploads/1/0/8/8/10887248/an_experimental_study_on_fear_of_public_speaking_using_a_virtual_environment.pdf
- Smith, C. D., & King, P. E. (2004). Student feedback sensitivity and the efficacy of feedback interventions in public speaking performance improvement. *Communication Education*, 53(3), 203–216. <https://doi.org/10.1080/0363452042000265152>
- Soureshjani, K., & Ghanbari, H. (2012). Factors leading to an effective oral presentation in EFL classrooms. *The TFLTA Journal*, 3, 34–47.
- Speyer, A. (2005). Topicalization and the trochaic requirement. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAVE 32*, 10(2), 243–256.
- Stapa, M., Murad, N. A., & Ahmad, N. (2014). Engineering technical oral Presentation: Voices of the stakeholder. *Procedia - Social and Behavioral Sciences*, 118, 463–467. <https://doi.org/10.1016/j.sbspro.2014.02.063>
- Stirling, L., & Manderson, L. (2011). About you: Empathy, objectivity and authority. *Journal of Pragmatics*, 43(6), 1581–1602. <https://doi.org/10.1016/j.pragma.2010.12.002>
- Stubbs, M. (1986). ‘A matter of prolonged field work’: Notes towards a modal grammar of

- English. *Applied Linguistics*, 7(1), 1–25.
- Swales, J. M. (1971). *Writing Scientific English*. London: Nelson.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press. Retrieved from http://adrastea.ugr.es/record=b1766123*sp
- Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills*. *College Composition and Communication* (3rd ed.). Ann Arbor: University of Michigan Press. <https://doi.org/10.2307/358319>
- Swanson, R. W., Spooner, H. J., Reeder, B. A., Haight, K. R., & Senthilselvan, A. (1992). Improving presentation skills of family medicine residents: A randomized controlled study. *Medical Education*, 26(3), 241–247. <https://doi.org/10.1111/j.1365-2923.1992.tb00160.x>
- Swati Puri, James S. Chickos, A., & Welsh, W. J. (2002). Three-Dimensional Quantitative Structure–Property Relationship (3D-QSPR) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): Enthalpies of fusion and their application to estimates of enthalpies of sublimation and aqueo. <https://doi.org/10.1021/CI0200164>
- Talebinejad, M. R., & Ghadyani, F. (2012). A contrastive rhetoric analysis of 'code glosses' in medicine academic research posters written in English by native and Iranian writers. *Journal of Language, Culture, and Translation*, 1(2), 81–95.
- Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18, S23–S39. [https://doi.org/10.1016/s0889-4906\(99\)00009-5](https://doi.org/10.1016/s0889-4906(99)00009-5)
- Tardy, C. (2011). Genre analysis. In K. Hyland & B. Paltridge (Eds.), *Continuum companion to discourse analysis* (pp. 54–68). London: Continuum.
- Taylor, K. L., & Toews, S. V. M. (1999). Effective presentations: How can we learn from the

- experts? *Medical Teacher*, 21(4), 409–414. <https://doi.org/10.1080/01421599979365>
- Taylor, P. (1992). Improving graduate student seminar presentations through training. *Teaching of Psychology*, 19(4), 236–238. https://doi.org/10.1207/s15328023top1904_12
- Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, 22(1), 58–78. <https://doi.org/10.1093/applin/22.1.58>
- Thompson, S. (1994). Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes*, 13(2), 171–186. [https://doi.org/10.1016/0889-4906\(94\)90014-0](https://doi.org/10.1016/0889-4906(94)90014-0)
- Thompson, S., & Collins, H. (1995). Dealing with face threats in oral presentations. *Trabalhos Em Linguística Aplicada*, (26), 81–99.
- Trevor. (2006). Top ten longest titles of research papers. Retrieved May 1, 2018, from <http://vocaro.com/trevor/blog/2006/09/12/top-ten-longest-titles-of-research-papers/>
- Tsai, S.-C. (2011). Courseware integration into task-based learning: A case study of multimedia courseware-supported oral presentations for non-English major students. *ReCALL*, 23(02), 117–134. <https://doi.org/10.1017/S0958344011000048>
- Tucker, M. L., & McCarthy, A. M. (2001). Presentation self-efficacy: Increasing communication skills through service-learning. *Journal of Managerial Issues*. Pittsburg State University. <https://doi.org/10.2307/40604346>
- Twentieth Century Arts Lecture: LEL320JU147. (2001). *Michigan Corpus of Academic Spoken English*. Retrieved from <https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;cc=micase;view=transcript;id=LEL320JU147>
- Tyler, A. (1994). The role of syntactic structure in discourse structure: Signaling logical and

- prominence relations. *Applied Linguistics*, 15(3), 243–262.
<https://doi.org/10.1093/applin/15.3.243>
- Tyler, A., Jefferies, A., & Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, 7(2), 101–110. <https://doi.org/10.1111/j.1467-971X.1988.tb00223.x>
- Umesaki, A. F. (1991). A comparison of scientific lectures and papers with regard to grammar and lexis: Differences between speech and writing. *Studies in Modern English*, (8), 37–53.
- Universidad de los Andes. (2016). Programa de Desarrollo Integral (PDI) 2016-2020. Bogotá.
Retrieved from
https://planeacion.uniandes.edu.co/dmdocuments/Programa_de_Desarrollo_Integral_PDI_2016-2020.pdf
- Valeiras, J. (2015). A multimodal approach to persuasion in conference presentations. In B. Crawford Camiciottol & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 108–130). New York, NY, NY: Routledge.
- Valero-Garcés, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish-English economics texts. *English for Specific Purposes*, 15(4), 279–294.
- van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review*, 14, 62–80.
<https://doi.org/10.1016/j.edurev.2015.02.002>
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82–93. <https://doi.org/10.2307/357609>
- Vassileva, I. (2002). Speaker-audience interaction: the case of Bulgarians presenting in English.

- In E. Ventola et al., *The Language of Conferencing*. Frankfurt am Main: Peter Lang (pp. 255–276).
- Ventola, E. (1999). Semiotic spanning at conferences: Cohesion and coherence in and across conference papers and their discussions. In W. Bublitz, U. Lenk, & E. Ventola (Eds.), *Coherence in spoken and written discourse* (pp. 101–124). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Ventola, E. (2002). Why and what kind of focus on conference presentations? In E. Ventola, C. Shalom, & S. Thompson (Eds.), *The language of conferencing* (pp. 15–50). Frankfurt: Peter Lang.
- Vergaro, C. (2004). Discourse strategies of Italian and English sales promotion letters. *English for Specific Purposes*, 23(2), 181–207.
- Wallwork, A. (2010). *English for presentations at international conferences*. New York: Springer.
- Ward, G., & Birner, B. (2004). Information structure and non-canonical syntax. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 152–174). Malden, MA: Blackwell Publishing Ltd.
- Wächter, B. (2008). Teaching in English on the rise in European education. *International Higher Education*, (52), 3–4. Retrieved from <https://ejournals.bc.edu/ojs/index.php/ihe/article/viewFile/8028/7179>
- Webber, P. (2002). The paper is now open for discussion. In E. Ventola, C. Shalom, & S. Thompson (Eds.), *The language of conferencing* (pp. 227–254). Frankfurt: Peter Lang.
- Webber, P. (2005). Interactive features in medical conference monologue. *English for Specific Purposes*, 24(2), 157–181. <https://doi.org/10.1016/j.esp.2004.02.003>

- Wecker, C. (2012). Slide presentations as speech suppressors: When and why learners miss oral information. *Computers and Education*, 59(2), 260–273.
<https://doi.org/10.1016/j.compedu.2012.01.013>
- Weissberg, B. (1993). The graduate seminar: Another research-process genre. *English for Specific Purposes*, 12(1), 23–35. [https://doi.org/10.1016/0889-4906\(93\)90025-J](https://doi.org/10.1016/0889-4906(93)90025-J)
- Wexelblat, A. (1998). Research challenges in gesture: Open issues and unsolved problems. In I. Wachsmuth & M. Fröhlich (Eds.), *Lecture Notes in Computer Science: International Gesture Workshop Bielefeld* (Vol. 1371, pp. 1–11). <https://doi.org/10.1007/BFb0052984>
- Widdowson, H. (1979). *Explorations in applied linguistics*. Oxford: Oxford University Press.
[https://doi.org/10.1016/0378-2166\(82\)90031-5](https://doi.org/10.1016/0378-2166(82)90031-5)
- Wiese, J., Varosy, P., & Tierney, L. (2002). Improving oral presentation skills with a clinical reasoning curriculum: A prospective controlled study. *The American Journal of Medicine*, 112(3), 212–218. [https://doi.org/10.1016/S0002-9343\(01\)01085-3](https://doi.org/10.1016/S0002-9343(01)01085-3)
- Wilson, A. (2013). Embracing Bayes factors for key item analysis in corpus linguistics. In M. Bieswanger & A. Koll-Stobbe (Eds.), *New approaches to the study of linguistic variability: Language competence and language awareness in Europe* (pp. 3–11). Frankfurt: Peter Lang.
- Wilson, J., & Brooks, G. (2014). Teaching presentation: Improving oral output with more structure. In I. Aishah M.K., Bhatt, S.K., Chan, W.M., Chi, S.W., Chin, K.W., Klayklung, S., Nagami, M., Sew, J.W., Suthiwan, T., Walker (Ed.), *Proceedings of CLaSIC 2014* (pp. 512–522). Singapore.
- Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, 37(3), 308–328. <https://doi.org/10.1177/0033688206071315>

- Wortham, S. (1996). Mapping deictics: A technique for discovering teachers' footing. *Journal of Pragmatics*, 25(3), 331–348. [https://doi.org/http://dx.doi.org/10.1016/0378-2166\(94\)00100-6](https://doi.org/http://dx.doi.org/10.1016/0378-2166(94)00100-6)
- Wright, L. J. (2008). Writing science and objectification: Selecting, organizing, and decontextualizing knowledge. *Linguistics and Education*, 19(3), 265–293. <https://doi.org/10.1016/j.linged.2008.06.004>
- Wu, Y.-H. (2010). *Oral Academic Presentations: Discourse Socialization Experiences of Eight EFL Graduate Students in TESOL*. National Chiao Tung University.
- Wulff, S., Swales, J. M., & Keller, K. (2009). “We have about seven minutes for questions”: The discussion sessions from a specialized conference. *English for Specific Purposes*, 28(2), 79–92. <https://doi.org/10.1016/j.esp.2008.11.002>
- Xiong, Y., Quek, F., & McNeill, D. (2002). Hand gesture symmetric behavior detection and analysis in natural conversation. In *4th IEEE International Conference on Multimodal Interfaces - ICMI 2002* (pp. 179–184). Pittsburgh: IEEE Computer Society. <https://doi.org/10.1109/ICMI.2002.1166989>
- Yaakob, S. (2013). *A genre analysis and corpus based study of university lecture introductions*. University of Birmingham. Retrieved from <http://etheses.bham.ac.uk/5241/1/Yaakob14PhD.pdf>
- Yang, A., Zheng, S. yuan, & Ge, G. chun. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes*, 38, 1–10. <https://doi.org/10.1016/j.esp.2014.10.005>
- Yang, L. (2010). Doing a group presentation: Negotiations and challenges experienced by five Chinese ESL students of commerce at a Canadian University. *Language Teaching Research*,

- 14(2), 141–160. <https://doi.org/10.1177/1362168809353872>
- Yeereem, C. (2013). *A comparative study of language use in oral presentations of Thai EFL learners and English-native Students*. Prince of Songkla University.
- Yeo, J. Y., & Ting, S. H. (2014). Personal pronouns for student engagement in arts and science lecture introductions. *English for Specific Purposes*, 34(1), 26–37.
<https://doi.org/10.1016/j.esp.2013.11.001>
- Young, M. R., & Murphy, J. W. (2003). Integrating communications skills into the marketing curriculum: A case study. *Journal of Marketing Education*, 25(1), 57–70.
<https://doi.org/10.1177/0273475302250574>
- Young, R. F., & Nguyen, H. T. (2002). Modes of meaning in high school science. *Applied Linguistics*, 23(3), 348–372+415. <https://doi.org/10.1093/applin/23.3.348>
- Yusoff, M. (2010). Analysing communication competence in oral presentations: Engineering students' experience. *Journal of Human Capital Development*, 3(1), 99–118.
- Zappa, S. (2001). *Academic presentations: Exploring the second language socialization of international graduate students across disciplines*. The University of British Columbia.
- Zare, J., & Keivanloo-Shahrestanaki, Z. (2017). Genre awareness and academic lecture comprehension: The impact of teaching importance markers. *Journal of English for Academic Purposes*, 27, 31–41. <https://doi.org/10.1016/j.jeap.2017.03.001>
- Zareva, A. (2009). Informational packaging, level of formality, and the use of circumstance adverbials in L1 and L2 student academic presentations. *Journal of English for Academic Purposes*, 8(1), 55–68. <https://doi.org/10.1016/j.jeap.2008.12.002>
- Zareva, A. (2012). Expression of stance and persuasion in student academic presentations. In G. Mininni & A. Manuti (Eds.), *Applied Psycholinguistics. Positive effects and ethical*

perspectives (Vol. 2, pp. 316–323). Milano, Italy: Franco Angeli.

- Zareva, A. (2013). Self-mention and the projection of multiple identity roles in TESOL graduate student presentations: The influence of the written academic genres. *English for Specific Purposes*, 32(2), 72–83. <https://doi.org/10.1016/j.esp.2012.11.001>
- Zareva, A. (2016). Multi-word verbs in student academic presentations. *Journal of English for Academic Purposes*, 23, 83–98. <https://doi.org/10.1016/j.jeap.2016.07.001>
- Zhang, Z. (2015). Disagreements in plenary addresses as multimodal action. In B. Crawford-Camiciottoli & I. Fortanet (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 27–48). New York / London: Routledge.
- Živković, S. (2015). Introducing students to the genre of presentations for professional purposes. *Mediterranean Journal of Social Sciences*. <https://doi.org/10.5901/mjss.2015.v6n2s2p201>

Appendixes

Appendix A: RUBRIC TO EVALUATE ESSAYS

Category of Evaluation	This category is absent from the text.	There is a hint of this category, but it's too faint; it needs much more attention.	This category has potential but clearly needs more attention.	Good job. Little attention is needed in this category.	Excellent job. No further attention is needed in this category.		
	(0.0 - 1.0)	(1.1 - 2.0)	(2.1 - 3.0)	(3.1 - 4.0)	(4.1 - 5.0)		
Structure and Organization							
The introduction includes contextualization and a focused thesis statement that allows the reader to predict the structure of the essay (subtopic division).	1	1	1	1	0,5	4,5	
Each body paragraph includes a topic sentence that is clearly stated and relates to the thesis statement, and details that support the topic sentence.	1	1	1	0,5		3,5	
The conclusion revisits all the key points of the essay and leaves the reader with a thought to consider.	1	1	1	1	0,5	4,5	4,2
Content and Development							
The essay addresses the prompt (a problem and/or possible solutions), demonstrates familiarity with the topic, and considers the audience's background knowledge. (first semester students)	1	1	1	1	0,5	4,5	
The essay presents a logical explanation with transitional devices that facilitate flow in ideas.	1	1	1	0,5		3,5	
All ideas displayed in the essay are concrete, and relevant, and supported by reliable examples or evidence.	1	1	1	1	0,5	4,5	4,2
Process and Linguistic Accuracy							
The student has fully participated in all 3 stages of the writing process: prewriting, drafting, and revision & editing.	1	1	1	1	0,5	4,5	
There are no one-sentence paragraphs, run-ons, stringy sentences, comma splices, missing subjects, unparallel structures or fragments. Articles and pronouns (<i>another the other, others</i> , etc) are properly used.	1	1	1	1		4	
Capitalization and punctuation are correct as related to phrase/clause combination; formatting is correct and appropriate for academic writing. The essay word count is within the 600-900 word limit.	1	1	1	1		4	4,2
						Final:	4,2

[illegible]

Appendix D. GENERAL TRANSCRIPTION CONVENTIONS

[fs]: false starts

Um, uh, er: hesitation marks

[reading 1]: sentences that were read either from a slide or a script

A: person speaking (presenter or member of the audience)

(word): words enclosed in parentheses refer to the transcriber's interpretation of words that were not completely understood and that are inferred either from how they sound or the general meaning of the speech

(xxx): used for words that were not understood or inferred

Appendix E. LIST OF ESSAY/OPs PARALLEL TEXTS

	ASSIGNED CODE	TITLE
1	S-H-ADMI-1	ENTITLEMENTS AND PROCEDURAL JUSTICE
2	S-H-ADMI-2	HIGH GROWTH FIRMS
3	S-H-ANTR-1	THE POROUS PROCESS IN THE SKULL: EVIDENCE FOR ANAEMIA IN ARCHAEOLOGICAL POPULATIONS
4	S-H-ANTR-2	A HEALTHY DIET FOR EVERYDAY
5	S-H-ANTR-3	DOMESTIC VIOLENCE AND LOVE IN COLOMBIA: A “MISTREATED” PROBLEM IN SOCIAL RESEARCH?
6	S-H-CBIO1	IMPACTS OF FUNGAL AND OOMYCETE PLANT DISEASE
7	S-H-CBIO2	IS IT POSSIBLE TO STOP THE SPECIES EXTINCTION IN THE WORLD?
8	S-H-CBIO3	PROCESS AND PATTERNS IN EVOLUTION AND DISTRIBUTION OF DINK FROGS (GENUS: DIASPORUS; FAMILY ELEUTHERODACTYLIDAE)
9	S-H-CBIO-4	VIBRATIONAL COMMUNICATION: THE CASE OF KISSING BUGS (TRITOMINAE-HETEROPTERA)
10	S-H-CBIO-5	GENETIC STRUCTURE OF BOTTLENOSE DOLPHINS IN BOCAS DEL TORO: IMPLICATIONS FOR CONSERVATION
11	S-H-CBIO-6	SEED DISPERSAL BY WOOLLY MONKEYS IN CUEVA DE LOS GUACHAROS NATIONAL PARK (COLOMBIA)
12	S-H-CBIO-7	ORCHID DIVERSIFICATION
13	S-H-CQUI-1	APPLICATIONS OF MOLECULAR RECOGNITION IN HEAVY METAL PROBLEMS
14	S-H-CQUI-2	ORGANIC SYNTHESIS: TOOL TO PREPARE BIOLOGICALLY ACTIVE COMPOUNDS
15	S-H-DERE-1	PROPERTY RIGHTS OVER THE URBAN LAND IN COLOMBIA
16	S-H-DERE-2	THE REBIRTH OF A POLITICAL VICTIM: THE RESTITUTION OF THE LEGAL STATUS TO THE UNIÓN PATRIÓTICA AS A COLLECTIVE REPARATION MEASURE IN COLOMBIA
17	S-H-ECON-1	DEINDUSTRIALIZATION IN COLOMBIA IN THE XX AND XXI CENTURY
18	S-H-ECON-2	EXPERIMENTAL ECONOMIC GAMES
19	S-H-ECON-3	RETURNS TO EDUCATION
20	S-H-EDUC-1	THE INTERCULTURAL COMMUNICATIVE COMPETENCE, A NEW PERSPECTIVE IN A GLOBAL WORLD
21	S-H-INGE-1	WIRELESS SENSOR NETWORKS
22	S-H-INGE-2	IMPORTANCE OF THE METHODOLOGIES FOR DECISION MAKING IN THE CONSTRUCTION OF PUBLIC INFRASTRUCTURE
23	S-H-INGE-3	SOFTWARE PATCHES AND STATIC PROGRAMS ANALYSIS
24	S-H-INGE-4	METAMODELS COMPOSITION
25	S-H-INGE-5	BAMBOO FIBRE AS A POTENTIAL REINFORCEMENT IN THE CEMENT INDUSTRY
26	S-H-INGE-6	3D LARGE-SCALE MODELS FOR SIMULATING EARTHQUAKE GROUND MOTION IN SEISMIC REGIONS
27	S-H-INGE-7	REDUCING OPERATING COSTS THROUGH THE IMPLEMENTATION OF A JUST-IN-TIME APPROACH AS OPPOSITE OF EOQ-BASED MODELS
28	S-H-LITE-1	GREEK AND LATIN NOVELS
29	S-H-MATE-1	DECIDABILITY’S TECHNIQUES IN MODEL THEORY
30	S-H-MATE-2	ALGEBRAIC TOPOLOGY
31	S-H-PSIC-1	HOW UNDERSTAND SELF-DECEPTION?
32	S-L-ADMI-1	ORGANIZATIONAL JUSTICE IN FAMILY CONFLICT RESOLUTION

- 33 S-L-ANTR-1 THE USES OF PHOTOGRAPHIES IN ANTHROPOLOGIE FROM 1845 TO 2006
- 34 S-L-ANTR-2 DNA ANALYSIS METODOLOGY FROM FAUNAL ARCHAEOLOGICAL REMAINS
- 35 S-L-ANTR-3 THE OPERATIONAL CHAIN OF ANCIENT COLOMBIAN METALLURGY
- 36 S-L-ANTR-4 THE GENESIS OF MESSIANIC MILLENARIAN MOVEMENTS
- 37 S-L-ANTR-5 DEATH AND CULTURE
- 38 S-L-CBIO-1 INTERACTION PLANT-PATHOGEN: XANTHOMONAS AXONOPODIS PV. MANIHOTIS (XAM) AND YUCCA
- 39 S-L-CQUI-1 CONTROLLED RELEASE DRUG
- 40 S-L-DERE-1 THE FUNDAMENTAL RIGHT TO PRIOR CONSULTATION
- 41 S-L-DERE-2 CIVIL LIABILITY OF STATE LEGISLATURE
- 42 S-L-EDUC-1 CYBER-CITIZENSHIP OR DIGITAL CITIZENSHIP?
- 43 S-L-EDUC-2 THE FEMINIZATION OF TEACHING IN A PRIMARY SCHOOL IN COLOMBIA
- 44 S-L-FILO-1 THE YOUNG HEIDEGGER AND HIS RELATION TO THEOLOGY
- 45 S-L-FILO-2 ASTHETIC OR PHILOSOPHICAL THEORY OF ART
- 46 S-L-HIST-1 THE SOCIABILITY: A CATEGORY FOR HISTORICAL STUDY
- 47 S-L-HIST-2 THE INSERTION OF THE CIVIL LAW IN THE SECOND HALF OF NINETEENTH CENTURY IN COLOMBIA
- 48 S-L-HIST-3 THE CONSTRUCTION OF THE MODERN MUNICIPAL SLAUGHTERHOUSE AT THE BEGINNING OF THE TWENTIETH CENTURY IN CALI
- 49 S-L-HIST-4 MADNESS AT THE END OF THE COLONIAL PERIOD
- 50 S-L-HIST-5 THE ROLE OF THE AMATEUR RADIO BROADCASTERS IN THE EARLY ORIGINS OF COLOMBIAN RADIO
- 51 S-L-INGE-1 ENERGY OPTIMAL CONSUMPTION FOR ELECTRICAL VEHICLES (EV)
- 52 S-L-INGE-2 STRATEGIES TO RESOLVE DESIGN CONTROL PROBLEMS IN AV
- 53 S-L-INGE-3 METHODOLOGIES FOR MAKING DECISIONS
- 54 S-L-INGE-4 FAULT—TOLERANT CONTROL
- 55 S-L-INGE-5 MATHEMATIC COMPREHENSION OF SOILS BEHAVIOR
- 56 S-L-MATE-1 ASYMPTOTIC BOUNDS IN THE NUMBER OF EDGES IN GRAPHS
- 57 S-L-MATE-2 THE SKEW PBW EXTENSION
- 58 S-M-ANTR-1 AGAINST THE MATERIAL VISION OF THE PHYSICAL BODY
- 59 S-M-ANTR-2 THE PFA COLOMBIAN CONDITIONAL CASH TRANSFER OR A GOOD SOCIAL POLICY
- 60 S-M-ANTR-3 THE PROBLEM OF POWER IN THE MUISCA SOCIETY
- 61 S-M-ANTR-4 REPARATION FOR INDIGENOUS PEOPLES IN THE INTER-AMERICAN COURT OF HUMAN RIGHTS
- 62 S-M-CBIO-1 GENETIC DIVERSITY IN COLOMBIAN POPULATIONS
- 63 S-M-CBIO-2 EFFECTS OF ANTIBIOTICS ON BACTERIAL CELLS
- 64 S-M-CBIO-3 AUTISM SPECTRUM DISORDER (ASD)
- 65 S-M-CPOL-1 SUBNATIONAL DEMOCRACY: STABILITY AND CHANGE
- 66 S-M-CPOL-2 ACCESS TO PUBLIC INFORMATION IN LATIN AMERICA: ANALYSIS OF INSTITUTIONAL DESIGN IN CHILE, COLOMBIA AND MÉXICO
- 67 S-M-CPOL-3 CURRICULAR MISALIGNMENT. A LOOK FROM THREE IMPLICATIONS IN THE EDUCATIONAL PROGRAMS OF THE UNIVERSITY SOCIAL RESPONSIBILITY
- 68 S-M-CPOL-4 PARAMILITARY GROUPS
- 69 S-M-CQUI-1 PYRAZOLES AS POSSIBLE IONOPHORES OF COPPER (II) FOR ITS USE AS ANTICANCER AGENTS
- 70 S-M-DERE-1 INVESTMENT TREATY ARBITRATION
- 71 S-M-EDUC-1 PEDAGOGY OF HISTORICAL MEMORY

72	S-M-EDUC-2	PROBLEM ABOUT SOCIAL RESPONSIBILITY TRAINING TROUGH COLOMBIAN UNIVERSITY'S CURRICULUM
73	S-M-EDUC-3	CURRICULUM PLANNING IN MATHEMATICS AT HIGH SCHOOLS IN COLOMBIA
74	S-M-FISI-1	FLUORESCENCE MICROSCOPE, FROM BIOLOGICAL SYSTEMS TO THE SINGLE MOLECULE
75	S-M-HIST-1	THE STUDY OF THE AFFECTIVE LIFE IN THE FIRST DECADES OF THE 20TH CENTURY
76	S-M-HIST-2	THE BOGOTA'S PANOPTICON
77	S-M-HIST-3	ANACHRONISM
78	S-M-INGE-1	DROUGHT IMPLICATIONS ON FOUNDATIONS STRUCTURES
79	S-M-INGE-2	MUDDY ROCKS FAILURE MECHANISMS AND ITS IMPORTANCE IN COLOMBIA'S INFRASTRUCTURE
80	S-M-INGE-3	SMART GRID AS A NATURAL CONVERGENCY OF AVAILABLE TECHNOLOGIES IN POWER SYSTEMS
81	S-M-INGE-4	STEPS FOR CONTROLLING SCHEDULE
82	S-M-INGE-5	ANALYSIS MODEL OF BUSINESS STRATEGY AND TECHNOLOGY FOR COLOMBIAN COMPANIES
83	S-M-INGE-6	MODELS AND METAMODELS IN AN ENTERPRISE ARCHITECTURE PROJECT
84	S-M-INGE-7	DISTRIBUTED GENERATION: ADVANTAGES AND CHALLENGES
85	S-M-INGE-8	BASIC PRINCIPLES FOR PLANNING OF ELECTRICAL DISTRIBUTION SYSTEMS
86	S-M-INGE-9	THE GLAUCOMA, TONOMETRY DIAGNOSIS PROBLEMS
87	S-M-INGE-10	ENVIRONMENTAL IMPACTS OF THE COAL MINING IN THE CESAR DEPARTMENT
88	S-M-MATE-1	GAUSS-BONNET THEOREM IN PRINCIPAL G-BUNDLES WITH SINGULARITIES

Appendix F. MULTIMODAL ANALYSIS TRANSCRIPTION CONVENTIONS (taken and adapted from Rendle-Short, 2006)³⁴

°okay°: talk is noticeably lower

OKAY: talk is noticeably louder


(): transcription doubt


(()): analyst's comments


[: overlapping utterances or actions

Pres: presenter

1. Gaze direction:

: at screen

: at computer

: at script

_____ : towards the audience

→→→: towards right (presenter's left)

←←←: towards left (presenter's right)

↑↑↑: upward

↓↓↓: downward

↘↘↘: right middle distance

↙↙↙: left middle distance

2. Hand movements:

LH: left hand

RH: right hand


BH: both hands

LH→: moves left hand to right

←RH: moves right hand to left

↘: right hand moved towards right lower corner of screen* presenter is on the left

³⁴ Conventions with an asterisk * are mine

RH : right hand interacts with screen

LH : left hand interacts with computer

3. Body alignment

_ _ _ : body facing audience

/ / / : body partially facing the audience

| | | : body turned away from the audience

→→→ : moves to the right (presenter's left)


←←← : moves to the left (presenter's right)

b b b b : backward position

f f f f : forward position


4. Position


L of : left of screen

LL of : far left of screen

R of : right computer

RR of : far right of computer

a : at the computer*

| | | : body turned away from screen

Appendix G. LOG LIKE LIKELIHOOD AND BIC (Bayes Factor Approximation) TESTS

EXCEL WORKSHEETS (Rayson, 2017)

Significance and Effect Calculator (for comparing 2 corpora)
<http://ucrel.lancs.ac.uk/wizard.html>
 Last Updated: 2008/08/08

Step 1. Insert the actual frequencies in columns B and C alongside Word1 to Word6
 Step 2. Insert the corpus sizes in column B and C row TOTAL
 Step 3. Read off the resulting LL values in column F

	observed frequencies		expected frequencies		Over/under-use	Log Likelihood	normalised frequencies		%DIFF	Bayes Factor	BIC/ELL	Relative Risk	Log Ratio	Odds Ratio
	corpus1	corpus2	corpus1	corpus2			corpus1	corpus2						
Word1	1500	15000	1500.00	15000.00	-	0.00	0.15000	0.15000	0.00	-11.61	0.00000	1.00	0.00	1.00
Word2	340	2500	258.18	2581.82	+	26.18	0.03400	0.02500	36.00	14.57	0.00004	1.36	0.44	1.37
Word3	200	2500	245.45	2454.55	-	9.83	0.02000	0.02500	-20.00	-1.78	0.00002	0.80	-0.32	0.80
Word4	7	0	0.64	6.36	+	33.57	0.00070	0.00000	*****	21.96	-0.00068	#DIV/0!	7.13	#DIV/0!
Word5	654	654	118.91	1189.09	+	1447.84	0.06540	0.00654	900.00	1436.23	0.00275	10.00	3.32	10.63
Word6	89	536	56.82	568.18	+	17.38	0.00890	0.00536	66.04	5.77	0.00004	1.66	0.73	1.67
TOTAL	10000	100000												

N.B. Don't edit columns D, E and F. They implement the formulae for Expected frequencies and LL itself.
 Degrees of Freedom = 1
 %DIFF Zero Frequency Adjustm 1E-18
 Log Ratio Zero Frequency Adjustm 0.5 (Note that this is the adjustment for observed frequency before it is normalised)

Significance and Effect Calculator (6 Corpora instead of 2)

Step 1. Insert the actual frequencies in columns B-G alongside Word1 to Word6
 Step 2. Insert the corpus sizes in columns B-G row TOTAL

Sheet1 Sheet2 Sheet3

Log Likelihood Calculator (3 Corpora instead of 2)

Step 1. Insert the actual frequencies in columns B-G alongside Word1 to Word6
 Step 2. Insert the corpus sizes in columns B-G row TOTAL
 Step 3. Read off the resulting LL values in column O
 Table 2. Significance analyses of deictics use by level of achievement

	observed frequencies			Totals	expected frequencies			log likelihood	Bayes Factor	ELL
	High	Medium	Low		corpus1	corpus2	corpus3			
Represent	38	31	23	92	34.16	33.00	24.84	0.68	-21.71	0.00000
Tourist	41	33	11	85	31.56	30.49	22.95	10.49	-11.90	0.00005
Architect	24	18	8	50	18.57	17.93	13.50	4.08	-18.31	0.00002
Novice	75	63	58	196	72.78	70.30	52.91	1.33	-21.06	0.00000
Opinion ev	1	4	0	5	1.86	1.79	1.35	5.18	-17.21	0.00024
Innovation	54	16	10	80	29.71	28.70	21.60	30.45	8.06	0.00014
TOTAL	27038	26117	19656	72811						

N.B. Don't edit columns D-O. They implement the formulae for Expected frequencies and LL itself.
 Note: cells are ignored in LL calculation if observed frequency is 0
 Degrees of Freedom = 2

Academic roles disciplines classroom roles discipline Hoja3 Academic roles achievement Classroom roles achiev ...

Appendix H. TABLES FROM PERSONAL PRONOUNS STUDY (Nausa, 2016)

Table 10.1. Raw (R) and normalised (N) frequencies, and percentages (%) of self-mention realizations by pronoun number

	<i>I, me, my</i>			<i>we, us, our</i>			<i>Total</i>		
	<i>R</i>	<i>N</i>	<i>%</i>	<i>R</i>	<i>N</i>	<i>%</i>	<i>R</i>	<i>N</i>	<i>%</i>
<i>Academic Roles</i>									
a. Representative	94	19.7	10.1	0	0.0	0	94	19.7	10.1
b. Guide	23	4.8	2.47	21	4.4	2.26	44	9.2	4.73
c. Architect	171	35.8	18.4	9	1.9	0.97	180	37.7	19.3
d. Recounter / announcer	50	10.5	5.37	185	38.8	19.9	235	49.2	25.2
e. Opinion-holder	35	7.3	3.76	22	4.6	2.36	57	11.9	6.12
f. Originator	43	9.0	4.62	22	4.6	2.36	65	13.6	6.98
subtotal	416	87.2	44.7	259	54.3	27.8	675	141.4	72.5
<i>Classroom Roles</i>									
<i>Knowledge Contribution Roles</i>									
g. Learner	8	1.7	0.86	1	0.2	0.11	9	1.9	0.97
h. Co-constructer	40	8.4	4.3	105	22.0	11.3	145	30.4	15.6
i. Provider	50	10.5	5.37	29	6.1	3.11	79	16.6	8.49
subtotal	98	20.5	10.5	135	28.3	14.5	233	48.8	25
<i>English Language Competence Roles</i>									
j. Learner	2	0.4	0.21	0	0.0	0	2	0.4	0.21
k. Independent user	6	1.3	0.64	0	0.0	0	6	1.3	0.64
l. Provider	15	3.1	1.61	0	0.0	0	15	3.1	1.61
subtotal	23	4.8	2.47	0	0.0	0	23	4.8	2.47
TOTAL	537	112.5	57.7	394	82.6	42.3	931	195.1	100

Table 10.2. Log likelihood and Bayes Factor (BIC) values by levels of achievement

		observed frequencies			Totals	log likelihood	Bayes Factor BIC
		High	Medium	Low			
a.	Representative	37	26	31	94	1.66	-19.88
b.	Guide	14	26	4	44	15.52	-6.03
c.	Architect	71	55	54	180	0.62	-20.92
d.	Recountner / announcer	97	92	46	235	9.65	-11.90
e.	Opinion-holder	24	11	22	57	6.04	-15.50
f.	Originator	50	7	8	65	39.71	18.16
g.	Learner	4	1	4	9	2.58	-18.97
h.	Co-constructor	91	36	18	145	37.67	16.13
i.	Provider	32	33	14	79	5.21	-16.34
j.	Learner	0	2	0	2	4.42	-17.13
k.	Independent user	2	4	0	6	4.99	-16.56
l.	Provider	12	3	0	15	14.39	-7.15
Corpora sizes		18473	15827	13428	47728		

Table 10.3. Log likelihood and Bayes Factor (BIC) values by disciplines

		Observed frequencies		log likelihood	Bayes Factor BIC
		Hard	Soft		
a.	Representative	47	47	0.03	-10.75
b.	Guide	32	12	8.78	-1.99
c.	Architect	84	96	1.25	-9.53
d.	Recountner / announcer	157	78	24.53	13.76
e.	Opinion-holder	4	53	51.69	40.92
f.	Originator	47	18	12.46	1.69
g.	Learner	6	3	0.92	-9.85
h.	Co-constructor	72	73	0.08	-10.69
i.	Provider	55	24	11.49	0.72
j.	Learner	1	1	0.00	-10.77
k.	Independent user	5	1	2.78	-7.99
l.	Provider	10	5	1.54	-9.24
Corpora Sizes		24259	23469		

Appendix I. KEYWORD LIST IN OPs SUBCORPUS / REFERENCE CORPUS: BROWN

(Francis & Kučera, 1964)

Keyword List Results 1

Types Before Cut: 5835			Types After Cut: 99	Search Hits: 0
Rank	Freq	Keyness (LL)	Keyword	
1	2046	1740.816	is	
2	1344	1590.955	this	Chapter 4 (deixis)
3	297	894.313	example	Chapter 7 (code glosses)
4	148	698.563	don	
5	121	670.888	colombia	
6	104	546.524	ok	
7	691	543.793	you	Chapter 5 (audience identity)
8	491	532.222	can	Chapter 6 (impersonal identity)
9	587	519.648	we	Chapter 8 (expression of modality)
10	200	478.562	problem	Chapter 6 (impersonal identity)
11	5997	460.252	the	
12	283	412.155	because	Chapter 7 (code glosses)
13	187	385.692	important	Chapter 8 (code glosses)
14	161	338.935	different	
15	79	333.868	yeah	Chapter 6 (impersonal identity)
16	2102	324.515	in	
17	54	304.527	metamodel	
18	62	296.762	topic	Chapter 6 (impersonal identity)
19	576	279.290	have	Chapter 6 (impersonal identity)
20	342	265.027	about	
21	100	264.998	theory	Chapter 6 (impersonal identity)
22	70	257.385	species	
23	1161	254.861	that	Chapter 4 (deixis)
24	76	249.321	solution	Chapter 6 (impersonal identity)
25	100	240.577	talk	
26	42	236.854	phd	
27	78	234.470	model	Chapter 6 (impersonal identity)
28	266	227.840	my	
29	105	226.173	process	Chapter 6 (impersonal identity)
30	55	225.638	okay	

Appendix J. CONCORDANCE SEARCH FOR DEICTICS

Concordance Hits 83		
Hit	KWIC	File
1	of the of the stridulation . And in you see here a laser that we use to record that eh	S-H-CBIO-4,
2	wanted to know what is going on close to here . A minute ? Oh ! Ok , or you can extend it ,	S-H-CBIO-6,
3	eh involve a complete universe . What is a universe here ? A universe here is that we are taking a	S-H-INGE-2,
4	't found match between them from there and from here , And occurrence of dolphins is very common , but like	S-H-CBIO-5,
5	and send the bill to the the treasury department , here , and the treasury department has to low the managers	S-H-INGE-7,
6	show that this population the populations of dolphins is here , and they compare de population from unintelligible to other	S-H-CBIO-5,
7	in the second year eh of the PhD programme here at at this university and the name of my	S-L-MATE-2,
8	my eh master eh research eh in education , specially here at the Universidad de los Andes and the Cooperative	S-M-CPOL-3,
9	high , it was small one but the problem started here because lot of foreign capital start to to go	S-H-ECON-1
10	reason , the nucleus of the salt is blue . Eh , here eh also is used eh the ADN eh prove .	S-M-FISI-1_t
11	eh have two main parts . The primary distribution system , here , eh and here the secondary distribution system . The difference	S-M-INGE-8
12	can see different eh organizations , structures , types of cells . Here , eh application for b for for small systems . Here	S-M-FISI-1_t
13	the study or the treatments of these problems . Eh , here eh I will show you only three . Eh , first ,	S-H-ANTR-3
14	to eh travel across the sample . The sample is here eh in the stage if the microscope . Eh , here	S-M-FISI-1_t
15	show eh the mathematical expression for the problem eh here eh is the expression mathematical for the optimization problem .	S-M-INGE-8
16	ha eh have a lot of differences , for example here eh the car has four wheels , has engine , many	S-H-INGE-4,
17	century , the last century eh we recognize that importance . Here eh we are took quotation that eh Paul Samuelson	S-H-ECON-2
18	here , is the her ubication into the cell . Eh , here eh we can see a a tissues as the	S-M-FISI-1_t
19	eh which often try to copulate with the flower . Here eh we have an example of species of genera	S-H-CBIO-7,
20	was 1994. And poor quality , I was trying to put here examples from the other parts different to Colombia , but	S-H-INGE-2,
21	. So eh in this case , as you can notice here for example in the university there are some exchange	S-H-EDUC-1
22	see examples of different proves used eh in cells . Here for example , the researcher use a blue eh prove	S-M-FISI-1_t
23	inflorescent shape , flower colour eh scent and nectar guides . Here , for example , this orchid eh no se ve , this	S-H-CBIO-7,
24	the compaction level of the hough highway . For example , here green is good compaction but eh red is bad	S-L-INGE-5,

Appendix K. CONCORDANCE SEARCH FOR *YOU*-IDENTITY CASES

Concordance Hits 691			File
227	, for example, the tayronas, the chibchas eh, eh did you compare this tradition with these other cultures that are		S-L-ANTR-5,
228	devices and consume consume energy, but what happen if you connect and connect devices, the the system of your		S-M-INGE-7,
229	, eh the asymmetric information in signal theory, eh if you contrast with the first theory eh only have the		S-H-ECON-3,
230	a an idea that you can save money if you control the compaction process. In these days, civil engineers		S-L-INGE-5,
231	important than the economy, try holding your breath while you count your money, by Dr. Guy McPherson. Thank you.		S-M-INGE-1
232	. Okay, refers to the to the rules eh when you define a metamodel, you can to eh define the		S-M-INGE-6
233	complete eh elements and for example here you can you define eh four clients with specific information and two		S-M-INGE-6
234	, the simulation parameter parameters and the source model. Here you define the characteristics of the material and how the		S-H-INGE-6,
235	kind of materials, the other process is how do you define the seismic hazard and the final process is		S-H-INGE-6,
236	analysis. What do you do in the seismic hazard? You define which is the level of hazard of the		S-H-INGE-6,
237	with eh use of the the the the exam, you diagnostic of the faulting system. And the se three		S-L-INGE-4,
238	the diversity. So that is why this star shows you different features that we need to include in our		S-H-EDUC-1
239	philosophy you can cut off all this process, because you do arrangement with a supplier that you can trust		S-H-INGE-7,
240	the final process is the structural analysis. What do you do in the seismic hazard? You define which is		S-H-INGE-6,
241	point is the vision. The vision is what do you do in the future, for example in three years,		S-M-INGE-5
242	, its impossible to find eh a good job if you do n't have at least eh high school		S-H-ECON-3
243	is attractive because if you trust of him, then you do n't have to do this operation, because		S-H-INGE-7,
244	have controls and then when you get the products you do n't have to do the process again,		S-H-INGE-7,
245	of just-in-time philosophy you can cut off you do n't need a warehouse, unintelligible, you do		S-H-INGE-7,
246	off you do n't need a warehouse, unintelligible, you do n't need a big warehouse, you do		S-H-INGE-7,
247	, unintelligible, you do n't need a big warehouse, you do n't need a technology to control all		S-H-INGE-7,
248	, I need to keep inventory because what happens if you do n't have inventory? What happens if a		S-H-INGE-7,
249	extension eh some tails but in the graded ring you do n't have you do n't have		S-L-MATE-2,
250	in the graded ring you do n't have you do n't have tails and eh the graded		S-L-MATE-2,

Appendix L. CONCORDANCE SEARCH FOR ADJ+THAT PATTERNS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List		
Concordance Hits 52		
Hit	KWIC	File
25	?_SENT Because eh the reach in economy are learning about eh that methodological eh vision _SENT Here I I show you	S-H-ECON-2
26	ng potential high growth firms is very difficult _SENT And second , that new firms in general deserve policy support due to	S-H-ADMI-2
27	book 3_CD of The Republic _SENT For Plato it was clear that on that only philosophers represent the gold , eh th	S-L-FILO-2_1
28	gy init initiatives , and personal objectives _SENT Is very important that point because the company is eh is conformada R _	S-M-INGE-5
29	The advantages eh for the use the resorcinarenes eh is eh that resorcinarenes no not are toxic , are excreted in u	S-L-CQUI-1_
30	company pay a lot of money for any software , is possible that software don't resolve the needs of the company _	S-M-INGE-5
31	ago _SENT M : yes , if you look look is so poor that that scenario and you eh watch the knows or	S-H-ECON-1
32	need education we need educate to our child because is necessary that the child know what is a specie and what	S-H-CBIO2_1
33	conditions of life of the other citizens , and maybe we unknown that the conflict produce pain , produce inequality , produce	S-M-EDUC-1
34	Eh , eh , we present eh a project that the hort horticultural that the fruit and vegetable sector _SENT That have eh	S-L-INGE-3_
35	romotion and finally the conclusions _SENT The the issue is eh that the household responsibilities are increa increasingly shared	S-L-ADMI-1_
36	eh the for this reason I conclude that this eh true that the PFA eh promote the corruption _SENT And the	S-M-ANTR-2
37	case of the skew PBW extension _SENT I tell us eh that the problem when you try to characterized characterize	S-L-MATE-2
38	the most important because eh they are the eh respon responsible that they are responsible of infection process in different h	S-H-CBIO1_1
39	occur the combination or union with copper _SENT Is very important that this molecules have electron non-bonding pair _SENT And	S-M-CQUI-1
40	team and is not one hundred percent eh sure or confident that this that this could eh predict all the bugs _	S-H-INGE-3_
41	NN about interculture or culture and language , so that is necessary that we as an English teacher include in our classes	S-H-EDUC-1
42	so _SENT Eh I can perceive this situation that is eh that we have a lot of messages about health care ,	S-H-ANTR-2
43	X by a variable Y , X by Y is the same that Y by X , but in the skew PBW extension	S-L-MATE-2
44	in specific some bays _SENT In some bays is particularly eh that you can find eh all with the same dolphins ,	S-H-CBIO-5_
45	does n't resolve eh the problem because it is possible that you can find another good road that solves the	S-H-MATE-2
46	A : Hi , I'm talking about self-deception _SENT Something different that you eh that you talk before _SENT Eh , what	S-H-PSIC-1_
47	in Bogotá a few years ago eh and that is possible that you have you have eh seen this on the	S-H-DERE-1
48	chool . havinn a male teacher ? _SENT Eh that 's nrnbahlv eh that you that that you had teach many teachers eh	S-I-FDUC-2

Appendix M. CONCORDANCE SEARCH FOR CODE GLOSSES (*IN THIS CASE*)

Tool Preferences Help			Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List
Concordance Hits 62			
Hit	KWIC	File	
1	is ve interesting . For example , if the surface or in this case a curve present a singular point . This	S-M-MATE-1	
2	end in the same point , then in this eh in this case , algebraic topology assigns not a number , algebraic	S-H-MATE-2	
3	presupposes a close relationship among different countries . So eh in this case , as you can notice here for example	S-H-EDUC-1	
4	eh appear that the selection with those methodologies . But in this case avocado is the product that we have	S-L-INGE-3_	
5	anthropologist didn't have to the colonialism enterprise And in this case _ B : Bernardo A : In this case this	S-L-ANTR-1_	
6	in a considerable way . You can see for example , in this case between the two thousand three and two	S-M-INGE-7_	
7	political crisis and third disappoint for the religion , eh in this case disappoint of traditional religion is very important	S-L-ANTR-4_	
8	can see this this signal . Eh in this eh in this case eh in this case eh was two	S-H-ECON-3	
9	signal . Eh in this eh in this case eh in this case eh was two people one of them	S-H-ECON-3	
10	have a couple of circles and another circle and in this case eh these two topological objects are the	S-H-MATE-2	
11	considered a fetishist . The fetishist is very very negative in this case . Eh Anthropologists have characterized the meeting in	S-L-ANTR-4_	
12	the second possibility inability for the default process legislature . In this case , eh the process , followed in the in	S-L-DERE-2_	
13	today . Eh the the most important eh change eh in this case eh was the separation the church of	S-L-HIST-2_1	
14	navigation of the displacement in a in in this in this case , eh in comparison , eh the same eh	S-L-INGE-4_	
15	coordination occur need electron non-bonding pairs . Be eh in this case eh the pyrazole have two electron non-	S-M-CQUI-1	
16	need to define who is the machine . The machines in this case eh is the student with all of	S-M-INGE-4	
17	reality is are discovered are discovery and in this in this case eh for example you need add information	S-M-INGE-6	
18	eh , eh can be eh formulate as formulas eh . In this case eh this is a surface eh suppose	S-M-MATE-1	
19	, about opinions , about other things that are not quantitative . In this case , for example , the criteria to choose a	S-H-INGE-2_	
20	the coefficients , is are in the in other rings . In this case , for example , the coefficients are in the	S-L-MATE-2_	
21	obtained eh through eh lineal synthesis . Eh I eh in this case I acompling two compounds and I will	S-H-CQUI-2_	
22	in specific , language . Ok , for example , if I in in this case I take a I I have a	S-H-MATE-1	
23	vehicle . Eh this vehides to eh know with names . IN this case I eh names for bluffing , spray , leader .	S-L-INGE-2_	
24	get the extremal graph . Eh ok , the problem is , in this case I have a forbidden configuration , what is	S-L-MATE-1	

Appendix N. MANUAL IDENTIFICATION OF PAIRS OF ESSAY-OPs SENTENCES

effect of the decision (the right to call to account is real). Second, involve a respectful and courteous treatment (personal feeling) and it is *includes* the right to speech, allowing the exchange of information on what is considered a fair deal. Furthermore Zitek et al. (2010) found that the entitlement serves as a mediator of selfish behavior, i.e., the victim of injustice acquires a sense of entitlement that leads to selfish behavior.

Indeed, the entitlement is a consistent construct across the studies, *broadly speaking, I propose the following definition of entitlement: it is a set of principles, informal and implicit rules, that constitute a psychological contract, it is related to the courteous treatment, respect and accountability. Besides the entitlement is based on the belief or perception by individuals who are deserving of a right to do or have something to aspire to something, or be someone in particular, without necessarily being linked to a legal right (can be informal) without that is related to a real effort or input; right as its basis is the dignity of persons.*

In the previous definition the entitlement is conceived positively, as a guarantee of a right, yet it is also perceived as a potential threat that can affect different people, and could be considered as a narcissistic trait that breaks societal rules. For example, when people claim rights to do or have something, but have not tried to get it or not society recognizes as worthy of that right. This is the case of extended academic rights, leading to students not to strive and claim a good note, simply by attending class, or because they pay tuition (Sparks, 2012). *The negative consequences of the entitlement related to narcissism, are related to an exaggerated self-esteem and an unwarranted demand for special treatment; this behavior is seen as pathological and socially undesirable.*

To conclude, we can define a macro framework to measure an entitlement. The entitlement has *expanded three-dimensional shapes or intensities: a sense of the appropriate entitlement (associated*

entitlements and many I in *good* present the measure of entitlements, because the measure is the problem to building a better eh public policy. *The theory of entitlements eh has two sides, one is the negative size that is related with narcissism eh that is exaggerated self-esteem eh and demand for special treatment, and the positive side, that is eh that I want to explore is a extended right.* Why eh I want to explore that eh concept? Eh because the victims, when become to the eh unit of victims eh, they don't know what *are the rights* eh that they have, but then, when he [fs] they receive a help they becomes a, [fs] they get beggars or beg for more rights that they don't have. Eh some questions that *illus*, [fs] eh made [fs] eh show what *is the entitlement* is what they deserve from others and what rights they have in relation with others.

Eh this is some scales that eh the theory have to eh measure the intensity of narcissism. The first one eh present a person that is self-deprecatd because they feel eh less than the others and the seven eh is a person narcissist, [fs] who is narcissist. *The definition of entitlement that, [fs] positive entitlement is a [reading 1] principles informal and implicit rules that constitute a psychological contract and is related to the courteous treatment and respects and accountability [reading 1] and besides is based on [fs] in the belief [fs] on the belief that the people perceive they deserve from the others eh and the basis of this kind of entitlement is the dignity.*

Eh the relation with the procedural justice eh *bueno*, well I need eh explain procedural justice first. Procedural justices eh refers a justice in the process [fs] in the process of made a decision and we need that this decision eh was taken without vice (bias?) and

ORAL PAIRS OF SENTENCES

[illegible]