



UNIVERSITAT DE
BARCELONA

Task-Modality Effects: A Study of Task Complexity Effects in Speech and Writing

Olena Vasylets

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE
BARCELONA

Task-Modality Effects: A Study of Task Complexity Effects in Speech and Writing

TESI DOCTORAL

presentada per

Olena Vasylets

2017

TASK-MODALITY EFFECTS: A STUDY OF TASK COMPLEXITY EFFECTS IN
SPEECH AND WRITING

TESI DOCTORAL

presentada per

Olena Vasylets

Directors:

Dr. Roger Gilabert Guerrero (Universitat de Barcelona)

Dra. Rosa M. Manchón (Universidad de Murcia)

Tutor: Dr. Roger Gilabert Guerrero (Universitat de Barcelona)

Programa de Doctorat: Lingüística Aplicada

Departament de Llengües i Literatures Modernes i d'Estudis Anglesos

Universitat de Barcelona

2017

To Lluís, Nina and Alex

ACKNOWLEDGMENTS

Completing this dissertation was a wonderful quest, which gave me a lot of joy and pleasure. I would not have been able to complete this quest without the help of wonderful people who accompanied me during this journey. First and foremost I would like to thank my supervisors Dr. Roger Gilabert and Dr. Rosa M. Manchón who accompanied me during this marvelous journey. I learned a lot from them, and I will be eternally grateful for their help and generosity. Without them, this journey would not be possible. I feel honored and privileged to be able to count on their support. Their help and the knowledge they shared with me are invaluable. I am extremely grateful to Dr. Roger Gilabert and Dr. Rosa M. Manchón for agreeing to supervise this dissertation and for being so supportive. I truly appreciate all the time and effort they invested in this project.

I would also like to express my gratitude to the head of the GRAL research group Dr. Carme Muñoz, who was so generous and supportive. I feel honored to be the member of the research group guided by Dr. Muñoz. I would also like to thank all my teachers from the Applied Linguistics MA program at the University of Barcelona who equipped me with all the necessary knowledge to complete this journey. I would like to thank Dr. Roger Gilabert, Dr. Joan Carles Mora, Dr. Elsa Tragant, Dr. Raquel Serrano, Dr. Imma Miralpeix, Dr. María Luz Celaya, Dr. Oliver Strunk and Dr. Marta Fernández-Villanueva for the knowledge they gave me. I am grateful to Dr. Julio Roca de Larios and Dr. Javier Marín for their valuable advice. I would also like to express my gratitude to Marina Ruiz Tada for the support and the encouragement she gave me.

ABSTRACT

One of the major areas of interest in task-based language learning and teaching (TBLT) has been the psycholinguistically-oriented strand, which explores the way in which task design may interact with second language (L2) learners' cognitive response in creating distinct opportunities for L2 learning and use. The idea that the controlled adjustment of task characteristics may produce predictable effects on L2 products and processes is very appealing to teachers, syllabus designers, testers as well as researchers. The main appeal of TBLT ideas is that a language task is presented as a manageable and, at the same time, powerful tool of instruction and research. What is necessary, however, is a clear understanding of the manner in which different task characteristics might affect learners' performance and development.

There are two major theoretical frameworks which provide an account of the purported effects of task features on L2 learning and use. These theoretical models have been developed by Peter Robinson (the Cognition Hypothesis, Robinson, 2011a) and Peter Skehan (the Trade-off Hypothesis, Skehan, 1998, 2009). The main construct of interest in both models is the construct of task complexity. Skehan and Robinson define this construct in a slightly different ways and they also make somewhat different predictions about the effects that task complexity might have on L2 products and processes. Robinson's and Skehan's predictions have been put to the empirical test and this research has produced mixed results. In addition, as the models have been created to account primarily for oral production, it is still an empirical question whether or not the predictions made by the two models can also be applied to written performance. In other words, it is still unknown whether or not task complexity produces the same effects in

speech and in writing. Moreover, mode (oral versus written) by itself may constitute an important task feature that can condition language use and learning opportunities.

At a more general level, it is also the case that current understanding of the idiosyncrasy of the language-learning potential of oral and written tasks is limited. Given this state of affairs, the objective of the current dissertation is twofold. Our first aim (operationalized as our first research question) is to explore whether or not L2 linguistic performance is different when the same task is performed in speech and in writing. Our second aim (corresponding to our second research question) is to investigate any potential effects of increasing task complexity in oral production as compared to written production. In terms of task complexity manipulations, we employed the Cognition Hypothesis as a theoretical framework and we specifically investigated the impact of the resource-directing variable of reasoning demands on L2 performance, which was operationalized as complexity, accuracy and time on task. To answer our research questions, we conducted a study with 78 participants who were Catalan/Spanish learners of English as a foreign language. Half of the participants performed the simple and complex versions of an argumentative, instruction-giving task orally, the other half did it in writing. In the analysis, we compared speakers' and writers' performance in terms of linguistic (lexical and structural) complexity, propositional complexity (operationalized as idea units), accuracy, and time on task. Our results revealed marked differences between oral and written production. Thus, we found that speakers produced more idea units, while writers achieved higher scores for subordination, mean length of analysis-of-speech units, lexical diversity, extended idea

units, and time on task. As for the effects of task complexity, the participants' written production showed more variation between the complex and the simple versions of the task. Changes in the written production also showed a better fit to the theoretical predictions advanced in the Cognition Hypothesis. These findings are interpreted in light of task modality effects in L2 learning and discussed in relation to task complexity theory and research.

RESUM

Una de les línies principals de l'àrea de l'ensenyament basat en les tasques pedagògiques (TBLT) és la línia psicolingüística. Aquesta línia explora com el disseny de les tasques interactua amb la resposta cognitiva dels aprenents creant diverses oportunitats per l'ús i aprenentatge de la llengua. La idea d'una controlada manipulació de les tasques amb els efectes controlats en l'ús i aprenentatge de les llengües és molt atractiu per a professors, dissenyadors dels continguts pedagògics, avaluadors i també investigadors. L'atractiu principal de la idea de TBLT és que la tasca es presenta com un instrument fàcil d'usar i potent per la instrucció pedagògica i recerca. Però el més essencial és una clara comprensió de com les característiques de les tasques influeixen en l'ús i aprenentatge de la llengua. Hi ha dos models teòrics principals que ofereixen prediccions de les possibles efectes de les tasques en l'ús i aprenentatge de la segona llengua. Aquests models han set desenvolupats per Peter Robinson (the Cognition Hypothesis, Robinson, 2011a) i Peter Skehan (the Trade-off Hypothesis, Skehan, 1998, 2009).

El concepte principal en els dos models és la complexitat cognitiva de les tasques. Skehan i Robinson defineixen la complexitat cognitiva des d'un enfocament divers i ofereixen diferents prediccions sobre els efectes de complexitat cognitiva. Els estudis empírics sobre les prediccions d'aquestes teories han produït resultats mixtes i contradictoris . Aquests models han set creats per l'aplicació en la producció oral però no està clar si aquestes prediccions es poden aplicar a les produccions escrites. Amb altres paraules encara no està del tot clar si els efectes potencials de la complexitat cognitiva són els mateixos en la producció oral i escrita.

Cal destacar que la modalitat (oral i escrita) també pot constituir un característica de la tasca que pot influir amb l'ús de la llengua i les oportunitats d'aprenentatge. És important destacar que hi ha una manca coneixement sobre la singularitat del potencial de l'aprenentatge de les tasques orals i escrites.

Donada aquesta situació, aquesta dissertació té dos objectius. El nostre primer objectiu és explorar si la producció lingüística és diferent en una tasca oral i escrita. El segon objectiu és comparar els efectes de la complexitat cognitiva de la tasca en la producció oral i escrita. Aquesta dissertació investiga específicament els efectes de la manipulació de la dificultat del nivell de raonament de la tasca en la complexitat precisió lingüística de la producció i també el temps utilitzat per l'execució de la tasca. En aquesta dissertació utilitzem the Cognition Hypothesis com a model teòric. Per complir aquests dos objectius s'ha dut a terme un experiment amb 78 participants (nadius catalans i castellans) dels quals estaven aprenent anglès com a segona llengua. La meitat dels participants han realitzat les versions simple i complexa de la tasca argumentativa de manera oral i la meitat restant ha realitzat les mateixes tasques per escrit. A l'anàlisi hem comparat la complexitat lèxica i sintàctica, el nombre i tipus d'idees produïdes, precisió lingüística i el temps per realitzar la tasca. Hem trobat que els participants que han fet la tasca de manera oral han produït més idees però cal esmentar que els participants que ho han fet de manera escrita han produït textos amb més subordinació, diversitat lèxica, idees més complexes, unitats lingüístiques més llargues i amb un temps més llarg per l'execució de la tasca. En relació amb els efectes

de la complexitat cognitiva de la tasca hem trobat que en la producció escrita havia més varietat entre la versió simple i complexa de la tasca. A més a més els efectes de la complexitat cognitiva en la producció escrita s'han ajustat més a les prediccions teòriques de la Cognition Hypothesis. Els resultats obtinguts van ser interpretats en el context dels marcs teòrics pels quals està guiada aquesta dissertació. Allà on va ser possible, els resultats són contrastats amb recerca prèvia.

TABLE OF CONTENTS

PART 1. THEORETICAL BACKGROUND

Chapter 1 **Overview of the dissertation** **2**

1.1 Introduction, 2

1.2 The intended contribution of the current dissertation, 3

1.3 The chapters in the current dissertation, 3

Chapter 2 **Task-based language teaching** **6**

2.1 Introduction, 6

2.2 Conceptual bases of TBLT, 6

2.3 Tasks from a psycholinguistic perspective, 9

2.4 Skehan`s “cognitive approach” to tasks, 11

 2.4.1 The rule-based and the exemplar-based systems, 11

 2.4.2 Skehan`s framework of task-based instruction, 13

 2.4.3 Skehan`s Trade-off Hypothesis, 17

2.5 Robinson`s Cognition Hypothesis, 20

 2.5.1 The Triadic Componential Framework, 21

 2.5.1.1 Task Complexity, 22

➤	Resource-directing variables, 22	
➤	Resource-dispersing variables, 26	
2.5.1.2	Task condition, 27	
2.5.1.3	Task difficulty, 28	
2.5.2	Implications for task design, 29	
2.6	Task demands and stages of speech production, 30	
2.7	Skehan`s and Robinson`s models and written task performance , 32	
2.8	Chapter summary, 38	
Chapter 3	Mode in theoretical TBLT agenda	40
3.1	Introduction, 40	
3.2	Language production in speech and writing, 41	
3.2.1	Models of oral and written production, 42	
3.2.1.1	L1 speech production: Levelt`s (1989) model, 43	
3.2.1.2	Bilingual and L2 speech production: Kormos`s (2006) model, 45	
3.2.1.3	L1 writing: Hayes and Flower`s (1980) model and subsequent revisions, 48	
3.2.1.4	L1 writing: Kellogg`s (1996) model, 51	
3.2.1.5	L1 writing: Van der Pool`s (1995) and van Wijk`s (1999) models, 53	

- 3.2.1.6 Models of L2 writing, 55
- 3.2.1.7 Comparison of oral and written production models, 57
- 3.2.2 Psycholinguistic comparison of speech and writing, 58
 - 3.2.2.1 Physical characteristics of the act of production and output, 58
 - 3.2.2.2 Relationship with the audience, 60
 - 3.2.2.3 Learners` beliefs and associations, 62
- 3.2.3 Comparison of speech and writing with planned and unplanned discourse, 63
- 3.2.4 Summary of the psycholinguistic comparison of speech and writing, 64
- 3.3 The language learning potential of L2 writing, 65
- 3.4 Juxtaposition of the language learning potential of oral and written tasks, 68
 - 3.4.1 Input and input processing, 70
 - 3.4.2 Intake and intake processing, 71
 - 3.4.3 L2 knowledge and knowledge processing, 72
 - 3.4.4 The language learning potential of oral and written output practices, 73
 - 3.4.5 Mode in TBLT research agenda: Conclusions and future research, 76
- 3.5 Chapter summary,78

Chapter 4 Mode and task complexity in empirical TBLT research 79

4.1 Introduction, 79

4.2 Mode effects in L2 performance, 80

4.2.1 Review of the empirical studies, 80

4.2.2 Critical analysis of the mode effects studies, 86

4.3 Task complexity effects in L2 oral and written production, 89

4.3.1 Reasoning demands, 91

4.3.2 The effects of reasoning demands on L2 oral production, 93

4.3.3 The effects of reasoning demands on L2 written production, 96

4.3.4 Critical analysis of the studies which explored
the effects of reasoning demands on L2 production, 103

4.4 Studies which compared task complexity effects in
L2 oral and written production within the same experimental design, 105

4.4.1 Review of the empirical studies, 105

4.4.2 Critical analysis of the studies which compared
task complexity effects in the two modes, 112

4.5 Chapter summary, 116

**Chapter 5 Propositional complexity: Definition and operationalization
in L1 and L2 research**

118

5.1 Introduction, 118

5.2 L2 complexity, 119

5.2.1 View on complexity in typological linguistics, 121

5.2.2 View on complexity in functional linguistics, 123

5.2.3 Functional-typological view on complexity, 128

5.2.4 Summary of the views on complexity, 131

5.3 Propositional complexity, 132

5.3.1 Conceptualizations of idea units in L1 English, 133

5.3.2 Summary of the theoretical positions on idea units, 140

5.3.3 Empirical research on idea units in L1 English, 141

5.3.4 Empirical research on idea units in L2 English, 147

5.3.5 Summary of the empirical findings on idea units, 149

5.4 Chapter summary, 150

PART 2. THE EMPIRICAL STUDY

Chapter 6 Aims and research questions 153

6.1 The focus of the study, 153

6.2 Research questions and hypotheses, 154

Chapter 7 Methodology 157

7.1 Introduction, 157

7.2 Participants and context, 158

7.3 Tasks and procedures, 162

7.3.1 The experimental task and the operationalization of task complexity, 162

7.3.2 Procedure, 165

7.3.3 Data coding and data analysis procedures, 167

7.3.3.1 Measures of linguistic complexity, 168

➤ Measuring lexical complexity, 168

- Measuring lexical diversity, 169
- Measuring lexical sophistication, 171

➤ Measuring structural complexity, 173

- Measuring overall structural complexity, 174
- Complexity via subordination, 175
- Complexity via phrasal elaboration, 176

7.3.3.2 Measuring propositional complexity, 177

- Semantic criteria, 178
- Intonational criteria, 179
- Syntactic criteria, 180
- *Non-extended* versus *extended* idea units, 184
- Relationships between idea units and other units of discourse, 185
- Measures of propositional complexity, 188

7.3.3.3 Assessment of accuracy, 189

7.3.3.4 Time on task measures, 192

7.3.3.5 Intra-rater reliability, 193

7.3.3.6 Summary of the measures used in the study, 193

7.4 Statistical analyses, 194

7.5 Chapter summary, 196

Chapter 8	Results	197
------------------	----------------	------------

8.1 Introduction, 197

8.2 Research question 1, 197

8.3 Research question 2, 201

8.4 Summary of results, 204

9.1 Introduction, 206

9.2 Research question 1: Oral versus written task performance, 206

9.2.1 Linguistic complexity, 208

9.2.1.1 Lexical complexity, 208

- Higher lexical diversity in writing, 209
- Similar levels of lexical sophistication in speech and writing, 210

9.2.1.2 Structural complexity, 212

- Higher overall complexity and complexity via subordination in writing, 213
- Similar levels of noun-phrase complexity in the two modes, 214

9.2.1.3 Interpretation of the results for linguistic complexity, 216

9.2.2 Propositional complexity, 216

9.2.2.1 Higher number of idea units in speech, 217

9.2.2.2 Higher ratio of extended idea units in writing, 219

9.2.2.3 Similar length of idea units in speech and writing, 221

9.2.2.4 Interpretation of the results for propositional complexity, 224

9.2.2.5 More versus different propositional complexity, 227

9.2.3 Similar accuracy in speech and writing, 227

9.2.4 Higher time on task in writing, 236

9.2.5 Summary and evaluation of the mode effects on L2 production, 237

9.3 Research question 2: The effects of task complexity in speech and writing,	241
9.3.1 Similar effects on linguistic complexity in speech and writing,	243
9.3.1.1 Similar levels of lexical diversity, but higher lexical sophistication in the complex task for writers and speakers,	244
9.3.1.2 Higher overall structural complexity, but no changes in subordination and noun-phrase complexity in the two modes,	245
9.3.2 More idea units in both complex oral and complex writing tasks, but higher ratio of extended ideas only in the complex writing task,	249
9.3.3 Higher accuracy in complex writing task, but no differences in speech,	251
9.3.4 Higher time on task in complex writing task, but no differences in speech,	256
9.3.5 Global interpretation of the findings,	258
9.4 Reconsidering models of task complexity,	262

Chapter 10 Conclusion, implications, limitations and suggestions for further research	264
References	270
Appendix “Fire-chief” task	302

LIST OF TABLES

Table 1	Definitions of “task” in TBLT literature, 10
Table 2	Studies which have investigated language performance in the oral and written mode, 86
Table 3	Empirical studies that have explored the effects of reasoning demands on L2 performance, 100-102
Table 4	Empirical studies that have compared task complexity effects in oral and written production, 114 - 115
Table 5	Descriptive statistics for the X_Lex and Y_Lex proficiency tests, 161
Table 6	Units of discourse analysis, 187
Table 7	Cohen`s kappa intrarater reliability coefficients for ratings of complexity and accuracy measures, 193
Table 8	Summary of task performance measures used in the study, 194
Table 9	Descriptive statistics for simple and complex oral and written tasks, 198
Table 10	Univariate test results for the main effect of mode, 200
Table 11	Univariate test results for the main effect of task complexity, 202
Table 12	Univariate test results for interaction effect between task complexity and mode, 204

LIST OF FIGURES

- Figure 1** Framework of task-based instruction (from Skehan, 1992), 15
- Figure 2** Revised framework of task-based instruction (Skehan, 1996), 16
- Figure 3** The Triadic Componential Framework for task classification - categories, criteria, analytic procedures, and design characteristics (from Robinson & Gilabert, 2007), 23
- Figure 4** Levelt's (1989) model of L1 speech production, 44
- Figure 5** Kormos's (2006) model of bilingual speech production, 46
- Figure 6** Hayes and Flower's model of L1 written production, 49
- Figure 7** Hayes's (2012) revised model of L1 written production, 50
- Figure 8** Kellogg's model of L1 written production, 52
- Figure 9** Model of written production inspired by Levelt's model (1989, adapted from van Wijk (1999), 54
- Figure 10** Zimmermann's (2000) model of L2 written production , 69
- Figure 11** Leow's (2015) model of the stages of learning processes in SLA, 74
- Figure 12** Effects of output in SLA processes (from Izumi, 2003, p.188), 74
- Figure 13** Taxonomic model of L2 complexity (Bulté & Housen, 2012), 120

PART I: THEORETICAL BACKGROUND

CHAPTER 1. OVERVIEW OF THE DISSERTATION

1.1 Introduction

Task-based language learning and teaching (TBLT) has received considerable attention from second language (L2) acquisition researchers as well as practitioners. Much of TBLT-oriented research endeavors have taken a psycholinguistic perspective, which sees tasks as pedagogic devices that can “predispose, even induce learners to engage in certain types of language use and mental processing that are beneficial for acquisition” (Ellis, 2000, p. 197). Notwithstanding notable theoretical and empirical achievements, the field of TBLT has an acknowledged limitation derived from privileging oral over written production (see, for instance, Byrnes & Manchón, 2014a, b; Gilabert, Manchón, & Vasylets, 2016).

This neglect of the written mode is problematic given that in a literate society both oral and written modes are indispensable and, accordingly, the command of both modes constitutes the hallmark of language proficiency (Berman & Ravid, 2009). In addition, both oral and written language use represent potential language learning opportunities, hence the relevance of looking into language learning associated with both speaking and writing tasks.

In line with these ideas, recent positions in TBLT have stressed the need for more mode-sensitive and integrative TBLT research and for a deeper understanding of the unique learning benefits that each mode can offer separately or in mutual interaction (Gilabert, Manchón, & Vasylets, 2016). The aim of this dissertation is to contribute to

these calls for more mode-balanced TBLT agendas. We do so with an analysis of L2 performance across modalities as mediated by task complexity, a central TBLT concept that has been recently problematized in its application to writing.

1.2 The intended contribution of the current dissertation

The intended contribution of this dissertation is envisaged as being theoretical, empirical, methodological and pedagogical in nature. From the theoretical and empirical perspectives, the potential contribution of this dissertation is linked to the disciplinary debates on (1) the already mentioned need for more mode-integrative TBLT research agenda, (2) the distinctive nature of the language learning potential of L2 writing and (3) the relevance of putting TBLT tenets to the empirical test across modalities. The methodological contribution of our study is closely linked to the analysis of learners' production for propositional complexity, a performance dimension that, as we shall explain in the relevant chapters, has been partially neglected in L2 research but that, nevertheless, is crucial to uncovering mode-related differences and idiosyncrasies. The study has also pedagogical and practical implications as it provides information for syllabus designers and practioners who employ tasks in the language classroom.

1.3 The chapters in the current dissertation

This dissertation is divided into ten chapters. Chapters 2 to 5 offer the necessary background to our own empirical study, whose main components (aims, method, results and their discussion) are reported in chapters 6 to 9. The final chapter (Chapter 10) synthesizes the main contribution of this dissertation, and presents the limitations, implications and suggestions for future research.

Chapter 2 presents the conceptual bases of task-based language teaching and introduces the tenets of the Trade-off Hypothesis (Skehan, 1998, 2009) and the Cognition Hypothesis (Robinson, 2001, 2011a) because these two theoretical positions are the main conceptual frameworks which inform this research. In this chapter we also highlight the privileged position of the oral mode in TBLT theoretical thinking, and review the positions of a number of researchers who problematize the applicability of the Trade-Off Hypothesis and the Cognition Hypothesis to writing.

Chapter 3 explores, from the theoretical perspective, the singularities of speech and writing as production modes and as language learning sites. For this purpose, we first offer an overview of relevant models of speech and writing production in L1 and L2, which is followed by a comparison of the two modes in terms of the physical act of production and the nature of output, relationship with the audience, and learners' beliefs. We then outline recent theoretical positions on the connections between second language acquisition (SLA) processes and oral and written tasks. Finally, we present the main tenets of the theorizing concerning the language-learning potential of L2 writing as our study is ultimately concerned with shedding light on the manner in which oral and written tasks can contribute to L2 development.

Chapter 4 offers a review of the research which is relevant to our study. Given the aims of our dissertation, we focus specifically on three groups of studies: (1) studies that have compared L2 oral and written production, (2) studies that have explored the effects of +/- reasoning demands in speech and writing, and (3) investigations that have compared the effects of task complexity on speech and writing within the same study. The analysis of previous empirical research makes evident the notorious neglect of the dimension of propositional complexity, which we identify as a gap in previous research that our study aimed at filling.

The analysis of discourse for propositional complexity is an additional methodological contribution of this dissertation. For this reason, before introducing the research questions and the hypotheses guiding this dissertation, Chapter 5 elaborates on the theoretical stance that this dissertation takes on the complexity of discourse and reviews early theorizing and empirical work of the construct of idea units. This review is important as it lays the basis for the elaboration of our own guidelines for the segmentation of oral and written discourse into idea units. These guidelines are considered to represent the specific methodological contribution of this dissertation.

Chapter 6 presents the research questions guiding this dissertation and the corresponding hypotheses. This is followed by a detailed description of the methodology employed in this study (including the detailed guidelines for the identification of idea units), which is the content of Chapter 7.

Chapter 8 reports the results obtained, accounting first for the results obtained in relation to each research question separately and then providing a summary of all results. Chapter 9 discusses the results obtained in this dissertation in a double direction: our results are compared to those of previous research, in addition to assessing them in light of the overall aims guiding this dissertation.

The final and concluding chapter, Chapter 10, synthesizes the main contribution of the research presented, acknowledges its limitations, draws implications deriving from the results obtained and finishes by putting forward suggestions for future research.

CHAPTER 2. TASK-BASED LANGUAGE TEACHING

2.1 Introduction

This chapter describes the general conceptual bases of task-based language teaching (TBLT) and reviews the tenets of two cognitive TBLT models created by Peter Skehan (the Trade-off Hypothesis) and by Peter Robinson (the Cognition Hypothesis). These influential frameworks represent different and occasionally competing accounts of the way in which task features may interact with learners' cognitive/behavioral response during task performance. The chapter summarizes the main assumptions and predictions advanced by Skehan and Robinson, and reviews critical evaluations of the frameworks' tenets in terms of their applicability to written task performance. The chapter also highlights the limitations of the two models in the account for the written mode of production, the scarcity of theoretical thinking attuned to writing and a notorious imbalance of modes in TBLT scholarship.

2.2 Conceptual bases of TBLT

Over the last decades, much prominence has been given to TBLT – a form of a teaching approach that places pedagogic tasks modeling real-life activities at the center of the L2 syllabus (Ellis, 2003; Long, 2014; Long & Crookes, 1992; Robinson, 2011b). Early forms of TBLT can be traced back to the 1980s, when the theoretical foundations as well as effectiveness of the forms-oriented language pedagogy started to be questioned and, as an alternative, a range of more communicative and learner-centered approaches emerged. The central idea, that shapes the idiosyncrasy of TBLT and also conditions its

methodology, is the view of language as a rich and flexible resource of meaning-making and communication – a resource which can be optimally acquired in the context of functional use. This conceptual basis of TBLT is linked to the ideas of prominent educational thinkers, such as Vygotsky (1978) or Wells (1987), who argued that meaningful tasks, resembling real-life activities, represented an optimal tool to develop complex functional abilities in students.

As a teaching approach, TBLT can be defined as being analytic. The *synthetic* and *analytic* distinction was first introduced by Wilkins (1976) to classify language teaching syllabi. In a synthetic approach, language is divided into linguistic units, such as words or grammar rules. These discrete units constitute the main blocks of instruction, and the learners' task is to *synthesize* these blocks into a linguistic whole. Examples of the synthetic approach are the Grammar Translation and Audio-Lingual Methods. In an analytic approach, like TBLT, language is treated as discourse in use during task performance rather than divided into discrete elements. In the classroom, learners are directly exposed to the genuine target language samples, and the learner's work is to *analyze* this authentic input, inducing knowledge about smaller linguistic units and grammar rules. The emphasis is also on meaningful and holistic language use in a variety of communicative situations. TBLT can, thus, be defined as communication-based instruction, in which focus-on-meaning is of primary concern. The most recent conceptualizations of TBLT also allow, and even encourage, timely focus-on-form interventions, as they are believed to optimize learning and increase effectiveness of language learning processes (Long & Robinson, 1998; Norris & Ortega, 2000). It is important to highlight that “focus-on-form”, which is attention to linguistic elements, is also TBLT specific, and it has to be distinguished from “focus-on-formS”

practices, which refer to the work with isolated linguistic structures (Long, 1998). As Long (1998) puts it:

Focus on form refers to how attentional resources are allocated, and involves briefly drawing students' attention to linguistic elements (words, collocations, grammatical structures, pragmatic patterns, and so on), in context, as they arise incidentally in lessons whose overriding focus is on meaning, or communication, the temporary shifts in focal attention being triggered by students' comprehension or production problems. (p. 40)

The purpose of focus-on-form is to induce noticing (Schmidt, 1990, 2001), i.e. registering of the forms in the input with the potential of further processing of these forms for learning. The key feature here is that focus-on-form implies attention to language forms in meaningful context, which potentially enhances the chances of noticing and learning.

Another defining feature of TBLT is that it represents a learner-centered approach, which entails a more prominent role for learners when determining the content of instruction or selecting linguistic resources during output production (Van den Branden, Bygate & Norris, 2009). To justify the effectiveness of TBLT from a psycholinguistic perspective, its proponents argue that holistic and meaningful communication can boost learners' motivation and engage more natural and implicit language learning mechanisms, resulting in a faster and more balanced interlanguage development (Ellis, 2003; Long & Crookes, 1992, 1993; Skehan, 1998; Skehan & Foster, 1999). As Skehan (1998) argues, 'instruction in which learners are given tasks to complete in the classroom makes the assumption that transacting tasks in this way will engage naturalistic acquisitional mechanisms, cause the underlying interlanguage system to be stretched, and drive development forward' (p. 95).

2.3 Tasks from a psycholinguistic perspective

As suggested in the previous section, in contrast to synthetic syllabi, in which discrete linguistic elements constitute the building blocks of the language syllabus, TBLT employs holistic, meaningful tasks as the central units of instruction and assessment. Definitions of the construct of a task are numerous. Table 1 provides a number of selected definitions of task available in the TBLT literature.

Despite the diversity of definitions, there are several common characteristics that can be summarized in the following way: a *task* represents a goal-directed, meaningful linguistic process; this linguistic activity bears a resemblance to a real-world communicative event and it can involve both oral and written language use. Another fundamental issue is that tasks provide a context for the provision of authentic L2 input and involvement of learners into the production of output. Tasks can, thus, be regarded as a major site for L2 performance and development. From the research perspective, tasks can also be conveniently employed as research units, which can be used to analyze production and explore L2 acquisition processes. As Bygate (2000) puts it:

Tasks can provide a focus and context for the study of a number of issues, such as how tasks can engage learners' cognitive processes, their discourse, their attention, their interaction around problem areas of language, and how tasks can be used in syllabus development. From this perspective, tasks can be usefully seen as multidimensional, enabling the study of learners' language use and development as a complex (i.e. socially contextualized, functional and hierarchically structured language) system. (p.187)

Much of the research into tasks has been done from a psycholinguistic perspective, which views a task as a device that can influence the performance and mental processing of learners. That is, the assumption is that certain task characteristics might have a direct and pre-determined effect on the quality of L2 production and learning. As Skehan, Foster and Menhert (1998) put it, "task properties have a

significant impact on the nature of performance” (p.245). Therefore, the central concern in the psycholinguistic strand of TBLT research has been to identify task features that can engage learners in effective language use and processing that are beneficial for acquisition.

Table 1. Definitions of "task" in TBLT literature

Author	Defintion
Long (1985)	A piece of work undertaken for oneself or for others, freely or for some reward. Thus, examples of tasks include painting a fence, dressing a child, filling out a form, buying a pair of shoes (...). In other words, by “task” is meant the hundred and one things people do in everyday life, at work, at play, and in between. Tasks are the things people will tell you they do if you ask them and they are not applied linguists.
Crookes (1986)	A piece of work or an activity, usually with a specified objective, undertaken as part of an educational course, or at work.
Bachman and Palmer (1996)	An activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation.
Skehan (1998)	A task is an activity in which meaning is primary, there is some communication problem to solve, there is some relationship to real-world activities, task completion has some priority, and the assessment of the task is in terms of outcome.
Bygate (2000)	The term ‘pedagogic task’ refers broadly to structured, bounded, purposeful activities involving the processing of language, which learners undertake in order to learn. They are used both inside and outside classrooms, and can vary widely in kind and purpose. They may involve listening, speaking, reading and writing. They may vary in their grammatical, lexical, or discourse focus. They may be undertaken by learners individually, or in groups, or with a teacher.
Bygate et al. (2001a)	An activity which requires learners to use language, with emphasis on meaning, to attain an objective.
Ellis (2003)	[task] involves a primary focus on meaning (...), it involves real-world processes of language use (...), it can involve any of the four language skills (...), it engages cognitive processes (...), and it has a clearly defined communicative outcome.
Van den Branden (2006)	A task is an activity in which a person engages in order to attain an objective, and which necessitates the use of language.
Samuda & Bygate (2008)	A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both.

Much of the psycholinguistic research into tasks has been inspired by two influential theoretical perspectives developed by Peter Robinson (The Cognition Hypothesis, 2001, 2011a) and Peter Skehan (the Trade-Off Hypothesis, 1998, 2009) respectively. Both of these perspectives view tasks as devices for influencing the way learners acquire and use language. However, Robinson and Skehan differ in their views about the kind of language use and opportunities for learning that arise during task performance. These discrepancies are the results of a somewhat different understanding that these researchers have of the way learners deploy their attentional resources during task performance. In his theorizing, Skehan emphasizes limitations in the learner's attentional capacity. On the other hand, Robinson adheres to the multiple-resource view of attention and puts emphasis on the flexibility of the attentional capacity rather than on its limited nature. In spite of the differences, Skehan's and Robinson's approaches are complementary along certain lines and both offer insights that are of value to cognitively-oriented task-based research. In the subsequent sections, Skehan's and Robinson's positions are reviewed in more detail.

2.4 Skehan's "cognitive approach" to tasks

2.4.1 The rule-based and the exemplar-based systems

Skehan (1992, 1996, 1998, 2009) has developed what he refers to as a "cognitive approach" to support his understanding of tasks. One of the central ideas in Skehan's theorizing is the distinction between *representation* and *processing* (this distinction is not task-specific, but can be applied to language learning in general). Representation is concerned with the nature of linguistic units, the way these units are organized and

stored in memory, and the way they can change. Processing, in contrast, relates to access to linguistic representations and the way cognitive pressures are handled during real-time performance. Skehan's main thesis is that representation can function by means of *rule-based* and *exemplar-based* system. The rule-based system, Skehan claims, is likely to be elegantly organized, with rules being compactly structured. This system is posited to be generative; rules can be applied in the creative fashion, which allows language users to express meanings with precision. The rule-based system is restructurable, with new rules subsuming and replacing the old ones. As the underlying system is restructured, it becomes more extended and efficient. All these gains, however, have a high cognitive cost and the implementation of the rule-based system in speech production is believed to pose a heavy processing burden on the learner. Skehan concludes that supportive conditions are necessary for the successful implementation of the rule-based system, although he does not specify these conditions.

The functioning of the exemplar-based system is less resource-demanding. This system is based on multiple and redundant representations of lexical elements. In consequence, this system lacks parsimony and has reduced generative power. The biggest advantage of the exemplar-based system is its low cognitive cost and high processing speed. As the use of the ready-made lexical chunks does not require excessive computation, the functioning of the exemplar-based system does not overtax cognitive resources, which provides learners with spare attentional resources. This system, however, is limited in its potential to express complex meanings with precision; additionally, the exemplar-based system is considered to be less appropriate to induce the restructuring of the learner's interlanguage.

In Skehan's view, neither the rule-based nor the exemplar-based system is perfect separately. The ideal situation is when the systems work together in a

harmonious way, with the learner calling upon one system or another in accordance with the user's communicative conditions and goals. When there is more time, and precision in the expression of the meaning is important, the rule-based system can be accessed. If the communicative pressure is high, the learner can draw on the exemplar-based system. In terms of language development, Skehan stresses that the use of both systems is desirable, because they have different and complementary effects on language use and interlanguage development. Thus, by engaging the exemplar-based system, the learner will make gains in terms of fluency and real-time language processing (i.e., control of the underlying system will be reinforced). The rule-based system, in contrast, stresses generativity and flexibility in language use and, accordingly, allows for the changes in the interlanguage to take place. As a result, the goal of the task-based instruction is to provide a balanced program which will maximize the chances of a balanced involvement of both rule-based and exemplar-based systems.

2.4.2 Skehan's framework of task-based instruction

According to Skehan, balance is also important in the prioritization of meaning over form and vice versa. Skehan highlights that for learners meaning is primary and the language forms have secondary importance. He further cautions against a potential disadvantage of the task-based approach which consists in the potential over-emphasis on meaning during the transaction of tasks, which might de-emphasize attention to form. So the objective of task-based instruction is to channel learners' attention through task design in order to obtain a balanced focus on the desired aspects of performance. To do this, Skehan claims, principled criteria for task design and implementation are needed. These criteria are necessary to engineer focus on form and enable interlanguage development, without compromising the authenticity and communicative value of the

task. In response to this need, Skehan (1992, 1996) proposed a framework for task-based instruction, which distinguishes between three major dimensions: *code complexity* (language required), *cognitive complexity* (thinking required) and *communicative stress* (performance conditions) (see Figure 1). *Code complexity* refers to the linguistic complexity of language, with some tasks inducing simpler language than others. *Cognitive complexity* concerns the content of what is communicated, and it is related to the “conceptualization” stage in production models (Levelt, 1989, 1993). Within *cognitive complexity*, Skehan makes a distinction between two aspects of cognition: *cognitive familiarity* (the capacity to access and retrieve ready-made chunks of linguistic material) and *cognitive processing* (cognitive effort related to the need to work out novel linguistic solutions on-line).

Cognitive familiarity is further broken down into topic familiarity and predictability, which is the extent to which relevant background knowledge is available. Familiarity of discourse genre is the extent to which ready-made genre schemas are available. Familiarity of task refers to the knowledge about task type. New (i.e., unfamiliar) tasks are considered as less predictable and, consequently, precluding the employment of the previously acquired communicative strategies. Cognitive familiarity refers to the processing load during task performance. Information organization is related to the logic or naturalness with which task-relevant information is structured – an example will be the extent to which a narrative task follows conventional temporal sequence. The amount of computation refers to the amount of information transformation or computation required during task performance. For example, a riddle would require more computational effort than a narrative task. Clarity and sufficiency of information concerns the directness with which information is presented to the learners and also the amount of inferences needed to be made during task performance.

Information type category is based on such contrasts as concrete-abstract, static-dynamic and contextualized-decontextualized information (Brown et al., 1984).

1. *Code complexity*

- linguistic complexity and variety
- vocabulary load and variety
- redundancy and density

2. *Cognitive complexity*

Cognitive familiarity

- familiarity of topic and its predictability
- familiarity of discourse genre
- familiarity of task

Cognitive processing

- information organization
- amount of "computation"
- clarity and sufficiency of information given
- information type

3. *Communicative stress*

- time limits and time pressure
- speed of presentation
- number of participants
- length of texts used
- type of response
- opportunities to control interaction

Figure 1. Framework of task-based instruction (from Skehan, 1992)

The third dimension, *communicative stress*, refers to performance conditions. Factors from this dimension are not related to meaning or form, but have implications for the processing pressures during task performance. Time limits and time pressure reflect the urgency of task completion and the perception, on the part of a learner, of the amount of pressure posed by the task. When a large quantity of material is presented at a high speed, processing is more taxing and the task completion becomes more demanding. If a more complicated type of response is required, or there is little opportunity to control the speed of an interactive task, the learner's cognitive resources

are assumed to be severely taxed, and less attention to form is available. It is important to point out that in the reformulation of his framework, Skehan (1996) revised the dimension of *communicative stress* by introducing the factors of *time pressure*, *modality*, *scale*, *stakes* and *control* (Figure 2).

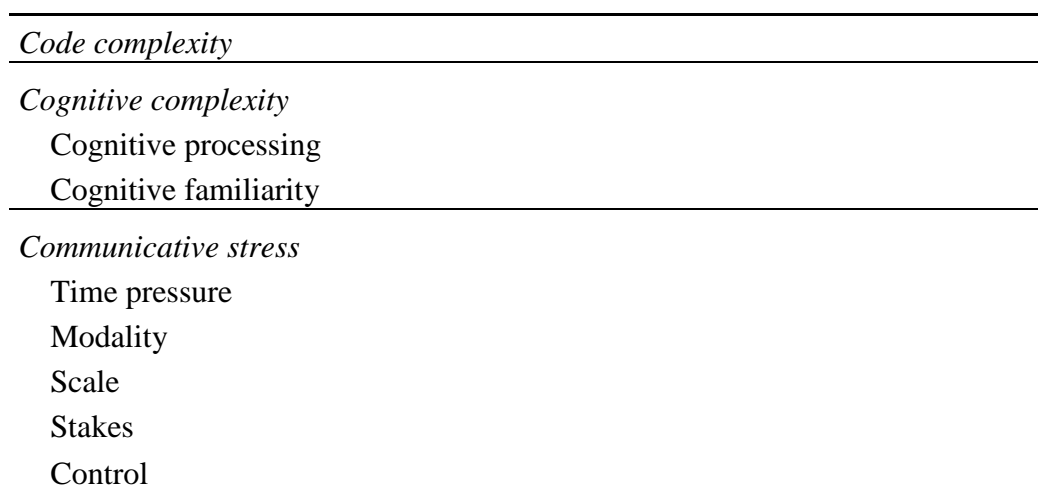


Figure 2. Revised framework of task-based instruction (Skehan, 1996)

In this new version, *time pressure* means how quickly the task has to be done and *scale* refers to the number of participants involved in task performance. *Stakes* reflects the importance of the outcome of the task. If consequences of the poor performance are few, then the stakes are low. Conversely, with negative consequences of poor performance, stakes are high, such as in an exam situation. *Control* refers to the extent to which the participants can influence the way the task is performed (e.g., if the task goal can be negotiated, if a learner can reduce the speed of the presentation of the input, if clarification questions can be asked, etc.). A completely new dimension is *modality*, which refers to the speaking/writing and listening/reading contrast. The *modality* dimension is not elaborated in detail, Skehan only posits that performing a task in the oral mode leads to more pressure than performing it in writing, and that listening poses more pressure than reading. Skehan (1996), however, also highlights the

importance of *communicative stress* factors in task design, and he points out that “perhaps the major area for adjustment while tasks are being completed is in the area of stress (or communicative pressure). Pressure manipulation can be based on [...] time, modality, scale, stakes, control [...]. The point is simply that they are susceptible to variation, with consequent impact on communicative pressure” (p.55).

Concerning the overall goal of his task-based framework, Skehan posits that his model can facilitate informed decisions during task design and implementation, as it can allow design and choose tasks in accordance with the proposed performance outcome and pedagogical goal. If tasks are well-chosen, the learning benefits are high, as the effective task implementation will allow for a balanced attention to meaning and form during task performance (see also Skehan & Foster, 1999). Additionally, the appropriate task choice can create optimal opportunities for restructuring (i.e., interlanguage development) to take place. Within the cognitive line of TBLT research, Skehan’s model represents an example of the approach which aims at establishing cause-and-effect connections between task characteristics and language performance and development. Skehan’s predictions of the effects of task demands on L2 performance, operationalized as complexity, accuracy and fluency (CAF), are discussed in the next section.

2.4.3 Skehan’s Trade-off Hypothesis

Another influential theoretical idea advanced by Skehan is the prediction about the effects of task demands on CAF of performance. Skehan’s main assumption is that real-time communication places high demands on attentional resources, which are limited, and, by extension, learners are induced to prioritize one aspect of production over another during language use. This linguistic compromise during task performance has

become to be known as the “Trade-off Hypothesis” (Skehan, 1996, 2009). Skehan posits that the performance areas that enter into competition with one another for attentional resources are fluency, accuracy and complexity. Skehan and Foster (1999) define these constructs as follows:

Fluency is the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems. Accuracy is the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as conservative orientation, that is, avoidance of challenging structures that might provoke error. Complexity/range is the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system. (pp. 96-97)

Because attention is limited, Skehan claims, it is unlikely that all three of these aspects of performance can be attended to simultaneously. As a result, a linguistic trade-off is likely to occur. Skehan’s predictions are influenced by current information processing perspectives on the nature of language learning. Van Patten (1990), for example, argued that limitations in cognitive resources have pervasive effects on what the learner can attend to during communication, with consequent implications for the quality of performance during tasks. As a result of the scarcity of attentional resources, learners cannot devote attention to each of the performance dimensions so that they can be met simultaneously. In accordance with his ideas of the primacy of meaning over form, Skehan (1996) posits that “performance is likely to prioritize fluency, and relegate restructuring and accuracy to lesser importance” (p.50). In other words, tension is likely to occur between fluency, which Skehan equates with meaning, and accuracy and complexity, which are equated with form. Then, within form, “one can contrast attention directed to using challenging language (complexity) relative to conservative,

less advanced language, but greater accuracy” (Skehan, 2009, p. 511). In other words, in Skehan’s view, complexity and accuracy compete for attentional resources by default. Allocating limited cognitive resources to one of the aspects of performance is likely to deplete performance quality in other areas. In terms of the precise way the trade-off is manifested linguistically, Skehan argues that high fluency during task performance can be accompanied by either enhanced complexity or accuracy of performance, but not by both. This trade-off is theorized to be most prominent in demanding tasks, in which learners reach the limits of their attentional resources.

From a pedagogical perspective, the goal of task-based instruction is to contrive balance in language teaching, avoiding excessive prioritization of anyone of the dimensions of fluency, accuracy, and complexity. As Skehan and Foster (1999) argue, “balance needs to be established among the three performance areas, in that one would like to see development in each of the areas, and without one area compromising development in the others” (p.97). Skehan (1998) grounds the case for balance on a number of premises. He argues that if fluency is prioritized, there can be an overreliance on the lexical exemplar-based system and the undesired fossilization of interlanguage can take place; consistent focus on accuracy can eventually hamper the development of fluency and induce avoidance of engagement with complex linguistic structures; consistent prioritization of complexity might deplete development in accuracy and control. The goal of task research is, therefore, to establish how balance can be achieved among the CAF areas and how task characteristics and task conditions can influence performance and influence balance among the CAF dimensions. In line with these ideas, the thrust of the empirical research framed within Skehan’s theories, was to establish the impact of task characteristics on L2 performance operationalized in terms of CAF.

2.5 Robinson`s Cognition Hypothesis

An alternative account on the effects of task demands on L2 performance and development was offered by Robinson (2001, 2011a), who summarized his theoretical ideas and predictions in the Cognition Hypothesis. The predictions of the Cognition Hypothesis, which will be described in detail below, are rooted in the related claims in L1 developmental psychology (e.g., Cromer, 1991; Slobin, 1993), functional linguistics (e.g., Givón, 1985, 1995; Talmy, 2000; Tomlin, 1990), theories of attention (Wickens, 1984, 2002) as well as SLA research (e.g., Doughty & Williams, 1998; Perdue, 1993; Schmidt, 2001). The hypothesis makes a number of predictions about the likely effects of task complexity on language learning and production. These claims concern output, uptake and interaction, learning from input, automaticity and individual differences (for the most recent formulation, see Robinson, 2011a). For output, Robinson predicts that increasing the cognitive complexity of a task along certain dimensions should push learners to greater complexity and accuracy (but lower fluency) in order to meet high functional demands posed by the complex task. In terms of development, Robinson claims that greater effort at conceptualization might induce learners to stretch and develop the L2 linguistic resources to express complex conceptual propositions. According to Robinson (2011a), “this should most clearly be evident during monologic task performance” (p.19). Cognitively complex tasks should also lead to more interaction and negotiation of meaning, heightened attention to and longer-term retention of input. Sequencing of tasks is also of paramount importance, as performing simple to complex sequences should lead to greater automaticity. Finally, the role of

individual differences in affective and cognitive abilities is posited to be most prominent in complex tasks.

The theoretical justification of the above predictions is specifically thorough for the effects of task complexity on L2 output and development. The central claim is that tasks with greater cognitive load will engage L2 learners in complex thinking, and, thus, create favorable conditions for language development (see also Cromer, 1974). The need to encode complex conceptual representations will also beneficially affect performance, as it will gear learners towards higher accuracy and complexity of production. The parallel increase in complexity and accuracy of production is the main singular feature that distinguishes Robinson's views from those of Skehan. In his theorizing, Skehan emphasizes the limited nature of attentional resources. Consequently, tasks can lead either to enhanced complexity or accuracy of production, but not to both. Robinson, in contrast, adheres to the multiple-resource view on attention (Wickens, 2002), which allows to hypothesize synchronous increases in accuracy and complexity of production. To guide research and pedagogy into the claims of the Cognition Hypothesis, Robinson (2003) proposed an operational taxonomy of task characteristics, known as the Triadic Componential Framework (see also Robinson & Gilabert, 2007).

2.5.1 The Triadic Componential Framework

As its name suggests, The Triadic Componential Framework distinguishes between three independent facets of task characteristics implicated in real-world task performance: *task complexity*, *task condition* and *task difficulty* (see Figure 3). To date, Robinson's taxonomy represents the most detailed framework which establishes the relationship between task and learner factors on the one hand, and linguistic performance on the other hand. Each group of factors is discussed in detail below.

2.5.1.1 Task Complexity

The *task complexity* dimension describes factors that contribute to the intrinsic cognitive demands tasks make on learners' memory, attention and reasoning. Within the cognitive factors associated with task complexity, an important theoretical distinction is made between cognitive/conceptual (*resource-directing*) and performative/procedural (*resource-dispersing*) variables (Robinson, 2001, 2003; Robinson & Gilabert, 2007). Increase in task demands along the resource-directing factors is believed to stimulate *analysis* and *development* of the existing interlanguage, while enhanced complexity along the resource-dispersing lines is theorized to promote *automatization* of and *access* to the current L2 knowledge. This distinction is similar to one made by Bialystok (1991), who differentiated between the dimensions of *analysis* and *control* in L2 learning. Subsequent sections discuss in detail the theorized effects of the resource-directing and resource-dispersing variables on learning and performance.

► Resource-directing variables

Resource-directing dimension embraces task characteristics related to the concepts that a task requires to express (e.g., causal relationships, intentionality, relative time, spatial location). Examples of resource-directing variables include [+/- here-and-now] dimension (events happening now, in a mutually shared context, versus events that occurred in the past, elsewhere), [+/- few elements] (few easily distinguished elements, versus many similar elements in a task), [+/- intentional reasoning] (simple information transmission, versus reasoning about other people's intentions), [+/- causal reasoning] (simple information transmission, versus reasoning about causal events and relationships) or [+/-perspective taking] (first person perspective on event, versus multiple persons' perspectives) (Robinson & Gilabert, 2007). Through the manipulation

of the resource-directing variables, learners can be engaged in lower or higher effort at conceptualization, with an effect on the way learners direct and use their attention and memory resources during task performance.

<i>Task complexity</i> (Cognitive factors)	<i>Task condition</i> (Interactive factors)	<i>Task Difficulty</i> (Learner factors)
(Classification criteria: cognitive demands) (Classification procedure: information-theoretic analyses)	(Classification criteria: interactional demands) (Classification procedure: behavior descriptive analyses)	(Classification criteria: ability requirements) (Classification procedure: ability assessment analyses)
(a) <i>Resource-directing</i> variables making cognitive/conceptual demands	(a) <i>Participation variables</i> making interactional demands	(a) <i>Ability variables</i> and task-relevant resource differentials
+/- here and now	+/- open solution	+/- working memory
+/- few elements	+/- one way flow	+/- reasoning
+/- spatial reasoning	+/- convergent solution	+/- task-switching
+/- causal reasoning	+/- few participants	+/- aptitude
+/- intentional reasoning	+/- few contributions needed	+/- field independence
+/- perspective-taking	+/- negotiation not needed	+/- mind/intention-reading
(a) <i>Resource-dispersing</i> variables making performative/procedural demands	(b) <i>Participation variables</i> making interactant demands	(b) <i>Affective variables</i> and task-relevant state-trait differentials
+/- planning time	+/- same proficiency	+/- openness to experience
+/- single task	+/- same gender	+/- control of emotion
+/- task structure	+/- familiar	+/- task motivation
+/- few steps	+/- shared content knowledge	+/- processing anxiety
+/- independency of steps	+/- equal status and role	+/- willingness to communicate
+/- prior knowledge	+/- shared cultural knowledge	+/- self-efficacy

Figure 3. The Triadic Componential Framework for task classification - categories, criteria, analytic procedures, and design characteristics (from Robinson & Gilabert, 2007)

As Robinson puts it, increasing task complexity along the cognitive/conceptual dimensions “has the potential to connect cognitive resources, such as attention and memory, with effort at conceptualization and the L2 means to express it, thereby promoting L2 development” (Robinson, 2011a, p.14).

In terms of language use, Robinson posits that the need to express more complex concepts would require the use of more linguistically elaborated structures and forms. This assumption is in line with the ideas of other theorists (e.g., Dipper, Blanc, & Bryan, 2005; Givón, 1985, 2009; Levelt, 1989), who posit an interactive relationship between conceptual and linguistic representations, which leads to “paring down of thought in language production” (Dipper et al., 2005, p.422). In other words, thoughts intended for expression must be organized in a language-appropriate way. Following these ideas, Robinson assumes that tasks requiring reference to events happening in the past could potentially redirect learners’ attention and memory resources to morphology conveying tense and aspect in the past. Similarly, tasks that require complex reasoning could be expected to promote the use of cognitive state terms for reference to other minds (e.g., “she suspects”, “he wonders”) or orient attention to the specific constructions (“suspected that”, “wonders whether”) and, thus, enable learners to use complex L2 syntax (Robinson, 2011a). The idea that more complex thoughts call for more complex language also comes from Givón’s (2009) work, who argued that “more complex mentally-represented events are coded by more complex linguistic/syntactic structures” (p. 12). Referring to Givón’s distinction between pragmatic and syntactic modes of production, Robinson claims that simple task demands would elicit simpler pragmatic mode of production, characterized by loose coordination of clauses and little use of grammatical morphology (i.e., low linguistic complexity). In contrast, conceptually demanding communicative task elicits syntactic mode of production, which is

morphologically richer and more complex structurally. Another central claim of the Cognition Hypothesis is that increasing the complexity of resource-directing task characteristics can potentially direct learners' attention "to the way L2 structures and codes concepts (e.g., Talmy, 2000, 2008), often in ways that differ from how they are structured and coded in the L1" (Robinson, 2011a, p.15). Understanding of the specificity of the linguistic coding of conceptual information in the L2 may draw learners' attention to the relevant L2 forms, with the consequent enhancement of accuracy in production.

In sum, Robinson's main thesis is that increasing task complexity along the resource-directing variables equals to greater conceptualization effort and, thus, stretches L2 learners' interlanguage, enhances noticing of the L2 forms and mobilizes cognitive and linguistic resources with a consequent qualitative increase in the linguistic complexity and accuracy of learners' output. As the Cognition Hypothesis contemplates the existence of multiple attentional pools for cognitive processing, this theoretical framework does not predict an obligatory trade-off effect between accuracy and complexity. Rather the prediction is that in tasks, which are complex along the resource-directing lines, there will be a parallel increase in complexity and accuracy of production, accompanied by a decrease in fluency. This is one of the main tenets of Robinson's Cognition Hypothesis. Importantly, the Cognition Hypothesis also claims that the beneficial complexity-accuracy enhancement will only hold as long as tasks are kept simple along resource-dispersing dimensions (e.g., when planning time is available). In case a task is made more complex along both resource-directing and resource-dispersing dimensions simultaneously (which is often the case in real-life task performance), the beneficial effects on speech production are "likely to be weakened or negated" (Robinson & Gilabert, 2007, p.167). Accordingly, to promote complex output,

accuracy and engage L2 developmental processes, a task with enhanced resource-directing demands must not be taxing in terms of resource-dispersing demands.

► **Resource-dispersing variables**

Resource-dispersing variables make performative and procedural demands on cognition. Examples of resource-dispersing factors include [+/- planning time] (giving or not planning time to learners), [+/- single task] (one thing to do or multiple things to do when completing a task), [+/-prior knowledge] (tasks that provide background knowledge or not). As their name suggests, the resource-dispersing variables disperse learners' attention and memory resources over many linguistic and non-linguistic features of a task. Under this effect, learners have fewer attentional resources left to allocate to particular language forms. As a consequence, increasing task demands along these variables will tax learners' attention and memory resources in task completion, adversely affecting accuracy, complexity and fluency of the L2 performance. However, increased complexity along the resource-dispersing dimensions is vital for creating the processing conditions of real-time language use. Accordingly, performing tasks with enhanced resource-dispersion means exercising real-time access to the current L2 knowledge, which will eventually result in faster and more automatic L2 access and use.

It is important to highlight that, like Skehan's (1998) Trade-off Hypothesis, the Cognition Hypothesis acknowledges the possibility of the degraded performance (i.e., low complexity, accuracy and fluency) in cognitively demanding tasks, but only in case of enhanced complexity along the resource-dispersing lines. Also, the cause of the performance decay is theorized differently. While Skehan (1998, 2009) links lower performance to the limitations in cognitive resources, Robinson (2003) claims that

performance suffers as a result of the loss of control over attention (see also, Navon, 1989). The involuntary dispersing of the learners' cognitive resources draw their attention away from the relevant linguistic aspects of the task, which results in the degraded linguistic complexity, accuracy and fluency of the L2 output.

The critical difference between the Cognition Hypothesis and the Trade-off Hypothesis, however, is that the latter does not make the distinction between the resource-directing and resource-dispersing dimensions. Consequently, Skehan's theory does not contemplate potential beneficial effects of the cognitive/conceptual variables on learners' performance.

2.5.1.2 Task condition

The second category in Robinson's taxonomy is the *task condition* dimension, which is divided into *participation* and *participant* variables. Participation variables concern the nature of the participation required in the task and contain (1) the number of participants (+/- few participants), (2) the degree of need of participants' contributions during task completion (+/- few contributions needed), (3) the need for meaning negotiation (+/- negotiation not needed), (4) the distribution and flow of information during task performance (+/- one way flow), (5) the degree of freedom in the task solution (e.g., +/- open solution; +/- convergent solution). Participant variables concern the participants' characteristics relevant to task performance, such as participants' gender, proficiency level, their status and role in the task, or the degree of mutually shared content or cultural knowledge. According to Robinson, the task condition variables should be "held constant are replicated each time more cognitively complex pedagogic versions are attempted" (Robinson, 2011a: p. 13), as this would create a

condition under which learners would transfer the same schemata of their performance from one version of the task to another.

2.5.1.3 Task difficulty

The third category, *task difficulty*, relates to learners' individual differences, which are purported to potentially affect (1) learners' perception of the complexity of the task, (2) task performance, and (3) L2 learning. The *task difficulty* dimension covers a number of relatively stable ability factors (e.g., working memory capacity, aptitude, individual reasoning abilities, etc.) and less stable affective variables that are subject to temporal change (e.g., learners' motivation, openness to experience, anxiety, emotional control, ability of self-regulation or willingness to communicate). Robinson claims that affective and ability variables can mediate the relationship between cognitive task demands and learners' task complexity perception. For example, more motivated learners, or learners higher in aptitude are likely to find the task easier than less motivated learners, or those with lower aptitude (Robinson, 2001). Besides, task difficulty variables can be responsible for between-learner variation when performing L2 tasks.

According to Robinson (2011a), the differentiated inter-learner performance, caused by the influence of the task difficulty factors, is likely to be particularly apparent in complex versions of tasks. Another issue highlighted by Robinson is the interrelationship existent between affective and ability factors. In his view, affective variables (e.g., anxiety or motivation) can temporally enhance or reduce learners' ability resources (e.g., aptitude), and, thus, influence L2 task performance. Task difficulty factors can be measured with the help of specific tests or questionnaires. For example, working memory or aptitude tests can be used to assess the ability variables. The

affective variables (e.g., interest, motivation, stress, etc.) can be measured after the completion of a task by means of a Likert-scale questionnaire. As Robinson highlights, such questionnaire is also instrumental to check if the low and high levels of task complexity conceptualized by the researcher match the learners' actual perceptions of low and high task demands.

2.5.2 Implications for task design

In Robinson's taxonomy, the special status is given to task complexity factors, which constitute the sole operational basis of task-syllabus. Robinson (2011a) claims that increasing cognitive demands of pedagogic tasks through one or another of the task complexity features "should help to explain within-learner variation in the language they use to perform them" (p.13), with more complex tasks purported to elicit more accurate and complex language as compared to simple tasks. Based on the resource-directing/dispersing distinction, Robinson (2010) proposed the following operational principles for sequencing tasks in task-based syllabus in order to promote learning: (1) task sequences should be based on increases in cognitive complexity, with the simple version of a task performed/practiced prior to the complex version, (2) resource-dispersing dimensions of complexity should be increased and practiced first to promote access and control of interlanguage; then, resource-directing dimensions of complexity should be increased and practiced to promote development of new form-function mappings and destabilize the current interlanguage system. Sequencing of tasks on the basis of increases in their cognitive complexity is the basic pedagogic claim of the Cognition Hypothesis (see Baralt, Gilabert & Robinson [2014] for the detailed theoretical rationale and the collection of empirical studies on task sequencing).

2.6 Task demands and stages of speech production

At some stage of the elaboration of their theories, both Skehan (2009) and, more recently, Robinson (2011a) made connections between their own theoretical predictions related to task performance and language production stages from Level's (1989) model of speech production. The establishment of this connection clearly evidences that both theories were elaborated to account primarily for oral production. Level's (1989) influential model identifies the following stages of speech production: (1) *conceptualization*, which entails generation of the ideational content involving the stages of microplanning (e.g., deciding if the utterance will be a request, apology, etc.) and microplanning (planning the linguistic realization of the content); (2) *formulation*, that is encoding of the preverbal message into linguistic form; (3) *articulation*, which is the overt production of the speech sounds, and (4) *monitoring*, which involves checking the correctness and appropriateness of the produced output.

Skehan (2009), in particular, theorizes that certain task characteristics or performance conditions are connected primarily to the stage of conceptualization, while others influence linguistic encoding. For example, the necessity to develop a more complex proposition, to deal with abstract information or with large quantity of information during task performance is theorized to strain conceptualization. Availability of planning time is also linked to the workings of the conceptual stage of message preparation. On the other hand, the real-time pressure under which speaking has to take place and the amount of input received by speaker before or during task performance have the effects on the stage of formulation. Skehan (2009) connects the construct of online pressures with Ellis's (2005) concept of online planning, and defines it as "an index of the amount of time the speaker has to access material, to build syntactic frames, and to regroup as necessary" (p. 526). Influencing the amount of time

available during task performance can thus strain or relax online pressures with consequent effects on the workings of linguistic encoding. Skehan also posits that absence or availability of pre-task and post-task activities, or the type of the task (monologic versus dialogic) can increase or reduce pressures on the speech production stages. Pre-task activity, for example, has favorable consequences for both conceptualization and formulation, as it gives an opportunity to prepare a pre-verbal message and to prime lexical and syntactic elements. Skehan also assumes that dialogue confers benefits for both conceptualization and formulation: while one's interlocutor has the floor, the other has more time for conceptual message preparation and linguistic encoding. In other words, availability of additional time is again posited as a factor that eases burden on speech production. Additionally, one's interlocutor can provide useful scaffolding and serve as a supply of lexis and vocabulary, which can help formulation. Skehan also links the stage of speech production to the CAF dimensions. He suggests that lexical and structural complexity is related to conceptualization, while accuracy and fluency are linked to linguistic encoding (i.e., formulation).

In one of his most recent publications, Robinson (2011a) also makes connections to the stages of language production from Levelt's (1989) model. Thus, Robinson relates the effects of task complexity to the stage of conceptualization with the subsequent influences on the lexico-grammatical encoding. In Robinson's terms, greater effort at conceptualization results in "creating the conditions for development and re-mapping of conceptual and linguistic categories" (Robinson, Cadierno & Shirai, 2009, p.537). The complexity of the preverbal messages determines the workings of the formulation stage, which pairs down the conceptual information with the relevant linguistic representation, with more complex conceptual representations encoded with the correspondingly complex linguistic forms.

As stated at the beginning of this section, both Robinson`s and Skehan`s theories were developed with the oral mode in mind. Such prioritization of the oral mode as the default mode for theory development is a reflection of the traditionally dominant and privileged role of oral production in SLA scholarship (Ortega, 2012). As proper Robinson (2011b) points out: “theoretical rationales for the influence of task demands on writing and both reading and listening comprehension currently lag behind rationales for their effects on speech production in articulating linkages between rationales proposed and explanatory psycholinguistic mechanisms” (pp. 15-16). Scarcity of theoretical accounts attuned to the singularity of written language production, resulted in a situation when theories, developed for oral production, are applied to explain learning and performance in writing tasks. This state of affairs has been critically examined in recent TBLT literature.

2.7 Skehan`s and Robinson`s models and written task performance

As Bygate, Van den Branden, and Norris (2014) point out “TBLT implicates centrally the language in its various modes” (p.xi). However, the implication of oral and written modes has not been at all equitable, as the overwhelming focus so far has been on tasks undertaken in the oral modality (Bygate et al., 2014; Byrnes & Manchón, 2014a, b). The centrality of the oral mode in the TBLT scholarship is a reflection of the traditional privileging of speech in the mainstream second language acquisition research (Carless, 2012; Ortega, 2012). Ortega (2012) traces the neglect of the written mode in the research to different reasons, starting with the ingrained view of writing as a “culture-dependent, secondary manifestation of human language, a derivation of the primary oral capacity for language that all healthy individuals of our species share” (p. 405). Additionally, the off-line, self-paced nature of writing has rendered written data, in the

view of many theorists, into an invalid reflection of tacit linguistic knowledge or of online ability for language use (Selinker, 1972). At the same time, spontaneity and the largely unmonitored quality of oral performance were praised in the exploration of the nature of language acquisition. In recent years, however, writing has started to play a more central role in SLA in general, and in TBLT in particular. This welcomed change of affairs can be attributed to a number of reasons, such as the recognition of writing as a vital skill to use language proficiently and also as a crucial tool to foster language development (see contributions to Manchón & Matsuda, 2016). A number of persuasive arguments for a more central role of writing in the TBLT concerns were provided by Manchón (2014a) who stated that:

...it makes theoretical and pedagogical senses to make writing a more central preoccupation for TBLT for two main reasons: first, because of the purported potential learning outcomes associated with literacy practices (see Manchón 2009, 2011a, 2011b, 2013a, 2014; Manchón & Roca de Larios 2007; Ortega 2012; Williams 2008, 2012); and, second, because communication through/in writing is indeed part of “the various domains of lifetime endeavor outside the language classroom” (Robinson 2011, p.11) and the “full complexity of real-world target-task performance” (Robinson & Gilabert, 2013, p.3) purported to be covered by the concepts and task and task learning. (p.29)

The traditional overemphasis on oral production has had a negative consequence – namely, the scarcity of the theoretical thinking attuned to the idiosyncrasy of written task performance. As a result, L2 researchers have to resort to speech-customized models as theoretical frameworks in the investigation of written production. In this regard, Byrnes and Manchón (2014b) have cautioned against the uncritical application of the speech-originated theories to writing and pointed out the urgent need to rethink and reconceptualize TBLT tenets and constructs in their application to writing. Byrnes and Manchón (2014b) further argue that theoretical reconsideration of such central

constructs as task complexity is not only necessary, but also highly beneficial as it will enlarge the theoretical scope of the cognitive frameworks.

In line with these ideas, Manchón (2014b) assessed Robinson's Triadic Componential Framework for its suitability for written mode. In her view, the model needs rethinking and expansion in all the groups of task factors, including *task complexity*, *task condition* and *task difficulty* dimensions. Concerning task complexity, both resource-directing and resource-dispersing variables should be reconsidered when applied to writing. Resource-directing variables are purported to draw learners' attention to the language code. In this respect, Manchón highlights that attention to L2 forms in writing is largely the question of the self-imposed goals pursued by the writer. The current formulation of the Triadic Componential Framework does not, however, include goal-setting among the resource-directing factors. Perhaps the major revision of the model concerns the resource-dispersing variable of [+/- planning]. In this regard, Manchón (2014b) points out that planning "is a uniquely distinctive phenomenon in writing" (p.32). Planning in writing is also of much more complicated nature than theorized in Robinson's model. First of all, it is necessary to consider that the implementation of planning is radically different in speech and writing: in oral production, planning is restricted by the on-line pressures, while writers can avail themselves of deeper planning activity during the entire task performance.

Task condition factors are also in need of rethinking to fully account for the written mode. In this regard, comprehensive theorizing of [+/- interaction] variable is needed, as the nature of performance in an individual writing task is distinct from that in collaborative writing. Because feedback constitutes an essential ingredient of learning and teaching in writing, Manchón proposes to include provision of feedback as a task condition factor. Finally, addition of a time-on-task variable is suggested on the

premises that text quality and time-on-task in writing are closely related (Hayes & Nash, 1996). Task difficulty factors, which, in Robinson's theorizing, include learner ability and affective variables, also need expansion to account for writing. Ability factors should minimally include genre knowledge and L2 writing expertise. Concerning affective factors, Manchón claims a more central role for motivation, because "whether people write, how long they write, and how much they attend to the quality of what they write will depend on their motivation" (Hayes 2012, p. 373).

Another critical evaluation of cognitive TBLT models was made by Tavakoli (2014), who highlighted that "little is known about whether cognitive complexity affects writing and speaking tasks in similar ways, or whether it has similar influences on L2 oral and written performance" (p. 217). Until recently, the predominant focus of the TBLT scholarship has been on oral tasks. From the scarcity of investigation of task performance in writing, Tavakoli argues, it is still an open question whether task complexity represents the same construct in oral and written modes. Tavakoli further overviews Skehan's and Robinson's models of task complexity and examines, from a theoretical perspective, the applicability of these frameworks to writing. One of the aims of her critical analysis is to reflect on whether task complexity, as theorized by the cognitive TBLT frameworks, could account for differences in both written and spoken tasks. Tavakoli (2014) points out that the initial analysis of the models immediately reveals their exclusive focus on oral production, as both frameworks "appear to take speaking, almost by default, as the mode in which the construct of task is examined, its complexity defined, and its criteria determined" (p. 222). This unique focus on the oral mode becomes more apparent in the discussion of what makes a task more demanding. For example, in both models, planning time is an important variable that can reduce or augment cognitive task demands. However, the two models do not discuss whether the

online planning time available to writers decreases the cognitive load of the task. Robinson, for example, in his discussion of how planning time may affect task complexity argues that tasks “without planning time, or prior knowledge support, requiring learners to simultaneously perform two steps in the task (for example, quickly *thinking up* and *describing* the direction from A to B on a map of an unfamiliar area) are more demanding than those with planning time and prior knowledge” (p. 294). In his theorization, however, Robinson never takes into account that in-built planning is an inherent characteristic of a writing task. In a similar vein, Skehan’s definition of communicative stress (i.e., production conditions under which the task is performed) is restricted solely to oral mode of production. Skehan (1998) maintains:

When larger quantities of material are presented quickly, and when little time is available for processing, when more complicated responses are required, and there is little opportunity to slow down interaction to take account of processing limitations, it is assumed that it will be more difficult to give attention to form. (p. 101)

Clearly, reference to “little time available for processing”, “little opportunity to slow down” or high speed of input presentation all suggest that the task under discussion is in the oral mode. Skehan (as well as Robinson), however, never explicitly points out that this theoretical description is intended solely for the oral mode of production and he never discusses how these conditions would change during writing task performance. In her analysis, Tavakoli concludes that both frameworks seem to be “limited in their implications for modelling task complexity in the written mode” (p. 221). But, at the same time, she cautions against automatically discarding Robinson’s and Skehan’s ideas from their application to writing. Clearly, more empirical research is required to determine if and how the theoretical ideas and predictions of these models could apply to writing.

Kormos (2014) similarly discusses the application of Robinson's Triadic Componential Framework (2001, 2003, 2011a) to writing performance. In Kormos's argument, one way in which this theoretical framework can be expanded is by incorporating *mode* as one of the task complexity factors. Building her argument on the comparison of the psycholinguistic mechanisms involved in oral and written language production, Kormos claims that mode fits as a task complexity dimension as it poses both resource-dispersing and resource-directing demands. The fact that speaking is a time-pressed activity and writing is a self-paced activity, allowing for more planning, constitutes a resource-dispersing constraint. At the same time, mode can also be regarded as a resource-directing factor. Kormos builds this claim on the purported propensity of the written mode to direct learners' attention to linguistic forms (Adams, 2006; Niu, 2009). While communicative pressures in speech can hinder attention to form, more relaxed production conditions in writing and the enduring nature of the written product benefit learners. As follows, writing conditions prove favorable for focus-on-form, as learners can attend to linguistic form both during and after text production.

This critical examination of the TBLT models suggests that Robinson and Skehan may not have considered the idiosyncrasy of writing conditions in their theorizing of the effects of task demands on L2 learning and performance. However, we consider that this gap in theoretical coverage should not automatically preclude the application of cognitive TBLT frameworks to written task performance. What is needed is the adjustment of the proposed tenets accounting for the singularity of writing. Changes in theory should also be informed by the empirical studies testing the main theoretical constructs, albeit developed for oral mode, in written performance. Findings

from such studies would be instrumental for the mode-sensitive rethinking of the cognitive TBLT accounts.

This overview of the mainstream TBLT theoretical thinking has revealed the marginal role of writing and the consequent imbalance of modes in the TBLT scholarship. This lack of balance is problematic given that “in a literate society, the functions of language are shared out between speaking and writing” (Halliday, 1990, p. 99) and, accordingly, the command of both modes constitutes the hallmark of language proficiency (Berman & Ravid, 2009). In addition, both oral and written language use represent potential language learning opportunities (Byrnes & Manchón, 2014a, 2014b; Manchón, 2011; Manchón & Williams, 2016; Ortega, 2012; Williams, 2012), hence the relevance of looking into language learning associated with both speaking and writing tasks (Bygate, Van den Branden, & Norris, 2014; Kuiken & Vedder, 2012). In line with these ideas, recent positions in TBLT (see Gilabert, Manchón, & Vasylets, 2016) have stressed the need for more mode-sensitive and integrative TBLT research and for a deeper understanding of the unique learning benefits that each mode can offer separately or in mutual interaction. The treatment of mode in recent TBLT theoretical thinking will be discussed in the next chapter.

2.8 Chapter summary

This chapter reviewed the conceptual bases of TBLT. The main tenets of the Trade-off Hypothesis (Skehan, 1998, 2009) and the Cognition Hypothesis (Robinson, 2001, 2011a) were also outlined. These frameworks offer competing accounts of the way task features may influence L2 performance and learning. The models' common feature is their orientation towards oral task performance, which is a reflection of the general tendency in the TBLT scholarship to privilege oral mode as a site of language learning and

research. The role of writing, however, is becoming more prominent. Burgeoning writing studies also employ Robinson`s and Skehan`s frameworks to interpret their findings. Uncritical application of the speech-oriented theories to writing is questionable, and a mode-sensitive reconceptualization of the TBLT tenets and constructs is highly desirable. Privileging of oral production has created imbalance of modes in TBLT, which is problematic.

CHAPTER 3. MODE IN THEORETICAL TBLT AGENDA

3.1 Introduction

In the previous chapter, we highlighted the absence of mode as an integral element in cognitive task-based learning theories (Robinson, 2001; Robinson & Gilabert, 2007; Skehan, 1998, 2009) as well as the relevance to redress this situation (Kormos, 2014; Manchón, 2014b; Tavakoli, 2014). We emphasized the importance of adjusting the tenets of cognitive TBLT theories to the singularity of writing, and we suggested that this adjustment should be informed by the results of empirical studies testing and comparing constructs and theoretical predictions in different modes of production. It was stressed that due to the idiosyncrasies of speech and writing, oral and written tasks may represent rather distinct language use and language learning environments and, accordingly, deeper understanding is needed of the unique contributions of oral and writing task performance to interlanguage development. We pointed out the importance of considering both modes because both oral and writing skills constitute integral parts of global language proficiency. It is also important to consider that research into oral-written contrasts is also relevant because oral and written languages coexist within the same mental space, constituting a synergistic system in which they “constantly feed and modify each other” (Ravid & Tolchinsky, 2002, p. 430). The blending of spoken and written modes also occurs in real-life language tasks, which frequently require the use of both oral and written language in the achievement of task-related communicative goals. It follows that in order to account for the complexity of real-life performance in

its full range, it is necessary to explore the learning opportunities potentially afforded by all language modalities (see also Manchón, 2014a).

The main aim of this chapter is to explore the singularities of speech and writing as language learning sites. To this end, we start with a comparative analysis of speech and writing as language production modes. We first present an overview of the models of oral and written production, followed by the juxtaposition of speech and writing in terms of the physical act of production and the nature of output, relationship with the audience and learners' beliefs and associations. On the basis of this analysis, we draw inferences about the manner in which mode can (similarly or differentially) influence L2 performance and, ultimately, L2 learning. The inferences that we make resonate with theoretical positions on the language-learning potential of L2 writing (Manchón, 2011a,b; Manchón & Williams, 2016) and with the recent theorizing on the connections between SLA processes and features of oral and written modes (Gilabert, Manchón & Vasylets, 2016). We outline the main tenets of these theoretical positions, and highlight the scarcity and need of the empirical research looking into the singularity of the language-learning potential of output practices in oral and written tasks.

3.2 Language production in speech and writing

Speech and writing are multifaceted constructs that defy straightforward definitions. Perhaps they can be best conceptualized as language production modes that exist on a continuum in the form of numerous, often overlapping variations, each being an adaptation of language to the particular circumstances of use (Chafe, 1994; Chafe & Danielewicz, 1987; Halliday, 1987; Horowitz & Samuels, 1987; Tannen, 1982). Casual conversation and academic writing would constitute the two modal points of this

continuum, representing the extremes of what speech and writing permit, and also embodying the prototypical features that are normally associated with one mode or the other (Chafe, 1985). For the purposes of this dissertation, we define prototypical oral performance as phonic language that is typically produced, processed and evaluated in the context of a physical presence of a listener, who can have a role of an interlocutor, responding with the language of her own. On the other hand, we define prototypical writing as a more planned, formal, conscious and deliberate discourse, which is produced in a context where audience is most commonly displaced in time and/or in place, and which in its final form of a written text represents an autonomous, relatively permanent object. Speech and writing are related activities as both represent expression of the inner language into linguistic form (Halliday, 1990). At the same, oral and written production are very different on many dimensions (Horowitz & Samuels, 1987), and the understanding of the singularity of speech and writing as language production modes is crucial to developing an understanding of the singularity of speech and writing as language learning sites.

3.2.1 Models of oral and written production

One way to delve into the idiosyncrasy of the two modes is to compare the psycholinguistic mechanisms involved in speaking and writing. This can be done through the juxtaposition of models of oral and written production (e.g., Kellogg, 1996; Levelt 1989). Speech and writing are intrinsically complex activities encompassing numerous interrelated mental operations (Hayes & Flower, 1980; Kormos, 2006). Models are useful as they represent a simplified outline of the production system. The asset of models is that they propose a relatively precise and analytic definition of oral

and written production activities, both in terms of the process architecture (arrangement of the processes and definition of the sub-processes or operations that compose the processes) and functioning (process management rules, control and activation in working memory, etc.) (see, for instance, Alamargot & Chanquoy, 2001). An additional advantage is that models allow to delimit and focus on the discrete processes of oral and written production without losing the perspective of the entire complex system within which these processes operate. In the next section, we will review and compare several models of oral and written production. It must be highlighted that models of L2 written production have gained less prominence than models of L1 writing, which explains why L2 writing researchers have frequently drawn on L1 models in L2 writing studies (e.g., Ellis & Yuan, 2005). Although this dissertation deals with L2 oral and written task performance, we consider it relevant to review both L1 and L2 production models, which will allow for a broad perspective on the processes involved in speech and writing generally speaking and, as a result, in L2 speech and writing.

3.2.1.1 L1 speech production: Levelt's (1989) model

Levelt's (1989) model is a very well-known model that describes L1 oral language production. The general structure of the model is modular, which means that it consists of a number of autonomous components responsible for different aspects of speech production. These components include the *conceptualizer*, *formulator*, *articulator*, *audition* and the *speech comprehension system*. The model also contemplates three monitor loops and includes different knowledge stores such as stores of lexicon (lemmas and forms), discourse knowledge, encyclopedic knowledge (Figure 4).

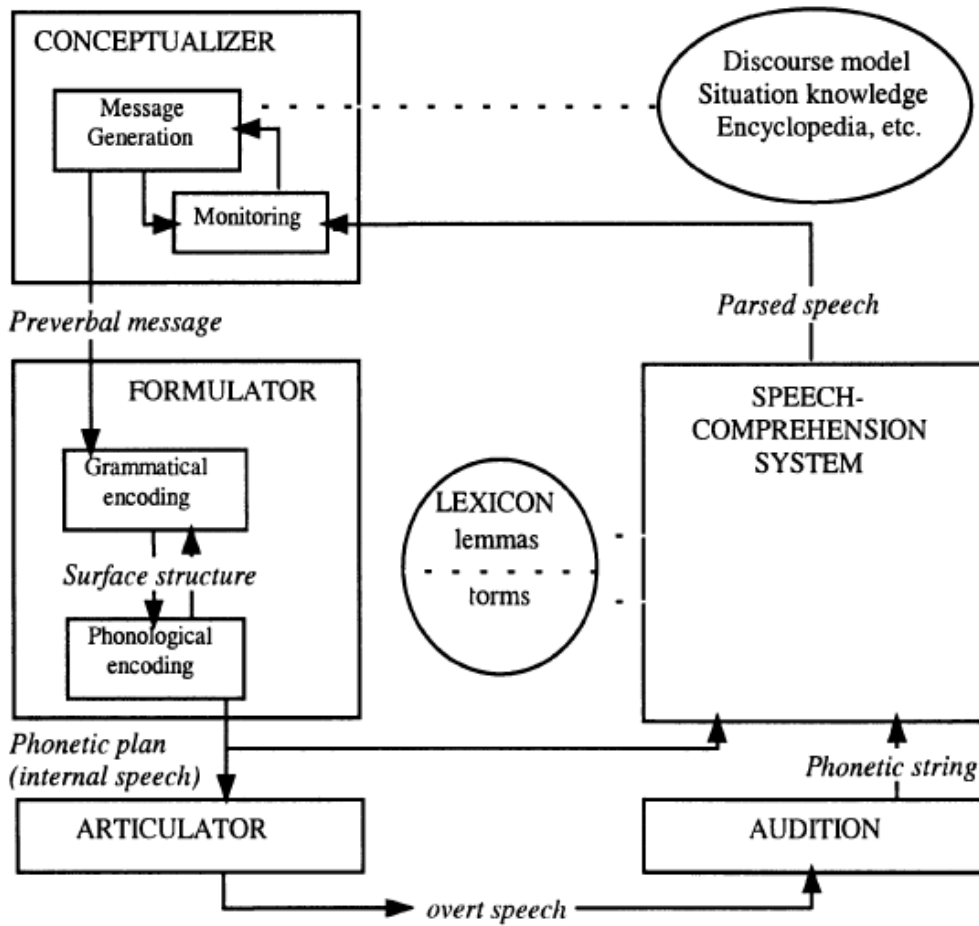


Figure 4. Levelt's (1989) model of L1 speech production

The *conceptualizer* is responsible for generating and monitoring messages, and its work requires conscious attention. The main objective of the conceptualizer is to generate conceptual content through macro- and micro-planning. *Macro-planning* involves the elaboration of communicative intention, which is expressed as a speech act (e.g., request, apology, etc.). *Micro-planning* involves planning of the linguistic realization of the content. The outcome of macro- and micro-planning is the production of a *preverbal plan*, which is not yet totally linguistic. The preverbal plan passes to the *formulator*, where it is given the linguistic form by means of lexico-grammatical and phonological encoding. The output is a *phonetic plan* (internal speech), which is

converted into *overt speech* in the *articulator*. Finally, the model contemplates a control process, assimilated to a *speech-comprehension system*. To avoid duplicating functions, Levelt assumes that the comprehension system attends to both one's own speech and other speakers' production, and that the interpretation of the perceived message is done in the conceptualizer, which is responsible for the generation of one's own message. There are three monitor loops in Levelt's model, whose function is to examine the quality of the outcome of the production processes. The first loop compares the preverbal plan with the original communicative intention before sending it to formulator. The second one checks the internal articulatory plan (internal speech) before articulation, which is also called covert monitoring. The third loop inspects the generated utterance after articulation. If any of these loops detects a problem, several options are available, such as ignoring the problem or revising the output by triggering the production mechanisms for a second time. Levelt posits that speech production modules work incrementally, that is, only a fragment of a module's specific input is needed to initiate the processes within the module. Another important assumption is that work in the *conceptualizer* requires conscious attention, while formulation and articulation are largely automatic and therefore the processing mechanisms can work in parallel.

Levelt's influential model has served as a basis for bilingual models of speech production (Kormos, 2006) and also for L1 written production (Van der Pool, 1995; van Wijk, 1999).

3.2.1.2 Bilingual and L2 speech production: Kormos's (2006) model

In 2006, Kormos proposed her model of bilingual speech production (see Figure 5).

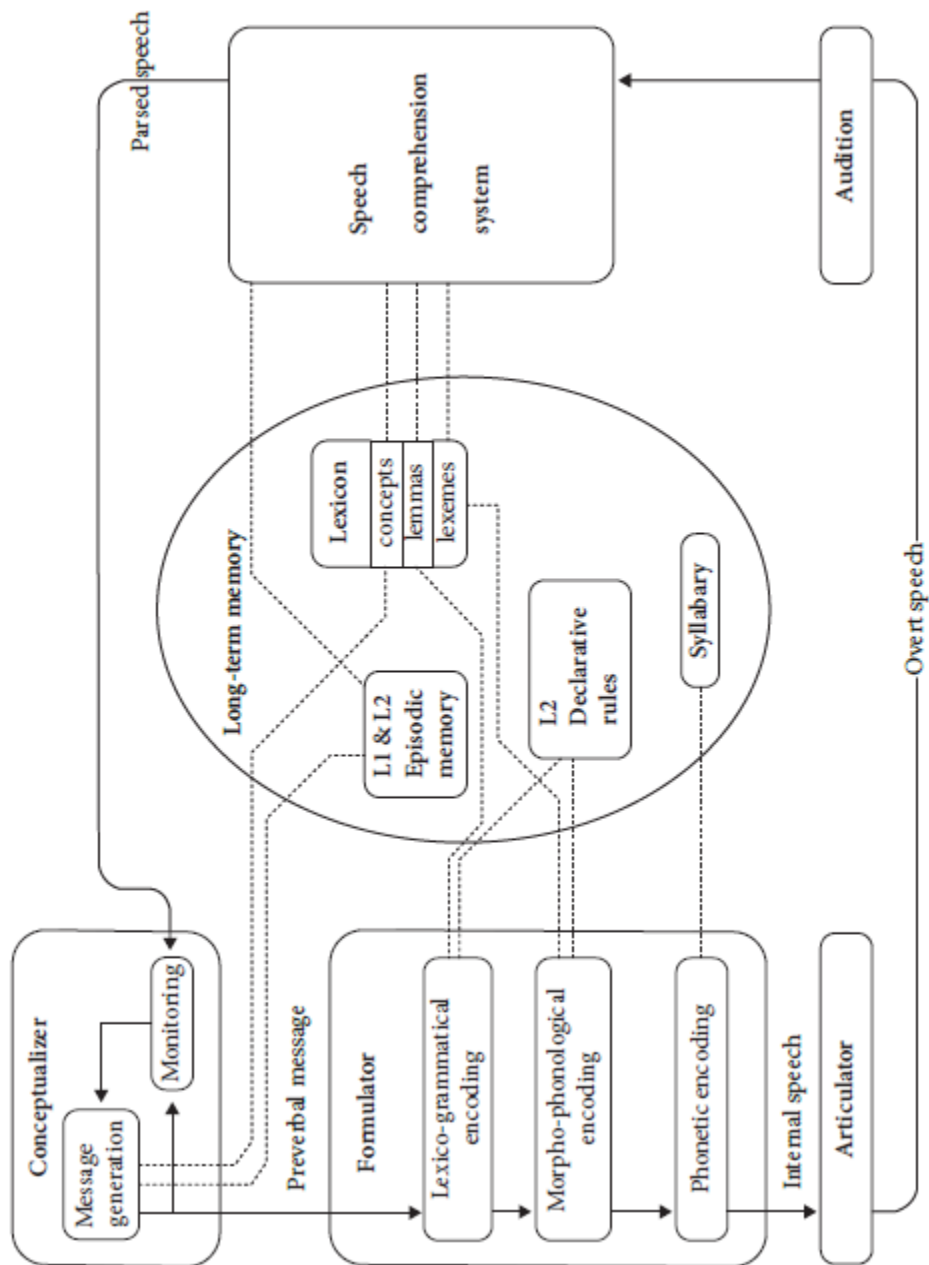


Figure 5. Kormos's (2006) model of bilingual speech production

This model is based on Levelt's (1989) model of L1 speech and it also draws on previous models of L2 speech production by de Bot (1992), Poulisse and Bongaerts (1994), and Towell, Hawkins, and Bazergui (1996). Following the principle of simplicity, which is prevalent in human cognition, Kormos assumes that the speech

production processes in bilingual speakers should not be significantly different from monolingual speakers (see also Kormos, 2011). The general structure and the operational principles proposed in the bilingual speech production model resemble those postulated in Levelt (1989). Accordingly, Kormos posits the existence of the *conceptualizer, formulator, articulator, audition* and *speech comprehension system*. As in L1 models, bilingual speech production works incrementally and processing can be serial or also parallel for L2 proficient speakers.

The model, however, fully accounts for the difference between L1 and L2 knowledge and the different nature of some L2 processing mechanisms. Kormos (2006, 2011) includes an additional store for declarative knowledge of syntactic and phonological L2 rules; also, episodic memory, lexicon and syllabary are conceptualized as enriched with L2 knowledge. The workings of different stages of production are theorized as being similar to L1 production. As in L1, conceptualization in the L2 requires conscious attention and it also involves an additional step to decide on the language in which the message will be spoken. During linguistic encoding, L1 and L2 lemmas compete for activation, which may require more intensive cognitive effort to resolve this conflict. For syntactic encoding, Kormos (2011) posits that there is “no fundamental difference between L1 and L2 production” (p.47); the same applies to phonologic encoding and monitoring, whose basic mechanisms are theorized as essentially the same as in L1. The workings of L2 production system, however, are conceptualized as proficiency-dependent. For low proficient learners, the production stages would require extra investment of cognitive resources, and the workings of the stages, which are purported to be largely automatized in L1 (e.g., formulation, articulation), would need linear processing and conscious attention in L2. With higher

proficiency, all the processes involved in L2 production are believed to level up to the smooth and largely automatic workings of L1 speech production.

3.2.1.3 L1 writing: Hayes and Flower`s (1980) model and subsequent revisions

Model-building of language production in writing started to develop even earlier than speech modeling. Hayes and Flower`s seminal model of L1 was introduced in 1980 and it served as a valuable basis for posterior models (e.g., Kellogg, 1996) which also gained prominence in writing scholarship. Hayes and Flower (1980) elaborated their model on the basis of a verbal protocol of a unique participant, who was asked to comment upon his mental activities during expository text writing. In terms of architecture, the model has three principal parts: the *task environment*, the *writer`s long term memory*, and the general writing process (see Figure 6).

The *task environment* comprises several writer-external factors that can influence performance: the type of the written assignment, communicative goal (audience), motivational factors (motivating cues). Another component, *text produced so far*, is the already written text, which can serve as an object for revision and as a reference to make future decisions during the writing process (i.e., all the parts of the text have to be coherently connected; thus, when a text starts taking the visible form, it starts influencing what will be written next). The *writer`s long term memory* contains the areas of *knowledge of topics* (domain knowledge), *knowledge of audience* (pragmatic knowledge), *stored writing plans* (knowledge about specific text types).

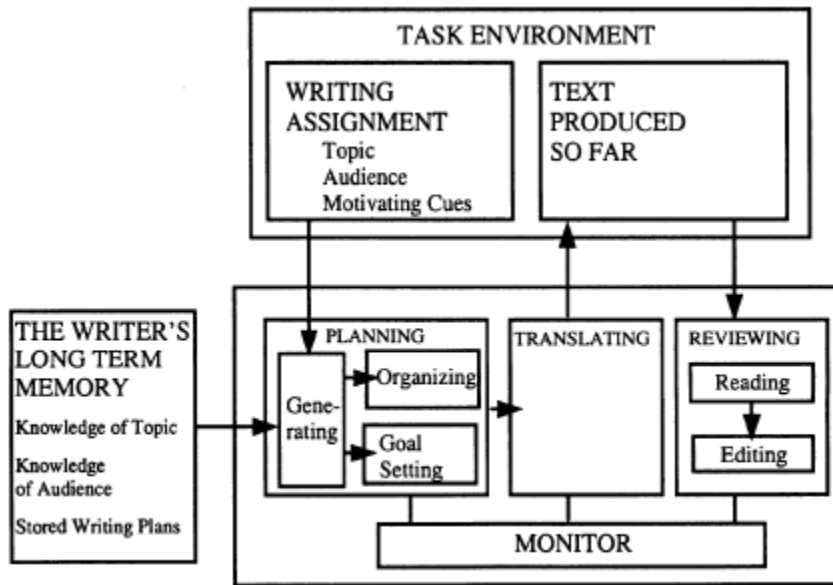


Figure 6. Hayes and Flower's (1980) model of L1 written production

The general writing process is composed of the processes of *planning*, *translating* and *reviewing* under the supervision of a *monitor*. The main function of *planning*, which comprises the sub-processes of *generating*, *organizing* and *goal-setting*, is to create conceptual content by using the domain knowledge from the long-term memory and information from task environment, if available. At this stage, the writer also creates the writing plan, which will guide and organize the subsequent writing stages. Planning work can be creative or writers can use the pre-defined writing plans stored in the long-term memory (e.g., writers can use a pre-defined schema about the structure and linguistic characteristics of a text in a certain genre). *Translating* gives conceptual content the corresponding linguistic form. *Reviewing* compares the written text with the goals set by the writer, and it entails reading of the produced text (*reading*) and its possible correction (*editing*). *Planning*, *translating* and *reviewing* are managed by a control process – *monitoring*.

The original Hayes and Flower's (1980) model has been subsequently revised by Hayes (1996, 2012) who made some substantial changes. Thus, in his most recent conceptualization of the writing activity, Hayes (2012) distinguishes between the control level, writing processes, task environment and the resource level (see Figure 7). Hayes (2012) highlights that “the addition of working memory repaired an obvious oversight in the original model” (p.370). Other important changes include the addition of the transcription process and motivation, and the removal of the monitor, the planning process, and the revision/reviewing process.

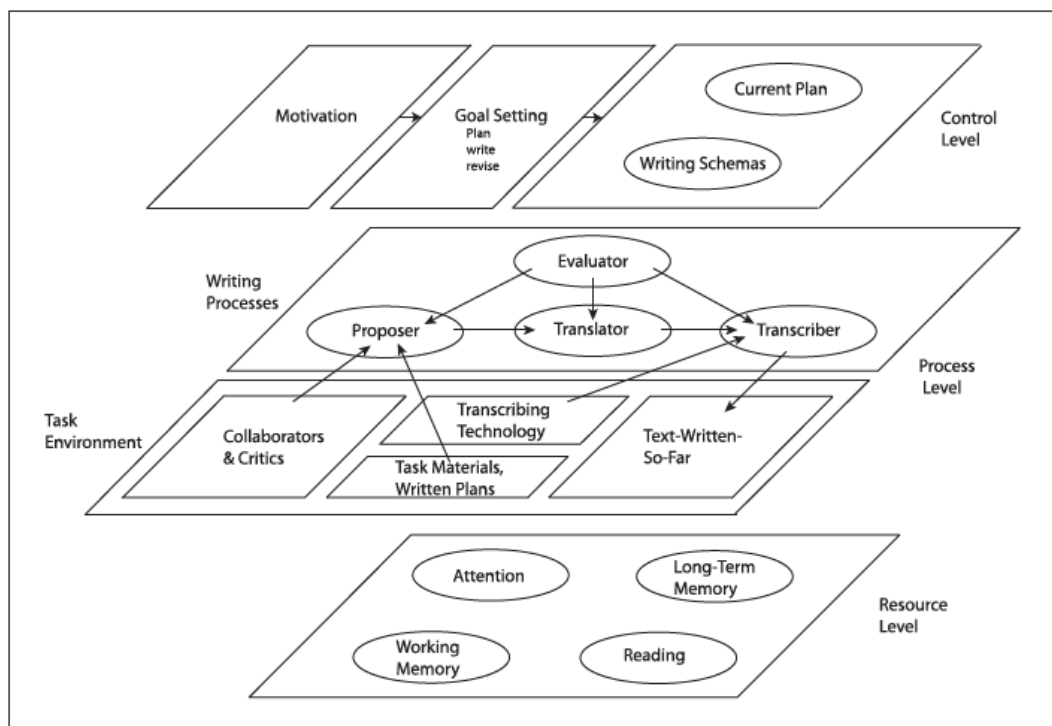


Figure 7. Hayes's (2012) revised model of L1 written production

Thus, in this new model, Hayes (2012) adds motivation because it influences “people's willingness to engage in writing” (p.372). Also, Hayes defines the stages of planning, transcription and reviewing in a rather different way. First of all, more

prominence is given to transcription which has been previously ignored in L1 writing modeling due to a shared belief that transcription in adults is highly automatized and, thus, does not have any significant impact on other writing processes. Hayes (2012), however, highlights that, as a result of his empirical work in the recent years, his view on the role of transcription has changed and his current view is that “transcription does compete with other writing processes for cognitive sources in both adults and children and must be accounted for in modeling all writers” (p.371-372). The most prominent change, however, concerns the new conceptualization of planning and revision. Hayes sees both planning and revision as specialized writing activities which involve the complete writing process. Hayes (2012) poses that:

Creating a written plan involves a complete writing process that produces a text designed to aid the author of the plan in producing another text. Viewed in this way, a separate planning process would simply duplicate an activity that can already be performed by the writing model. (p.376)

In Hayes`s (2012) view, the same applies to revision: “revision, like planning, is seen not as a separate writing process parallel to the other writing processes[...] but rather as a special application of the writing model”(p.376). Although this new conceptualization of the writing processes is, without doubt, original and valuable as it provides an updated vision on the nature of writing, we consider that such globalized view has also a disadvantage as it precludes from discerning the specific characteristics of the processes of planning and revision.

3.2.1.4 L1 writing: Kellogg`s (1996) model

Another classical model of L1 written production was advanced by Kellogg (1996). This model has gained much prominence and it is still widely used as a framework in

L1 and L2 writing studies. The popularity of this model can be accounted by the fact that it represents a convenient heuristic which comprehensibly defines the principal writing stages and links them to the component of working memory as defined by Baddeley (1986). Thus, in his model, Kellogg distinguishes three main components: formulation, execution and monitoring, each one entailing two sub-processes. *Formulation* is composed of *planning* (conceptual preparation of the message) and *translating* (transformation of ideas into linguistic structures) (see Figure 8).

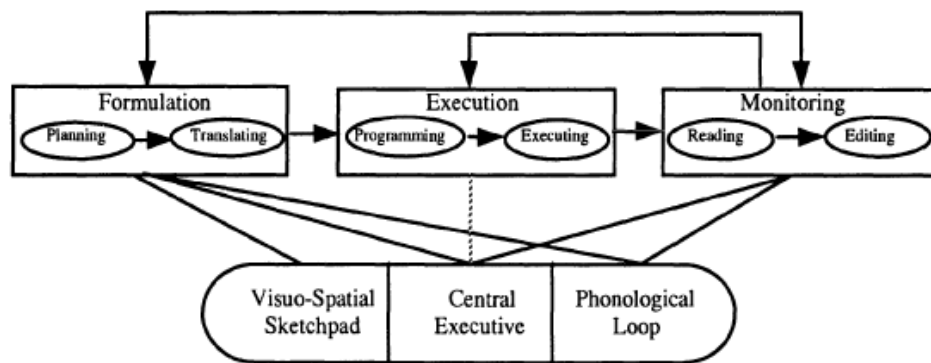


Figure 8. Kellogg's (1996) model of L1 written production

The output of *formulation* is the input for *execution*, which entails the sub-processes of *programming* (preparation of the motoric system) and *executing* (physical realization of the message). *Monitoring* implies *reading* the produced text and *editing* of the message. *Reading* of the message can take place during and after text elaboration. The aim of *editing* is to detect, diagnose and correct problems in the produced text. *Editing* also compares the writer's intention with the output of different systems, with the discrepancies leading to a feedback to the preceding process and re-starting it again.

Kellogg posits that writing processes can be implemented both in a serial fashion and recursively.

Besides the definition of the global architecture and functions of each sub-process, Kellogg also describes the interactions between writing processes and the components of working memory (Baddeley, 1986). In Baddeley's view, the working memory system consists of a central executive, globally dedicated to complex processes, and two slave registers, the visuo-spatial sketchpad and the articulatory loop, which allow to maintain visual and phonological representations. *Formulation* is the most taxing process, as it draws upon all the sub-components of working memory. *Monitoring* is also demanding, as it needs the resources of the central executive and phonological loop. The least taxing process is *execution*, which, in Kellogg's view, needs only the central executive. By establishing the theoretical links between writing processes and working memory, Kellogg highlighted relevance of the limitations in cognitive resources for writing.

3.2.1.5 LI writing: Van der Pool's (1995) and van Wijk's (1999) models

Van der Pool's (1995) and van Wijk's (1999) model is of interest because it represents an adaptation of Levelt's (1989) model to writing (Figure 9). This model has the same structure as in Levelt, but it shows in somewhat greater detail the operations of each component. The main point in this adaptation consists in the inclusion of the processes and representations specific to writing, such as *graphemic encoding*, *graphemic plan*, *graphemic decoding* or *text record*. Such a straightforward adaptation of the speech model to writing has been, however, criticized, as it was considered to be shallow in reflecting the idiosyncratic nature of written production. It is precisely for this reason,

that Alamargot and Chanquoy (2001) consider that this model is inadequate for the comparison of the speech and writing, as “many questions arise as soon as it is necessary to explain, on a functional viewpoint, the very deep differences between speaking and writing activities” (p.14).

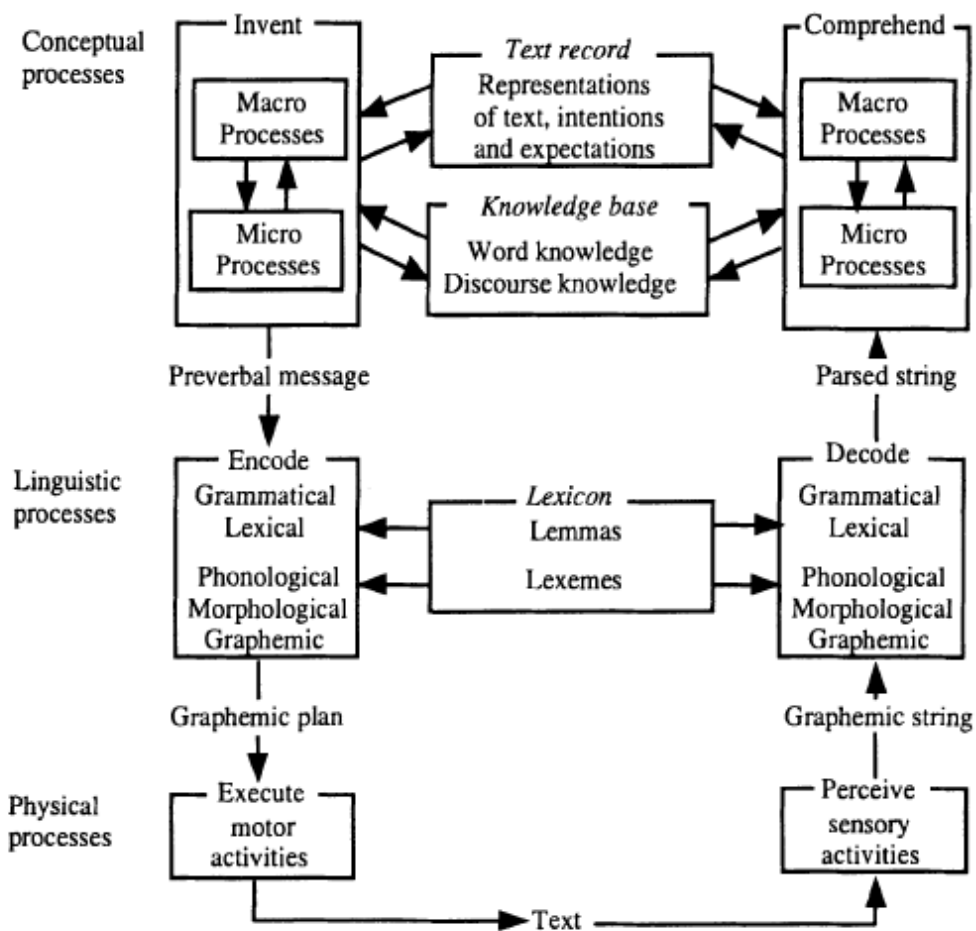


Figure 9. Model of written production inspired by Levelt’s model (1989, adapted from van Wijk, 1999)

3.2.1.6 Models of L2 writing

There have been several attempts to model L2 written production (e.g., Börner, 1987; Krings, 1992; Zimmermann, 2000). These proposals have not had much repercussion, probably because they were mere reflections of L1 writing models and did not specify the singularity of L2 writing in sufficient detail. Zimmermann's (2000) model, for example, proposed the same stages of planning, linguistic encoding and reviewing as in L1 writing models (see Figure 10).

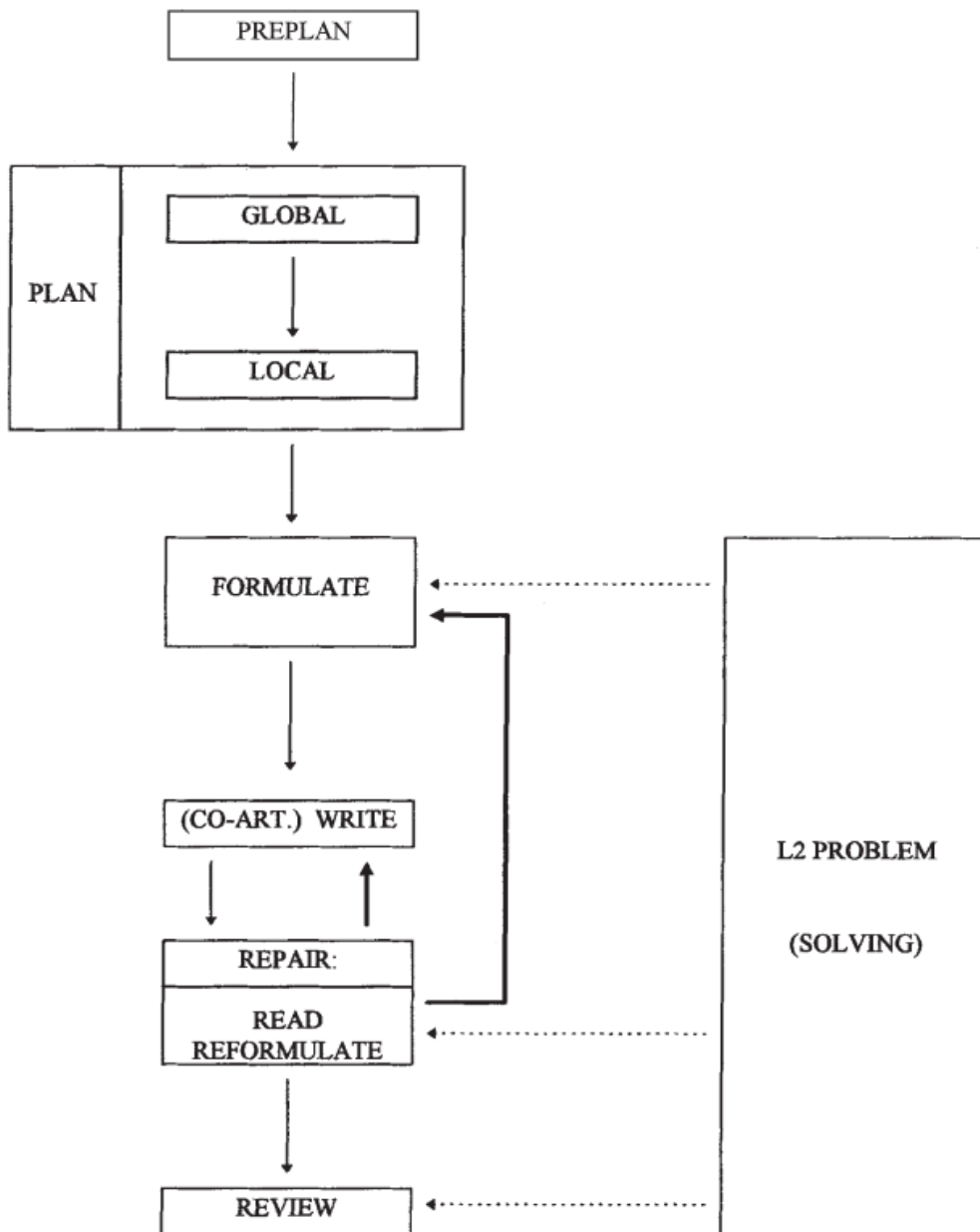


Figure 10. Zimmermann's (2000) model of L2 written production

Zimmermann (2000) represents the components in a linear fashion, but emphasizes that “the stages and processes of writing are overlapping and recursive and [...] in principle almost any sub-process can occur at any time” (p. 54). The limitations of this proposal is that it does not specify in detail the working of the model's components and ignores the role of text or knowledge stores (e.g., discourse knowledge,

world knowledge, etc.). The specific contribution of Zimmermann (2000), however, is the special status assigned to *formulation* (i.e., linguistic encoding), as he posited that *formulation* was the stage “where L2 problem solving activity appears most typically” (p. 85). The special status of formulation has been borne out in the empirical L2 writing research. Thus, when reporting the results of a research program, which focused on the exploration of L2 writing processes, Manchón, Roca de Larios, and Murphy (2009) highlight the specific nature and specific status of formulation in L2 writing. In this regard, they stress “the more labor-intensive nature” of formulation in L2 writing (p.110). The workings of formulation (as of other writing processes) has also been found to be proficiency dependent, because, as suggested by Manchón et al (2009), “the likelihood of attending to higher-level concerns while writing increases as writers become more capable of using the L2, and this applies to the problems writers pose themselves while planning, formulating and revising their texts” (p.116).

In sum, the most recent conceptualizations of writing activity in L2 assign a more prominent role to linguistic formulation, which is, in fact, “the only nonoptional writing process” (Manchón et al., 2009, p.116), provided that writers may or may not engage in substantial planning or revision when creating a writing text.

3.2.1.7 *Comparison of oral and written production models*

The juxtaposition of speech and writing production models shows that, irrespective whether L1 or L2 is involved, both speech and writing consist of similar macro-stages subordinated to an instance of control. The macro-stages that can be found in both speech and writing are: (1) planning the content, (2) transforming this content into linguistic form, (3) physical execution of an auditory signal or visible sign, and (4)

revision of production. Also, in both modes, language production is viewed as constrained by the limitations in the processing capacity of the speaker or writer. The comparison of the models also reveals differences in the workings of oral and written production, as writing models emphasize the recursive and cyclical nature of the writing processes, although without providing the specific details how this recursiveness is implemented (Alamargot & Chanquoy, 2001). The overall conclusion that can be drawn from the comparison of the models is that the global architecture of the production processes is similar in the two modes. This, however, does not reveal much about the idiosyncrasy of speech and writing. To deepen our understanding of the modes singularity, further analysis is needed, which goes beyond the comparison of the processes structure and functions.

3.2.2 Psycholinguistic comparison of speech and writing

Despite obvious similarities in the global architecture of speech and writing, the two modes also display marked differences, whose consideration is crucial in ascertaining the language-learning potential that may derive from L2 oral and written task performance. To disclose the idiosyncratic features, we will compare speech and writing along three dimensions: (1) physical characteristics of the act of production and of the resulting output, (2) relationship with the audience, and (3) beliefs and associations that language users might have about speech and writing.

3.2.2.1 Physical characteristics of the act of production and output

The first dimension deals with the nature of the language signal and of the final product in the two modalities. Speech involves producing language with the help of auditory

signals, which is physically less effortful and a faster way of production as compared to production of written signs (Horowitz & Samuels, 1987). Speaking takes place under the constraints of on-line processing, but, at the same time, it has an additional advantage of prosodic (pitch, intonation, etc.) and paralinguistic resources (gestures, facial expressions) which convey part of the information and, thus, allow speakers to be less explicit. Articulation of sounds in speaking is inherently faster than in writing, so speakers are able to produce more linguistic material per unit of time. However, online pressures and the evanescent character of oral output put constraints on speech production processes that may limit speakers' variation in production, use of complex or non-fully automatized linguistic structures, and deployment of explicit or metalinguistic knowledge (Williams, 2012). Time pressure during speaking also results in hesitations and pauses, which are more evident in oral production than in writing, as the former exists in the form of a process, while the latter is typically available in the form of the final polished product, where the production interruptions become invisible (Halliday, 1987).

Writing, in contrast, is offline (with the exception of synchronous computer-mediated written communication) and self-paced and its output is visible and relatively permanent, which eliminates certain constraints and pressures inherent to speech. Thus, writers have more time to exploit and strategically distribute their cognitive resources, which promotes the implementation of the appropriate metacognitive strategies, and also benefits language production processes (Kormos & Trebits, 2012). In writing, language users do not feel the tension to prioritize one process over the other, and they can pay attention to both content and linguistic form (Chafe, 1985, 1988). In other words, the offline and self-paced nature of written production facilitates retrieval/generation of the ideational content and allows for exhaustive lexical and

syntactic searchers, as well as enhanced monitoring (Chafe, 1985; Chafe & Danielewicz, 1987). As a consequence, the encoding of the messages in writing can be more precise and linguistically more complex as opposed to fast-flowing speech.

Production processes in writing are also influenced by the characteristics of the final output. Thus, visible texts serve as an extension of working memory, facilitating in this way on-line planning and the recall of the information from long-term memory (Grabowski, 2007). The static written text can also be repeatedly scanned and re-analyzed, which supports revision (Allal, Chanquoy, & Largy, 2004; Piolat, Roussey, Olive, & Amada, 2004). At the same time, visibility of the text calls for more precision and accuracy, which adds to the difficulty of writing (Tannen, 1982). The slower pace of writing, on the other hand, makes written production a more time-intensive activity. Recent theorizing on the language learning potential of writing (Manchón & Williams, 2016) has emphasized the precise language learning benefits that derive from the time availability, the visibility, and the permanence of written texts. Finally, it must be noted, that the spoken-written distinctions, attributable to the physical characteristics of the language signal and the output are difficult to override, and, consequently, they remain present even when the uses or goals of writing and speech overlap (Chafe, 1987, 1992; Chafe & Danielewicz, 1987).

3.2.2.2 *Relationship with the audience*

Another aspect to be considered is the relationship with the audience (Horowitz & Samuels, 1987; Nystrand, 1987; Ong, 1982). Oral production typically entails a face-to-face contact between the communicator and the recipient of the message. In this way, spoken utterances are produced in the environment where background information is

shared and there is a possibility to employ modulatory possibilities of the voice (pitch, intonation) and also gestures to convey part of the meaning. This allows speakers to be less linguistically explicit and to construct cohesive discourse differently as compared to writing, by employing, for example, deixis or prosodic devices (Chafe, 1985). Face-to-face contact also means that a speaker can receive instant feedback about the comprehensibility or accuracy of production (Bereiter & Scardamalia, 1987). This allows speakers to make quick adjustments to ensure the correct functioning of the communicative channel. Immediacy of face-to-face communication poses pressures on both speaker and listener, as the former has to produce speech on-line and the latter has to decipher/comprehend speech on-line. For efficient communication, speakers have to adapt to the processing capacities of listeners. Repercussions of the on-line pressure and adjustment to the listener might be manifested in the quality of linguistic performance in the form of simplification of lexis and grammar or use of short idea units (Chafe & Danielewosz, 1987). Speakers also tend to adapt their language to the level of proficiency of the recipient/s of the message, which also conditions their linguistic choices (Bereiter & Scardamalia, 1987). Additionally, in order to prevent breakdowns in communication, speakers try to avoid long pauses in their discourse, which enhances the pressures of the on-line production and can complicate the processes of planning, formulation and monitoring, resulting in simplified or non-accurate linguistic choices (Chafe, 1987; Olson, Torrance & Hildyard, 1985). If oral production is purely monologic and there is no reciprocal communication with the listener, influences stemming from the immediate presence of audience will not be present, but speakers will still be under the pressures posed by real-time production (Nystrand, 1987).

In contrast to oral communication, “the writer`s audience is always a fiction” (Ong, 1977, p. 53), which means that written production is more often a solitary activity,

addressing a hypothetical and sometimes unknown reader, who is displaced in time and in space. Under such conditions, written messages do not function communicatively at the time of their creation (Nystrand, 1987), and it becomes an important task for a writer to facilitate the correct decipherment of the texts they produce. Thus, in order to ensure that their message is comprehensible and communicatively effective, writers have to aim at precision in their lexical and grammatical choices, and they have to pay special attention to the coherence and completeness of their discourse (Chafe, 1985, 1994; Tannen, 1982). It must be noted, however, that the processing constraints, as well the linguistic consequences attributable to the presence or absence of audience, can be easily overridden if the conditions of production change, as for example in internet chatting - a form of written language production, in which audience is displaced in place, but not in time, and the communication, thus, becomes highly interactive and spoken-like.

3.2.2.3 *Learners' beliefs and associations*

Finally, the third key dimension distinguishing oral and written discourse concerns different histories, as well as divergencies in the traditional uses of writing and speech (Ong, 1982). As compared to oral language, writing appeared much more recently in human society. For centuries, the use of the written word was confined to a narrow circle of privileged language users, and the goals that writing served in the society significantly diverged from those of the oral language (Ong, 1982). Nowadays, writing is widely practiced by the vast majority of the language users in literate communities, and the contexts of use of the written language have become much more diversified and overlapping with the uses of the oral language. In spite of these important changes, the

written word has not lost its normative power and authority, and the traditional associations tied to writing are still deeply interiorized in the psyche of the modern society (Ong, 1977, 1982). As a consequence, today`s language users still attribute to written language its traditional normative power and authority, and readily associate writing with more formal and planned discourse, which is acquired explicitly and with effort (Flower, 1990). This is important since these associations might condition writers` choices, implicitly inducing them to employ more sophisticated lexis or more complex grammar.

3.2.3 Comparison of speech and writing with planned and unplanned discourse

Differences between oral and written performance can also be seen as conceptually analogous to the differences between unplanned and planned performance. Real-time oral production would be more associated with unplanned discourse, while writing, which inherently entails in-built planning opportunities, would conceptually correspond to planned discourse. In her oft-cited publication, Ochs (1979) defined unplanned discourse as characterized by greater word repetition, less complex syntax, and also preference for linguistic structures acquired earlier in life. In contrast, planned discourse was described as displaying more complex structures as well as drawing upon knowledge acquired in formal contexts. In a similar line of thinking, L2 acquisition researchers have regarded planning as a crucial variable that may impact both L2 development and production (Ortega, 1999). Availability of planning time has been posited to lessen the cognitive load imposed by the task and free up attentional resources, inducing focus-on-form processes and the engagement of controlled

processing as well as access and use of L2 forms from the upper limits of the interlanguage (see contributions to Ellis, 2005).

3.2.4 Summary of the psycholinguistic comparison of speech and writing

This comparative analysis reveals a number of singularities of speech and writing in terms of the physical characteristics of the act of production and output, relationship with the audience and learners' beliefs and associations. We would like to suggest that these singularities determine the idiosyncrasy of linguistic behaviors and learning possibilities in the two modes. The idiosyncrasy of the physical attributes of the act of production and output is plausibly the factor which contributes most to the singularity of speech and writing in general (Chafe, 1987). It could be suggested that the same attributes might, in great part, define the idiosyncrasy of speech and writing as language learning sites (although, the relationship with the audience and mode associations, undeniably, play their roles as well). Speech represents phonic language, characterized by a high (as compared to writing) speed of delivery and the evanescence of the resulting output. In relation to language learning, rapidity and non-permanence of speech may limit the processes of online planning, linguistic encoding, and monitoring or editing during oral performance.

Writing is graphic language, which, as opposed to speech, is slow and self-paced in production, but visible and permanent in output. The slow pace and visibility convert a writing task into a self-regulated and recursive process, which affords learners time and possibilities to strategically distribute cognitive resources as well as extensively plan (including during production), carefully encode, and monitor or edit their language. Williams (2012), for example, singled out pace and permanence of writing as the main

features which can facilitate learning. In Williams's view, these characteristics may allow more learner control over attentional resources, as well as a greater need and a better opportunity for focus-on-form and use of explicit knowledge both during and after task performance. In a similar line, Polio (2012) assumed that even under time pressure, any writing task "can tap into both explicit and implicit knowledge, in contrast to speaking tasks, which more often require implicit knowledge to complete fluently" (p. 322). Specific affordances of written production associated with linguistic processing that may derive from the greater time availability and visibility of output form the basis of the theorizing on the language learning potential of writing theorizing (Manchón, 2011, 2014, 2016; Manchón & Williams, 2016; Williams, 2012), which will be discussed in more detail in the next section.

3.3 The language learning potential of L2 writing

In recent SLA-oriented writing scholarship, writing is seen much more than an important literacy skill to acquire. Writing is also regarded as a tool to promote L2 development through engaging a number of processes associated with language learning (Manchón & Williams, 2016; Williams, 2012). Positive effects for L2 learning are believed to derive from individual as well collaborative writing, and also from processing of written feedback (Manchón, 2011, 2014a, b; Bitchener, 2012; Bitchener & Knoch, 2010; Polio, 2012). Language learning through writing is not, however, uniform for different kinds of writing activities. Thus, as argued by Manchón (2014a), any discussion of writing as tool of language learning should take into consideration that L2 writing encompasses:

Individual and collaborative writing, in time-constrained and time-unlimited conditions, in both pen-and-paper and computer-mediated environments, totally

or partially performed within and/or outside the confines of the language classroom, with and without the availability of (printed or electronic) external sources, and, importantly, with and without the availability of (different types of) feedback provided at different points in the composing process, which may serve different functions. (pp. 29–30)

Of specific relevance to this dissertation is the language learning potential of individual L2 writing tasks. According to Manchón (2014b), the role of individual writing in language development can be attributed to the combination of a number of factors: (1) the availability of time, (2) the visibility and permanence of written text, (3) the challenging, problem-solving nature of some writing tasks, and (4) the possibility of metalinguistic reflection during writing. These ideas are in line with the early research by Cumming (1990), who was a pioneer in advancing the argument for the case of L2 learning through individual writing. According to Cumming, “composing might function broadly as a psycholinguistic output condition wherein learners analyze and consolidate second language knowledge that they have previously (but not yet fully) acquired” (p.483). In his view, the learning-beneficial effects stem primarily from the challenging, problem-solving nature of the writing task, absence of time-pressure and from the “natural disjuncture between written text and the mental processes of generating and assessing it” (p. 483). These attributes should facilitate reflective metalinguistic thinking, form-meaning connections and also monitoring, with the resulting consolidation and enhanced control of L2 knowledge. Later theoretical proposals (Manchón, 2014; Manchón & Williams, 2016; Williams, 2012) are in line with Cumming’s early ideas, and Manchón (2014) points out:

The pace and permanence of writing make it possible for L2 writers to be more in control of their attentional resources, more prone to prioritize linguistic concerns (in contrast to oral production) and, accordingly, more likely to attend to focus on

language during both their composing activity and their processing of the feedback received. Similarly, the problem-solving activity engaged in during writing requires decision-making (at various levels) and deep linguistic processing with potential beneficial effects on learning. (p.99)

Additionally, writing activity is posited to advance linguistic competencies by engaging learners in the formulation and testing of hypotheses about the L2 and developing fluency and automaticity (Manchón, 2014). In their most recent conceptualization of the arguments for the potential of writing as a means to reinforce L2 knowledge, Manchón and Williams (2016) posit that the pace and permanence of writing allow writers to reflect on the linguistic demands of the task, engage in deeper planning, draw on different knowledge stores, and employ all these resources to revise and edit the output.

Writers can access both implicit and explicit knowledge, but it is still an empirical question how exactly these types of knowledge are deployed during writing. As argued by Manchón and Williams (2016), there are different possibilities. Thus, writers may access explicit knowledge and compare that knowledge with the produced output, or they can also access implicit and unanalyzed knowledge. Both suggest enhanced attention to linguistic forms during writing, which can potentially result in more accurate and complex performance than in tasks in which this access is not available. In other words, writing may result in more accurate and complex output than speaking. Empirical evidence for these claims comes in part from the few empirical studies that have compared oral and written production in terms of accuracy and complexity (Kormos, 2014; Tavakoli, 2014). However, because of the scarcity of empirical data to substantiate claims about the advantages of writing, these claims still remain largely speculative, and more research is required to develop a convincing

empirical base for the purported language learning potential of writing. Compelling evidence for the singularity of writing as a language learning site can come from the studies comparing written and oral production on the same task. Juxtaposition of modes will contribute to a more rigorous conceptualization of the language learning potential of writing tasks and will allow to better reveal the singularity of task-based performance and learning in the two modes.

3.4 Juxtaposition of the language learning potential of oral and written tasks

As Bygate et al. (2014) have pointed out, “an adequate version of TBLT must enable its learners to engage with it through both oral and written media” (p.x). In light of the demands of a more mode-sensitive and mode-balanced theoretical and empirical research agenda, Gilabert, Manchón and Vasylets (2016) explored, from the theoretical point of view, the way oral and written modes may differentially influence processes involved in second language acquisition in the context of TBLT. To this end, Gilabert et al.(2016) established links between the characteristics of oral and written tasks and language learning processes as theorized in Leow`s (2015) model, which brings together the commonly agreed-upon constructs and phases of L2 development (see Figure 11).

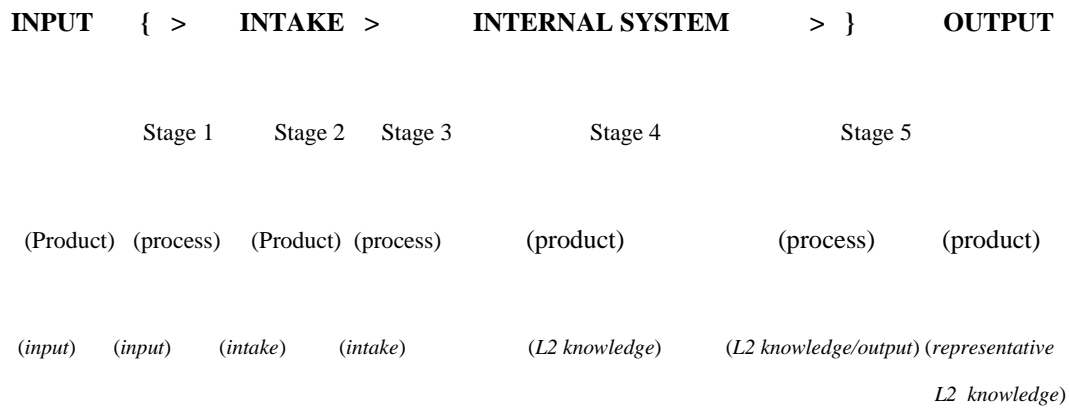


Figure 11. Leow`s (2015) model of the stages of learning processes in SLA

In Leow`s (2015) model, input and output are external products, while learning is conceptualized as encompassing both products and processes. Learning as a *process* is internal and comprises the stages of (1) *input processing* (the initial stage of the learning process, which contains the phases of perception, detection, and noticing of both content and linguistic data found in the input), (2) *intake processing* (form-meaning connections, hypothesis formation and testing, as well as hypothesis modification, rejection or confirmation), and (3) *L2 knowledge processing* (integration and analysis of L2 internal representations, and also learning resulting from the learners` manipulation of L2 knowledge). Learning as a *product* is presented both internally (L2 knowledge) and externally (representative L2 knowledge). TBLT tasks are generators of input and output and are susceptible of potentially engendering all the internal processes and products included in Leow`s model.

3.4.1 *Input and input processing*

There is a general consensus that input is necessary for learning (Gass, 1997), hence the attention paid to input in SLA research. The amount, quality, and access to input can vary considerably depending on teaching approaches and practices (Long, 2015; Long & Robinson, 1998). As Gilabert et al. (2016) point out, input in TBLT takes the shape of *positive evidence* of the target language or is presented as *corrective feedback*, and it is typically part of a dynamic, goal-oriented, input-output-feedback cycle in both the oral and written modes. The authors also remind that input is external to the learner, and how it is processed depends on both learner characteristics (e.g., learner internal syllabus, communicative needs, developmental readiness, felt task demands, prior language knowledge in terms of L2 proficiency or L1, processing capabilities, motivation, task construal and agency, etc.) and input characteristics (e.g., saliency, communicative value of the form or input mode) (see also Han, Park, & Combs, 2008 for a review). When contrasting the two modes in terms of their affordances for learning from input, Gilabert et al. (2016) indicate that oral input poses considerable attentional demands. Because of the evanescent nature of the oral input, linguistic forms stay available for noticing for just a fraction of a second. The challenge is even greater for further processing following noticing (e.g., form–function mapping or hypothesis formation). On the contrary, the permanence and self-paced nature of written input may liberate attentional resources that may facilitate noticing and learning from input. In terms of empirical evidence, the bulk of the research has investigated written input processing, and no systematic agenda is available for comparing the effectiveness of learning from oral versus written input presented as positive evidence.

Input can also be presented as corrective feedback. In this regard, Gilabert et al. (2016) find certain advantages of visible and permanent written feedback as compared

to non-visible and transient oral feedback. As Gilabert et al. (2016) point out, oral feedback has to be attended under on-line pressures, while corrective written feedback is usually explicit and visible, and there is more time to process it. Although these factors do not guarantee learning, they undoubtedly facilitate feedback noticing, which is a prerequisite for uptake and further processing. In terms of empirical evidence, feedback studies in the oral domain, for example, have shown us that under certain circumstances, oral feedback may not be noticed and therefore is not processed any further (Goo, 2012). At the same time, a number of recent studies which employed immediate and delayed post-tests as a measure of learning, have demonstrated that written, corrective feedback improved accuracy of performance and retained it over time (Bitchener & Knoch, 2010; Sheen, 2007). At the same time, the authors highlight that in spite of the positive evidence for the role of written feedback in SLA, it is not yet clear how extensive this role might be, or which types of learners (e.g., in terms of L2 proficiency or cognitive abilities) can benefit from it most (see also Bitchener, 2012; Bitchener & Storch, 2016).

3.4.2 *Intake and intake processing*

The sole exposure to input and noticing of linguistic features is not enough for learning to take place. Input has to become *intake*, which Reinders (2012) defines as “a subset of the detected input (comprehended or not), held in short-term memory, from which connections with long-term memory are potentially created or strengthened” (p. 28). According to VanPatten (2011), the process of the conversion of input into intake involves establishing form–meaning connections, which is the process of associating these two aspects of language. Creating form-meaning connections is conceptualized as

an incremental and recursive process. This process which does not automatically imply acquisition, but its product (i.e., *intake*), which is stored in working memory, has a potential of being further processed and incorporated into the L2 system. Establishing form–meaning connections can be mediated by different factors, such as the learner characteristics, learning context or nature/mode of the input (VanPatten, Williams, & Rott, 2004). As far as mode is concerned, Gilabert et al. (2016) claim that input in both modes has the potential to engage learners with the transformation of noticed input into intake through form–meaning connections. Because of its fixed nature, written output may be facilitative in the input-to-intake transition. Gilabert et al. (2016) theorize that access to written input during task performance may assist intake processing mechanisms because the written mode allows for working memory processes to be activated following noticing (Robinson, 2011a). The potential step-by-step nature of the written mode may promote recursiveness, that is the learners’ going back to unknown, new items in the input, and can also facilitate available resources for engaging in hypothesis formation and cognitive comparisons. Empirical investigations of intake and intake processing are scarce in both modes, which can be attributed to the difficulties in the measurement and operationalization of the internal processes

3.4.3 L2 knowledge and knowledge processing

Knowledge processing includes internalization, modification, and consolidation of knowledge (Leow, 2015). Associated with knowledge processing are reconceptualization or restructuring leading to grammatization and syntactization in the L2 (Robinson, 2011a). Automatization and consolidation of memories are also considered dimensions of knowledge processing. From the perspective of the

idiosyncrasies of oral and written modes, Gilabert et al. (2016) suggest that such processes, which require deeper level of processing, would be facilitated in the written rather than in the oral mode. Lack of empirical evidence, however, precludes definitive conclusion in this regard and, undeniably, more research is required. In terms of empirical investigation, Gilabert et al. (2016) point out that this is the area where TBLT research has advanced the least, which can be explained by the fact that knowledge processing is not open to direct introspection.

3.4.4 The language learning potential of oral and written output practice

As Leow (2015) puts it, output is both a product (what is learned) and a learning process within the knowledge processing stage. As Gilabert et al. (2016) point out, investigation of output as a product (i.e., task performance studies) has been the main preoccupation of the TBLT research agenda since the mid-1990s, and considerable research efforts have been put into measuring the effects of manipulating task design features on both *L2 performance* (typically operationalized as complexity, accuracy, and fluency, or CAF; see, e.g., Robinson, 2011a; Skehan, 2009) and *L2 acquisition*.

From an acquisitional perspective, the Output Hypothesis (Swain, 1995, 1998, 2005) posited that “the act of producing language (speaking or writing) constitutes, under certain circumstances, part of the process of second language learning” (Swain, 2005, p. 471), moving learners from semantic processing in comprehension to more syntactic processing in production. Syntactic processing demands higher attention to linguistic forms and deeper language analysis, with potential consequent effects on language development. The production of output is postulated to trigger the whole range of beneficial processes, such as noticing and focus on form, hypothesis testing,

metalinguistic reflection, and automatization (DeBot, 1996; Erlam, Loewen, & Philp, 2009; Uggen, 2012; Swain, 1995, 2005) (see Figure 12).

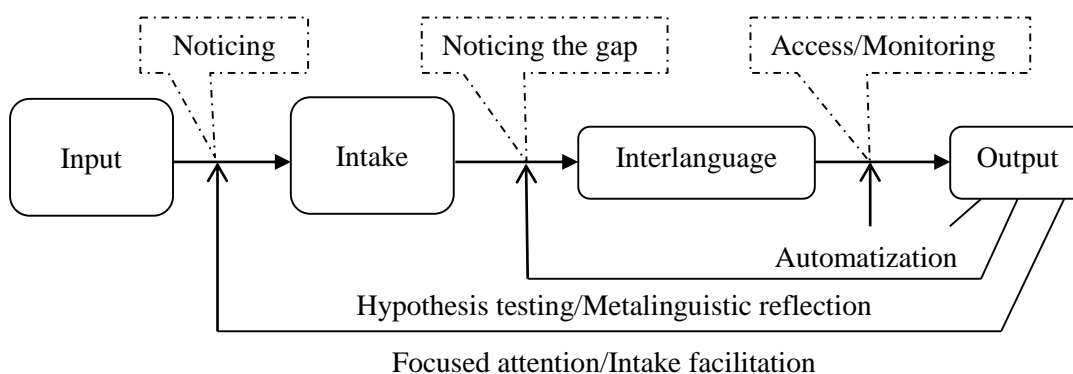


Figure 12. Effects of output in SLA processes (from Izumi, 2003, p.188)

Gilbert et al. (2016) argue that TBLT paradigm offers an optimal context for the sustained and context-embedded type of output practices that theorists consider vital for L2 development. The crucial point here is that the learning possibilities stemming from output practices may work differently in oral and written tasks. The *noticing* function during output production refers to both holes and gaps—in other words, becoming aware of the holes in interlanguage and also “noticing the gap between the interlanguage and the target language” (Muranoi, 2007, p. 57). Such detection is expected to engage learners in the analysis of their existing interlanguage and may promote a more form-focused, syntactic analysis of the incoming input, this input including the whole range of options mentioned in an earlier section. Noticing of linguistic problems can occur in both oral and written tasks, although it is in principle more likely to take place in writing. This is so because of the evanescent nature of oral output, learners may register linguistic inconsistencies only transiently, with the result

that even if noticed, the noticed elements can fade away from the speaker's working memory without any further processing. In contrast, the permanence and self-paced nature of writing may provide more of a facilitative context for both noticing and further processing. As Doughty (2001) put it, the ability to notice the gap and perform cognitive comparisons demand "sufficient and coordinated working and long-term memory resources" (p. 225). As the visibility of the written output provides certain relieving effects for working memory, we could suggest that cognitive comparison is also facilitated during the performance of the written task.

Output can also serve as a way for learners to try out new linguistic structures (i.e., *hypothesis testing*). Learners can test their hypotheses against external feedback or by relying on their own internal devices. The permanent record of written output also facilitates this crucial learning process, as writers can conveniently review and contrast their L2 hypotheses against future input (Schoonen, Snellings, Stevenson, & van Gelderen, 2009). Relief for working memory resources through the visible text and availability of time can also facilitate hypothesis testing during writing tasks, as learners can devote more time and more cognitive resources to this mental operation. As Williams (2012) expressed it, "the cognitive window is open somewhat wider and learners have a richer opportunity to test their hypotheses when they write than when they speak" (p. 328).

Another learning function of output is the *metalinguistic* one. The claim is that "as learners reflect upon their own target language use, their output serves a metalinguistic function, enabling them to control and internalize linguistic knowledge" (Swain, 1995, p. 126). In line with the theory of mind, Swain posited that the use of language produced by other parties or by the learners themselves, mediates L2 learning. The idea is that learners externalize their thinking and thus have an object to reflect

upon that helps them to crystallize their ideas and detect problematic issues in their interlanguage. Writing tasks again appear to function differently as far as metatalk is concerned. In an individual writing task, learners cannot benefit from the metatalk of an external party, and therefore metalinguistic dialogue is restricted to dialoguing with oneself. In any type of writing, the writer's thoughts are externalized into a visible text, which supplies the primary basis for the metalinguistic discussion.

Another function of output proposed by DeBot (1996) is that of developing *automaticity*. Drawing on Levelt's (1989) model of speech production, DeBot argued that output practice enhances fluency and thus contributes to converting declarative knowledge into procedural with a consequent increased control over and consolidation of learners' interlanguage (also suggested in more recent work on the impact of practice on learning—see DeKeyser, 2007). In his theorizing, DeBot seemed to take the oral mode as the default mode in which the function of automaticity is deployed. Gilabert et al. (2016) indicate that it is a matter for future research to compare the way automaticity is fostered by oral versus writing modes in tasks. In terms of empirical evidence, little research has tried systematically to compare oral and written output practices within one experimental design (cf., Ellis & Yuan, 2005; Kormos, 2014). Scarcity of empirical investigation of mode effects on L2 performance and development suggests a broad agenda for future work in this area.

3.4.5 Mode in TBLT research agenda: Conclusions and future research

The major conclusion that Gilabert et al. (2016) draw is that mode as a task-design option has been under-researched (see also Byrnes & Manchón, 2014a,b). Globally

considered, the oral mode has received more attention than the written mode (e.g., in interaction studies) in TBLT theory and research, and it is only recently that the contrast between the two modes has been approached in an emerging line of research on task-modality effects (García Mayo & Azkarai, 2016; Kuiken & Vedder, 2011; Tavakoli, 2014). Another point highlighted by Gilabert et al. (2016) is the tendency to separate modes and treat them individually in TBLT scholarship. This separation is in sharp contrast with what we know about real-life and classroom task performance, in which oral and written modes are intertwined in complex ways (see also Long, 2015). Scarcity of knowledge on the singularity of production and learning opportunities in different modes does not only impoverish task design, but it also precludes understanding of how and why the interaction of modes may contribute to language learning in tasks. Understanding of oral and written modes idiosyncrasy is vital because, as suggested Gilabert et al. (2016), acquisition may proceed in the mingling and interweaving of modes, where, for example, new forms in written or oral input foster acquisitional processes, which are later supported by input-output-feedback cycles, which in turn result in written or oral products.

In their conclusions, Gilabert et al. (2016) call for a TBLT research agenda that is integrative and sensitive to the hybrid nature of discourse and accounts for the idiosyncrasy of performance and learning in different modes. In terms of suggestions for future empirical research, Gilabert et al. (2016) particularly highlight the need to look into the way in which mode may influence the language learning potential of output practices.

3.5 Chapter summary

Comparison of speech and writing reveals similarities in terms of the general architecture of the production processes in the two modes. However, important differences exist between speech and writing in terms of the nature of act of production and output, relationship with the audience and learners' beliefs and associations. High speed and non-permanence of speech versus the slow pace and visibility of writing are suggested as the main features determining singularity of performance and learning in different modes, although other factors (audience, learner beliefs) also play a role. Writing is purported to possess a number of features that may confer learning advantages in L2 written performance. However, scarcity of empirical studies render these claims speculative and more empirical research comparing oral and written task performance is required.

CHAPTER 4. MODE AND TASK COMPLEXITY IN EMPIRICAL TBLT RESEARCH

4.1 Introduction

The objective of this chapter is to present the review and critical analysis of previous empirical studies which are relevant to this dissertation. The studies to be reviewed are grouped into three main categories. We first analyze the empirical investigations which have compared L2 oral and written performance on the same (or equivalent) task within the same experimental design. These empirical studies are relevant to the first aim of this dissertation, which is to explore mode effects on L2 performance. We then look at the empirical work on the effects of task complexity on L2 oral and written performance. We specifically focus on the resource-directing variable of reasoning demands. We review the studies which investigated the effects of reasoning demands on either L2 oral or written production, and we also discuss the few available investigations which explored and compared the effects of reasoning demands on speech and writing within the same experimental design. These empirical studies are relevant to the second aim of this dissertation, which is to explore whether increases in task complexity affect L2 speech and writing in the similar or different ways

In the review of the studies, we focus exclusively on monologic L2 oral and written production. The chapter offers a critical perspective on the reviewed studies in terms of methodological characteristics (participants, task design, procedures, outcome measures) and their results. The main characteristics of the studies are also presented in the tables to facilitate reading of the chapter. In the final part of the analysis, we

specifically focus on such methodological issues as the selection of the outcome measures, highlighting a notorious absence of the assessment of propositional complexity in task-based studies. This leads us to the identification of a gap in the measurement practices in task-based research, which this dissertation will aim to fill.

4.2 Mode effects in L2 performance

One of the goals of this dissertation is to explore the way in which mode (oral versus written) influences L2 performance. Crucial to this dissertation is the review of the empirical studies which compared, within one experimental design, oral and written performance on an L2 task. In what follows, we will provide the details of the relevant empirical studies, followed by the critical analysis of the overall results.

4.2.1 Review of the empirical studies

In his early study, Ellis (1987) compared accuracy of past-tense forms (regular past, irregular past and past copula) in written and oral narrative performance. The participants were 17 L2 English learners from different L1 backgrounds and with early-intermediate level of L2 proficiency (level two of Cambridge Proficiency). The learners performed a picture description task in writing, then they performed the same task orally, and finally the participants performed orally another narrative task, which was based on a different set of pictures. The participants were given one hour for writing; during oral task performance, the learners were given two minutes of planning time. The analysis showed that all three past tense forms were more accurate in the written task than in oral performance.

In a later study, Ellis and Yuan (2005) broadened the scope of measures and compared fluency, complexity and accuracy in oral and written task performance. The main focus of this study was to explore the way different planning opportunities might affect oral and written production, but the authors also compared overall performance in the two modes. The participants were 42 L1 Chinese undergraduate learners of L2 English, aged 18-20, who were at the upper-intermediate level of L2 proficiency. In the experimental part, the participants were required to tell or write a story based on a set of pictures. Although the picture prompts were different for the oral and written tasks, Ellis and Yuan (2005) argued that the tasks in the two modes were equivalent, as they involved the same amount of pictures, had the same amount of elements and depicted similar situations. Furthermore, both tasks were intended to elicit monologic narrative discourse. In the analysis of performance, Ellis and Yuan (2005) measured fluency by calculating the number of syllables per minute and the percentage of disfluencies (i.e., words that were reformulated); ratio of clauses to T-units was calculated for syntactic complexity; syntactic variety was assessed by calculating the total number of different grammatical forms; for lexical complexity, Mean Segmental Type/Token Ratio (MSTTR) (Malvern & Richards, 2002) was computed; for accuracy, the proportion of error-free clauses and correct verb forms was used. The overall comparison of oral and written task performance showed that learners were less fluent but more complex (in grammar and lexis) and more accurate in writing than in speech. These results corroborated Ellis's (1987) earlier findings of higher accuracy of past-tense forms in written production.

In Granfeldt's (2008) study, Swedish university learners of L2 French performed orally and in writing two tasks of different genres. During the first data collection session, the participants had to produce two expository texts. Half of the learners spoke

before writing, and the other half performed the writing task first. During the second session, the same procedure was repeated with two narrative tasks. In line with Ellis and Yuan (2005), Granfeldt found that lexical complexity was higher in writing. However, contrary to Granfeldt's hypothesis, writers were less accurate than speakers and exhibited lower levels of grammatical complexity, as measured by the sub-clause ratio and the ratio of advanced syntactic structures.

Ferrari and Nuzzo (2009) compared grammatical complexity and accuracy in oral and written performance of young learners of L2 Italian from different L1 backgrounds; the authors also compared oral and written production of L1 Italian speakers. The oral task was a retelling task and the written task was a narrative. Performance was assessed by means of general measures of syntactic complexity (length of clause, ratio of dependent clauses per AS-unit or T-unit) and specific measures of syntactic complexity (type of dependent clauses). The authors also assessed the use of cohesive devices. The linguistic behaviors of native speakers and L2 learners were different in the two modes. The production of native speakers was more syntactically complex in writing than in speech, while for learners there was only a tendency towards higher complexity in writing. Like in Granfeldt's (2008) study, L2 learners were more accurate in speech, but the reverse pattern was found for the native speakers. On the basis of these findings, Ferrari and Nuzzo (2009) concluded that the superiority of either the written or the oral mode might be moderated by the level of linguistic competence.

Studies by Yu (2009) and Bulté and Housen (2009) focused on lexical complexity. Bulté and Housen (2009) explored the development of lexical complexity in oral and written production in L2 French by native speakers of Dutch in Belgium. Lexical complexity was assessed by a number of measures, including D-value, Guiraud

Index, Uber and lexical profile measures. The results showed that the development of lexical complexity did not run in parallel in speech and in writing. In line with Ellis and Yuan (2005) and Granfeldt (2008), Bulté and Housen (2009) found that the scores on the written tasks were generally higher than the scores on the oral tasks, for all the measures, except for the lexical profile measures.

A different pattern of findings was obtained by Yu (2009), who compared lexical diversity in written compositions and oral interviews of 25 learners of English from various L1 backgrounds (the majority were L1 Tagalog/Filipino and Chinese speakers) and of different levels of L2 proficiency. Yu employed D-value to assess lexical complexity of oral and written production. The comparison of the production in two modes showed that the lexical diversity of oral and written outputs was of the similar level. On the basis of these findings, Yu suggested that lexical diversity might not be affected by task type (written narratives versus spontaneous interviews) or other task features such as time pressure and pre-task planning. The results obtained by Yu (2009) run counter to the findings in Ellis and Yuan (2005), Granfeldt (2008) and Bulté and Housen (2009), who obtained higher lexical complexity in the written task. A possible explanation to the results obtained by Yu (2009) is that the comparison was made between two different types of tasks: written compositions and spontaneous narratives. Written compositions represented monologic discourse on a number of different topics, while interviews represented interactive unplanned discourse (the exact topic was not specified). Genre and topic of the task could also have an influence on the level of lexical diversity. Thus, it would be reasonable to suggest that the effects of mode on lexical complexity could be confounded with the effects of genre and type of discourse (monologic versus interactive/dialogic).

More evidence about linguistic differences between L2 oral and written production comes from Kormos (2014). In this study, 44 L1 Hungarian learners of L2 English performed orally and in writing a cartoon description task and a picture story task. First, participants performed the two tasks orally. The cartoon description task consisted in describing a comic strip of six pictures which were presented in the correct order and formed a coherent storyline. In the picture narration task, the learners were presented with six unrelated pictures, and they had to invent the content of the story based on these pictures. Although the visual prompts for the writing tasks were different from the prompts in oral performance, Kormos argued that oral and written tasks were equivalent as they involved the same number of pictures, had a similar topic and were designed to elicit narrative discourse. The oral tasks were administered individually, and the participants were given two minutes of planning time for each task. After one month interval, the participants performed the equivalent tasks in writing. The participants had 45 minutes to do the two writing tasks and they were instructed to write a minimum of 150 words in each task. For the analysis of oral and written production, both general and specific measures were used. For lexical complexity, D-value (Malvern & Richards, 1997) was employed to tap into lexical diversity and vocabulary range was measured with the help of Nation's Range program (Heatley, Nation, & Coxhead, 2002). Syntactic complexity was operationalized as the ratio of subordinate clauses to the total number of clauses, length of clauses and the mean number of modifiers per noun phrase, which was calculated with the help of the Coh-Metrix program (MacNamara, Louwse, & Graesser, 2002). The general measure of accuracy used was the ratio of error-free clauses, which was supplemented by a number of specific measures, including the ratio of error-free verbs, past-tense verbs and relative clauses. For cohesion, Coh-Metrix program was employed to calculate the indices of spatial,

temporal, and causal cohesion as well as the frequency of different types of connectives. The analysis showed that, in writing, learners used more varied vocabulary, exhibited higher noun-phrase complexity, and were more accurate than in speech.

Kormos's findings for accuracy are similar to Ellis's (1987), who also found higher accuracy in past-tense forms in writing, but contradict Granfeldt (2008) who found higher accuracy in oral production. Higher lexical diversity in writing found by Kormos was corroborated in a number of previous studies, including Ellis and Yuan (2005), Bulté and Housen (2009) and Granfeldt (2008). In terms of subordination, there were no differences between the modes, which resonates with the results in Granfeldt (2008). Participants, however, used more modifiers per noun phrase in writing, which indicates higher levels of noun-phrase complexity. Concerning cohesion, the participants' writing exhibited higher levels of temporal and spatial cohesion, while the use of connectives displayed a reverse pattern, with participants using more connectives in speech than in writing. On the basis of these results, Kormos (2014) concludes that availability of time and monitoring resources during writing allow learners to achieve higher lexical complexity in writing. Different patterns of findings for subordination and noun-phrase complexity (only the latter was higher in writing) suggest that phrasal elaboration might be specifically revealing about the nature of performance in speech and writing (see also Norris & Ortega, 2009). When interpreting the findings of higher accuracy in writing (indices of accuracy varied depending on the linguistic structure/form), Kormos (2014) suggests that "writing can direct learners' attention to grammatical form, but that the structures that benefit from this focus on form are those in which students already demonstrate high mastery" (p.210). Finally, the results for cohesion are interpreted as evidence that, in speech, L2 learners might rely more on

connectives to create coherent discourse, while, in writing, L2 learners might employ lexical means of cohesion to a greater extent.

4.2.2 Critical analysis of the mode effects studies

Several conclusions can be drawn from the review of the studies which compared L2 oral and written production. First of all, we must point out the relatively small number of studies on the topic (see Table 2).

These studies have certain common characteristics, but they also differ on a number of features. One of the common features is the predominant use of adult L2 learners (a rare exception, Ferrarri & Nuzzo [2009] who involved young learners).

Table 2. Studies which have investigated language performance in the oral and written mode

Dimension	Higher scores in the written mode	Higher scores in the oral mode	No effect of mode
<i>Syntactic complexity</i>	Ellis and Yuan (2005) Kormos (2014) (noun-phrase complexity)		Granfeldt (2008) Kormos (2014) (subordination) Ferrari and Nuzzo (2009)
<i>Lexical complexity</i>	Granfeldt (2008) Bulté and Housen (2009) Kormos (2014) Ellis and Yuan (2005)		Yu (2009)
<i>Accuracy</i>	Kormos (2014) Ellis and Yuan (2005) Ellis (1987)	Granfeldt (2008) Ferrari and Nuzzo (2009)	
<i>Fluency</i>		Ellis and Yuan (2005)	

Also, the bulk of the investigations made within-learner comparisons of performance in speech and writing. Another common aspect is that the majority of the studies employed picture description task (Ellis, 1987; Ellis & Yuan, 2005; Kormos, 2014) or some other kind of a narrative task (Ferrari & Nuzzo, 2009) to elicit performance.

On the other hand, this group of studies shows a great variety in terms of the characteristics of the participants, research focus, methods of data collection and analysis. Thus, the studies exhibit a variety of L1 and L2 languages, cf. L1 Chinese (among others) – L2 English (Ellis, 1987; Ellis & Yuan, 2005; Yu, 2009), L1 Swedish – L2 French (Granfeldt, 2008), L1 Dutch – L2 French (Bulté & Housen, 2009), L1 Hungarian – L2 English (Kormos, 2014), various L2 backgrounds – L2 Italian (Ferrari & Nuzzo, 2009). The majority of studies employed learners of upper-intermediate level of proficiency. However, the level of proficiency was determined in different ways: some studies employed standardized tests (Ellis, 1987; Ellis & Yuan, 2005), while others relied on teachers' ratings (Kormos, 2014). Some studies do not provide enough information about the proficiency of the participants.

Also, the research focus of the studies ranged from a narrow focus on accuracy (Ellis, 1987) or lexical complexity (Bulté & Housen, 2009; Yu, 2009) to a broader scope on accuracy and complexity (Ellis & Yuan, 2005; Granfeldt, 2008; Ferrari & Nuzzo, 2009; Kormos, 2014). The study by Ellis and Yuan (2005) was the only one to compare fluency in speech and writing, and the studies by Kormos (2014) and Ferrari and Nuzzo (2009) have additionally looked at cohesion. There is also a lot of variation in the way the tasks were administered. The majority of studies involved within-learner comparisons of oral and written production (Ellis; 1987; Kormos, 2014), while the study by Ellis and Yuan (2005) involved between-learner comparison of the

performance in the two modes. The studies with the within-learner design also differed between themselves in terms of task administration: some studies were cross-sectional and required learners to perform oral and written task during one experimental session (Ellis, 1987), while in other studies there was a gap between oral and written data collection (e.g., Kormos, 2014; Yu, 2009). It must be highlighted, however, that time-on-task was not limited in the majority of studies. Finally, there is considerable diversity in the choice of performance measures. Some studies employed only one or two measures (e.g., Ellis, 1987; Yu, 2009), other investigations employed a wide range of measures, but focused only on one performance dimension (e.g., Bulté & Housen [2009] examined only lexical complexity). Finally, a number of studies looked at different performance dimensions and employed a wide range of measures (e.g., Ellis & Yuan, 2005; Kormos, 2014), but the choice of measures largely varied from study to study.

These numerous differences between the studies in terms of the participants' characteristics, methods of data collection and measures employed, make it problematic to make a straightforward comparison among the studies, and preclude from drawing unambiguous conclusions concerning performance differences in writing and speech. Nevertheless, certain patterns can be detected. With respect to accuracy, a rather mixed picture emerges, with some studies showing higher accuracy in writing (e.g., Kormos 2014) and others in speech (Granfeldt, 2008). Inconclusive results have also been obtained for syntactic complexity, as some studies found higher values in writing, while others did not find any differences between the two modes. However, certain consistency in findings can be seen for lexical complexity, which tends to be higher in writing. Fluency is predictably higher in oral production. Undoubtedly, more research is

required to determine whether, and if so, how L2 learners' performance varies depending on the mode of production.

4.3 Task complexity effects in L2 oral and written production

This analysis of the available research does not only demonstrate the difficulty in establishing unequivocal mode effects on L2 linguistic behaviors but also emphasizes the relevance of problematizing the applicability to writing of TBLT theoretical tenets originally intended to account for the oral mode. As mentioned previously, several voices (Kormos, 2014; Manchón, 2014a; Tavakoli, 2014) have recently pointed out that current cognitivist TBLT positions are somewhat limited in their ability to account for mode-related singularities and that more research addressing the role of mode in task-based performance and learning is required. In an attempt to advance research in the field, the second aim of this dissertation is to explore L2 performance differences between the two modes, comparing task complexity effects in speech and writing. This dissertation will specifically focus on the factor of reasoning demands, a decision which will be explained in detail in the Method section. The factor of reasoning demands belongs to the group of the resource-directing factors. To date, the available empirical evidence for Robinson's prediction concerning the effects of resource-direction on L2 performance comes mainly from studies of oral production, and the results have been mixed and inconclusive.

Thus, Jackson and Suethanapornkul (2013) meta-analyzed the available studies to investigate the overall effects of raising resource-direction task demands on learner output during monologic oral tasks. This meta-analysis included the studies which manipulated the factors of +/- here-and-now, +/- few elements and +/- reasoning

demands, all of which are comparable as they are claimed in the Cognition Hypothesis to raise complexity and accuracy, but lower fluency, in complex task versions. Findings in Jackson and Suethanapornkul (2013) provide only partial support to the Cognition Hypothesis, as the meta-analysis revealed small positive effects for accuracy, small negative effects for fluency, but no conclusive evidence was obtained for complexity. The authors cautioned that the results of their investigation should be considered preliminary, given the small sample of the analyzed studies (N=9).

Vasylets and Gilabert (2013) synthesized the primary studies which manipulated the factors of +/- here-and-now, +/- few elements and +/- reasoning demands to explore the effects on L2 written performance. The results of the synthesis revealed a cause-and-effect link between increases in resource-direction and quality of L2 written performance. The results showed that, in the complex condition, written texts were longer, more accurate and more syntactically complex as compared to the production in the simple writing task. It is interesting to note that the results of the synthesis of writing studies are more supportive of the Cognition Hypothesis than the meta-analysis of oral investigations, as only in writing a parallel increase in accuracy and complexity of production was reported.

As mentioned earlier, one of the aims of this dissertation is to explore how task complexity effects compare in L2 oral and written production. More specifically, this dissertation will investigate the way increases in the resource-directing dimension of reasoning demands affect speech as compared to writing. For this reason, previous investigations which explored the effects of +/- reasoning demands on L2 oral and written production are of specific relevance. In the following sections, we will first provide a more detailed definition of reasoning demands in TBLT, which will be followed by the review of the available investigations that explored the effects of

reasoning demands on L2 speech and writing separately and within the same experimental design. As in the previous section, the review will be restricted to the studies exploring monologic task performance.

4.3.1 Reasoning demands

In his taxonomy, Robinson (2001, 2011a) distinguishes between spatial, intentional and causal reasoning. Spatial reasoning is associated with a relative ease or difficulty with which speakers must make references to physical location. Studies which have manipulated spatial reasoning typically have employed a “map task” (see Gilabert, 2007), which requires to make reference to the physical location of task components and describe how to move, and in what manner, from one point to another. In Robinson’s view, such tasks can potentially promote awareness of L2 linguistic structures for describing motion events. Intentional reasoning is related to the understanding of the intentional states that motivate others to perform actions. Intentional reasoning can be manipulated, for example, by requiring a learner to interpret the intentions, motives and mental states of a character in a story. Engagement in tasks requiring intentional reasoning is expected to orient learners’ attention to the cognitive state terms for reference to other minds – *she suspected, wonders*, and so on – and, by extension draw attention to the complement constructions – *suspected that, wonders that* – so promoting an attempt to use complex L2 syntax.

The specific focus of this dissertation is on causal reasoning. Low demands on causal reasoning can be conceptualized as straightforward information transmission, requiring, for example, to narrate a story based on the set of pictures presented in the correct order (Kormos, 2014; Tavakoli, 2014). On the other hand, a complex reasoning

task would require understanding of people`s intentions, beliefs and relationships and awareness of the associations and cause-and-effect links between events and phenomena. In the picture narration, for example, high reasoning demands would correspond to a task in which the pictures are presented in the randomized order and the requirement is to establish the correct sequence of the events, creating a coherent story. Another example of a reasoning task is a “Fire-chief” task (Gilbert, 2007) or a “University task” (Vasylets & Gilbert, 2014), which require understanding of the roles and the cause-and-effects links between the elements of the task (the detailed description of these tasks will be provided in the following sections).

On the basis of the above-mentioned definitions and examples, reasoning demands can be broadly defined as the number and intensity of mental operations a task requires from a learner. Accordingly, the higher the intensity of the mental computations that the successful completion of a task requires – the more complex a task is along the dimension of reasoning demands. In Robinson`s taxonomy, the variable of reasoning demands falls into the category of resource-directing dimensions. Accordingly, increases in the reasoning demands of a task is expected to be detrimental to fluency, leading, at the same time, “to greater accuracy and complexity of L2 production when compared to performance on simpler task versions that require little or no reasoning” (Robinson, Mackey, Gass, & Schmidt, 2012, p. 255). In what follows, we will review the empirical studies that have explored the effects of causal reasoning demands on L2 oral and written production.

4.3.2 The effects of reasoning demands on L2 oral production

In a study by Gilabert (2007), 42 Spanish learners of L2 English (intermediate proficiency) performed a decision-making “Fire-chief” task. In this task, learners are presented with a visual prompt which depicts a building on fire and where a number of people have to be rescued. The learners have to decide and justify which characters should be given priority when saving them. In this task, reasoning demands were operationalized as specific roles assigned to the characters and availability of the resources (e.g., fire-trucks, fire extinguishers, etc.). In the simple task, there were plenty of resources, the characters were placed in relatively safe places, etc. In the complex task, the characters had an assigned role (e.g., there was an elderly man, a pregnant woman, etc.), resources were limited and the factors in the task were intricately related, which was thought to require more intensive engagement of reasoning abilities for the successful task resolution. The data were collected during one single session. The participants had one minute of planning time. In the analysis, Gilabert (2007) specifically focused on accuracy, and employed both general measures and measures of self-repairs, which were taken as measures of accuracy since they denote both attention to form and an attempt at being accurate. The results showed that both the simple and the complex tasks generated a similar number of errors and repairs, with two measures (all repairs per 100 words and corrected ratio repaired/unrepaired errors) pointing in the direction of the advanced hypothesis which predicted fewer errors and higher rate and number of self-repairs in the complex task. Gilabert interpreted these findings as suggesting that, as there were no pre-task planning time, learners might have directed their attention towards the justification of their actions, with more resources employed in structural complexity.

This supposition was confirmed in the study by Shia and Adams (2011), who partially replicated Gilbert`s (2007) investigation with 15 learners of L2 English from various L1 backgrounds. Shia and Adams (2011) employed a “Fire-chief” task to elicit L2 oral production which was assessed for accuracy and linguistic complexity. Each participant carried out the simple and the complex version of the task individually, with five minutes of planning time for each task. The analysis revealed higher lexical and syntactic variety in the complex task. As in Gilbert (2007), there were no substantial differences between simple and complex task in terms of accuracy, thus providing partial support to the Cognition Hypothesis.

Another oral study in which task complexity was manipulated along the reasoning demands was carried out by Choong (2011). In the study, 22 Japanese learners of English, of advanced L2 proficiency, were required to perform orally a picture description task under four different conditions. The conditions were defined as [+/- contextual support], which was operationalized as whether or not the participants could see the picture set as they narrated, and [+/- reasoning demands], which was operationalized as whether or not the picture series were in sequence. Oral production was analyzed for propositional and syntactic complexity. In his operationalization of idea units, Choong followed Carrell (1985), who provided the following definition of an idea unit:

Each idea unit consisted of a single clause (main or subordinate, including adverbial and relative clauses). Each infinitival construction, gerundive, nominalized verb phrase, or conjunct was also identified as a separate idea unit. In addition, optional and/or heavy prepositional phrases were also designated as separate idea units. (p. 737)

Choong (2011) also provides an example of a sentence divided into idea units from Carrell: “serious nuclear accidents...have lead/ in the past/ to the implementation of

strict safety rules/ regarding the construction/ and operation of nuclear power stations” (p. 752). The results showed that the impact of the contextual support variable on production was negligible. Reasoning demands did not have a statistically significant effect on syntactic complexity, but it affected propositional complexity, prompting learners to provide more content (i.e., more idea units) in the complex condition.

Similar operationalization of reasoning demands can be found in a recent study by Fukuta and Yamashita (2015) who explored the way two types of cognitive demands (reasoning demands and dual-task demands) influenced L2 performance and orientation of attention during production. The participants in the study were 36 under- and post-graduate L1 Japanese learners of English, with upper-intermediate level of English (B2) proficiency according to the Common European Framework of Reference (CEFR). The participants performed three oral monologic tasks that required to describe stories using the pictures. There were three conditions: (1) control condition (i.e., simple task), in which the four-frame cartoons formed a coherent story; (2) reasoning demands condition, in which the sequence of the pictures was randomized and the participants had to decide on the correct sequence to create a story; (3) dual-task condition, in which learners were engaged in parallel finger tapping while narrating the story. The control condition was considered to be a simple task, while reasoning demands and dual-condition were considered to be complex tasks. The participants were provided with one minute of planning time for each task. Immediately after performing the oral tasks, the learners were engaged in stimulated recall, conducted in L1, during which the participants watched their recorded production and commented on what they were thinking at that moment. The oral performance was analyzed for syntactic complexity, accuracy and fluency (see Table 3). The protocols of stimulated recall were coded for *Content*, *Linguistic form*, or *Others*. The episodes related to form were coded into three

sub-episodes: (1) syntactic encoding, which related to word order, sentence structure, etc.; (2) lexical encoding, which encompassed attention to lexis; (3) phonological encoding, which involved attention to phonological features. The analysis of the performance revealed lower fluency in the complex tasks and absence of changes in the complexity of production. Accuracy was highest in the reasoning demands task and lowest in the dual-condition task. The protocols analysis revealed higher frequency of the syntactic encoding episodes in the simple task, while there were no differences between the tasks in terms of the lexical encoding episodes; the number of the phonological encoding episodes was negligible in all three conditions. Fukuta and Yamashita (2015) concluded that increases in reasoning demands of a task might boost accuracy but decrease fluency of L2 production. Another conclusion was that increases in task complexity might be taxing for cognitive resources, as a consequence, attention to syntactic encoding can be inhibited; at the same time, lexical encoding is not necessarily affected in demanding tasks.

4.3.3 The effects of reasoning demands on L2 written production

The study by Kuiken and Vedder (2008) involved 91 Dutch learners of L2 Italian and 76 Dutch learners of L2 French. The participants were first-, second- and third-year university students, which ensured the involvement of the learners with different levels of L2 proficiency. The participants completed in writing two tasks in which cognitive complexity was manipulated. In both conditions, the learners were presented with a prompt in L1 (Dutch), which explained they had to write a letter to a friend recommending a holiday destination out of five options. In the simple version of the task, there were three requirements to take into account, while in the complex task there were six requirements. It was expected that the necessity to take into account a higher number of requisites would require a more intensive involvement of reasoning abilities.

The tasks were completed during the writing classes and there was a time limit of 40 minutes per each task. Written compositions were assessed for accuracy, syntactic complexity, and lexical variation (see Table 3). In addition to the general measures of accuracy, Kuiken and Vedder (2008) used a specific measure dividing errors into first-degree (i.e., minor) errors, second-degree (more serious errors) and third-degree errors (which made the text nearly incomprehensible). The results were very similar for the students of Italian and French: the learners made less mistakes in the complex task, while there were no significant differences in syntactic and lexical complexity. Another finding was an absence of interaction between task complexity effects and L2 proficiency level. As the Cognition Hypothesis (Robinson, 2001) predicts a parallel increase in accuracy and complexity in complex reasoning task, Kuiken and Vedder's (2008) findings provided only a partial support to Robinson's theorizing.

Rather different pattern of results was obtained in the study by Salimi and Dadashpour (2012) in which the "Fire-chief" task from Gilabert (2007) was performed in writing by 29 Turkish learners of L2 English. During the first session of data collection, the participants were given 45 minutes to perform the simple version of the task. After two weeks, the learners performed the complex version of the task. Written compositions were analyzed for fluency, syntactic complexity and accuracy. In contrast to Kuiken and Vedder (2008) who found higher accuracy in the complex task, Salimi and Dadashpour (2012) found that task complexity did not affect accuracy of production, but the learners were more fluent and produced more syntactically complex output in the complex task. It must be taken into account, that the authors assessed fluency by calculating the length of T-unit, which can also be regarded as a measure of syntactic complexity (see Norris & Ortega, 2009).

In the study by Vasylets and Gilabert (2014), 32 Russian/Ukrainian and 19 Spanish/Catalan learners of L2 English performed in writing the simple and the complex versions of the “University task”, which asked participants to write a recommendation of a university taking into account a number of requisites. In the simple task, the participants had to consider three simple requirements, which were not contradictory and were rather easy to accommodate. By contrast, in the complex task, the three requisites were more contradictory, assuming that in this way the learners had to employ their reasoning abilities more intensively in order to justify their choice in a convincing manner. There were two sessions of data collection with a week`s interval in-between. Half of the participants performed the simple task during the first session and completed the complex version during the second session. Accordingly, the other half performed the tasks in the reversed order. During each session, the subjects had 40 minutes to complete the writing task. Written compositions were assessed for accuracy, complexity and cohesion (see Table 3). The results showed that, in the complex task, the production was more syntactically complex and lexically sophisticated, and there were more task-relevant connectives. In comparison to Salimi and Dadashpour (2012), the study by Vasylets and Gilabert (2014) employed a wider range of production measures. However, the results of both studies concur in that increases in reasoning demands did not affect accuracy of production, but it boosted syntactic complexity.

Ruiz-Funes (2015) explored the influence of task complexity on L2 written output of L1 English undergraduate L2 Spanish learners of advanced and intermediate L2 proficiency. Advanced learners (N=8) and intermediate learners (N=24) were assigned two tasks of different levels of cognitive complexity in relation to familiarity of topic, genre, and reasoning demands. Advanced learners completed an analytical essay, which required to compare history or traditions in the students` home culture

with those in a Spanish-speaking country, and an argumentative essay, which required to read a text about illegal immigration in Spain and argue, in writing, in favor or against the main thesis of the text by providing supporting arguments. An analytical essay was considered to be a simple task with lower reasoning demands as it required students to rely on their own personal experience. An argumentative essay was regarded to be a complex task since it required students to develop their own arguments, which demanded a high level of reasoning processing. Intermediate learners wrote a personal essay about themselves and their interests, which was a simple task. The complex task for this group of participants was an expository essay about the benefits and challenges of studying abroad. Both groups completed the tasks in two class sittings (one task on each day). Advanced learners had 90 minutes per task; intermediate learners were allowed 50 minutes per task. Written compositions were assessed for syntactic complexity, accuracy and fluency (see Table 3). Due to the small number of participants, only descriptive statistics was used for data analysis. The descriptive statistics for CAF measures showed an identical pattern for advanced and intermediate learners, who showed lower accuracy and coordination in the complex task. At the same time, there were higher values for syntactic complexity in the complex condition, which concurs with the findings in Vasylets and Gilabert (2014) and Salimi and Dadashpour (2012). Ruiz-Funes (2015) explained the obtained results in light of the language learning potential of writing theorizing, concluding that the potential of writing to learn a language is related to the degree of cognitive complexity, as learners exploited the facilities of written mode in different ways, with the simple task allowing to pay more attention to form (hence, higher accuracy of production) and a complex task boosting syntactic complexity.

Table 3. Empirical studies that have explored the effects of reasoning demands on L2 performance

Study	Participants				Task	Measures	Findings
	N	Age	L1	L2			
Gilabert (2007)	42 speakers	18-40	Spanish/ Catalan	English	argumentative task based on the visual prompt and written instructions	accuracy: (1) number of errors per AS-unit; (2) ratio of errors to words; (3) number of error repairs per AS-unit; (4) ratio of error-repairs to words; (5) total number of repairs; (6) ratio of error-and non-error repairs to words; (7) percentage of self-repairs; (8) ratio of repaired to unrepaired errors; (9) corrected ratio of repairs	no significant effects of task complexity
Shiau & Adams (2011)	15 speakers	20-93	Mandarin Korean Japanese	English	see Gilabert (2007)	lexical complexity: Guiraud Index syntactic complexity: (1) clauses per AS-unit; (2) length of clause; (3) number of different grammatical verb forms accuracy: (1) errors per AS-unit; (2) ratio of total errors to total words; (3) self-repairs per AS-unit; (4) ratio of self-repairs to total words	higher lexical and syntactic variety in the complex task

Table 3 continued.

Study	Participants				Task	Measures	Findings
	N	Age	L1	L2			
Choong (2011)	22 speakers	mean 38	Japanese	English advanced	narrative task based on picture description	propositional complexity: (1) number of idea units syntactic complexity: (1) length of T-unit	more idea units in the complex task
Fukuta & Yamashita (2015)	40 speakers	18-60	Japanese	English B2 level	narrative task based on picture description	syntactic complexity: (1) clauses per AS-unit; (2) length of AS-unit accuracy: (1) % of error-free AS-units fluency: (1) pruned words per minute	lower fluency in complex tasks, the highest accuracy in the reasoning demands task
Kuiken & Vedder (2008)	91 L2 Italian writers & 76 L2 French writers	not indicated	Dutch	Italian (low & high prof.) French (low & high- prof)	argumentative task based on the written prompt	lexical complexity: (1) TTR1; (2) alternative type-token ratio syntactic complexity: (1) clauses per T-unit; (2) dependent clause ratio accuracy: (1) errors per T-unit; (2) 3 degrees of error seriousness	higher accuracy in the complex task

Table 3 continued.

Salimi & Dagashpour (2012)	29 writers	not indicated	Turkish	English not indicated	see Gilabert (2007)	fluency: (1) length of T-unit syntactic complexity: (1) S-nodes per T-unit accuracy: (1) error-free T-units per T-unit	higher syntactic complexity and fluency in the complex task
Vasylets & Gilabert (2014)	51 writers	18	Spanish/ Catalan; Russian/ Ukrainian	English upper-interm.	argumentative task based on the written prompt	lexical complexity: (1) D-value; (2) Lexical Frequency Profile syntactic complexity: (1) length of T-unit; (2) ratio of subordinate clauses to total number of clauses accuracy: (1) errors per 100 words; (2) errors per T-unit cohesion: (1) indexes of causal, temporal, spatial cohesion (2) incidence of connectives	higher syntactic complexity and lexical sophistication in the complex task
Ruiz-Funes(2015)	8 writers (Study 1) 24 writers (Study 2)	21-25 (Study 1) 19-25 (Study 2)	English	Spanish advanced (Study 1) interm. (Study 2)	writing tasks of different genres: analytical essay and argumentative essay (Study 1) personal essay and expository essay (Study 2)	syntactic complexity: (1) length of T-unit; (2) T-units per sentence; (3) dependent clauses per T-unit accuracy: (1) errors per 100 words; (2) errors per T-unit; (3) total number of errors fluency: (1) words per minute	higher syntactic complexity and lower accuracy and fluency in the complex task

4.3.4 Critical analysis of the studies which explored the effects of reasoning demands on L2 production

The review of the studies comparing the effects of reasoning demands on L2 oral and written performance allows for a number of conclusions. In terms of the participants' characteristics, both oral and written studies involved adult learners from a variety of L1 backgrounds (e.g., Spanish, Russian, Japanese, Turkish, etc.) and of predominantly intermediate or advanced level of L2 proficiency. Oral studies have exclusively looked at L2 English, while the studies in writing involved other L2 besides English, such as L2 Italian and French (Kuiken & Vedder, 2008) or L2 Spanish (Ruiz-Funez, 2015). L2 proficiency in the studies was assessed in different ways, and in some cases there are not sufficient details on the learners' level of proficiency or proficiency tests employed.

The construct of reasoning demands was operationalized in different ways in both modes, including picture sequencing (oral production: Choong, 2011; writing: Fukuta & Yamashita, 2015), necessity to accommodate a number of requisites (Kuiken & Vedder, 2008; Vasylets & Gilabert, 2014) or as related to the genre of the task (Ruiz-Funez, 2015). Gilabert's (2007) "Fire-chief" task, which operationalized reasoning demands as the roles assigned to the characters and availability of resources, was replicated in both oral production (Shiau & Adams, 2011) and in writing (Salimi & Dagashpour, 2012). All oral studies were similar in terms of the procedure employed, as the data were collected on the individual basis and during one single session; there were some variations in terms of planning time: for example, up to one minute in Gilabert (2007), and five minutes for planning in Shiau and Adams (2011). There was not much variation in the procedure in the writing studies, as the learners completed the writing

tasks in two sittings, there was a gap of at least several days between the two tasks, and the learners had minimum 40 minutes to complete each task. Perhaps the biggest variation can be found in the selection of the measures, which differed from one study to another. The majority of studies, however, assessed both accuracy and complexity of production (Fukuta & Yamashita, 2015; Vasylets & Gilabert, 2013), but there were also studies with a more narrow focus, such as Gilabert (2007) who explored only accuracy or Choong (2011) who exclusively focused on complexity. It must be pointed out that the study by Choong (2011) is the only study which explored propositional complexity (operationalized as idea units) in task performance.

Concerning the effects of reasoning demands on L2 oral and written production, overall results in both modes are of mixed nature. According to the predictions of the Cognition Hypothesis, enhanced reasoning demands should produce a parallel increase in accuracy and complexity of production, while negatively affecting fluency. None of the oral or writing studies have obtained a simultaneous improvement in complexity and accuracy of production. Overall, the findings were rather mixed in both modes, which precludes from drawing definitive conclusions about the effects of reasoning demands on L2 oral and written performance. Thus, in the oral mode, some studies have reported higher values for accuracy (Fukuta & Yamashita, 2015), lexical and structural variety (Shiau & Adams, 2011), or propositional complexity (Choong, 2011) in complex reasoning tasks. However, in other studies, no task complexity effects were found for accuracy (Gilabert, 2007; Shiau & Adams, 2011) or linguistic complexity (Choong, 2011; Fukuta & Yamashita, 2015). For writing, the expected higher accuracy in complex reasoning tasks was reported in Kuiken and Vedder (2008). Higher linguistic complexity in complex writing tasks was obtained in Salimi and Dadashpour (2012) and Vasylets and Gilabert (2014), but no changes in this dimension were found by Kuiken

and Vedder (2008). Differences in the participants' characteristics, operationalization of reasoning demands or the selection of measures can explain the variability in the obtained findings. Inconclusiveness of the results in both modes also precludes from comparing the nature of the effects of reasoning demands on L2 performance in different modes. To make this comparative analysis, in the next section we will review the available investigations which have compared task complexity effects in writing and speech within one experimental design.

4.4 Studies which compared task complexity effects in L2 oral and written production within the same experimental design

As mentioned in the previous sections, crucially relevant for this dissertation are those few investigations that have compared task complexity effects on L2 oral versus written production within the same study (see Table 4). In all these studies, task complexity was operationalized, albeit in different ways, in terms of the reasoning demands.

4.4.1 *Review of the empirical studies*

To our knowledge, Kuiken and Vedder (2011) were the first to explore the effects of task complexity on L2 performance in relation to mode of production (oral versus written). The theoretical basis of this study was the Cognition Hypothesis (Robinson, 2011, 2011a). Additionally, the study investigated the role of L2 proficiency in relation to mode. In other words, the authors explored whether there was a difference between low- and high-proficiency learners with respect to the influence of task complexity in written versus oral tasks. The study adopted a between-learner design and involved two groups of participants. The participants in the oral mode were 44 learners

of L2 Italian, with Dutch as the native language. Their performance was compared with the written production of 91 Italian learners with L1 Dutch. The proficiency level was determined by means of a cloze-test and the learners were subsequently divided into low-proficiency and high-proficiency groups. Kuiken and Vedder (2011) employed an argumentative task which consisted in producing an oral or written advice to a friend concerning a holiday destination. Each writer and speaker performed the simple and the complex version of the task. In the simple task, a learner had to take into account three requirements when choosing a destination; the complex task contained six requirements. The learners had to provide arguments for their choice. There were no time limits for oral production and the majority of the learners completed the task in 3-5 minutes. Writers were given 40 minutes to complete the task.

Learners' written and oral production was coded in terms of accuracy and syntactic and lexical complexity (see Table 4). On the basis of the findings in their previous study (Kuiken & Vedder, 2007), in which task complexity specifically affected accuracy of L2 performance, the authors focused in detail on the dimension of accuracy in this study and employed both general and specific measures in its assessment. In addition to calculating the total number of errors per T-unit (for writing) or per AS-unit (for speech), Kuiken and Vedder (2011) made a division between three degrees of errors: first-degree errors included minor errors that did not interfere with the comprehensibility; second-degree errors contained more serious deviations; third-degree errors made the text incomprehensible. Additionally, in the written mode the total number of errors per T-unit was calculated with respect to Appropriateness (i.e., pragmatic errors), Grammar and Lexicon. In the oral mode, the total number of grammatical and lexical errors per AS-unit was counted. The authors also counted the number of orthographic and pronunciation errors.

The results showed that task complexity effects were largely the same in written and oral performance. Thus, in both writing and speech learners made fewer errors in the complex task as compared to the simple task. The results also showed that, in both oral and written production, the overall increase in accuracy was mainly attributed to the decrease in Lexical errors. Concerning lexical variation, task complexity did not affect this dimension neither in writing nor in speech. However, the effects of task complexity played out somewhat differently in writing and speech for syntactic complexity: while this dimension was not affected in writing, speakers produced fewer dependent clauses in the complex task. For both modes, there were no interaction between task complexity and proficiency level. This finding allowed to conclude that the influence of task complexity on L2 performance does not seem to differ according to the level of L2 proficiency, neither in writing nor in speech. In spite of the differential impact of task complexity on syntactic complexity in the two modes, Kuiken and Vedder (2011) surprisingly conclude that “the influence of task complexity on linguistic performance is not substantially constrained by mode. This mean, that, on the basis of our findings, it does not seem necessary to include “mode” on Robinson`s Triadic Componential Framework” (p.103).

A very different opinion in this respect was expressed by Kormos and Trebits (2012) and also Tavakoli (2014) who reported different task complexity effects on L2 production in writing and in speech. In their study, Kormos and Trebits (2012) explored, among other issues, if task demands affected similarly or in a different way oral and written performance. For this purpose, Kormos and Trebits employed two narrative tasks that differed with regard to whether the coherent plot of the story was provided or not. Inspired by the Cognition Hypothesis (Robinson, 2001, 2011a), the authors defined the provision of more or less plot as a task feature that can pose separate

demands on conceptualization and linguistic formulation. The first task was a cartoon description task which involved a description of a comic strip consisting of six pictures, which were presented in the correct order and formed a coherent story line. This task was considered to pose low cognitive load on conceptualization, as it provided a ready-made plot; at the same time, the task was considered to be potentially taxing in terms of the demands on formulation, as it required the use of potentially complicated linguistic features. The cartoon description task had two versions. These versions were devised to prevent repetition of the same task in oral and written production. Kormos and Trebits (2012) argue that the two versions were equivalent in terms of the cognitive load they posed, as they involved the same amount of actors, key narrative events and both had an element of surprise. The second task was a picture narration task. This task required to tell a story based on six unrelated pictures. The learners had to create a plot of the story based on the pictures and they had to include all the pictures in the narration. This task was considered to posit high demands on conceptualization, as it required to create a coherent story; at the same time, the task was considered to place low demands on linguistic encoding, as the learners could adapt the plot to their linguistic choices, and, thus, avoid using language which was difficult for them. Again, two versions of the same task were created for oral and written performance, which were considered to be identical in all the aspects, except for the visual cues.

The tasks were administered to 44 Hungarian learners of English, who were students in a Hungarian-English bilingual secondary school. The participants' age was between 16 and 17 years and their level of proficiency corresponded to B1/B2 in the Common European Framework of Reference (Council of Europe, 2001). During the first session of data collection, the participants performed the cartoon and the picture description task orally. The oral tasks were administered to the students individually;

the learners were given 2 minutes to prepare for each task. The order in which the learners performed the oral tasks was randomized. In one month, the learners completed the cartoon description and the picture narration tasks in writing during a regular English class. The learners were given 30 minutes to complete both tasks, and they were required to write a minimum of 150 words for each task. The obtained oral and written production was analyzed for accuracy and complexity (see Table 4); oral production was also rated for fluency (speech rate was calculated as the mean number of syllables per minute). The comparison across modes showed an interaction between task type (cartoon vs. picture narration) and mode (oral vs. written) for some dimensions of lexical and syntactic complexity, as well as for accuracy. Thus, in oral production, the cartoon description task elicited more varied language than the picture narration; also, in speech, learners were more accurate in the picture narration, as measured by the ratio of error-free verbs. At the same time, there were no significant differences between the cartoon and the picture description task in terms of lexical variety and accuracy in writing. The pattern of findings was reversed for syntactic complexity. While there were no changes in this dimension in oral performance, the results indicated that in written performance, learners produced longer clauses and more relative clauses in the picture narration task as compared to the cartoon description task.

These findings can be interpreted as pointing to a potential interaction between task complexity and mode of production: written production appeared to be more susceptible to the impact of cognitive task demands along the dimension of syntactic complexity, while speech was more sensitive in the areas of accuracy and lexical variety. On the basis of these results, Kormos and Trebits (2012) conclude that “tasks with different cognitive and linguistic demands seem to elicit different patterns of

performance in writing than in speech” (p.25), suggesting that mode has to be taken into consideration when interpreting the effects of task complexity on L2 production.

Similar conclusions were drawn by Tavakoli (2014), who also compared the effects of cognitive task complexity in L2 writing and speech. Tavakoli adopted Skehan`s Trade-off Hypothesis as a theoretical basis for her study and operationalized cognitive task complexity as storyline complexity. In this operationalization, a distinction is made between *foreground* and *background* storyline complexity. A narrative task with foreground storyline complexity has only one storyline, while a task with background complexity has more than one storylines, which have to be interweaved to create coherent discourse. It follows that a narrative task with background events is more complex than a narrative with foreground events. In her study, Tavakoli (2014) employed a narrative picture description task with foreground events (simple task) and background events (complex task) to elicit L2 written performance. The participants in the study were 40 intermediate-level (B2 according to CERF) learners of L2 English. The participants came from a range of L1 backgrounds and they were aged between 19 and 35. During data collection, the participants were given the simple and the complex task in succession (presentation was counterbalanced). For each set of pictures, the learners had to write a story as if they were relating the story to someone who had not seen the pictures; the learners had 20 minutes per each task.

The obtained written texts were analyzed for syntactic complexity. Tavakoli employed two measures: length of T-unit and the ratio of clauses to T-units as a measure of subordination. The results of the analysis of the written texts were compared to the findings in Tavakoli and Foster (2008), and Foster and Tavakoli (2009) who employed the same tasks to elicit oral performance from L2 learners and from L1

English speakers respectively. The comparison showed that, regardless of the mode, a more complex task with background events encouraged more subordination and longer grammatical units. However, the effects of storyline complexity on syntactic complexity were stronger in speech than in writing. As the results indicated, the differences in syntactic complexity between simple and complex version of the task were statistically significant in oral production, but the results did not reach a statistically significant level in writing, although the descriptive values were higher in the complex writing task. These results suggest that the impact of the storyline complexity seems to be in the same direction in the two modes, but the strength of the impact may vary according to the mode of production. Tavakoli tentatively attributes the lack of significant differences between simple and complex writing tasks to the nature of written production and the learners' associations and beliefs about written production. The self-paced nature of writing allows for extensive on-line planning, careful linguistic encoding and monitoring, which can give rise to syntactically complex language irrespective of the demands of the task. Another interpretation that Tavakoli offers is that in the written environment learners may automatically take a "writerly" stance and construct syntactically elaborate discourse by default.

In sum, the study by Tavakoli (2014) found that the effects of task complexity played out somewhat differently in speech and writing. On the basis of the results obtained, Tavakoli (2014) concluded that task design may have a differential impact on L2 performance across the two modes, which may imply that "task complexity does not operate in isolation" (p.232) but rather interacts with the mode in which the task is performed.

4.4.2 Critical analysis of the studies which compared task complexity effects in the two modes

In sum, the overall findings of the studies, which have compared tasks complexity effects in writing and speech, indicate that task complexity seems to have a differential impact on oral and written production (see Table 4). We must also point out that the task complexity and mode interaction revealed itself in a rather different manner in the three studies. Thus, in Kuiken and Vedder (2011) differences in task complexity effects in speech and writing were found for syntactic complexity, with speakers showing lower values of subordination and writers showing the same levels of syntactic complexity in simple and complex tasks. Task complexity and mode interaction along the dimension of syntactic complexity was also found by Kormos and Trebits (2012) and Tavakoli (2014). Similar to Kuiken and Vedder (2011), Tavakoli (2014) found that syntactic complexity of oral performance was more susceptible to task demands than writing. In contrast to Kuiken and Vedder (2011), who found lower syntactic complexity in the complex oral task, Tavakoli (2014) obtained a reversed pattern of higher complexity in the complex oral task. Kormos and Trebits (2012), however, found that syntactic complexity of written production was more sensitive to task demands: in their study, syntactic complexity increased in the complex writing task, but there were no changes in oral production.

Concerning other performance dimensions, the study by Kormos and Trebits (2012) revealed task complexity–mode interaction also in the areas of accuracy and lexical variety, with speech showing more susceptibility to task demands and writing exhibiting similar values across task types. In contrast, Kuiken and Vedder (2011) found that task complexity affected accuracy and lexical variety in the similar way in

speech and writing, as accuracy increased in the complex task in both modes and there were no changes in lexical variety neither in written nor in oral production. As Tavakoli (2014) focused exclusively on syntactic complexity, we do not have evidence for other performance dimensions.

Variations in the pattern of the results can be attributed to the numerous differences between the studies, which differ in terms of the operationalization of task complexity, characteristics of the participants, tasks, procedures, and also measures employed. However, overall results seem to be indicative of an interaction between task complexity and mode of performance. Provided the small number of previous investigations and the above-mentioned variations in the obtained findings, the conclusion about mode and task complexity interaction is highly tentative. Undoubtedly, more research is required to elucidate this issue.

Table 4. Empirical studies that have compared task complexity effects in oral and written production

Study	Participants				Task	Measures	Findings
	N	Age	L1	L2			
Kuiken & Vedder (2011)	44 speakers 91 writers	not provided	Dutch	Italian (low & high proficiency)	argumentative task based on the written prompt	<p>lexical variety: (1) alternative type-token ratio</p> <p>syntactic complexity: (1) clauses per T-unit or AS-unit; (2) dependent clause ratio</p> <p>accuracy: (1) total number of errors per T-unit or AS-unit; (2) division into 1st,2nd, 3rd degree errors; (3) division into error types, such as Appropriateness, Grammar, Lexicon, Orthographic, Pronunciation, Other</p>	<p>largely the same task complexity effects in the two modes: (1) both writing & speech: effects on accuracy (higher in the complex task), but no effects on lexical complexity; (2) effects on syntactic complexity (lower in the complex task) only in speech</p>
Kormos & Trebits (2012)	44	16-17	Hungarian	English B1/B2 (CERF)	cartoon description task: a comic strip consisting of 6 pictures presented in the correct order (simple task) or scrambled (complex task)	<p>lexical variety: D-value</p> <p>syntactic complexity: (1) clause length; (2) ratio of subordinate clauses; (3) ratio of relative clauses</p> <p>accuracy: (1) ratio of error-free clauses; (2) ratio of error-free relative clauses; (3) ratio of error-free verbs; (4) ratio of error-free past-tense verbs</p>	<p>task complexity affected writing & speech differently: (1) speech: effects on accuracy & lexical variety; (2) writing: effects on syntactic complexity</p>

Table 4 continued.

Study	Participants				Task	Measures	Findings
	N	Age	L1	L2			
Tavakoli (2014)	40 writers	19-35	varied	English B2 (CERF)	narrative task based on the description of pictures with foreground events (simple task) & background events (complex task)	syntactic complexity: (1) mean length of T-unit; (2) ratio of clauses to T-unit	task complexity affected speech & writing differently: effects on syntactic complexity (higher in the complex task) were greater in speech
Tavakoli & Foster (2008) (employed in Tavakoli [2014] for comparison)	60 speakers	19-45	Farsi	English intermed. proficiency	intermediate (according to the locally administered placement test of English proficiency)	see Tavakoli (2014)	
Tavakoli & Foster (2008) (employed in Tavakoli [2014] for comparison)	40 speakers	18-60	English			see Tavakoli (2014)	

4.5 Chapter summary

This chapter has reviewed and critically analyzed the empirical investigations which are relevant to the aims of this dissertation. We first reviewed the studies that have compared L2 oral and written performance on the same (or equivalent) task. The findings of these studies are mixed, which can be attributed to the differences in terms of the participants' characteristics, methods of data collection and measures employed. However, there seems to be a trend towards higher complexity (specifically, lexical complexity) in the written mode.

The studies looking into the effect of reasoning demands on L2 speech and writing have also produced mixed findings, which we again can attribute to methodological differences among the studies. The global conclusion that can be drawn is that none of the studies reviewed (either oral or written) have found a parallel increase in the accuracy and complexity of the output in the complex task. This means that none of the studies on reasoning demands have provided evidence to the main tenet of the Cognition Hypothesis concerning the effects of resource-direction on L2 monologic production.

Concerning the nature of the effects of task complexity in different modes, the few available investigations seem to indicate that task complexity has a differential impact on L2 speech as compared to writing. However, the scarcity of research (there are only three empirical studies available) and the fact the task complexity and mode interaction revealed itself in a rather different manner in each study, precludes from drawing definitive conclusions in this regard. In sum, this review of the relevant empirical literature has clearly shown a need for further enquiry into the mode and task complexity effects on L2 performance.

Another conclusion that can be drawn from the preceding literature review concerns methodological practices. We observed great variability in the outcome measures and, very importantly, none of the available studies (with the exception of Choong [2011]) have analyzed propositional complexity, which is an important, albeit neglected, dimension in L2 performance. Propositional complexity refers to the amount of information or idea units a speaker/writer encodes to convey content in a task (Bulté & Housen, 2012). Crucially, L1 researchers consider idea units to be a feature par excellence for discriminating between speech and writing (Chafe, 1982, 1994). In the L2 literature, the analysis of production for idea units has been presented as providing the degree to which learners engage in conceptualization during the performance of the task (Ellis & Barkhuizen, 2005). Given that this dissertation is concerned with mode and task complexity, and that task complexity is hypothesized to affect primarily the conceptualization stage of language production (Robinson, 2011a), we consider that the analysis of production for idea units is pertinent and necessary because it could potentially offer additional insights on the idiosyncrasy of the performance and task complexity effects on different modes. The next chapter will discuss the intended methodological contribution of this dissertation, which is the analysis of L2 performance for idea units.

CHAPTER 5. PROPOSITIONAL COMPLEXITY: DEFINITION AND OPERATIONALIZATION IN L1 AND L2 RESEARCH

5.1 Introduction

The review of literature in the previous chapter has made patent the scarcity of L2 investigations which include measures of propositional complexity among the outcome measures of L2 performance. The measurement of propositional complexity (operationalized as idea units) is ultimately relevant for this dissertation, which aims to explore the way mode (oral versus written) mediates L2 performance and task complexity effects. There is a lack, however, of the reliable and valid guidelines for the segmentation of discourse into idea units. The primary aim of this chapter is to provide the definition and operationalization of idea units, which will lay the basis for our guidelines for the segmentation of discourse into idea units.

The chapter is structured as follows: we first provide the general definition of L2 complexity and define the theoretical stance on L2 complexity adopted in this dissertation. We then briefly define our view on linguistic complexity, which is another dimension of L2 complexity that we deal with in this dissertation. This will be followed by the sections in which we provide the definition of propositional complexity, review early theorizings of idea units in L1 English, and summarize the available empirical findings. This will lay the basis for our guidelines for the segmentation of discourse into idea units, which constitute the specific methodological contribution of this dissertation.

5.2. L2 complexity

Complexity in L2 scholarship is viewed as a multi-componential construct, consisting of a number of dimensions which can be independently evaluated (Bulté & Housen, 2012, 2014; Housen & Kuiken, 2009; Norris & Ortega, 2009; Pallotti, 2009, 2015). A descriptive-analytic framework of complexity proposed by Bulté and Housen (2012) reflects the multidimensionality of complexity as it has been interpreted in L2 research (see Figure 13). The first distinction is made between *relative* and *absolute* complexity. In Bulté and Housen's (2012) view, relative complexity, or *difficulty*, implies "mental ease or difficulty with which linguistic items are learned, processed or verbalized" (p.23). The *absolute* complexity implies inherent properties of linguistic units and systems. Absolute complexity is further divided into *linguistic* complexity, *discourse-interactive* complexity and *propositional* complexity.

Given that this dissertation deals with monologic L2 oral and written performance, of specific relevance to us is linguistic and propositional complexity of discourse. According to Ortega (2012b), L2 researchers employ complexity measures "with at least three main purposes in mind: (1) to gauge proficiency, (2) to describe performance, (3) to benchmark development" (p.128). In this dissertation, we will measure complexity, in addition to accuracy and fluency, in order to describe L2 performance. As mentioned above, we will focus on propositional and linguistic complexity in particular.

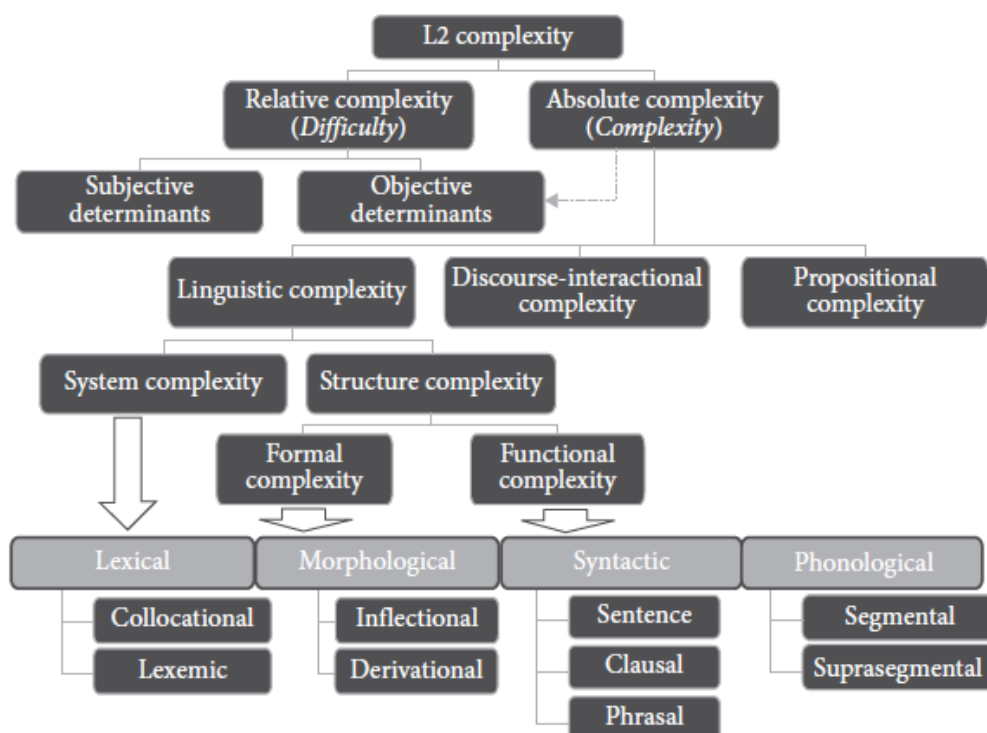


Figure 13. Taxonomic model of L2 complexity (Bulté & Housen, 2012)

According to Bulté and Housen`s (2012) taxonomy, linguistic complexity can be explored at the level of the language system as a whole (*system* complexity) and at the level of the individual linguistic features (*structure* complexity). This dissertation focuses on the system complexity, which “refers to the degree of elaboration, the size, breadth, width, or richness of the learner`s L2 system or “repertoire”, that is, to the number, range, variety or diversity of different structures and items that he knows or uses” (Bulté & Housen, 2012, p.25). We will further refer to the system complexity as *linguistic* complexity. Propositional complexity “refers to the number of information or idea units which a speaker/writer encodes in a given message content” (Bulté & Housen, 2012, p. 24).

Propositional complexity is a relatively new notion in the L2 literature (Bulté & Housen, 2012) and the main aim of this chapter is to expose our definition and treatment of propositional complexity in this dissertation. Before proceeding with propositional complexity, we will first briefly define our view on and treatment of linguistic complexity, which is another dimension of L2 complexity that we deal with in this dissertation.

Concerning linguistic complexity, we primarily adhere to the view offered by typological linguistics, which takes an objective and quantitative perspective on complexity. However, we additionally take into account functional (Halliday, 1987, 2003) and functional-typological (Givón, 1985, 1995) views on complexity, as we consider that these perspectives are highly instrumental in the understanding of the nature of complexity (both linguistic and propositional) in different modes of production. In the followings sub-sections, we will briefly review typological, functional and functional-typological views on complexity. Then we will proceed with the definition and operationalization of propositional complexity in this dissertation.

5.2.1 *View on complexity in typological linguistics*

As mentioned in the previous section, linguistic complexity (which would correspond to the *linguistic system* complexity in Bulté and Housen`s [2012] taxonomy) refers to the dynamic property of the learner`s interlanguage system. Linguistic complexity is manifested in learner`s language production and it is commonly interpreted as “the size, elaborateness, richness, and diversity of the learner`s linguistic L2 system” (Housen & Kuiken, 2009, p.464). We also adhere to the view that linguistic complexity is composed of two interrelated components: lexical complexity and syntactic complexity.

The idea of the lexicon-syntax interface implies that lexical items and syntactic constructions conspire to make a piece of discourse more or less complex (Ravid & Berman, 2010). Following this idea, we consider that both lexical and syntactic dimensions of performance have to be assessed in order to gain thorough understanding of the linguistic complexity of production.

To assess the lexical and syntactic components of linguistic complexity, we, first of all, adhere to the objective and quantitative perspective from typological linguistics (e.g., Dahl, 2004; Miestamo, 2008), which defines the complexity of linguistic units and features in terms of (1) the number and nature of discrete parts that feature consists of and (2) the number and the nature of the interconnections between the parts. In other words, for linguistic complexity, we assume that more components signify more complexity and that a higher density in the relationships between the components signifies higher complexity. Thus, in the assessment of linguistic complexity, we follow the five assumptions advanced in Bulté and Housen (2014): (1) more is more complex (e.g., more words, more idea units represent higher complexity), (2) longer linguistic units are more complex (e.g., greater word, clause, idea, etc. length is assumed to index higher complexity), (3) more and/or more deeply embedded is more complex (e.g., more recursion means more complexity), (4) more varied or diverse is more complex (e.g., more different types of grammatical or lexical forms indicate higher complexity), (5) more marked, infrequent, sophisticated, cognitively difficult or later acquired features are more complex.

This perspective on linguistic complexity is, along certain lines, in accordance with what Pallotti (2015) calls “a simple view of linguistic complexity” (p.117). In this view, complexity is defined in a purely structural way, as arising from the number of linguistic elements and their interrelationships. Undoubtedly, such a definition of

linguistic complexity is very useful as it allows to quantify objectively a number of defined properties and permits to establish a higher or lower degree of complexity in discourse. At the same time, restricting the meaning of complexity to the purely structural and formal aspects may be a too narrow approach if the aim is to assess and compare linguistic complexity in the two modes. As this dissertation deals with the comparison of production in speech and writing, we think it necessary to consider other than purely objective perspectives on linguistic complexity in discourse. Such a perspective is offered by functional linguistics, which proposes that the linguistic complexity of writing and speech is of different nature and achieved by different means. The following section will briefly summarize the treatment of complexity in the two modes in functional linguistics.

5.2.2 View on complexity in functional linguistics

The view of functional linguistics is that complexity of speech and writing is of different nature (Fang, Schleppergrell & Cox, 2006; Halliday, 1987, 200; Halliday & Mathiessen, 1999). According to Halliday (1987, 2003), the critical variable that distinguishes spoken and written language is that of consciousness: in its essence, writing is a more conscious process than speaking. Another critical feature is the differences in the mode of existence of writing and speech: in Halliday's definition, speech exists as a process, while writing exists as a product. This distinction between process-as-mode-of-existence and product-as-mode-of-existence is crucial because it determines different ways of knowing associated with speaking and writing. The written language, presents a *synoptic* view. In this view, the universe is defined as product rather than as process and, consequently, meaning is encoded as things that exist. As a

consequence, written language will display a strong tendency to encode meaning in a nominal (i.e., objectified) form: in head nouns, nouns and adjectives in the nominal group, and nominalized clauses. On the other hand, spoken language presents a *dynamic* view. It defines its universe as process and, thus, encodes it primarily by means of verbs and verbal constructions.

This idiosyncrasy of meaning-making in the synoptic and dynamic views determines that spoken and written language achieve complexity in different ways. When defining and comparing the complexity of oral and written production, Halliday considers that none is more complex than the other, rather the two modes “tend to display different kinds of complexity” (Halliday, 1987:66). Halliday uses the metaphor of *movement* versus *structure* to explain the differences in the nature of complexity in speech and writing respectively. He defines the complexity of spoken language as choreographic, dynamic and intricate, while the complexity of written language is described as crystalline, static and dense. Halliday (1987) continues:

The complexity of spoken language is in its flow, the dynamic mobility where each figure provides a context for the next one, not only defining its point of departure but also setting the conventions for reference to which it is to be interpreted. With the sentence of written language, there is solidarity among its parts such that each equally prehends and is prehended by all the others. It is a structure and not essentially violated by being represented synoptically, as a structural unit. With the clause complex of spoken language, there is no such solidarity, no mutual prehension among all its parts. Its mode of being is a process, not as product. (pp.66-67)

It follows that interpretation and assessment of complexity of production in speech and writing have to be carried out with the awareness of the fact that spoken and written language differ in their preferred patterns of lexico-grammatical organization and, thus, exploit different features of linguistic system to achieve complexity.

To describe how complexity can be achieved in different ways, Halliday (1987, 1990, 2003) distinguishes between *grammatical intricacy*, *lexical density* and *grammatical metaphor*. Halliday considers that a higher degree of grammatical intricacy is to be found in spoken discourse. With grammatical intricacy, Halliday refers to the amount and organization of clauses. In Halliday's view, a prominent characteristic of the spoken language is that it tends to accommodate into a syntagm several related clauses. In other words, spoken discourse is grammatically intricate, as it tends to be formed of multi-componential clause complexes, in which clauses are bound by means of both paratactic (equal status) and hypotactic (unequal status) relationships. Halliday (1987) does not claim the average number of clauses per clause complex would be always greater in speech, rather his point is that this particular kind of syntactic complexity (i.e., grammatical intricacy) is more characteristic of the most unconscious spontaneous uses of language: "it would be better to say that the greater intricacy of a clause complex the more likely it is to be a product of spontaneous speech" (p.71).

In relation to writing, Halliday considers that written language tends to display high lexical density, which is the proportion of lexical items (content words) to the total discourse. Lexical density can be measured in various ways: the ratio of lexical items either to total running words or to some higher grammatical unit (e.g., clause), with or without weighting for relative frequency (in the language) of the lexical items themselves. Importantly, Halliday relates lexical density to the density of information in production, but he does not specify if higher lexical density would signify higher informational load.

In addition to high lexical density, written language also tends to display a high degree of grammatical metaphor, which in Halliday's view, is the most distinctive characteristic of writing. For grammatical metaphor, nominalization plays a significant

part: “nominalizing is the single most powerful resource for creating grammatical metaphor” (Halliday, 1985, p. 352). By means of nominalization, processes (worded as verbs) and properties (worded as adjectives) are reworded metaphorically as nouns, instead of functioning in the clause. Halliday proposes that the tendency of writing to favor grammatical metaphor stems from the fact that, in the synoptic view, the world is objectified and treated as if it was a thing, hence nominalization is needed. Halliday (1985) also indicates that, while not inherently value-laden, grammatical metaphor is “nevertheless inherently complex, and that the least metaphorical wording will always be the one that is maximally simple” (p.350).

In sum, the position of functional linguistics is that spoken and written language favor different kinds of complexity. Written language presents a synoptic view, in which a nominal group is widely employed to pack lexical items at high density; here the complexity is more static. On the other hand, spoken language presents a dynamic view and achieves complexity primarily through grammatical intricacy, by building up elaborate clause complexes out of parataxis and hypotaxis. Here the complexity is more dynamic. Halliday also makes clear that these are general tendencies, and not every particular instance of spoken and written language will conform. He highlights that spoken and written language is not a simple dichotomy but there are many mixed and intermediate types. There are types of writing which are relatively spontaneous, and there are types of speech which are self-conscious and monitored. Thus, the manifestation of grammatical intricacy, lexical density and grammatical metaphor would also depend on other variables such as register, genre or if the performance is monologic or dialogic. Thus, a proper understanding of the complexity of spoken and written language requires due consideration of these variables as well.

Findings from the multidimensional studies of register variation, first undertaken in the 1980s (Biber, 1985, 1986, 1988) provide empirical evidence to Halliday's propositions. The study of register variation entails corpus-based analysis, which is a research approach that focuses on the analysis of a large body of texts (mainly in L1 English) that have been systematically collected to represent different registers, such as conversation, fiction, newspaper language, textbooks, and academic prose. Summarizing the overall findings of these studies, Biber (2009) points out that spoken registers exhibit a heavy reliance on finite clausal syntax, while the written registers show a dense use of complex noun-phrase structures. Biber attributes these findings to the production circumstances of speech and writing. Spoken texts are usually produced in real time, which limits the production of complex noun-phrase structures. Conversely, writing has less time constraints, hence, writers can exploit nominalization to a greater degree. It must be highlighted that the usefulness of the application of the constructs of functional theory to TBLT has been also stressed by Byrnes (2014), who conceptualizes language learning as the diversification of the choices to construct discourse complexity:

Complexity is thus fundamentally about learners being challenged to move back and forth in the intricate weaving together of lexicon and grammar. They make choices regarding greater or lesser cross-clausal and intraclausal complexity, realizing those choices at the phrasal and group level; and they become able to do so with ever greater facility, purposiveness, and creativity toward intended meanings in major text types [...] complexity is that gradual, yet developmentally well motivated increase in the ability to use different register types as they are realized in the textual genres that a particular language system and its users privilege. (p.94)

In addition to the functional linguistics view on complexity, it is also of relevance to consider the functional-typological perspective on complexity formulated in the work by Givón (1975, 1981, 1985).

5.2.3 *Functional-typological view on complexity*

Functional-typological syntactic analysis (Givón, 1985, 1995, 2009), which was originally formulated in the study of diachronic syntax, now is also employed to account for L1 and L2 acquisition. The basic claim of this theoretical position is that “syntax cannot be explained or understood without reference to its use in communication” and that syntactic structure can be shown to “emanate from the properties of human discourse” (Givón, 1979, p.49). Of specific interest to this dissertation is Givón’s (1985) distinction between *paratactic/pragmatic* and *syntactic* communicative modes. Givón’s posits that the discourse-pragmatic communicative mode develops first, and then it gives rise to the syntactic communicative mode. In his own words:

Syntactic structure – characterized by tight constructions under extended intonation contours with attendant grammatical/inflectional morphology – arise ontogenetically, diachronically, and probably also phylogenetically by syntacticization of another communicative mode, the pragmatic/paratactic mode. (Givón, 1985, p.140)

Givón (1979) defines syntacticization as a tendency for “loose, paratactic, “pragmatic” discourse structures [to] develop over time into tight, “grammaticalized” syntactic structures” (p.208). While the term *parataxis* is usually employed to refer to the relationship holding between clauses, Givón uses the term in a broader sense: he distinguishes between pragmatic/paratactic and syntactic modes of communication, the two modes being characterized by pairs of contrasting features:

Pragmatic mode	Syntactic mode
a. topic comment structure	a. subject-predicate structure
b. loose coordination	b. tight coordination
c. slow rate of delivery	c. fast rate of delivery

d. small chunks under one intonation contour	d. large chunks under one intonation contour
e. lower noun/verb ratio discourse, with more simple verbs	e. higher noun/verb ratio discourse, with more complex verbs
f. no use of grammatical morphology	f. extensive use of grammatical morphology

(Givón, 1985, p. 1018)

While the pragmatic mode is a slow means of processing, it is also a more transparent communicative system. The syntactic mode is a faster, semi-automated mode of processing, but is also less transparent, which means that the relationship between and message (content) and code (linguistic form to express the content) is more intricate. Givón (1979, 1985) posits that children acquire first the pragmatic mode of communication, then they gradually syntacticize it. On the other hand, adults dominate all the modes, and use them appropriately under appropriate conditions.

This distinction between pragmatic and syntactic modes is useful for this dissertation as it will help us characterize the nature of complexity of production in the oral and written modes. Of specific interest to our work is also Givón's proposal the way information is encoded in the two modes: pragmatic mode is characterized by the "small chunks under one intonation contour", while syntactic mode exhibits the "large chunks under one intonation contour", with the *chunks* referring to the propositional content or *idea units* in our definition. Also, highly relevant to this dissertation is Givón's theorizing on the definition and measurement of functional (message) complexity (which, in our view, is comparable to what we understand as *propositional* complexity). First of all, Givón (1985) states that functional complexity should be defined independently of structural complexity and proposes intonational phrase as a

unit of functional complexity. Givón (1985) proposes to define the intonational phrase as “a sequence of speech produced under a single intonation contour without a break/pause” or as “a sequence of speech containing an X number of predications/verbs” (p.148); as for functional complexity, Givón (1985) proposed to express it “in terms of number of predications/verbs per unit of time” (p.148). The syntactic mode then is theorized as more complex than the pragmatic mode by virtue of packing more information into the same unit of time. He also asserts that some syntactic constructions are more complex than others in terms of information because they pack in more information per time as compared to other “less-syntacticized” constructions.

At the same time, Givón illustrates that the assessment of lower or higher levels of functional complexity is not straightforward. He provides the following example:

- | | |
|-----------------|--|
| (1) PARATACTIC | (a) ...John TALKED to Mary
sweetly,

(b) so she finally SLEPT with him... |
| (2) SYNTACTIC | ...John SWEET-TALKED Mary
into SLEEPING with him... |
| (3) LEXICALIZED | ...John SEDUCED Mary... |
- (Givón, 1985, p.151)

In these examples, the same information is presented as two separate propositions (i.e., two separate idea units) in (1), as a single proposition with two verbs – main and subordinate – in (2), and as a single proposition with a single verb in (3). In other words, two events in (1) are condensed in one event in (2) and then lexicalized in (3). Per time-unit, (3) is more complex informationally than (2), and (2) more than (1).

However, according to Givón, this parameter is not sufficient to compare functional complexity of these three examples, as the same unit of information is never the same when construed in the context of a different communicative mode, which renders the interpretation of “lower” or “higher” functional (i.e., propositional) complexity complicated.

5.2.4 Summary of the views on complexity

As mentioned previously, this dissertation focuses on linguistic and propositional complexity of production as the measures of L2 performance. In the assessment of linguistic complexity in this dissertation, we will primarily adhere to the objective and quantitative view offered by typological linguistics, which defines complexity in terms of the number of constituents and the nature of the relationships between these constituents (Bulté & Housen, 2014). However, the idea of functional linguistics (Halliday, 1985, 1987) and functional-typological linguistics (Givón, 1979, 1985) are also relevant to us, as they offer an alternative way of interpretation of linguistic and also propositional complexity in L2 discourse. Functional linguistics, in particular, makes a useful distinction between synoptic and dynamic views on complexity, positing that the complexity of spoken and written language is of different nature and achieved by different means (Halliday, 1985, 1987). Functional-typological linguistics (Givón, 1979, 1985) makes a distinction between pragmatic and syntactic mode of communication, and also theorizes about the measurement of functional (i.e., propositional) complexity. In the following sections, we will define the construct of propositional complexity and review the empirical findings in L1 and L2 English.

5.3 Propositional complexity

As defined in Bulté and Housen (2012), the notion of discourse complexity comprises at least three components: propositional, discourse-interactional and linguistic complexity. While linguistic (grammatical and lexical) complexity has been investigated widely (Bulté & Housen, 2014; Ortega, 2015), and discourse-interactional complexity has been investigated occasionally (Gilabert, Barón & Llanes, 2009), the area of propositional complexity has rarely been explored (for the rare exception, see Choong, 2011). Propositional complexity refers to the amount of information (expressed as the number of idea units) which a speaker or writer encodes to convey the intended message (Ellis & Barkhuizen, 2005). It is generally acknowledged that the performance of a speaker/writer who encodes more idea units on a given task is more propositionally complex than that of a speaker/writer who encodes less idea units on the same task (Bulté & Housen, 2012).

The specific relevance of idea units for this dissertation is mainly accounted by two factors. Firstly, the aim of this dissertation is to explore differences between oral and written production, and idea units are considered to be a par excellence feature that discriminates between the two modes, at least in L1 English (Chafe, 1985). Secondly, our dissertation is concerned with task complexity, which is hypothesized to affect primarily the conceptualization/planning stage of language production (Robinson, 2011a). In their turn, idea units represent a notion which is also linked to the planning processes (Butterworth, 1975; Chafe, 1980, 1982), and the analysis of production for idea units is believed to provide the degree to which learners engaged into conceptualization during the performance of the task (Ellis & Barkhuizen, 2005). Thus, the use of idea units could offer us additional insights into the potential effects of task complexity on the conceptualization in L2 production. We will first review the

available theoretical conceptualizations of idea units, which all come from L1 English research. This will be followed by the review of the empirical findings in L1 and L2 English. This will lay the basis for our own guidelines for the identification and classification of idea units in oral and written monologic discourse, which will be presented in detail in the description of the empirical study (chapter 7).

5.3.1 Conceptualizations of idea units in L1 English

The construct of the idea unit has been extensively theorized in early L1 English research. Different theorists provided slightly different definitions for this construct, which was also given different names. Halliday (1967, 1985) was one of the first to introduce a concept of *tone unit* or *tone group*. In Halliday's definition, a tone group expresses a unit of information, a single message block defined by a tone contour. Halliday (1985) also devises a connection between a tone group and a clause:

In English, there is a pattern whereby the clause (a grammatical unit) is associated with the tone group (a phonological unit, characterized by intonation: one melodic movement, or tone contour): other things being equal, each clause is spoken as one tone group. (p.36)

The association between a clause and a tone group is not fixed, and one clause can contain more than one tone unit. At the same time, one tone unit can contain more than one clause. Halliday also considers that punctuation in writing can be used to identify the limits of the tone units.

Tone unit (sometimes called *tone group*) has also been employed in grammatically-oriented research into spoken English. Tone unit is particularly associated to the work by Quirk, Crystal, and their associates (Crystal, Fletcher, &

Garman, 1976; Quirck et al., 1985). Quirk and Crystal (1964) provide their own definition of a tone unit:

For us, the tone-unit is a stretch of speech ... in which there is a climax of pitch prominence which takes the form of 'nuclear' pitch movement of-in the case of level tones-pitch sustention...the nucleus is generally realised on a single syllable, though the pitch movement or sustention may be continued on one or more further syllables which constitute the 'tail' of the tone-unit. . . Little more need be said here . . . because the basic typology does not differ for the most part from that described in considerable detail in the standard works that have treated English intonation in terms of contours rather than phonemic levels. (pp. 50-51)

An example of a separation of a stream of speech into tone units is the following:

But I personally // have never found // a boffin or a statistician // who worked things out in theory which ever came out in practice. (after Crystal, 1969, p. 260)

There is an obvious similarity between the *tone unit* as defined by Quirck and Crystal (1964) and the *tone group* of Halliday (1967). However, there is a difference in the two conceptualizations, as the tone unit in Halliday is considered to have a closer connection to the grammatical clause, while Quirck and Crystal (1964) rely primarily on the pitch-related criteria.

Kroll (1977) was one of the first to employ the term of *idea unit*. It is interesting to note that Kroll developed the construct of idea unit as she was analyzing a corpus which contained oral monologues and written versions of the same. For this contrastive analysis of oral and written production, Kroll needed a unit which could account for both speech and writing. In Kroll's view, T-unit and c-unit, which were the primary units of analysis at the moment, accounted well for written discourse, but not for oral production. Thus, a different unit was necessary which could be used in both modes. In Kroll's (1977) view, an idea unit, defined as "chunk of information which is viewed by

the speaker/ writer cohesively as it given a surface form...related...to psychological reality for the encoder” (p.85) can be used in both speech and writing. In addition to the semantic definition, Kroll (1977) provides detailed syntactic criteria for the segmentation of discourse into idea units:

- (1) a subject and verb counted as one idea unit together with (when present) a (a) direct object, (b) prepositional phrase, (c) adverbial element, or (d) mark of subordination
- (2) full relative clauses counted as one idea unit when the relative pronoun was present
- (3) phrases which occurred in sentence initial position followed by a comma or which were set off from the sentence with commas were counted as separate idea units
- (4) verbs whose structure requires or allows a verbal element as object were counted with both verbal elements as one idea unit
- (5) reduced clauses in which a subordinator was followed by a non-finite verb element were counted as one idea unit
- (6) post-nominal *-ing* phrases used as modifiers counted as one idea unit
- (7) other types of elements counted as idea units were (a) absolutes, (b) appositives, and (c) verbals (p.90)

An example of the segmentation of written text into idea units is as follows (Kroll 1977):

Sue roared all the harder.// She claimed I looked funny,// clinging there, // screaming

[4 idea units] (p. 91)

Much theoretical and empirical work on idea units has been done by Chafe (1980, 1982, 1985, 1994) who employed the term of *idea unit* and *intonation unit* to refer to the same phenomenon.

In his early work, Chafe (1980) equates idea units and tone-units of Crystal:

The “tone-unit” of Crystal...is essentially the same...Most idea units end with an intonation contour that might appropriately be called clause-final: usually either a rise in pitch... or a fall...A second factor is pausing. Idea units are typically separated by at least a brief pause. (pp.13-14)

In his later work, Chafe (1985) has extensively theorized on the nature of this construct, defining an idea unit as a clause-like “spurt of discourse”, which verbalizes the information active in the speaker`s mind at its onset (p. 106). Another way to put it is to say that an idea unit expresses what is held in short-term memory at a particular time. In Chafe`s understanding, short-term memory contains approximately the amount of information that can comfortably be expressed with about seven English words, and as the content of short-term memory changes about every two seconds, an average idea unit in oral discourse is also about seven words long and takes about two seconds to produce.

Chafe offers a number of criteria to identify idea or intonation units in oral discourse. In the most recent development of his theorizing, Chafe (1994) employs the term of *intonation unit* as he considers intonational features as major cues to intonation unit boundaries. First of all, intonation units are preceded and followed by some of kind of hesitation, ranging from a momentary break in timing to a filled or unfilled pause that can last for a few seconds. Chafe also admits that pauses may also occur within intonation units, so additional identifying features are necessary. In Chafe`s (1994) view, other features that characterize intonation units may also involve changes in duration (shortening or lengthening of syllables or words), changes in intonation (rising or falling patterns), changes in voice frequency (perceived as pitch), changes in intensity (perceived as loudness) or other changes in voice quality. A major feature to identify

intonation unit boundaries is change in duration: the intonation unit ends with a word of extended length (i.e., deceleration of the speed of production occurs), while the beginning of the intonation unit is characterized by acceleration. Chafe considers that this pattern of deceleration-acceleration as the primary evidence of the intonation unit boundaries. Another consistent indicator is a terminal (typically falling) pitch contour at the end of each intonation unit. When it comes to voice quality, intonation units often end and sometimes begin with creaky voice, which is also a cue for idea unit delimitation. In sum, the main distinguishing features of an intonation unit are the (1) pauses preceding and following it, (2) pattern of acceleration-deceleration, (3) overall decline in pitch level, (4) falling pitch contour at the end, and (5) creaky voice at the end.

Chafe (1994) also distinguishes between different types of idea units. He terms unfinished or abandoned units as *fragmentary*, and subcategorizes successful units as *substantive* (those that convey information about events, states, or referents) and *regulatory* (their function is to regulate information flow). Regulatory units tend to be one word long and the majority represent particles. A prototypical substantive idea unit has a form of a clause, that is, it contains one verb phrase along with whatever noun phrases, prepositional phrases, adverbs, and so on are appropriate. Not all of the substantive idea units contain an entire verb-complement construction, and some of the units represent a prepositional phrase, a noun phrase or a syntactic fragment of some other kind. The following conversational excerpt illustrates Chafe's division of spoken discourse into idea units:

...Cause I had a...a thick patch of barley there, [1 substantive idea unit = IU]

...mhm, [1 regulatory IU]

...about the size of the...kitchen and living room, [1 substantive IU]

...and went over it, [1 substantive IU]

...and then, [1 regulatory IU]

...when I got done, [1 substantive IU]
I had a little but left, [1 substantive IU]
...so I turned around, [1 substantive IU]
and I went and sprayed it twice. [1 substantive IU]
...and it`s just as yellow as...can be. [1 substantive IU]

(Chafe, 1994, pp. 66-67)

Chafe (1994) is positive about the feasibility of segmenting speech into idea or intonation units and he observes that “intonation units emerge from the stream of speech with a high degree of satisfying consistency” (p.62). At the same time, he admits that problematic cases are unavoidable, and that sometimes it is not easy to read off the boundaries of intonation units in the acoustic data as “the physical manifestations of psychologically relevant units are always going to be messy and inconsistent” (Chafe, 1994, p.58).

In Chafe`s (1985) view, an idea unit is a unit “of cognitive processing and memory” in speech, but they can also be identified in writing (p.111). In fact, Chafe (1985) considers that “some of the differences between written and spoken language can be understood only with reference to the notion of idea units” (p.106). Due to the on-line pressures inherent in speech, speakers do not have time or mental resources to compose long idea units with complex syntax. On the other hand, the increased amount of time available to writers and the editability of writing allow to create “super-idea units” (Chafe, 1985, p.111) or “expanded idea units” (Chafe & Danielewicz, 1987, p.86), which are longer and more complex than those in speech. As Chafe (1982) describes it:

In writing, it would seem, our thoughts must constantly get ahead of our expression of them in a way to which we are totally unaccustomed when we speak. As we write down one idea, our thoughts have plenty of time to move ahead to others. The result is that we have time to integrate a succession of ideas into a single linguistic whole in a way that is not available in speaking. In

speaking, we normally produce one idea unit at a time [...] In writing, we have time to mold a succession of ideas into a more complex, coherent, integrated whole, making use of devices we seldom use in speaking [...] the abnormal quickness of reading fits together with the abnormal slowness of writing to foster a kind of language in which ideas are combined to form more complex idea units and sentences. (pp.37-38)

In sum, Chafe views spoken language as typically consisting of chains of relatively brief, relatively independent idea units, while written language is viewed as presenting longer idea units placed in various relations of dependence.

It is also interesting to note that Chafe (1985) views written language as “historically secondary to spoken language” (p.107), which entails that writing is founded on the resources of spoken language. This means that an idea unit is a natural unit of oral production, while, in writing, it represents a carryover from speech. Chafe considers that writers intuitively organize their discourse into the format of idea units, using punctuation marks to show the same intonational and hesitational patterns that in oral production would indicate the limits of idea units. As Chafe and Danielewicz (1987) put it, “written language has a covert prosody which is analogous to that of spoken language: both writers and readers assign pitch, stress, and pauses to language as they write and read it” (p.95), as a result it can be assumed that boundaries of idea units “are at least reasonably well indicated by punctuation” (p.96).

Another point stressed by Chafe is that idea units represent a significant unit of both language production and comprehension. Chafe theorizes that listeners assimilate information in chunks of idea unit size, and, in the similar way, presentation of written discourse in the format of idea units enhances the readability of written language.

5.3.2 Summary of the theoretical positions on idea units

The preceding review shows that the construct of the idea unit has been extensively theorized in L1 English research. Different theorists proposed different names for this unit, and they also defined this construct in slightly different ways. Thus, the units described in this review are similar, but not identical, as for some theorists this unit is primarily phonological (Halliday, 1985; Quirk & Crystal, 1964), while others have conceptualized this unit as a primarily semantic unit of discourse (Chafe, 1980, 1985; Kroll, 1977) and, importantly, as a unit of planning (Chafe, 1985, 1988, 1994). It is interesting to note that in his early work Chafe (1980, 1985) employed the term *idea unit*; however, in his most recent publications (e.g. Chafe, 1994) he started using the term *intonation unit*. This stress on the “tone” or “intonation” is the recognition of the importance of the prosodic factors in delineating idea units. Concerning writing, Halliday (1985) and Chafe (1985) coincide that punctuation in written production reflects the covert prosody, and, thus, can be used as a cue to the boundaries of idea units. The relationship between the idea unit and the clause has also been charted: Halliday (1985), Kroll (1997) and Chafe (1985) emphasized the close (but not fixed) association between the idea unit and the clause, positing that a prototypical idea unit would typically have a form of a clause. The biggest contribution to the theorizing about idea units has been, undoubtedly, done by Chafe (1980, 1985, 1994). Of specific relevance to this dissertation is Chafe’s (1985, 1988) proposition that idea units differ in writing and in speech; additionally, Chafe (1988) indicated the existence of supra-clausal, more complex ideas, although he did not provide the exact definition of this construction. In sum, L1 English research provides a valuable starting point for the definition of the construct of the idea unit for the purposes of the L2 discourse analysis. In the following section, we will also review the empirical findings on idea units in L1

and L2 English research. In contrast to the theoretical propositions, the empirical research on idea units has been rather scarce.

5.3.3 Empirical research on idea units in L1 English

One of the first studies on idea units was by Horowitz and Samuels (1964), who performed two experiments in which they compared lexical complexity and idea units in oral and written production. In the first experiment, the participants were 40 L1 English university students. The participants were offered two topics to discuss and their task was to say (or write) for two minutes all they could about the first topic, and then to write (or say) all they could about the second topic. Half of the participants spoke on one and wrote on the other topic, and the other half reversed the procedure. The authors admit, however, that in the first experiment they did not take into account the time-intensiveness of writing. For this reason, they set a second experiment with longer experimental periods for written expression. In the second experiment, three groups of subjects participated. Group 1 (N=20), Group 2 (N=16) and Group 3 (N=13) were tested under identical conditions to the first experiment with the exception: in the instructions, Group 1 was told that they would be allowed 4 minutes for writing and were stopped; Group 2 was told that they would be allowed 6 minutes for writing; Group 3 were given no time limitation, but they were stopped after 12 minutes had elapsed.

Written and oral production was analyzed for lexical complexity (type-token ratio), fluency (words per minute) and idea units (number of idea units, number of ideas per 100 words, efficiency of the production of ideas). For Horowitz and Samuels (1964) analysis of discourse in terms of idea units represents psychological analysis and they

define an idea as “an utterance that expresses a thought in a meaningful, relevant, and unique way” (p.642). By *meaningful* the authors mean an exposition of material that makes sense, cognitively, to the experimenters; *uniqueness* refers to non-repetitiveness. The authors also distinguish between *ideas*, *subordinate ideas*, *ancillary ideas*, *communicative signals* and *orientation signals*. An *idea* is a meaningful, relevant and unique utterance, which expresses a concept, thought or notion and which is relevant to the topic of the discussion. *Subordinate ideas* are not unique, they are restated or polished ideas which have been already expressed. *Ancillary ideas* are the ideas which are non-relevant to the topic under discussion; *communicative signal* is a linguistic introduction to the expression of an idea (e.g., “to my mind”); *orientation signal* is an expression which is irrelevant to the topic under discussion (e.g., “Do I have any time left?”).

Horowitz and Samuels (1964) provide the following examples of the segmentation of discourse according to the advanced criteria:

Example 1

To my mind, a good teacher must have a basic foundation of eagerness to learn – a desire for intellectual attainment.

Communicative signal: *To my mind.*

Idea: *A good teacher...eagerness to learn.*

Subordinate idea: *A desire for intellectual attainment.*

Example 2

A good citizen is a person who will take responsibility upon him such as voting and will affiliate with political organizations. Political parties are very serious things in this modern world.

Ideas: (a) *A good citizen...such as voting,*

(b) *and will affiliate...organizations.*

Ancillary idea: *Political parties...this modern world.* (p.642)

The analysis of the data from the first experiment showed that lexical variety was higher in writing, while speakers produced more words and more idea units in general. For the analysis of idea units in the second experiment, the authors counted the number of different types of ideas and they also calculated Specific Idea Impulse, which is the average of ideas produced per unit of time per word multiplied by 100 (N of idea units/unit time/word \times 100). According to the authors, Specific Idea Impulse indicates the efficiency of the production of idea units. The results showed that in the second experiment, speakers produced more linguistic material in general (more words, more idea units), they were more fluent (i.e., more words per minute), but lexical variety was lower in speech. On the other hand, writers were more efficient in the production of ideas. Another relevant finding was that the ratio of production of different types of ideas was changing depending on the mode of production and on the time period of task performance. For example, during the first two minutes of production, the ratio of major ideas was higher in writing than in speech, while the ratio of subordinate ideas (elaborative material) and ancillary ideas (irrelevant material) was higher in oral production.

In the general discussion, Horowitz and Samuels (1964) attribute their findings mainly to the differences in the act of production and output in the two modes. Speech is inherently faster, and a direct result of this ease is prolificacy – speaking produces more material (words, ideas) per unit of time. But, while speech is more prolific, writing is more efficient in terms of ideas produced per unit of time. The authors consider that

speech might be more repetitive also by the virtue of the fact that a speaker does not remember precisely what he said a minute ago and, thus, he might repeat the same. Another assumption that Horowitz and Samuels (1964) advance is that the differences between spoken and written expression can be explained by the fact that speakers “could not tolerate silence” (p.647), which induced them to fill the pauses with the repetitious material and also produce new ideas, including some non-relevant, ancillary ideas. The authors advance a conjecture that it is this elaborative material, with the repetition sometimes of nearly whole sentences and words that makes for the significant differences in the lexical variety in the two modes of production. Horowitz and Samuels (1964) also relate their findings to the fact that speaking and writing represent different strata of the person: speech includes more first thoughts and impulsivity, “whereas writing samples more of the intellectualized and rational and deliberate person” (p.647). In sum, the main conclusion was that the specific conditions of the two modes induce writers and speakers to convey the propositional content in a somewhat different manner, with speakers producing more ideas and writers being more efficient in employing their time when conveying the propositional content of the message.

Another study of interest is the investigation by Butterworth (1975). This study is interesting because it belongs to the work exploring planning in speech production (see also Beattie, 1980, 1983; Garrett, 1982; Roberts & Kirsner, 2000). In this line of research, the temporal structure of discourse is used as an index of cognitive planning during language production. It is assumed that hesitation pauses (of 200 milliseconds or longer) during speech production reflect the initiation of planning of a new chunk of language. In terms of the idea units, the study of Butterworth is interesting as it intends to capture the semantic units (i.e., idea units) independently from the temporal properties of discourse and also aligns the semantic units with the temporal units and

with grammatical units (clauses). The speech samples used in the study were collected from eight L1 English speakers, all male, mean age 24 years. During individual sessions of data collection, the subjects were given a set of social and political topics and were asked to select the one they most agreed with and the one they most disagreed with. The task was to make out the best case for the chosen topic. There were two conditions: an “agree” condition and a “disagree” condition. During oral production, the experimenter (the author) at times participated in the conversation like an interlocutor.

The recorded production was analyzed in terms of (1) temporal cycles (stretches of speech between pauses), (2) clauses and (3) ideas. The transcripts were divided into ideas by human raters. An idea was globally defined as a semantic unit of discourse and there were no restrictions on the size of the unit. The raters were given the transcripts with normal punctuation, and apart from the punctuation, there was no other information about the temporal structure of the text. In spite of certain variation between the raters in dividing the text into ideas, the divisions were found to coincide considerably with the boundaries of the temporal cycles and with the boundaries of the clauses. The author also reports that many raters used brackets to group ideas into a larger (i.e., compound) idea, even though the instructions did not ask them to do so. In the final analysis, Butterworth (1975) reports that the larger were the units the better was the coincidence between the temporal cycles, ideas and syntactic units. Thus, there was a better match between temporal cycles, compound ideas and sentences. The author interprets the correspondence between the temporal cycles and semantic units (i.e., idea units) as evidence to the fact that temporal cycles represent planning units of discourse.

Perhaps the most oft-cited investigation of idea units is that by Chafe and Danielewicz (1987). The authors employed samples from a project containing data from 20 L1 English adult speakers who were university professors and graduate

students. Four different types of language were collected, including conversations, lectures, letters and academic papers. The conversation samples were taken from the recordings of casual talk which took place during an informal dinner in one of the homes. The lecture samples were taken from the same professors and students speaking in class (that is, the samples were not the formal lectures presented to large audiences). Letters consisted of informal letters written by the same participants to relatives, friends, or colleagues. Finally, the academic writing samples were excerpts from articles that had been written for academic journals or books.

For each participant, a sample of 100 idea units was selected from each type of language samples. The data were analyzed for syntactic and lexical complexity, including the analysis of idea units. The division into idea units in speech was carried out in accordance with the intonation features indicative of the unit boundaries (e.g., clause-final intonation, pauses, etc.; see Chafe, 1985). Idea units in writing were identified by relying, in great extent, on the punctuation marks. Chafe and Danielewicz found that idea units were shorter in speech (conversations: mean of 6.2 words; lectures: 7.3 words) than in writing (letters: 8.4 words; academic papers: 9.3 words). The authors attribute these results to the differences in the conditions under which oral and written language are produced: while the on-line pressures in speech keep down the size of an intonation unit, writing allows for extensive planning and editing, which results in longer idea units. Chafe and Danielewicz also identified the most common linguistic devices employed by the writers to expand the idea units. These devices included prepositional phrases, nominalizations, attributive adjectives and nouns. On the basis of these results, Chafe and Danielewicz (1987) conclude that “a principle way in which written language differs from spoken is in the greater length of written intonation units” (p.102).

5.3.4 Empirical research on idea units in L2 English

As mentioned previously, the construct of idea unit is relatively new in L2 literature (Bulté & Housen, 2012) and empirical research in this realm is very scarce. In the previous chapter we reviewed the study by Choong (2011), who found that L2 learners produced more idea units in the picture description task when the pictures were presented in the random order as compared to the condition when the pictures were given in the correct sequence. In other words, Choong (2011) found that learners produced more idea units in the task with enhanced reasoning demands as compared to the simple reasoning task.

Another pertinent study was conducted by Larsen-Freeman (2006). Although this study operationalizes idea units in a rather different manner, it offers very interesting insights and gives an example of a qualitative analysis of idea units. In her study, Larsen-Freeman (2006) examined the development of L2 English proficiency in five L1 Chinese learners. The participants were all female, with ages ranging from 27 to 37. Their proficiency level was impressionistically determined as high intermediate. The learners were asked to carry out the same task 4 times over a six-month period. The participants had to write a narrative about a past episode that they wanted to share. The learners were not allowed to consult a dictionary and they were also told not to worry about whether or not their writing was in perfect English. Three days after writing each story (which was the same story each time), the learners had to tell the story orally. Both written and oral production was untimed and the learners did not receive feedback on their production.

The study was exploratory in nature and both quantitative and qualitative analyses were employed. Quantitative measures were used to account for the macro-level perspective and they included the measures of fluency (length of T-unit), syntactic complexity (clauses per T-unit), accuracy (ratio of error-free T-unit to T-units) and lexical complexity (type-token ratio adjusted for the length of production). At the micro-level, the author examined learner performance from a qualitative angle, which was concerned with the details of how the use of language changed with time. For the qualitative analysis, both the oral and the written narratives were analyzed into idea units. Idea units were defined as “a message segment consisting of a topic and comment that is separated from contiguous units syntactically or intonationally” (Ellis & Barkhuizen, 2005, p.154). In the analysis, Larsen-Freeman arrayed the corresponding idea units side by side to explore how the narrative was constructed each time and how it differed over time. For comparability, the most syntactically succinct idea unit (i.e., anchor idea unit) was entered into the table first, and then the other idea units were mapped against this one. This entailed that the anchor unit in one narration could be realized as several idea units in another version of the narrative. However, to ensure comparability, the single idea unit and the multiple idea units could be entered side by side in single rows of the table. If an idea, which was present in one of the tellings of the story, was not conveyed in another story, the cells were left empty. Arraying the data in this way allowed to explore the way learners changed the use of their linguistic resources over time.

The results of the quantitative analysis showed that over the six-month period, learners were writing more fluently and accurately. There was also an increase in the syntactic and lexical complexity of written production. Also, individual variability in the developmental path of complexity and accuracy was detected. Idea units were

analyzed qualitatively. Because of the space limitations, the author draws selectively on the data to illustrate her approach to idea unit analysis, and to call attention to certain performance features which illustrated language development as of complex, dynamic system. In the analysis of idea units, Larsen-Freeman, for example, reports that later narratives contained idea units that were not present in earlier writing, also the time line of narration exhibited a more sophisticated patterns; at the same time, there were also many segments of idea units that persevered from one narration to the next. Although Larsen-Freeman (2006) indicates that the identified idea units were “mostly full clauses” (p.598), the examples that she provides show that each idea constitutes a chunk of discourse corresponding to one or more than one full sentence.

5.3.5 Summary of the empirical findings on idea units

This review reveals, in the first place, the scarcity of empirical research on idea units. Moreover, idea units were defined in rather different ways in all these studies. Also, the majority of the investigations have been done in L1 English, thus, the application of the researched findings to L2 has to be done with caution. To date, the major empirical findings are as follows: speakers seem to produce more idea units on the same task as compared to writers, but writers use their time more efficiently when conveying the propositional content (Horowitz & Samuels, 1964), idea units are longer in writing than in speech (Chafe and Danielewicz, 1987), there is a high degree of coincidence of idea units with clauses and with the bursts of speech between pauses (i.e., planning units) (Butterworth, 1975). In L2 research, Choong (2011) has demonstrated that speakers produce more ideas in the complex reasoning task as compared to the simple task, and Larsen-Freeman (2006) has shown the way idea units can be employed for the

qualitative analysis of the content of discourse from the developmental perspective. Findings from these studies indicate that the idea unit represents a promising construct, which can reveal new facets of language performance and acquisition. However, to employ the construct of idea units for research purposes, a clear theoretical definition of the construct as well as the operational guidelines are needed. For the purposes of this dissertation, we elaborated the guidelines for the segmentation of discourse into idea units. These guidelines, which are based on previous theoretical and empirical research into idea units, represent the specific methodological contribution of this dissertation. The guidelines will be presented in chapter 7.

5.4 Chapter summary

In this chapter, we presented the views on linguistic and functional complexity in different schools of linguistics. Concerning the view on complexity in this dissertation, we primarily follow typological linguistics which takes an objective and quantitative view on complexity. However, we additionally draw on the ideas from functional and functional-typological linguistics which provide useful insights on how complexity can be interpreted in different modes of production. Of specific interest to us are Halliday's (1987) ideas concerning different nature of complexity in speech and writing, and also Givón's (1979) distinction between pragmatic and syntactic modes of communication and his views on the assessment of functional complexity in discourse. Additionally, the chapter reviewed early conceptualizations and empirical research into the *tone unit* or *idea unit*. This review reveals that there are numerous definitions of the *idea unit* (or similar type of units) in research. Previous theorizations converge on the following points: idea units are clearly identifiable, they are closely associated with the speech

burst among pauses (i.e., planning units) and with grammatical clauses. There also seem to be differences in the way speakers and writers construct idea units. Empirical research on idea units is scarce, but the scant available empirical evidence provides certain support to these theoretical claims. Previous theoretical discussions and empirical findings provide a valuable basis for our own definition of the idea unit and the operational guidelines, which will be presented in detail in the Method section in chapter 7.

PART II: THE EMPIRICAL STUDY

CHAPTER 6. AIMS AND RESEARCH QUESTIONS

In this chapter we first start by defining the issues under research, and then contextualize our study in previous research. This is followed by the research questions and hypotheses that guided our study.

6.1. The focus of the study

In this section, we contextualize our research in light of previous work and provide the outline of the issues under investigation, highlighting, at the same time, the specific intended methodological contribution of this study.

As mentioned above, the first aim of this dissertation is to provide further insights on the effects of mode and task complexity on L2 performance. In other words, our objective is to explore if linguistic performance is different when the same task is performed in speech and in writing. Previous empirical research on this issue has produced mixed results, and it is still unclear how exactly task performance differs in the two modes. Our second aim is to investigate any potential effects of increasing task complexity in oral production as compared to written production. In terms of task complexity manipulations, we specifically investigate the impact of the resource-directing variable of reasoning demands on L2 performance, which is operationalized as complexity, accuracy and time on task. As mentioned previously, the Cognition Hypothesis (Robinson, 2001, 2011a) predicts that increases in task complexity along the resource-directing lines should lead to a parallel increase in complexity and accuracy of production. However, it must be mentioned that although Robinson's theorizing was

primarily concerned with the oral mode of production, his predictions are not mode-sensitive, which means that the same hypothesis can be applied to speech and writing. Empirical testing of Robinson's predictions has produced mixed results, and it is still not clear if the nature of the effects of task complexity is the same in speech and writing.

Regarding the study design, this research resembles those few studies which explored the effects of mode and task complexity on L2 performance within the same experimental design. The experiment in this dissertation, therefore, resembles the design of the studies by Kormos (2014), Kuiken and Vedder (2011) and Tavakoli (2014), all of which have been reviewed previously. The original contribution of our study is that it explores the effects of mode and task complexity on propositional complexity, in addition to other dependent variables. This performance dimension has been previously neglected in task-based research (but see, Choong, 2011). We expect that exploration of propositional complexity would provide additional insights on L2 task-based performance. In the following section, we will provide the research questions and hypotheses which guided this research.

6.2. Research questions and hypotheses

Based on the above considerations, the following research questions and hypotheses guided our research:

1. Are there any differences in terms of complexity (lexical, syntactic, and propositional), accuracy, and time on task when performing a L2 task orally compared to performing the same task in writing?

Hypotheses 1: For lexical and syntactic complexity, we hypothesized higher values in written production. This hypothesis was based on previous empirical research, as the majority of the available studies found higher linguistic complexity in written production. Thus, higher syntactic complexity in writing was obtained in Ellis and Yuan (2005) and also Kormos (2014), who found higher noun-phrase complexity in written production. Higher lexical complexity was found in Granfeldt (2008), Ellis and Yuan (2005), Kormos (2014) and Bulté and Housen (2009).

Given the inherent time intensiveness of writing, we predicted larger values for time on task in the written mode. To the best of our knowledge, this was the first empirical study comparing propositional complexity in L2 speech and writing; for this reason, a null hypothesis was assumed for propositional complexity. To date, the findings in studies comparing the levels of accuracy in the two modes have been mixed. Thus, Kormos (2014), Ellis and Yuan (2005) and also Ellis (1987) found higher accuracy in writing, while Granfeldt (2008) and Ferrari and Nuzzo (2009) found higher accuracy in speech. Consequently, we assumed a null hypothesis for accuracy.

2. Does task complexity affect lexical complexity, syntactic complexity, propositional complexity, accuracy, and time on task in L2 oral performance in the same way as in L2 written performance?

Hypothesis 2: To our knowledge, there are only three empirical studies which compared task complexity effects in L2 oral and written production within the same experimental design. These studies have arrived to somewhat different conclusions. Thus, Kuiken and Vedder (2011) found that task complexity affected lexical complexity and accuracy in

the same way in writing and speech, but there were differences in the effects of task demands on the syntactic complexity of production in the two modes. Somewhat different results were obtained in Kormos and Trebits (2012) who found that written production was more susceptible to the impact of cognitive task demands along the dimension of syntactic complexity, while speech was more sensitive in the areas of accuracy and lexical variety. Finally, Tavakoli (2014) found greater changes in the syntactic complexity of oral production as compared to writing. The conclusion that can be made on the basis of these findings is that the effects of task complexity seem to be of somewhat different nature in the two modes. For this reason, a non-directional hypothesis was assumed for this research question, and we hypothesized that task complexity, operationalized as +/- reasoning demands, would produce different effects in L2 oral and written task performance.

In the next chapter we provide a detailed account of the methodology that we employed to answer the research questions guiding this dissertation.

CHAPTER 7. METHODOLOGY

7.1 Introduction

The aim of this chapter is to provide the description of the experimental part of this study. In the following sections, we describe the participants and the experimental task. In the description of the experimental task, emphasis is put on the operationalization of task complexity and its independent measurement. We then present the procedure and data analysis. We provide the detailed definition of the performance dimensions under investigation and the measures that we employ. In data analysis, we also provide the detailed guidelines for segmenting discourse into idea units. These guidelines constitute the specific methodological contribution of this dissertation. We finally present the statistical analysis employed in this study.

We tested our hypotheses in a repeated-measures counterbalanced design, with task complexity (simple vs. complex) as a within-subjects factor and language mode (oral vs. written) as a between-subjects factor. Task complexity was operationalized in terms of the resource-directing variable of +/- reasoning demands. The choice of this variable in particular was based on a number of considerations. Following Robinson's (2007) suggestions for task design, we opted for a task variable that would be: (1) effective in promoting learning, (2) useful for real-world task performance, and (3) operationally feasible. We considered that the variable of +/- reasoning demands complied with these requirements as it was supposed to engage learners in argumentative discourse. Ability to engage in argumentation is regarded as a valuable skill in everyday communicative contexts (Kuhn, 1991). However, argumentation skills

often pose a challenge to L2 learners because of the complexity of the structure of the argumentative text (Connor, 1990; Schiffrin, 1985). Practice in combination with reasoning demands could therefore be a useful cognitive and linguistic exercise intended to promote the valuable skill of argumentation in a L2. In terms of operational feasibility, the design of a reasoning task is rather straightforward, and in our case it was based on needs analysis. The details of the operationalization of task complexity will be provided in the further sections.

7.2 Participants and context

Two groups of participants took part in the study: an oral group ($n = 39$) and a writing group ($n = 39$). The oral group data came from the study by Gilabert (2007), in which the participants were university learners of English from two different institutions in Barcelona: the private Ramon Llull University and the state University of Barcelona. All of the participants were Spanish and Catalan bilinguals, and their ages ranged from 18 to 40 years old, with the majority being in their 20s. The students from Ramon Llull University, who were in the first year of the university studies related to the advertising sector, followed a task-based program of learning English through the performance of advertising-related tasks, such as campaign presentations, production or focus group meetings. Students from the University of Barcelona were in the first year of their undergraduate degree in English Studies. In the English Studies degree at a public university, learners follow a program which includes topic-based individual and collaborative linguistic tasks. In the two institutions, students perform tasks both orally and in writing. Also, in these universities, English is not only part of the program but also of the final evaluation. Given that the students from both institutions were in the first year of their university degree studies, it was considered that the differences in the

curricula of the two institutions did not play a significant role in determining the profiles of the learners.

As the students in both institutions had been learning English for a similar number of years in an instructed setting ($M_s = 16\text{--}17$ years) the learners were thus considered to be fairly homogeneous in terms of English learning history. The data for the writing group ($n = 39$) were specifically collected for this study. All the participants in the writing group were the students from the University of Barcelona in the first year of their undergraduate degree in English Studies. Care was taken to choose the participants with the same L1 background and English learning history as in the oral group. The oral and writing groups were also similar in terms of age and gender distribution (oral group: 11 men, 28 women, 18–40 years old; writing group: 10 men, 29 women, 18–37 years old).

It is important to emphasize that the oral group data were analyzed in Gilabert (2007) only for accuracy. This dissertation, however, employs a wider set of measures in the analysis of oral and written performance. Thus, although the oral data were not originally collected for this dissertation, the analysis that we perform in this study is original and was carried out for the specific purposes of this dissertation. All participants in this dissertation were administered the X_Lex (Meara & Milton, 2003) and Y_Lex (Meara & Miralpeix, 2006) vocabulary tests as a proxy of general L2 proficiency. These tests have been shown to correlate strongly with proficiency (Milton, 2009; Miralpeix, 2012) and have been used as an indicator of L2 proficiency in a number of previous empirical studies (Gilabert, Barón, & Llanes, 2009; Gyllstad & Wolter, 2016). X_Lex and Y_Lex assess receptive knowledge of vocabulary at 1,000–5,000 and 6,000–10,000 frequency bands, respectively. In each test, participants are presented with a list of 120 words and are asked to indicate, in a yes/no format, if they

know the word. In order to compensate for the possibility of a learner's claim to know a word that they, in reality, do not know, a considerable number of pseudo-words made up from Greek and Latin roots are included. The final score of the test is mathematically determined based on the accuracy of the learners' responses. For reliability, scores are adjusted when pseudo-words are identified as real words.

We added up the adjusted scores of the X_Lex and Y_Lex tests to obtain a global vocabulary size score, which we used to estimate the level of L2 proficiency (see Table 5). According to Meara (1992), learners with a vocabulary score of about 5,000 words would normally be classed as intermediate level, but 7,000–8,000 words would be needed to achieve the Cambridge Proficiency level. Given that the mean scores of our speakers and writers were above 5,000 words, we considered our participants as intermediate learners. This estimation of the proficiency level for our participants was also in line with Milton (2009), who indicated that the knowledge of 2,500–3,250 words assessed with X_Lex would correspond to the intermediate Common European Framework of Reference for Languages level (B1 and B2; see Table 5). Finally, we ensured that speakers and writers did not differ in terms of their global vocabulary size, $F(1, 76) = 2.55, p = .11, \eta_p^2 = .01$.

We fully acknowledge the limitations that may result from our choice of a vocabulary test as a measure of L2 proficiency. However, lexical knowledge is purported to have important weight in both speaking and writing proficiency, and there is empirical evidence showing that growth in general L2 proficiency correlates with an increase in lexical knowledge (Meara & Jones, 1988). Thus, various studies have found that vocabulary knowledge was one of the principal factors that allowed discrimination between different levels of L2 global proficiency (cf. Zareva, Schwanenflugel, & Nikolova, 2005).

Table 5. Descriptive statistics for the X_Lex and Y_Lex proficiency tests

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	95% confidence interval	
				Lower limit	Upper limit
X_Lex					
Speakers	39	3,114.46	696.03	2,888.83	3,340.08
Writers	39	3,543.07	570.51	3,358.13	3,727.01
Y_Lex					
Speakers	39	1,890.76	935.54	1,587.50	2,194.03
Writers	39	1,939.20	646.68	1,729.57	2,148.83
X_Lex + Y_Lex					
Speakers	39	5,005.23	1,548.59	4,503.23	5,507.22
Writers	39	5,482.28	1,040.78	5,144.89	5,819.66

Vocabulary knowledge has also been found to be one of the dimensions that had the greatest impact on holistic scores given by raters of L2 oral production (Iwashita, Brown, McNamara, & O'Hagan, 2008) or of L2 written production (Engber, 1995). Even more compelling evidence has come from those studies that took a componential approach and assessed the role of lexical knowledge in relation to other language abilities. Thus, speaking proficiency, in its most recent conceptualization, is viewed as a multicomponential construct consisting of linguistic components and language-processing components (i.e., ability to use linguistic knowledge with sufficient speed and accuracy; Hulstijn, 2015). A study by De Jong, Steinel, Florijn, Schoonen, and Hulstijn (2012) employed structural equation modeling to assess the weight of linguistic

and processing skills in L2 oral communicative ability and found that vocabulary knowledge was one of the best predictors of speaking proficiency. The components of writing proficiency have been shown to be even more numerous, consisting of linguistic, metacognitive, and processing skills plus additional elements such as writing expertise (Schoonen et al., 2003). Studies by Schoonen et al. (2003) and Baba (2009) showed that vocabulary knowledge was an important factor in L2 writing proficiency, although its contribution was not unique. In sum, based on the evidence of the importance of lexical knowledge in both speech and writing, we considered the X_Lex and Y_Lex tests to be an appropriate proxy of L2 proficiency.

7.3 Tasks and procedures

7.3.1 The experimental task and the operationalization of task complexity

To define the operationalization of cognitive task complexity in the experimental task, we adopt a top-down approach, which means that we will discuss complexity at three levels, starting with the most general theoretical level, followed by the observational level and, finally, by the operational level. The experimental task employed in this dissertation is the “Fire-Chief” task from Gilabert (2005, 2007), which has been used in a number of previous investigations (e.g., Gilabert, 2007; Salimi & Dagashpour, 2012; Shiao & Adams, 2011).

At the theoretical level, the Cognition Hypothesis defines task complexity as “the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner” (Robinson, 2001, p.28). Task complexity is further broken down into the resource-directing and resource-dispersing factors, which represent researchable variables that can be deliberately

manipulated to achieve the predetermined effects on L2 performance and development. In this dissertation, we investigate the effects of reasoning demands (in particular, causal reasoning), which pertains to the group of the resource-directing variables in Robinson`s framework. At a more specific level, we operationally defined reasoning demands as the amount and intensity of mental computation required by the task. Linking this design feature to the task performance, it is expected that a more complex version of the reasoning task would engage more intensively such cognitive processes as comparing, classifying, induction, deduction, error analysis, constructing support, abstracting or analyzing perspectives. Therefore, at the observational, performance level, we expect qualitative enhancement of complexity and accuracy of L2 production, with a possible detriment in fluency.

Drawing on these ideas, Gilabert (2005, 2007; see Appendix) created a “Fire-Chief” task, consisting of simple and complex version with salient differences in cognitive complexity. Both versions of the task represent a visual prompt with the short instructions in L2. The visual cues represent a schematic depiction of a building on fire, with several human characters situated in different parts of the building, and also resources such as fire-trucks. The instructions are the same in both task versions: the task requires to save as many people as possible, and the requirement is to specify and justify the sequence of the actions to take. The distinction between the simple and the complex versions of the task lies in the visual cues, which represent, accordingly, simple and complex scenarios of a problem-solving situation. The design of this task sprang from a needs analysis (Gilabert, 2005) in which domain experts described how scenario planning works within the area of crisis management in public relations. In crisis management, teams of experts and non-expert emergency teams gather to predict potential institutional crises and their related actions and communication plans. During

scenario planning, the experts also analyze what makes a problem-solving task more or less difficult (e.g., different risk levels, combinations of events and people involved). As defined in ergonomics literature, a complex problem-solving task is: “(1) dynamic, because early actions determine the environment in which subsequent decisions must be made and features of the task environment may change independently of the solver’s actions; (2) time-dependent, because decisions must be made at the correct moment in relation to environmental demands; and (3) complex, in the sense that most variables are not related to each other in a one-to-one manner” (Quesada, Kintsch & Gómez-Millán, 2005, p.6).

In line with this specification, Gilabert (2005) characterized task complexity at the operational level in the following way: in the simple version of the task, the involved human characters have no particular roles and they are exposed to the similar level of risk. There are also a plenty of resources (two fire-trucks, a helicopter) and the situation is not critical. Thus, in the problem-solving scenario represented in the simple task, there is a loose connection between the involved factors, and the decisions required by the task are not time-dependent or specifically constrained by the cause-and-effect links. On the other hand, in complex task the learners have to deal with specific type of people (e.g., an injured person, an elderly man, a pregnant woman with children), which is expected to create complex relations between the factors and make the task dynamic and time-dependent. There are also less resources (only one fire-truck) and the situation is critical, as there are different fires moving towards the people. All these factors are expected to induce learners to prioritize and justify their actions, and, thus, make a greater effort at mental operations related to reasoning.

The choice of the “Fire-Chief” task for this dissertation was based on a number of reasons. First, this task qualified as a TBLT task because it resembles a real-world

language activity in which meaning is primary and there is a communicative goal to resolve (Skehan, 1998). Second, the design of the task allowed its administration in either the oral or written mode without compromising its authenticity or meaningfulness. The possibility of using the same task in the two modes was particularly relevant because of the desirability of comparing equivalent communicative tasks in order to be able to isolate mode-related idiosyncrasies in performance (Chafe & Danielewicz, 1987; Tannen, 1982). Also, by employing dual-task methodology, self-ratings, and expert judgments, Révész, Michel, and Gilabert (2016) showed that the complex version of the “Fire-Chief” task indeed imposed a higher cognitive load, compared to the simple version of this task. In order to obtain an independent measure of the cognitive load imposed by the task, this dissertation employed the Affective Questionnaire (Robinson, 2001), which asks participants to self-assess the mental effort experienced during task performance on a 9-point Likert scale. As in Révész et al. (2016), it was found that speakers’ ratings of the cognitive load increased significantly from the simple task ($M = 5.08$, $SD = 2.20$) to the complex task ($M = 5.87$, $SD = 1.98$), $F(1, 38) = 5.56$, $p = .024$, $\eta_p^2 = .13$). The same was found for writers, whose ratings increased from the simple ($M = 3.64$, $SD = 1.58$) to the complex condition ($M = 5.03$, $SD = 1.86$), $F(1, 38) = 19.04$, $p < .0001$, $\eta_p^2 = .33$). The obtained results confirmed our expectation that both writers and speakers would perceive the complex version of the task as imposing a higher cognitive load, compared to the simple task.

7.3.2 Procedure

As mentioned previously, the oral data for the present investigation came from the pool of data from the study by Gilabert (2007), which explored the effects of task complexity on self-repair behavior. It is important to reiterate that the research questions in

Gilbert's investigation were completely different from those in this dissertation. We also employed a wider range of performance measures tapping into the dimensions that had not been analyzed previously.

The oral data were collected individually in a single session with each participant. Each participant was given the visual prompt and the instructions for the simple and complex tasks. With an aim of ruling out any potential influences of pre-task planning on L2 production, a very short pre-task planning time (up to 1 minute) was given to the speakers. After reading the instructions, each participant performed orally both simple and complex versions of the task (counterbalanced) and then rated the cognitive load for performing the tasks. The time allowed for task performance was not limited. The majority of participants completed the simple oral task in less than 2 minutes. Similar times were obtained for the complex task.

The written data were collected specifically for this study. Care was taken to choose participants with a profile maximally similar to those who had performed the task in the oral mode. Data collection procedures were also replicated. The written data were collected from all participants together in one group session. The participants received the prompt for either the simple or complex task (distribution of the prompts was counterbalanced) and were given the same amount of pre-task planning time (up to 1 minute) as had been given for the oral task. As soon as a participant finished one version of the task, they rated the cognitive load for that task. Immediately afterward, the participant completed the other version of the task, followed by cognitive load rating. Documenting the amount of time spent on production was an important aspect of the experiment; thus, care was taken to ensure that all participants took notes about their time on task as accurately as possible. As in the oral data collection, the amount of time

allowed for performing the writing task was not limited. Most writers took about 10 minutes to complete the simple task and 13 minutes for the complex task.

7.3.3 Data coding and data analysis procedures

The transcriptions of the oral productions and written compositions were analyzed in terms of complexity (linguistic and propositional), accuracy, and time on task. The CA mode of CHILDES (MacWhinney, 1995) was used for the transcription of the oral compositions. The written compositions were transcribed in the Word format. The measures that we employ in the analysis tap into various CAF dimensions, which have been claimed to constitute valid descriptors of L2 performance (Housen, Kuiken, & Vedder, 2012; Pallotti, 2009).

We were guided by several principles in the choice of the performance measures. First of all, we strived to choose measures that would be relevant for our purposes and adequate for the learner population and the task employed (Norris & Manchón, 2012). Another important consideration was to assess L2 performance multidimensionally (Housen et al., 2012; Norris & Ortega, 2009), which means that a wide range of measures has to be used to tap into various aspects of performance. At the same time, we strived to avoid redundancy in the measurement, which entails the use of the measures tapping into distinct aspects of production. As pointed out by Norris and Ortega (2009), redundant measures create multicollinearity in the analysis, which is highly undesirable (Tabachnik & Fidell, 1996). Conversely, measures that are distinct and complementary can provide a valid assessment of L2 performance (Norris & Ortega, 2009). An important distinction must be also made between general and specific measures. General measures are the ones that look at an aspect of performance by using global calculations (e.g., number of errors per 100 words), while specific measures take

a narrow approach by focusing on a selected discrete feature (e.g., number of lexical errors per 100 words). In this dissertation we primarily employed general measures, as “these have been argued to be the most sensitive to differences between groups of learners in experimental studies” (Skehan & Foster, 1999, p.229). An equally important reason for using general measures was to ensure the comparability of findings between the current dissertation and previous (and future) research. In response to the recent calls for specific production measures (Norris & Ortega, 2009; Robinson, Cadierno, & Shirai, 2009), we also employed a specific measure of propositional complexity (ratio of extended idea units), as we considered that this measure was particularly suitable for revealing the idiosyncratic nature of complexity in the two modes. In the following sections, we will provide a detailed description of the measures employed in this dissertation.

7.3.3.1 *Measures of linguistic complexity*

As stated by Bulté and Housen (2014), “lexical and grammatical complexity constitute separate, independent dimensions of L2 performance and L2 proficiency, rather than being different aspects of the same L2 performance-proficiency area” (p.53). Thus, in this dissertation we assessed both lexical and grammatical (structural) complexity.

► Measuring lexical complexity

Read (2000) conceptualized lexical complexity, which he terms as lexical richness, as a multidimensional feature of a learner’s language, consisting of the following interrelated components: lexical density, lexical variation, lexical sophistication, and number of errors in vocabulary use (see also Lu, 2012). Although we acknowledge the relationship between lexical errors and task performance quality (e.g., Engber, 1995;

Wolfe-Quintero, Inagaki & Kim, 1998), we consider that the measure of lexical errors constitutes a measure of accuracy rather than complexity. Lexical density, which is the ratio of the number of lexical (as opposed to grammatical) words to the total number of words in a text, has been reported to differentiate oral and written production (Halliday, 1985). However, we could not employ this measure in this dissertation because of the issue of multicollinearity. Because of multicollinearity, we also had to discard the use of the *Guiraud* index (Guiraud, 1959) and *Advanced Guiraud index* (Daller, Van Hout, & Treffers-Daller, 2003), which Bulté and Housen (2014) employed as the measures of vocabulary richness and sophistication respectively.

Thus, in this dissertation, we measured lexical variation (i.e, lexical diversity) by means of the D value (Malvern et al., 2004) and lexical sophistication by means of the Lexical Frequency Profile index (Van Daele, 2007). We consider this assessment to be valid: Wolfe-Quintero et al. (1998), for example, posited that lexical complexity is primarily manifested in terms of the range (lexical variation) and size (lexical sophistication) of learners` productive vocabulary. The same opinion was expressed by Tonkyn (2012) who considered that out of the four aspects of “lexical richness” identified by Read (2000), lexical variation and lexical sophistication could be considered as “especially complexity-related” (p.224).

- **Measuring lexical diversity**

Lexical variety (or lexical diversity) refers, in its classical definition, to the range of different words used in spoken or written production, with a greater range indicating a higher diversity (McCarthy & Jarvis, 2010). It has been shown that the index of lexical diversity serves as a useful predictor of other important constructs, such as lexical proficiency and general language proficiency (Crossley, Salsbury, McNamara, & Jarvis,

2011; Housen et al., 2011; Muñoz, 2010). An index of lexical diversity has also been found to serve as a useful gauge of, inter alia, non-nativeness (Gui, 2010) or aphasia (MacWhinney, Fromm, Holland, Forbes, & Wright, 2010). In one of the most recent theorizations of lexical variety in language, Jarvis (2013) draws on the conceptualization of diversity by ecologists who view it as a multidimensional phenomenon including the properties of richness (number of species), evenness (proportional distribution of individuals across species), disparity (the amount of difference between species), importance (centrality and uniqueness), density and dispersion (spatial distribution of individuals and species). In line with the ideas in Jarvis (2013), Mazgutova and Kormos (2015), for example, tapped into the dimensions of rarity, variability and disparity of written production. Although this new conceptualization of lexical diversity is undoubtedly worth exploring, we consider that more research is necessary to validate measures that can reliably assess these purported sub-dimensions of lexical variety. For this reason, in this dissertation we restrict the measurement of lexical diversity to the range of different words in production.

The most common measure of lexical diversity used in SLA research is Type-Token Ratio (TTR) (Templin, 1957). TTR can be defined as the ratio of the number of lexical items to the number of words. However, TTR has been criticized for two reasons (Engber, 1995; Wolfe-Quintero et al., 1998). First of all, it is difficult to decide on the fine-grained definition of a lexical item. Secondly, TTR is considered to be sensitive to text length: “the ratio of tokens tends to decrease as essay length increases” (Engber, 1995, p.145). To overcome this, a number of alternative measures of vocabulary diversity have been developed, among them – the diversity index D (Malvern et al., 2004). D represents a mathematical transformation of the standard TTR intended to reduce the intervening effects of text length. As stressed by Malvern et al. (2004), D

represents “a robust measure of lexical diversity which is not a function of sample size in the way TTR and its simple transformations are” (p.60). A full discussion of the rather complex derivation of the *D* measure is beyond the scope of this chapter (for details, see Malvern et al., 2004; McKee et al., 2000). Basically, *D* provides an indication of the degree of words` repetition in a text. The fewer words are repeated – and thus the more different words that are used in a text – the higher score for *D*. To calculate *D*, we employed D_Tools software (Meara & Miralpeix, 2006) available at http://www.lognostics.co.uk/tools/D_Tools/D_Tools.htm.

- **Measuring lexical sophistication**

According to Meara (1996), the construct of lexical sophistication involves both the depth and breadth of lexical knowledge. Lexical sophistication, also known as lexical rareness, measures “the proportion of relatively unusual or advanced words in the learner`s text” (Read, 2000, p.203). In other words, lexical sophistication has to do with the use of words that are not among the most frequent in the language (e.g., sad vs. forlorn, ask vs. request) (Jarvis, 2013). Frequency as a measure of lexical sophistication is based on the notion that words that are more frequent in natural language are usually learned earlier and are used more often than less frequent words (Kyle & Crossley, 2015). Consequently, the use of rare lexical items would point to a larger and more advanced vocabulary (Jarvis, 2013; Linnarud, 1986). Also, from the information-theoretic points of view (Dahl, 2004), frequent words can be argued to be less complex, than rare words because frequent items are purported to carry lower amounts of information (Juola, 2008). Evidence for the higher complexity of rare words also comes from the lexical decision time studies that have shown that less frequent words are processed slower than high-frequency words (Kuperman et al., 2012). Concerning the

relationship between lexical sophistication and language proficiency, Laufer and Nation (1995), for example, found that more proficient L2 writers employed more rare words in task performance. Lexical sophistication of words can also be related to the performance perception and evaluation: Crossley, Cobb, and McNamara (2013) and Laufer and Nation (1995) found that speech samples and essays that had a higher rate of frequent words tended to earn lower holistic assessment or proficiency scores. Cumulative findings on the relationship between lexical sophistication and language proficiency are mixed, which is attributed to the variability in the way sophisticated words are defined in different studies (Kyle & Crossley, 2015).

In terms of the assessment of lexical sophistication, perhaps the best-established method is to use a corpus-derived frequency counts which allow to examine the frequency of words in language. For example, the Range program (Heatley, Nation, & Coxhead, 2002), available at www.lextutor.ca, makes it possible to identify what percentage of words used in a text belongs to the most frequent 1000 words (Base-list 1 or K1) and 2000 words (Base-list 2 or K2) based on “A General Service List of English Words” by West (1953). This program also measures the percentage of words from “The Academic Word List” (AWL) by Coxhead (2000). Function words, proper nouns and uninterpretable words are classified as “Off-list” words. The Range program has been employed in a number of empirical studies. Kormos (2011b), for example, explored the effects of task complexity on L2 written performance by assessing, among other dimensions, changes in the lexical sophistication of production. In the analysis, Kormos reported the percentages of words in different frequency bands. Following this approach, however, makes it difficult to perform statistical analysis. For this reason, in this dissertation we employed the output from the Range program to calculate the weighted Lexical Frequency Profile index (LFP) (Van Daele, 2007), using the following

formula: $(\text{tokens K1} \times 1) + (\text{tokens K2} \times 2) + (\text{tokens AWL} + \text{Off-list} \times 3) / \sqrt{\text{tokens}}$
(Van Daele, 2007). There are a number of advantages in using the index. First, LFP index is easy to interpret, with higher value indicating higher lexical sophistication. Also, LFP facilitates statistical analysis.

► Measuring structural complexity

According to Norris and Ortega (2009), structural complexity is a multifaceted construct. As follows, the selection of metrics should measure different dimensions of complexity with as little overlap as possible to avoid redundancy. Similarly, Bulté and Housen (2014) have stressed that the multidimensional and multilayered nature of L2 complexity demands “a sufficiently wide range of judiciously chosen complexity measures” for its assessment (p.56). Ortega (2015) also pointed out that in the assessment of structural complexity it is important to take into account, *inter alia*, the purpose for which complexity is assessed, the level of proficiency of the learners or the mode of performance.

Concerning the assessment of complexity, Norris and Ortega (2009) recommended assessing minimally the following sub-dimensions of structural complexity: (1) overall or general complexity, (2) complexity via subordination, (3) complexity via phrasal elaboration. In line with this recommendation, we employed three complexity measures, with each tapping into one of these three sub-dimensions. According to Norris and Ortega (2009), such an approach to measuring structural complexity would reflect a variety of ways in which speech may get complexified. Such choice of measures also goes in line with the recommendations of Bulté and Housen (2014) to employ at least two sets of complexity measures: one for measuring complexity in dynamic styles, typically oral, and one that captures complexity in

synoptic styles, which are typically written (p.56). In this regard, we considered that the measure of subordination and phrasal elaboration would be accordingly related to the dynamic and synoptic styles of production. Similarly, Biber, Gray and Poonpon (2011) considered that measures of subordination were appropriate to assess structural complexity in oral production, while the assessment of complex noun phrases was necessary for writing. Given that in this dissertation we deal with both oral and written samples, we employed both measures of subordination and noun phrase complexity, in addition to the measure of general complexity. In what follows, we discuss each measure in detail.

- **Measuring overall structural complexity**

According to Norris and Ortega (2009), overall or general complexity can be measured by means of any length-based metric with a potentially multiple-clausal unit of production in the denominator. In this dissertation, we employed the mean length of AS-unit as a measure of overall structural complexity. This measure was calculated by dividing the total number of words by the number of the AS-units. The selected syntactic unit of segmentation, “AS-unit”, was chosen due to its suitability for both spoken and written discourse. The definition of AS-unit which guided the subsequent coding process was taken from Foster et al. (2000): “An AS-unit is a single speaker’s utterance consisting of *an independent clause, or sub-clausal unit*, together with any *subordinate clause(s)* associated with either (...) *An independent clause* will be minimally a clause including a finite verb” (p. 365, italics in the original). Length-based measures are common in TBLT research. Mean length of AS-unit, in particular, was used in Tavakoli and Foster (2008), Tavakoli and Skehan (2005), Fukuta and Yamashita (2015).

- **Complexity via subordination**

Norris & Ortega (2009) argued that complexity via subordination can be measured by any metric with clause (or subordinate or dependent clause) in the numerator. In this dissertation, we employed S-nodes per AS-unit as a measure of subordination. The definition of S-node was taken from Ellis et al. (1994): “S-nodes are indicated by tensed or untensed verbs” (p.483). According to Foster et al. (2000), “clause” and “S-node” are interchangeable and both units “provide a relatively unproblematic measure to subdivide units into smaller segments” (p.359). S-nodes per AS-unit is similar to other subordination indices, such as S-nodes per T-unit, clauses per T-unit, clauses per AS-unit which have been extensively used in previous research (e.g., Fukuta & Yamashita, 2015; Kuiken & Vedder, 2008; Salimi & Dagashpour, 2012).

The measure of subordination was specifically relevant for various reasons. First, subordination is considered to be specifically reflective of the way complexity is achieved in oral production (Biber et al., 2011) or dynamic style (Halliday, 1985). Provided that in this dissertation we compared oral and written samples, it was instrumental to explore the way subordination was deployed in the two modes. Also, subordination was appropriate considering the intermediate level of proficiency of the participants in this dissertation. According to Norris and Ortega (2009), subordination measure “is of great value” when measuring structural complexity at intermediate and upper-intermediate levels (p.573). Another motivation was related to the task design feature of reasoning demands. In particular, the task required to justify actions, which would naturally elicit subordination markers such as “because”. This is also consistent with Matthiessen and Thompson (1988) who posited that English subordinate clauses typically indicate reason or cause.

- **Complexity via phrasal elaboration**

To assess complexity via phrasal elaboration, we calculated the mean number of modifiers per noun-phrase; to obtain this measure, we used the Coh-Metrix 3.0 program (MacNamara, Louwrese, & Graesser, 2002). This metric gives information about the complexity of noun-phrases, with higher indices indicating higher noun-phrase complexity.

The basic structure of a noun-phrase is:

Determiner + (premodification) + headnoun + (postmodification and complementation)
Huddleston & Pullum (2002)

Thus, a noun-phrase can be complexified by incorporating pre-head and post-head modifiers. Pre-modifiers include adjectives (*black cat*), participles (*living creatures*) and nouns (*safety measures*). Post-modifiers include relative clauses (*people who work at this place*), -ing clauses (*people using this service*), -ed clauses (*problems associated with the lack of financing*), to-infinitive clauses (*bricks to build a wall*), prepositional phrases (*the cost of living*), noun-phrases in apposition (*the opera "Carmen"*), and complement clauses (*the fact that cats eat mice*) (Huddleston & Pullum, 2002).

With regard to the function of extended noun groups, Halliday (1985, 2003) posits that a nominal expression is a powerful resource of meaning-making as it makes text informationally denser. There is evidence to suggest that nominal complexity is particularly reflective of synoptic style of discourse, which is associated with written

production, while clausal complexity is viewed as characteristic of oral production (Biber, 1988; Biber & Gray, 2011; Halliday, 1985). In Halliday's view (1985, 1987) clausal and nominal complexity are different types of complexity and none is more "complex" than the other (see also, Ortega, 2015).

7.3.3.2 Measuring propositional complexity

For the purposes of this dissertation, we elaborated guidelines for the segmentation of oral and written language into idea units. In these guidelines, semantic criteria are the primary ones, and intonational and syntactic criteria are additionally employed for greater validity and reliability. These guidelines allow to segment oral and written monologic discourse into idea units; the guidelines also allow to distinguish between two main types of idea units: *non-extended* (i.e., simple) idea units and *extended* idea units. The guidelines can be applied to both L1 and L2 English.

Our general definition of an idea unit is as follows: an idea unit is a meaningful chunk/segment of language production; intonationally, an idea unit is typically preceded by a pause and is characterized by a clause-final intonation at its boundary; syntactically, an idea unit can be a part of a clause, a clause (in a prototypical case), or can form a larger, above-clause syntactic construction, which will always be below or maximally equal to an AS-unit (Foster et al., 2000).

In what follows, we provide more details on the semantic, intonational and syntactic criteria employed, and we also explain the distinction between *non-extended* and *extended* idea units.

► Semantic criteria

An idea unit is primarily a semantic unit of discourse. For this reason, the semantic criterion is the principle one in the definition of idea units. According to the semantic criterion, an idea unit constitutes a meaningful, semantically integral chunk of information. “Semantically integral” means that an idea unit conveys a message which constitutes an undivisible block of meaningful information. By “meaningful”, we mean linguistic material that makes sense, cognitively, to the speaker/writer who produces an idea, and also to the listener/reader who interprets the idea. “Meaningful” also implies that an idea conveys information about an event, state, referent or location in space or time. An event can be defined as something that happens: someone or something does an action, or there is change in the state; a state can be defined as a situation a property that exists for a certain period without change. Most events and states would include within them one or more referents, which are the participants in states and events. Location in time and space, which is expressed by a propositional phrase, indicates the situation of an object, abstraction, state or event in time or in the physical realm.

This definition of an idea unit is similar to what Chafe (1994) calls a *substantive* idea. It can thus be stated that these guidelines distinguish only between substantive ideas. These guidelines do not recognize as independent idea units the chunks of discourse which have been defined in other frameworks (cf., Chafe, 1994; Horowitz & Samuels, 1964) as *fragmentary* idea units (i.e., abandoned idea units, false starts), *regulatory* ideas or *communicative signals* (e.g., expressions “to my mind”, “I think”, “in this view”). In our guidelines, such fragmentary, regulatory ideas or communicative signals are subsumed by the idea units which convey information about an event, state, referent or location in space or time.

Irrelevant to our guidelines is also the fact of whether an idea expresses new or old information. In other words, we do not distinguish between idea units that introduce a new topic, and idea units that repeat, restate or elaborate already mentioned information. Also irrelevant is the fact of whether an idea unit conveys the essential content of the message or relates some unessential or even off-topic details.

► **Intonational criteria**

In addition to the semantic criterion, we employ the intonational one. Thus, we assume that, in speech, an idea unit is typically preceded by a pause and is characterized by a clause-final rising or falling pitch at its boundary. Although the clause-final intonation has been reported to be the most consistent signal of an idea unit boundary in an L1 (Chafe 1994), this criteria must be applied with caution to L2 discourse because of the vagaries of intonation patterns and pausing behaviors in L2 production. Pauses in L2 performance do not necessarily occur at unit boundaries, and it is sometimes difficult to distinguish between pauses that result from the formulation of a message or lexical/syntactic search. The intonational criteria would be most suitably applied to the production of proficient and fluent L2 users, but it might be problematic to apply these criteria to the production of low-proficient dysfluent speakers, who tend to produce a fairly large number of inter-clausal pauses (see also Foster et al., 2000). For this reason, intonational criteria are secondary, and they are additionally complemented with the syntactic criteria, which will be discussed in the following section.

Intonational criteria for the segmentation of discourse into idea units can be applied to written production as well. This is based on the assumption that, in written production, language users express prosody which is analogous to that of speech. This means that writers also express pauses, stress or pitch. We assume that writers employ

punctuation marks to denote prosody (see also, Chafe & Danielewicz, 1987; Chafe, 1988; Halliday, 1985). As follows, we assume that punctuation marks, such as full stop, comma, semicolon, exclamation and question mark, parentheses, quotes or hyphen, can be employed as indicators of the boundaries of idea units in written production.

► Syntactic criteria

We additionally employ syntactic criteria to delimit the boundaries of idea units. Syntactically, idea units represent constructions that are smaller or bigger than a clause. Thus, structurally defined, an idea unit can be: (1) a word, a phrase or syntactic fragments of some other kind, (2) a clause, (3) a unit above clause.

Following Chafe (1982, 1987) and Halliday (1985), we assume that a prototypical idea unit would have a form of a clause. That is, an idea unit would contain one verb phrase with whatever noun phrases, adverbs, prepositional phrases, and so on are appropriate (Huddleston & Pullum, 2002). However, the association between an idea unit and a clause is not fixed and, as indicated above, an idea unit can represent a construction below or above a clause. An idea unit, however, cannot be bigger than an AS-unit, which is defined as “a single speaker’s utterance consisting of an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster et al., 2000, p.365).

As mentioned above, an idea unit can be a construction other than a clause. Independent noun phrase satellite units illustrate this type of idea units. In the transcript, the boundary of an idea unit is marked by a double slash (/ /), and the independent noun phrase satellite unit is underlined.

Example 1

There are people on the roof// but not the old man and the woman// they are inside the building [3 idea units]

When bound by a coordinate relationship, which entails a status of equality (Huddleston & Pullum, 2002), the segments of discourse are regarded as separate ideas, as in Example 2.

Example 2

I'm volunteered for the university's fire emergency team// and I must make sure that everyone is saved [2 idea units]

In the case of a subordinate relationship between the clauses, it must be taken into account that a monolithic conception of subordination would be misleading when identifying idea units. First of all, we must consider that different sub-classes of subordinate clauses may be planned differently, and, thus, exhibit different degrees of semantic and syntactic integration (Holmes, 1995; Huddleston & Pullum, 2002). According to Huddleston (1984), in the most central cases, the main and subordinate clauses are closely related, however, this might not apply to all the types of subordinate clauses:

Where a subordinate clause is a constituent of the superordinate clause, we say that it is embedded...However, contrary to what is assumed in many modern grammars, there is a good reason to doubt whether all cases of clause subordination should be handled in terms of embedding. (p. 379)

Holmes (1988, 1995) holds a similar position and considers that it is necessary to make a distinction between *combined* clauses, which would include nonrestrictive relatives, independent clauses and adverbials, and *embedded* clauses, which would include restrictive elements and complements. Holmes (1995) also claims that combined clauses are planned more independently than embedded clauses, and provides evidence from an empirical study (Holmes, 1988), in which he found that combined clauses were preceded by more hesitations than embedded clauses. In line with these positions, Chafe (1980) has proposed that embedded restrictive relative and complement clauses are produced as part of an idea unit containing the preceding main clause.

Thus, in the case of subordination, it is essential to determine the instances when the main clause and its attached subordinate clause(s) represent separate idea units and when the main clause is integrated with its attached subordinate clause(s) to form a single idea unit. In our guidelines, whether or not a subordinate clause constitutes a separate idea unit depends on the conceptual dependency between the main clause and subordinate clause(s) (Schilperoord & Verhagen, 1998). If there is as strong conceptual dependency between the main and the subordinate clause(s), we assume that the subordinate clause(s) constitute(s) an integral part of the conceptualization of the main clause, and therefore they are not available as separate discourse segments. The central cases of conceptual dependency are the cases when a subordinate clause provides information that the main clause is conceptually dependent on. In other words, the information contained in the subordinate clause(s) forms an integral part of the message conveyed by the larger construction, and deletion of the subordinate clause(s) would result in a discourse segment that cannot be properly comprehended. The majority of such cases would entail (1) restrictive relative clauses; (2) clauses introduced by *that*; (3)

complement clauses introduced by *to*; (4) final adverbial clauses, as in Examples 3, 4, 5, 6:

Example 3

I must save the people who are on the roof [1 idea unit]

Example 4

I know that he is wounded [1 idea unit]

Example 5

I decided to take up the stairs [1 idea unit]

Example 6

Everything is easy when there is a helicopter [1 idea unit]

In the case of a loose conceptual relationship between the main and subordinate clause(s), the clauses are considered to represent separate idea units. The majority of such cases would entail (1) nonrestrictive relative clauses; (2) *while, because* clauses; (3) *although, as for, since* clauses; or (4) adverbial clauses in the initial position, as illustrated in Examples 7, 8, 9, 10.

Example 7

The person who is on the ground floor //can be saved later [2 idea units]

Example 8

While they are going down the stairs// I will go to save the others [2 idea units]

Example 9

Because he is wounded// I will save him first [2 idea units]

Example 10

Since they are children// they must be saved [2 idea units]

► ***Non-extended versus extended idea units***

Our data coding guidelines further distinguishes between two main types of idea units: *extended* and *non-extended* idea units. An *extended* idea unit is a higher-level construction than a *non-extended* idea unit. An extended idea unit represents an aggregate of semantically and prosodically-related clauses that are held together by means of embedded subordination, which centrally includes (1) restrictive relative clauses; (2) clauses introduced by *that*; (3) complement clauses introduced by *to*; (4) final adverbial clauses. All other idea units are considered as *non-extended* or simple idea units. We consider extended idea units to be informationally dense ideas given that they incorporate by default more than one semantically meaningful constituent. Example 10 illustrates an extended idea unit:

Example 10

Secondly I would tell everybody not to take the elevator [extended idea unit] // because they can get stuck [non-extended idea unit]

We considered that the distinction between clause-based (i.e., non-extended) and supra-clausal (i.e., extended) idea units is important as it offers greater validity to our guidelines. According to Foster et al. (2000), the clause-based units are useful as they enable easier (in many cases) analysis; however, the supra-clausal units offer greater validity, as they “allow the analyst to give credit to performers who can embed clauses and hence construct chunks of speech which reflect more sophisticated planning processes” (p. 362).

In sum, our data coding guidelines defined an idea unit as a primarily semantic unit of discourse, which can also be identified on the basis of a number of intonational and syntactic criteria. The use of three different types of criteria is claimed to offer greater reliability and validity to our guidelines. Additionally, we established a distinction between clausal and supra-clausal idea units, which allowed us to account for performance with complex planning behavior.

► Relationships between idea units and other units of discourse

In our guidelines we define an idea unit as a primarily semantic unit, which can also be characterized by a number of intonational and syntactic features. It is also instrumental to chart the relationships between idea units and other units of discourse analysis. Table 6 provides the definitions and criteria for the most widely used units of discourse analysis. All these units have been developed to address different research objectives. Some of these units are applicable primarily to written discourse (e.g., T-unit), while others (e.g., clause, AS-unit, utterance) can be applied to both oral and written production. As our guidelines are intended for monologic production, we ignore the

interactive discourse units (e.g., turns); we also ignore the sentence, as this unit has been shown to be problematic in the analysis of spoken and written data (see, for example, Foster et al., 2000; Quirk et al., 1985).

The first point to highlight is that the majority of the units appear to rely mainly on syntactic criteria. An exception is the utterance, which also relies on semantic and intonational criteria. As indicated by Foster et al. (2000), “proponents of [...] the utterance, such as Crookes (1990), do not seem always to distinguish it clearly from the idea unit” (p. 359). In fact, our own definition of an idea unit is rather similar to the definition of utterance provided by Shewan (1988) and Crookes and Rulon (1985), with the difference that our guidelines additionally specify detailed syntactic criteria, which are also instrumental for differentiating between non-extended and extended idea units. Concerning the relationship between utterance and idea units, it could be suggested that there would be systematic co-occurrences between idea units and utterances, although a certain vagrancy in the definition of an utterance does not allow us to describe this relationship in a more precise manner. It is interesting to note that the definition of the utterance provided by Shewan (1988) comes from clinical linguistics, which is an area of research quite separate from L1 and L2 language acquisition. The fact that clinical linguistics employs utterance (which resembles an idea unit) as a base unit provides additional evidence for the fact that this type of unit is psycholinguistically valid (see also, Horowitz and Samuels, 1964; Chafe, 1982, 1985).

Table 6. Units of discourse analysis

Unit	Definition	Criteria
Idea unit	simple or non-extended idea unit: "an idea unit is a meaningful chunk/segment of language production; intonationally, an idea unit is typically preceded by a pause and is characterized by a clause-final intonation at its boundary; syntactically, an idea unit can be a part of a clause, a clause (in a prototypical case), or can form a larger, above-clause syntactic construction, which will always be below or maximally equal to an AS-unit" (definition adopted in this dissertation)	primarily semantic, also intonational and syntactic
	extended idea unit: "an aggregate of semantically and prosodically-related clauses that are held together by means of embedded subordination" (definition adopted in this dissertation)	mainly semantic and syntactic
Clause	a construction containing one verb phrase with whatever noun phrases, adverbs, prepositional phrases, and so on are appropriate (Huddleston, 1984)	syntactic
S-node	"S-nodes are indicated by tensed or untensed verbs" (Ellis et al., 1994, p.483)	syntactic
T-unit	"one main clause plus whatever subordinate clauses happen to be attached or embedded within it" (Hunt, 1966, p.765)	syntactic
C-unit	"grammatical independent predication(s) or...answers to questions which lack only the repetition of the question elements to satisfy the criterion of independent prediction..." "Yes" can be admitted as a whole unit of communication when it is an answer to a question such as "Have you ever been sick?" (Loban, 1966, p.5-6)	syntactic
	"utterances, for example, words, phrases and sentences, grammatical and ungrammatical, which provide referential or pragmatic meaning" (Pica et al., 1989, p.72)	mainly semantic
	"an independent grammatical predication, the same as a T-unit except that in oral language elliptical answers to questions also constitute predication " (Chaudron, 1988, p.45)	syntactic
AS-unit	"an AS-unit is single speaker`s utterance consisting of an <i>independent clause or sub-clausal unit</i> , together with any <i>subordinate clause(s)</i> associate with either" (Foster et al., 2000, p.365)	syntactic
Utterance	"a stream of speech with at least one of the following characteristics: (1) under one intonation contour; (2) bounded by pauses; and (3) constituting a single semantic unit" (Crookes & Rulon, 1985, p.9)	mainly intonational, also semantic
	"a complete thought, usually expressed in a connected grouping of words, which is separated from other utterances on the basis of content, intonation contour, and/or pausing. (i) <i>Content</i> . A change in content is used as one criterion for segmenting utterances ... (ii) <i>Intonation Contour</i> . A falling intonation contour signals the end of an utterance. A rising intonation signals the end of an utterance if it is a question. . . (iii) <i>Pauses</i> . Pauses are used in conjunction with the above two criteria to segment utterances"(Shewan, 1988, p. 124)	

Conceptually, idea units, which are primarily semantic units, are distinct from the syntactic units. However, some connections can be established. The idea unit is closely related to the clause (or S-node): in our guidelines, a prototypical non-extended idea unit is posited to be a clause. The T-unit, the AS-unit and the C-unit can be defined as super-ordinate categories, as they can contain one (minimally) and more than one idea unit.

In sum, the detailed definition and various types of criteria that our guidelines provide allowed us to chart the relationship between idea units and other units of discourse. This comparative analysis shows that there is a traceable relationship between idea units and other relevant units of discourse. At the same time, this analysis provides evidence that an idea unit constitutes a distinct and relevant unit.

► **Measures of propositional complexity**

For the purposes of this dissertation, we calculated: (1) the number of all idea units, (2) the mean length of idea units (tokens/all idea units), and (3) the ratio of extended idea units (extended idea units/all idea units). Of these measures, two metrics were general and one was specific. Concerning the total number of idea units, Bulté and Housen (2012) suggested that “the L2 performance of a speaker who encodes 55 idea units in narrating a story or in describing a picture will be propositionally more complex than that of a speaker who only encodes 25 idea units” (p.24). Given that in our study writers and speakers performed the same task, we considered that the total number of idea units represented a useful general metric to employ in the analysis of propositional complexity. Another general metric used was the mean length of an idea unit, which was calculated by dividing the number of all tokens by the number of all idea units. We

considered this length-based metric as a general measure, similar to the length of an AS-unit and T-unit. As our study compared oral and written production, we considered this metric as specifically relevant, provided Chafe and Danielewicz`s (1987) claims that the main way in which written discourse differs from speech is in the greater length of idea units in writing. Finally, we employed a specific measure of propositional complexity, which was the ratio of the extended idea units. This measure was calculated by dividing the number of extended idea units by the number of all idea units. This measure was also considered as specifically relevant in the study because of the Chafe`s (1982, 1985) claim about the idiosyncrasy of idea units in speech and writing, with written language presenting more complex idea units.

7.3.3.3 *Assessment of accuracy*

Accuracy refers to the degree of conformity of L2 learner`s performance to a norm, where norm is usually defined as the rule system of the target language (Housen & Kuiken, 2009; Housen et al., 2012; Pallotti, 2009; Skehan, 1996). Accuracy, as well as complexity of production, is believed to be linked to the current state of the learner`s (partly explicit and partly implicit) interlanguage knowledge; thus, accuracy, like complexity is seen as “relating primarily to L2 knowledge representation and to the level of analysis of internalized linguistic information” (Housen & Kuiken, 2009, p.462).

Concerning the measurement of accuracy, Housen et al. (2012) highlighted that it can be problematic. The primary problematic issue concerns the criteria for evaluating accuracy and identifying deviations. Housen et al. (2012) argue that it is necessary to define whether the criteria should be restricted to the standard forms or also take into account non-standard and non-native usages, which are acceptable in some registers or contexts. In light of these considerations, Housen et al. (2012) propose the interpretation

of performance in terms of *acceptability* and *appropriateness*, in addition to accuracy. Another issue concerns the choice of the type of measurements, which represent great diversity (for a review, see Van Daele, Housen, Kuiken, Pierrard, & Vedder, 2007). The available options include holistic and analytic scales (e.g., Knoch, 2009; Shaw & Weir, 2007), weighted ratio of accuracy (Foster & Wigglesworth, 2016; Wigglesworth & Foster, 2008), calculating the ratio of error-free units (e.g., AS-units, clauses) (Fukuta & Yamashita, 2015; Kormos & Trebtsis, 2012), counting different types of errors (e.g., severe vs. less severe errors, grammatical vs. lexical errors) in relation to the number of words, T-units, AS-units or clauses (Kuiken & Vedder, 2008, 2011) or counting the overall number of errors in relation to the number of words, T-units, AS-units or clauses (Gilabert, 2007; Shia & Adams, 2011).

All of these measures are not without problems. The scales, for example, are subject to the raters' varying levels of subjectivity. As point out by Schneck and Daly (2012), "criteria within rubrics have a great deal of ambiguity and disparity, which make determinations of validity and reliability more problematic. Raters are compelled to interpret criteria differently, leading to the introduction of personal biases during the writing evaluation process" (p.1321). There has been also criticism of the weighted ratios of accuracy. This method consists in dividing discourse into clauses and assigning to each clause one of the four weights based on the guidelines; the weight depends on the severity of an error: with higher severity, less weight is assigned (Foster & Wigglesworth, 2016; Wigglesworth & Foster, 2008). The rationale behind this approach is that different types of errors affect comprehensibility to varying degrees (Nas, 1975). This type of measurement, however, has been considered as overlapping with what Pallotti (2009) has referred to as "communicative adequacy" or the extent to which learner performance "is more or less successful in achieving the task's goals efficiently"

(p.596). In Pallotti`s (2009) view, communicative adequacy should be defined and assessed as a separate construct. In terms of the empirical validation of the weighted clause ratio, Evans, Hartshorn, Cox, and de Jel (2015) compared this measure with the ratio of error-free T-units and the ratio of error-free clauses. The conclusion was that a weighted clause ratio could be generally considered a valid linguistic accuracy measure, but it discriminated between different levels of proficiency less well than the ratio of the error-free clauses. The use of ratios of error-free units has also been criticized for various reasons. For example, Bardovi-Harlig and Bofman (1989) pointed out that measures such as the ratio of error-free T-units do not reveal how the errors are distributed within the T-unit (or any other unit of analysis), because a T-unit containing a single error is treated identically to a T-unit containing multiple errors. Similarly, Skehan and Foster (2007) pointed out that the unit-based measures of accuracy do not take into account the length of a unit. As a consequence, very short clauses can emerge as disproportionately accurate when compared with longer ones.

Taking into account all these considerations, this dissertation employs the measure of the number of errors per 100 words: $(\text{all errors/words}) \times 100$. Other studies that have used this measure include Mehnert (1998), Fortkamp (1999), Sangarun (2005), and Ruiz-Funez (2015). The advantage of this measure is that it is objective and global, and it allows comparability of our findings to previous research. The criteria to define an error were set against the criteria of Standard English. An error was defined as a “linguistic form or combination of forms, which in the same context and under similar conditions of production would not, in all likelihood, be produced by the speaker’s native speaker counterparts” (Lennon, 1991, p. 182). To prevent multicollinearity, we did not distinguish between error types, such as lexical errors, grammar errors or pragmatic errors (but see Kuiken & Vedder, 2008). Thus, we labeled as *errors* all types

of errors, which basically included, according to the guidelines provided in Polio (1997), word choice errors, word order errors, omissions of obligatory constituents, errors in articles, prepositions, pronouns, quantity words, modals, agreement, number, tense, aspect, voice or negation and pragmatic errors. Borrowings (L1 words without any phonological or morphological adaptation to L2 [Poulisse & Bongaerts, 1994]) and lexical inventions (morpho-phonologically adapted to the target language lexemes that are never used by the native speakers [Dewaele, 1998]) were also counted as errors. Pronunciation errors in speech and spelling and punctuation errors in writing were not taken into account. In the identification of the errors, all cases of ambiguity were given the most favorable interpretation, granting the learners the benefit of the doubt.

7.3.3.4 *Time on task measure*

Rather than calculating fluency measures (e.g., words per minute), we decided to employ time on task (in minutes) as a dependent variable. This decision was the result of a number of considerations. First, we could not use several time-based measures, such as words per minute and total time, within one design as it would create multicollinearity issues. As speech is faster than writing by default, the measure of the speed fluency would not be particularly revealing. Time on task, however, offered a slightly different perspective on task performance. Time on task constitutes an important variable in task performance in general and certainly in writing, where time on task has been shown to be strongly and positively related to text quality (cf. Hayes & Nash, 1996). Thus, the use of this dependent variable would allow to explore: (1) if speakers spent significantly more time on task than writers, and (2) if time on task changed in simple and complex conditions in the two modes. To our knowledge, this is the first empirical study that has explored this aspect.

7.3.3.5 Intra-rater reliability

A few months after the initial data analysis, 35% of the data for the measures that required manual calculations was recoded, which included the number of AS units, S-nodes, idea units, and errors. The obtained Cohen's kappa coefficients (see Table 7) indicated an acceptable degree of intra-rater reliability ($M = .90$, $SD = .012$, $range = .80-.96$).

Table 7. Cohen's kappa intrarater reliability coefficients for ratings of complexity and accuracy measures

Measure	<i>n</i>	Speaking task		Writing task	
		Simple	Complex	Simple	Complex
AS-units	14	.897	.856	.917	.928
S-nodes	14	.942	.921	.932	.938
Idea units	14	.958	.938	.945	.952
Errors	14	.801	.822	.834	.845

7.3.3.6 Summary of the measures used in the study

Table 8 summarizes the task performance measures employed in this dissertation. As mentioned previously, we strived to choose the measures in order to be able to assess the performance multidimensionally, but avoiding, at the same time, redundancy. The measures that we chose are objective, which allows for their quantification and rather straightforward interpretation. All the measures are general, except for one measure of propositional complexity (ratio of extended idea units). Objective measures are thought

to facilitate comparison of findings with other studies; at the same time, the use of the specific measure was expected to reveal the idiosyncrasy of encoding propositional content in writing and speech. In the next section, we will provide the details of the statistical analyses employed in this study.

Table 8. Summary of task performance measures used in the study

Lexical complexity	Structural complexity	Propositional complexity	Accuracy	Time-on-task
D-value	Length of AS-unit	Mean length of IU	Errors/100 words	Total time in min.
LFP	S-nodes/AS-units Modifiers/noun phrase	Total N of IU Ratio of extended IU		

Note. IU = idea unit; LFP = Lexical Frequency Profile.

7.4 Statistical analyses

Repeated measures, two-factor multivariate analysis of variance (MANOVA) was performed with task complexity (simple vs. complex) as a within-subjects factor and mode (oral vs. written) as a between-subjects factor. We considered MANOVA to be appropriate for our purposes, as this test is used to compare groups on a number of dependent variables. To do this, MANOVA creates a new summary dependent variable, which is a linear combination of each of the dependent variables. Essentially, MANOVA tells if there is a significant difference between the groups on the composite dependent variable, and it then performs an analysis of variance using this new combined dependent variable. There is a significant advantage of using an omnibus test (i.e., MANOVA) instead of a series of individual ANOVAs for each dependent variable. The advantage of MANOVA is the following: this test is designed to help protect

against the likelihood of a Type 1 error, which consists in rejecting the null hypotheses (e.g., there are no differences among the groups) when it is actually true. However, the adjustment for the Type 1 error makes MANOVA a more stringent test, which means that it is more difficult to obtain statistically significant differences, especially when the sample is small. Another advantage of MANOVA is that within the same test we can obtain the main effects and the interaction effects. The main effects are effects of the independent variables (mode and task complexity) on the dependent variables (L2 production). The interaction effects show if the impact of an independent variable (task complexity) is the same for the two populations (writers and speakers). The MANOVA test represents a complex set of procedures; thus, it has a number of additional assumptions that must be met. Accordingly, before proceeding with the main analysis, we performed a number of preliminary tests to check if our data conformed to the assumptions of MANOVA (Field, 2013; Pallant, 2010).

As a first step, we checked for the univariate outliers and eliminated the performances that were above or below three standard deviations from the mean. In terms of the distribution of scores, the Shapiro-Wilk test showed that there were no major deviations from normality. By calculating the Mahalanobis distance, we confirmed that our data did not contain multivariate outliers. We also controlled for multicollinearity by examining correlations between the dependent variables and ensuring that the indices were only moderately related. Additionally, we performed Levene's test for homogeneity of between-group variances, conducted Mauchly's test for the sphericity of within-group variances, and used Box's M test values to explore the equality of group sizes (Field, 2013). No serious violations were noted, and (on the basis of these preliminary tests) we concluded that, overall, our data conformed to the assumptions of MANOVA. The effect sizes (η_p^2) were calculated and classified as .01 =

small, .06 = medium, and .14 = large (Cohen, 1988). To report the results of the MANOVA, we used Pillai's trace as it is deemed to be robust for small sample sizes (Hair, Black, Babin, & Anderson, 2009).

7.5 Chapter summary

In this chapter we have explained the methodological decisions taken in our empirical study. We first accounted for the characteristics of the participants and the way their proficiency in L2 was assessed. Then we described the experimental task, with a special focus on the operationalization of task complexity. We also explained the way we performed the independent measurement of the cognitive load posed by the simple and complex versions of the task. This was followed by the explanation of the procedure and the way data were analyzed. Then, we defined the measures employed in this dissertation, with a specific focus on the guidelines for segmenting discourse into idea units. In our guidelines, an idea unit is defined as primarily a semantic unit of discourse; additionally, there are a number of intonational and syntactic criteria to delimit an idea unit in production. We also distinguish between non-extended (i.e., simple) and extended (more informationally dense) idea units. Finally, we established connections between the idea units and other established units of discourse analysis. Definition and operationalization of idea units constitutes the specific methodological contribution of this dissertation. The last sub-section of the chapter presented the statistical test employed. The next chapter will present the results obtained in the experiment.

CHAPTER 8. RESULTS

8.1 Introduction

This chapter reports the results obtained in this dissertation. We present the results for each research question separately. We present, first, the results for the differences between oral and written production and, second, the results for the comparison of the effects of task complexity in speech and writing. We provide the overall MANOVA results, the descriptive statistics, and the univariate test results. We conclude the chapter with the summary of all findings.

8.2 Research question 1

Our first research question asked if there were any differences between oral and written production in terms of complexity (lexical, structural and propositional), accuracy and time on task. For this research question, we predicted higher linguistic (lexical and structural) complexity and higher time on task in writing. For accuracy and propositional complexity we assumed a null hypothesis. The overall MANOVA showed large and statistically significant effect for mode, $V = .92$, $F(10, 54) = 62.46$, $p < .001$, $\eta_p^2 = .92$, observed power = 1.00. This result indicates that there were differences between oral and written production on the combined dependent variable. Table 9 provides the descriptive statistics for the simple and complex speaking and writing tasks.

The comparison of the descriptive statistics across modes shows that, for linguistic complexity, all the scores had higher values in the written performance, except for the mean number of modifiers per noun phrase and the measure of lexical sophistication (Lexical Frequency Profile), which showed similar values in the two modes.

Table 9. Descriptive statistics for simple and complex oral and written tasks

Measure			Simple oral task	Complex oral task	Simple writing task	Complex writing task
Structural complexity	Mean length of ASU	<i>n</i>	38	38	38	39
		<i>M</i>	12.20	12.78	13.44	14.89
		<i>SD</i>	2.66	2.59	2.36	3.27
		<i>CI.95</i>	11.33-13.08	11.76-13.81	12.55-14.33	13.85-15.93
	S-nodes/ASU	<i>n</i>	39	39	38	39
		<i>M</i>	1.58	1.59	1.90	2.04
		<i>SD</i>	.31	.32	.38	.43
		<i>CI.95</i>	1.46-1.70	1.45-1.71	1.77-2.02	1.91-2.18
	Modifiers/Noun phrase	<i>n</i>	39	38	38	39
		<i>M</i>	.76	.77	.78	.82
		<i>SD</i>	.11	.17	.18	.17
		<i>CI.95</i>	.71-.82	.71-.84	.72-.84	.76-.89
Lexical complexity	D-value	<i>n</i>	39	39	39	38
		<i>M</i>	29.76	31.28	44.79	43.93
		<i>SD</i>	10.69	9.11	13.41	11.60
		<i>CI.95</i>	25.55-33.97	27.66-34.-91	40.51-49.06	14.72-16.28
	LFP	<i>n</i>	38	39	38	39
		<i>M</i>	14.11	15.50	14.07	15.45
		<i>SD</i>	2.07	2.32	1.83	2.15
		<i>CI.95</i>	13.43-14.80	14.72-16.28	13.43-14.80	14.72-16.28
Propositional complexity	Total N of IU	<i>n</i>	38	39	37	39
		<i>M</i>	14.42	17.06	12.87	14.71
		<i>SD</i>	4.56	4.78	3.79	4.84
		<i>CI.95</i>	12.96-15.88	15.38-18.73	11.39-14.35	13.01-16.42
	Ratio of extended IU	<i>n</i>	39	39	39	38
		<i>M</i>	.19	.19	.32	.41
		<i>SD</i>	.11	.13	.14	.14
		<i>CI.95</i>	.15-.24	.14-.24	.27-.37	.36-.47
	Mean length of IU	<i>n</i>	39	38	38	38
		<i>M</i>	11.00	10.80	10.99	11.30
		<i>SD</i>	1.79	1.77	1.60	1.89
		<i>CI.95</i>	10.41-11.59	10.16-11.44	10.39-11.59	10.66-11.95
Accuracy	N of errors/100 words	<i>n</i>	39	39	39	39
		<i>M</i>	7.69	7.71	7.93	6.55
		<i>SD</i>	4.83	5.22	2.76	2.12
		<i>CI.95</i>	6.32-9.07	6.31-9.10	6.53-9.32	5.14-7.97
Time-on-task	Time spent on task, min.	<i>n</i>	37	38	39	39
		<i>M</i>	1.52	2.22	10.06	13.15
		<i>SD</i>	.47	.71	2.41	3.31
		<i>CI.95</i>	.92-2.12	1.08-2.55	9.05-10.27	12.41-14.30

Note. ASU = Analysis of speech-unit; CI95 = 95% Confidence Intervals; IU = idea unit; LFP = Lexical Frequency Profile.

When the results for dependent variables were considered separately (see Table 10), we confirmed our initial observations for linguistic complexity. Thus, we found that there were no statistically significant differences between writers and speakers in terms of the mean number of modifiers per noun phrase, which is the measure of complexity via phrasal elaboration ($p = .363$, $\eta_p^2 = .01$). Similar values in writing and speech were also obtained for Lexical Frequency Profile, which is the measure of lexical sophistication ($p = .919$, $\eta_p^2 = .00$). However, there was a statistically significant difference between writers and speakers for two measures of structural complexity and for the measure of lexical diversity, with higher scores in written production. Thus, writers produced longer AS units ($p < .002$, $\eta_p^2 = .14$), showed higher indices of subordination ($p < .0001$, $\eta_p^2 = .32$), and obtained higher value for lexical diversity ($p < .0001$, $\eta_p^2 = .33$). All the effect sizes were high.

For propositional complexity, significant results were obtained for two measures out of three. Thus, speakers obtained a higher value for the total number of idea units ($p < .038$), while writers obtained a higher score for the ratio of extended idea units ($p < .0001$). It is interesting that the effect sizes were different for these two measures. While for the number of idea units the effect size was medium ($\eta_p^2 = .07$), the effect size for the ratio of extended idea units was large ($\eta_p^2 = .40$). Finally, there were no differences between writers and speakers in terms of the idea unit length ($p < .491$, $\eta_p^2 = .01$).

Concerning accuracy, the descriptives showed similar values across modes. The follow-up ANOVA results indeed confirmed that there was no statistically significant differences between writers and speakers in terms of accuracy ($p < .617$, $\eta_p^2 = .00$). Finally, the descriptive data for time on task clearly showed higher values in writing,

which was statistically significant ($p < .0001$), with large effect size ($\eta_p^2 = .89$). In sum, for the majority of measures we obtained higher scores in written production; only one metric (overall number of idea units) was significantly higher in oral production.

Table 10. Univariate test results for the main effect of mode

Dependent measure	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2	Observed power
Structural complexity					
Mean length of AS-unit	9.939	1	.002**	.14	.874
S-nodes/AS-units	29.692	1	.0001***	.32	1.000
Modifiers/noun phrase	0.839	1	.363	.01	.147
Lexical complexity					
D-value	31.517	1	.0001***	.33	1.000
Lexical frequency profile	0.010	1	.919	.00	.051
Propositional complexity					
Total idea units	4.480	1	.038*	.07	.550
Extended idea unit ratio	42.021	1	.0001***	.40	1.000
Mean length of idea unit	0.484	1	.491	.01	.051
Accuracy					
Errors/100 words	0.253	1	.617	.00	.079
Time on task, minutes	534.091	1	.0001***	.89	1.000

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

8.3 Research question 2

Our second research question asked whether the effects of task complexity were different in speech and writing. The overall MANOVA showed large and significant effect for task complexity, $V = .61$, $F(10, 54) = 8.49$, $p < .001$, $\eta_p^2 = .61$, observed power = 1.00, which means that there were significant differences between simple and complex task performance. We also obtained a large and significant effect for the interaction between mode and task complexity, $V = .54$, $F(10, 54) = 6.37$, $p < .001$, $\eta_p^2 = .54$, observed power = 1.00, which means that task complexity affected writing and speech differently.

When the results for the dependent variables were considered separately, we found that the main effect of task complexity was obtained for the mean length of AS unit ($p < .021$, $\eta_p^2 = .08$), Lexical Frequency Profile ($p < .0001$, $\eta_p^2 = .27$), total number of idea units ($p < .0001$, $\eta_p^2 = .16$), extended idea unit ratio ($p < .001$, $\eta_p^2 = .16$), and time on task ($p < .0001$, $\eta_p^2 = .53$) (see Table 11). As can be observed, the effect sizes for all the measures were high, except for the effect size for the mean length of AS unit, which was medium. These results showed that, for these measures, there was a significant change from simple to complex task performance.

Further examination of the data revealed that for some variables there was a significant main effect of task complexity, but the interaction between task complexity and mode was not significant, which means that, for these variables, the effects of task complexity were of the same nature in the two modes (see Table 12). These variables include the mean length of AS unit, Lexical Frequency Profile and the total number of idea units. For these variables, both writers and speakers showed higher values in the complex task.

Table 11. Univariate test results for the main effect of task complexity

Dependent measure	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2	Observed power
Structural complexity					
Mean length of AS-unit	5.650	1	.021*	.08	.648
S-nodes/AS-units	1.690	1	.198	.03	.249
Modifiers/noun phrase	1.231	1	.272	.02	.194
Lexical complexity					
D-value	0.062	1	.804	.00	.057
Lexical frequency profile	22.941	1	.0001***	.27	.997
Propositional complexity					
Total idea units	12.171	1	.001***	.16	.930
Extended idea unit ratio	5.246	1	.025*	.08	.616
Mean length of idea unit	0.051	1	.822	.00	.056
Accuracy					
Errors/100 words	3.277	1	.075	.05	.430
Time on task, minutes	70.833	1	.0001***	.53	1.000

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Thus, for the mean length of AS unit, speakers obtained higher scores in the complex condition (speakers simple task: $M = 12.20$, $SD = 2.66$; speakers complex task: $M = 12.78$, $SD = 2.59$), and the same pattern of findings was obtained for writers (writers simple task: $M = 13.44$, $SD = 2.36$; writers complex task: $M = 14.89$, $SD = 3.27$). For Lexical Frequency Profile, we also found that speakers obtained higher values in the complex task (speakers simple task: $M = 14.11$, $SD = 2.07$; speakers complex task: $M = 15.50$, $SD = 2.32$). Again, the same pattern of findings was obtained for writers (writers

simple task: $M = 14.07$, $SD = 1.83$; writers complex task: $M = 15.45$, $SD = 2.15$). Finally, speakers produced more idea units in the complex task (speakers simple task: $M = 14.42$, $SD = 4.56$; speakers complex task: $M = 17.06$, $SD = 4.78$). Writers also produced more idea units in the complex task (writers simple task: $M = 12.87$, $SD = 3.79$; writers complex task: $M = 14.71$, $SD = 4.84$).

We also found that for some variables the interaction between task complexity and mode was significant (Table 12). These variables included the ratio of extended ideas ($p < .023$, $\eta_p^2 = .08$) and time on task ($p < .0001$, $\eta_p^2 = .44$). The interaction between task complexity and mode narrowly missed the benchmark for statistical significance for the measure of accuracy ($p = .07$, $\eta_p^2 = .05$). These results mean that task complexity effects were different in the two modes for the ratio of extended ideas, time on task and, possibly, accuracy. To explore further, we conducted follow-up ANOVAs, with the alpha level set at .01 after Bonferroni correction. We indeed confirmed that task complexity affected these dimensions of performance differently in the two modes.

Thus, we found that for writers, the ratio of extended ideas was higher in the complex task, compared to the simple task, $F(1,38) = 7.45$, $p < .01$, $\eta_p^2 = .16$, but there were no changes for speakers across tasks, $F(1, 38) = 0.24$, $p = .88$, $\eta_p^2 = .00$. Also, writers spent more time on task in the complex task, $F(1, 38) = 47.59$, $p < .001$, $\eta_p^2 = .56$, whereas speakers spent the same amount of time in each task, $F(1, 35) = 5.49$, $p = .025$, $\eta_p^2 = .13$. The follow-up analysis for accuracy showed that writers produced fewer errors in the complex condition, $F(1, 38) = 9.95$, $p < .01$, $\eta_p^2 = .21$. For speakers, however, the levels of accuracy were the same in the simple and complex tasks, $F(1, 38) = 0.04$, $p = .85$, $\eta_p^2 = .00$.

Table 12. Univariate test results for interaction effect between task complexity and mode

Dependent measure	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2	Observed power
Structural complexity					
Mean length of AS-unit	1.043	1	.311	.02	.171
S-nodes/AS-units	1.546	1	.218	.02	.232
Modifiers/noun phrase	0.528	1	.470	.00	.110
Lexical complexity					
D-value	0.800	1	.374	.01	.143
Lexical frequency profile	0.000	1	.988	.00	.050
Propositional complexity					
Total idea units	0.381	1	.539	.01	.093
Extended idea unit ratio	5.401	1	.023*	.08	.629
Mean length of idea unit	1.040	1	.312	.02	.171
Accuracy					
Errors/100 words	3.407	1	.070	.05	.444
Time on task, minutes	50.411	1	.0001***	.44	1.000

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

8.4 Summary of results

The empirical study presented in this dissertation was designed with a dual aim in mind: (1) to explore if there were differences in terms of complexity, accuracy and time-on-task in L2 oral and written task performance; (2) to explore if task complexity affected L2 oral and written production in the same or different way. Our major findings were

that: (1) oral and written production differed on a number of performance characteristics, and (2) task complexity affected speech and writing in a different way along certain dimensions.

Specifically, our results show that oral and written performance were manifestly different in terms of a number of measures, with higher scores in writing for the general measure of syntactic complexity, subordination, lexical diversity, ratio of extended idea units, and time on task. In contrast, speakers produced more idea units than writers. Task complexity affected speech and writing in a manifestly different way along various dimensions, including the ratio of extended ideas, time on task and accuracy. We also found that task complexity affected writing and speech in the similar way along the dimensions of overall structural complexity, lexical sophistication and one of the measures of propositional complexity (total number of idea units), with higher values in the complex condition.

In the next chapter we will provide the interpretation of these results.

CHAPTER 9. DISCUSSION

9.1 Introduction

In the previous chapter we described in detail the results and the statistical analysis employed in this study. The aim of the current chapter is to provide interpretations to the results obtained in light of the research questions and hypotheses that guided the study.

This dissertation had a dual objective. Our first aim was to explore if and how L2 production differs depending on the mode (oral or written) in which a task is performed. Our second aim was to compare task complexity effects in L2 oral and written production. To answer these questions, we conducted an experiment in which 39 Spanish/Catalan learners of L2 English performed orally the simple and complex versions of an argumentative task, and another set of 39 learners, with a similar profile, performed the same tasks in writing. Our main findings were that (1) our participants' L2 oral and written performance was manifestly different on a number of dimensions, and (2) task complexity affected L2 speech and writing in different ways on a number of dimensions. In what follows, we will discuss the results for each research question separately.

9.2 Research question 1: Oral versus written task performance

The first research question in this dissertation was concerned with the comparison of L2 oral and written task performance. To re-establish the context of this research question,

it should be recalled that, until recently, the predominant focus of the TBLT scholarship has been on oral production. As a result, little is known about the idiosyncrasy of L2 performance and learning opportunities across language modalities. In order to contribute to the knowledge on the way mode may moderate L2 performance and learning opportunities, we posed the following research question:

Research question 1: Are there any differences in terms of complexity (lexical, syntactic, and propositional), accuracy, and time on task when performing a L2 task orally compared to performing the same task in writing?

For this research question, we advanced a number of hypotheses, which were (1) theoretically motivated, and (2) based on previous empirical findings. More precisely, for lexical and structural complexity, as well as for time on task, we predicted higher scores in written production. A null hypothesis was assumed for accuracy and propositional complexity.

The results obtained largely confirmed our directional hypotheses, and the overall findings were as follows:

1. Linguistic complexity (lexical and structural): writers obtained higher scores than speakers on a number of dimensions, including lexical diversity, mean length of AS units and subordination. There were no differences between writers and speakers in terms of lexical sophistication and phrasal complexity.
2. Propositional complexity: we found that speakers produced more idea units overall, while writers employed extended idea units to a greater extent than speakers.

3. Accuracy: we found similar levels of accuracy in oral and written production.
4. Time on task: as predicted, we found that writers spent more time on task than speakers.

In what follows, we will discuss findings for each performance dimension separately.

9.2.1 *Linguistic complexity*

In the assessment of linguistic complexity, we explored various sub-dimensions of structural and lexical complexity. Based on the findings from previous studies, we could formulate a directional hypothesis for this performance dimension. Rather than advancing specific predictions for each individual sub-dimension, we advanced a general prediction that linguistic complexity would be higher in the written mode. Overall, our findings confirmed this hypothesis, as the scores on the majority of measures were higher in written production. In what follows, we will discuss in detail findings on linguistic complexity.

9.2.1.1 *Lexical complexity*

To assess lexical complexity, we employed the measure of lexical diversity (*D*-value), which is considered to assess the breadth and width of vocabulary, and the measure of lexical sophistication (Lexical Frequency Profile), which is believed to gauge the depth of the vocabulary. We found that lexical diversity was higher in writing, while there were no differences between the two modes in terms of lexical sophistication. In the following sub-sections, we provide the discussion of findings for each lexical measure.

► Higher lexical diversity in writing

Previous studies looked primarily at the diversity of vocabulary, and the consistent findings were of higher lexical diversity in writing (Bulté & Housen, 2009; Granfeldt, 2008; Ellis & Yuan, 2005; Kormos, 2014). Yu (2009), however, found similar levels of lexical diversity in writing and speech, but this study has been criticized for employing oral and written samples belonging to different types of discourse, which precludes their straightforward comparison.

In line with the majority of previous investigations, we found higher lexical diversity in writing, with large size effects. To explain this finding, we could suggest that the time pressure conditions in speech might tax the stage of conceptualization, making it difficult for speakers to prepare a complex preverbal plan, requiring a rich and varied lexicalization. Also, at the stage of formulation, speakers have to make their choices very quickly, which may keep lexical variety low. The evanescent character of oral output may also limit the process of monitoring and, as a result, a certain amount of unwanted word repetition or redundancy might stay undetected and/or uncorrected. In contrast, the self-paced conditions in writing provide time for conceptual work and also allow for more relaxed and exhaustive lexical searches as well as monitoring at the level of word selection. Absence of immediate audience might additionally induce writers to employ richer vocabulary in order to transmit their message with maximal precision. Our findings for lexical diversity can also be taken as empirical confirmation of the theoretical predictions concerning the characteristics of planned discourse as well as the language-learning benefits associated with the deeper linguistic processing that characterizes writing, as discussed in earlier sections (see also Manchón, 2014a; Manchón & Williams, 2016).

► **Similar levels of lexical sophistication in speech and writing**

Very few previous investigations have explored differences between oral and written production in terms of lexical sophistication. Kormos (2014), for example, found that writers used more words from the Academic Word List than speakers, which indicates higher lexical sophistication in written production. This finding runs counter to the results in this study, as we found that the levels of lexical sophistication were similar in the two modes. This discrepancy in the findings could be attributed to the methodological differences between Kormos's and our own study, especially concerning the tasks employed. In both studies, picture cues were employed to elicit oral and written performance. However, while in our study the visual prompts were exactly the same, the prompts in Kormos differed in the oral and written tasks. To elicit oral production, Kormos employed two tasks. The first oral task required to describe a comic strip which was a story of a shipwrecked man on a desert island. The second oral task required to narrate a story based on six unrelated pictures which included a number of objects. The tasks were different in writing. The first task was a picture narration task which required to write a story based on the comic strip. This comic strip represented a car in the middle of the desert and a wizard who transforms this car into a carriage without a horse. The second writing task required to narrate a story based on a number of unrelated pictures, which also included objects, but these objects were different from the ones in the oral task. According to Kormos, the oral and written versions of the picture narration task were comparable because they included the same number of actors and narrative events. In the task based on the unrelated pictures, oral and written versions were also considered similar, as both contained an object (book vs. ring), a picture depicting an adverse weather condition (a storm with lightning vs. a storm at

sea), a means of transport (boat vs. airplane), a geographical location (mountains vs. and island), a house (in the middle of a forest vs. in a town) and a door (locked vs. open). According to Kormos, as the prompts for the writing tasks were equivalent to the prompts for the oral tasks, differences in oral and written task performance were attributed to the effect of mode. However, it must be highlighted that the picture cues used to elicit oral and written discourse samples called for the use of different lexical items. For example, while the writing task called for the use of such words as *wizard*, *carriage*, *desert*, the oral task called for the use of *man*, *island*, *ship*. It would be feasible to hypothesize that writers used more sophisticated words in their production because the visual prompt induced them to do so.

In our study, however, this confound was avoided as we used the same picture cues to elicit oral and written performance and our finding was that the level of lexical sophistication was the same in writing and speech. To explain this finding we would like to suggest that the use of the words from different frequency bands is not so much determined by the mode of production, but rather by the type of genre of the task employed. In a similar line, Chafe and Danielewicz's (1987) suggested that the choice of the level of vocabulary is determined according to the "judgments of appropriateness" depending on the context or objective of production (p. 94). As the tasks used in our study were the same in the two modes, our writers and speakers may not have had special reasons to choose lexis from different frequency band levels.

In relation to our findings on lexical complexity, we would like to suggest that mode seems to influence diversity and sophistication of vocabulary in different ways. While the lexical diversity of language might directly depend on the mode of performance, with the written mode eliciting more diverse language, lexical sophistication seems to be more related to type of task or genre, which may determine

the stylistic choices of the language users. Similarly, Chafe and Danielewicz (1987) suggested that the differences in the diversity of vocabulary represent “relatively pure difference between speaking and writing” (p.91), while “there is nothing in the nature of speaking which prevents a speaker from using literary vocabulary, and nothing in the nature of writing which prevents a writer from using colloquial vocabulary” (p.93). In other words, the use of the words from different frequency bands might not be imposed by the cognitive limitations and language production pressures, but by the stylistic decisions taken by the language users in response to a particular task.

9.2.1.2 *Structural complexity*

In the assessment of structural complexity, an effort was made to take into account the multidimensional nature of this construct. Following Norris and Ortega's (2009) recommendations, we chose measures that would tap into (1) overall structural complexity, (2) complexity via subordination, and (3) complexity via phrasal elaboration. This choice of measures was also in accordance with the recommendation to employ the metrics gauging complexity associated with both dynamic (typically oral) and synoptic (typically written) styles of production (Biber et al., 2011; Bulté & Housen, 2014). We found that overall complexity and complexity via subordination were higher in writing, while there were no differences in terms of noun phrase complexity. The comparison of our findings with previous studies is limited because of the overreliance of previous research on the measures of subordination - the caveat which has been repeatedly highlighted by Ortega (2003), Norris and Ortega (2009), and, recently, by Bulté and Housen (2014). Our literature review reflects the same tendency: all the studies that have compared oral and written performance have employed some type of a subordination measure, only Kormos (2014) assessed phrasal elaboration, and none of

the studies employed a general metric of structural complexity. The findings of previous studies were mixed: Granfeldt (2008) found higher subordination in speech, while Ellis and Yuan (2005) found that it was higher in writing. In Kormos (2014), there were no differences in subordination between the two modes, while the clauses were longer in speech and noun phrase complexity was higher in writing. In what follows, we will discuss in detail the findings for structural complexity in this study.

► **Higher overall complexity and complexity via subordination in writing**

In terms of the general structural complexity, we found that writers produced longer AS-units compared to speakers. Based on Norris and Ortega's (2009) consideration of mean length of multicausal production (AS unit in this case) as a general metric of complexity, we conclude that the production of our writers was more syntactically complex overall, compared to that of speakers. However, as Norris and Ortega also noted, a length-based measure with a multicausal denominator can become longer in various ways, for example, through the addition of subordinate clauses or prepositional phrases.

In this sense, further exploration of the data showed that the writers in our study exhibited higher subordination in their production, a finding similar to those in some previous studies (cf. Ellis & Yuan, 2005; Kuiken & Vedder, 2011) but different from the results in Granfeldt (2008) and Kormos (2014). The theories of oral and written production outlined in previous chapters can help us in the interpretation of our findings. Syntactic encoding, which starts at the stage of microplanning and concludes during the stage of formulation, is thought to be largely automatic in a L1 (Levelt, 1989). However, this is not necessarily the case for L2 learners, who might require additional time and attentional resources to activate relevant syntactic information and apply syntactic

encoding mechanisms (Kormos, 2006). The recursiveness of writing processes, together with the visibility of written output, may facilitate the processes of syntactic encoding and, thus, enable writers to produce longer linguistic units or exploit subordinate structures. In contrast, L2 speakers might need to resort to simpler syntax in order to be able to cope with the pressures of online communication. Our results for syntactic complexity therefore represent empirical confirmation of the purported potential advantages of linguistic processing (at least at the syntactic level) in the written mode (see Manchón & Williams, 2016; Williams, 2012) and also corroborate assumptions about why planned discourse can be more propitious for complex syntax (Ellis, 2005; Ochs, 1979).

► **Similar levels of noun-phrase complexity in the two modes**

To assess complexity via phrasal elaboration, we calculated the mean number of modifiers per noun phrase. We found that the indices of noun-phrase complexity were very similar in the two modes. This finding runs counter to the results in Kormos (2014), who found higher noun phrase complexity in written production. Although Kormos employed the same metric to assess noun phrase complexity, caution is required when comparing her findings with the results in this study. As mentioned previously, Kormos (2014) employed different picture cues to elicit oral and written discourse. It is plausible that the writing prompt called for a more intensive use of adjectives, participles and other devices which are used to complexify a noun-phrase, hence higher values for the mean number of modifiers in written production.

In contrast to Kormos (2014), the prompts in this study were identical in the oral and written production tasks and, therefore, our findings of similar levels of noun-phrase complexity cannot be said to have been influenced by the task. As reviewed in

previous sections, extended noun groups are considered to be a characteristic feature of written production (Biber, 1988; Biber et al., 2011; Halliday, 1985). Hence, it would be logical to expect higher noun-phrase complexity in the written samples. However, we found similar indices for this performance dimension. It must be taken into account, however, that the ability to elaborate language by means of nominalization develops later than hypotaxis, which is the use of subordination (Halliday & Mattiesen, 1999). The same idea is conveyed in the theorizing by Byber et al. (2011), who proposed the developmental stages for the ability to complexify discourse in L1 and L2. The rationale behind Biber et al.'s (2011) stages is as follows: oral conversation is acquired first, and the grammar of writing is acquired later, and not always successful. Grammatical structures that are acquired at early stages and are frequently produced in conversation by all native speakers "are obviously not difficult; therefore these structures do not represent a high degree of production complexity" (p.29). On the other hand, complex phrasal embedding is more characteristic of formal writing, and formal writing has to be learned explicitly and it fully develops in adulthood. As a result, Biber et al. (2011) suggest that complex nominal constructions "represent a considerably higher degree of production complexity than the conversational complexity features" (p.29). Following this logic, Biber et al. (2011) proposed the developmental patterns for L1 and L2 learners of English, reflecting the progression from conversational competence to competence in academic writing. The proposed stages are as follows:

The stages generally progress from finite dependent clauses functioning as constituents in other clauses, through intermediate stages of nonfinite dependent clauses and phrases functioning as constituents in other clauses, and finally to the last stage requiring dense use of phrasal (nonclausal) dependent structures that function as constituents in noun phrases. (Biber et al., 2011, pp. 29-30)

The empirical evidence for these theoretical predictions comes from developmental studies in L1 and in L2 that have shown that the ability to employ expanded nominal groups is characteristic of the registers of advanced literacy (Biber et al., 2011; Byrnes, 2009; Byrnes, Maxim, & Norris, 2010; Fang, Schleppegrell, & Cox, 2006). In other words, the ability to pack meaning into the noun phrase, making the text nominally complex, develops later than the ability to make the text clausally complex. In line with these ideas, Norris and Ortega (2009) posited that subordination could be “a useful and powerful index of complexification at intermediate levels,” and the phrasal-level elaboration “should become the pervasive means by which syntactic complexity is achieved at the most advanced levels of language development and maturity” (p. 563). In relation to the findings in this dissertation, it is plausible that our participants were at a point in their L2 development when they had not mastered nominalization sufficiently to use it as a device to make their discourse more complex, hence the similar levels of noun-phrase complexity in the two modes.

9.2.1.3 *Interpretation of the results for linguistic complexity*

In sum, we found that lexical diversity, overall structural complexity and complexity via subordination were higher in writing than in speech. There were no differences between the modes in terms of lexical sophistication and noun-phrase complexity. As the majority of metrics exhibited higher scores in writing, we conclude that, overall, linguistic complexity was higher in written production.

9.2.2 *Propositional complexity*

An important intended contribution of this dissertation is the comparison of oral and written production in terms of propositional complexity, which relates to the encoding

of the semantic content of the task. We operationalized propositional complexity in terms of the idea units, which is a semantic unit of discourse. For the purposes of this dissertation, we defined an idea unit as a meaningful, semantically integral chunk of information. For the analysis of propositional complexity, we elaborated coding guidelines that allowed to segment discourse into idea units and distinguished between *non-extended* (i.e., simple) and *extended* (i.e., complex, informationally packed) idea units. For the assessment of propositional complexity, we employed three measures: (1) the number of idea units, (2) the ratio of extended idea units, and (3) the mean length of idea units. To our knowledge, this is the first empirical study comparing propositional complexity in L2 oral and written production. For this reason, a null hypothesis was advanced for this performance dimension.

Our results showed that speakers produced more idea units overall (medium effect size), while the ratio of extended idea units was higher in writing (large effect size). We also found that the length of idea units was similar in speech and in writing. It is somewhat problematic to compare our findings with L2 studies because the few available investigations have defined this construct very differently (e.g., Larsen-Freeman, 2006) or pursued different research questions (Choong, 2011). A certain point of reference, however, could be found in early research into idea units conducted in L1 English. In what follows, we will discuss our findings for each measure of propositional complexity in detail.

9.2.2.1 Higher number of idea units in speech

Our first finding was that speakers produced more idea units overall than writers. We consider that this finding could be attributed to the differences in the act of production in speech and writing. As mentioned in the introductory chapters, the physical act of

production constitutes one of the main idiosyncratic features distinguishing writing and speech. Speech represents phonic language. In order to produce speech, a vocal tract is employed to produce auditory signals. Writing entails encoding of linguistic representations into visible signs, which are produced by means of handwriting or by means of software. Articulation of phonic signals is inherently faster and physically easier than the production of visual signs. Consequently, written production constitutes the slowest mode of language production: according to Fayol (1997), writing is five to seven times slower than speaking in an individual. Differences in the act of production in the two modes are considered to be consequential for the quality and quantity of oral and written discourse, and it has been well established that speakers are able to produce more linguistic material per unit of time as compared to writers (Horowitz & Samuels, 1987). Another important assumption is that spoken-written distinctions, attributable to the physical characteristics of the act of production, are permanent and impervious to influences, which means that they remain present even when the goal, register or genre of writing and speech overlap (Chafe, 1987, 1992; Chafe & Danielewicz, 1987). In line with these ideas, we would like to suggest that inherent advantage of speakers at the execution stage (i.e., higher speed fluency) allowed speakers to produce more linguistic material (more idea units) as compared to writers.

In terms of the comparison of our results to previous investigations, similar findings were obtained in the early study by Horowitz and Newman (1964). In this investigation, 40 L2 English university students produced oral and written discourse elicited by means of various topics offered for discussion. Written and oral samples were analyzed for lexical complexity and idea units. The authors found that speakers produced more idea units than writers. Horowitz and Samuels (1964) also attributed the prolificacy of speech to the ease and inherently faster speed of oral production. In spite

of the similarity of findings, caution is required when comparing our study to the study by Horowitz and Samuels (1964). Firstly, it must be taken into account that our experiment involved L2 learners, while the study by Horowitz and Samuels (1964) involved L1 speakers. However, the most important point to mention is that the segmentation of discourse into idea units is only broadly comparable in the two studies. Thus, when segmenting discourse into idea units, Horowitz and Samuels (1964) distinguished between *ideas*, *subordinate ideas*, *ancillary ideas*, *orientation signals* and *communicative signals*. While the first four categories broadly coincide with our definition of an idea unit, *communicative signals* (e.g., expressions like “to my mind”, etc.) do not represent independent idea units in our guidelines. This means that the segmented data in Horowitz and Samuel’s study contained a number of short idea units (e.g., “in my view”, “I think”, “in my opinion”), which, according to our guidelines would be subsumed by a larger, fully fledged ideas. Difficulty in comparing our findings to previous investigations precludes us from affirming that speakers would necessarily produce more idea units than writers when performing the same task. Undoubtedly, more research in this area is required.

9.2.2.2 Higher ratio of extended idea units in writing

Our second main finding was that the ratio of extended idea units was higher in writing. In our guidelines, an extended idea unit is defined as a higher-level construction than a non-extended idea. The distinction between non-extended and extended ideas in the guidelines was prompted by the mention in the early L1 studies of “super-idea units” (Chafe, 1988) or “expanded ideas units” (Chafe & Danielewicz, 1987), which were presented as more characteristic of written discourse. One of the first mentions of supra-clausal compound ideas was made by Butterworth (1975), in whose study the coders

were instructed to segment L1 English discourse into ideas. However, no definitions or guidelines were provided, and the coders had to follow their own intuition and criteria when analyzing discourse samples. When presenting the results, Butterworth highlighted that several coders, by their own initiative, grouped ideas into larger units. Interestingly, there was a better coincidence between the boundaries of these large ideas, temporal cycles (bursts of speech between pauses) and syntactic units (clauses and sentences). Also Chafe (1985, 1988) posited an existence of supra-clausal, compound ideas which represented larger chunks of discourse than prototypical clause-like ideas. In Chafe's view, these complex ideas would be more characteristic of written production, as the slowness of writing might induce writers to construct complex ideas, which will contain smaller clause-like ideas bound by means of various relations of dependence. Chafe, however, does not provide any concrete definition or guidelines for the identification of such "super" ideas.

In line with Chafe (1985, 1988), we considered that the use of a specific metric of the ratio of extended idea units would be specifically relevant to tap into the idiosyncrasy of the construction of idea unit in speech and writing. We operationalized an extended idea unit as an aggregate of semantically and prosodically-related clauses, held together by means of embedded subordination. As the extended idea units contain, by default, more than one semantically meaningful constituent, we consider this type of idea units as more complex and informationally dense as compared to non-extended idea units. In line with the theoretical suppositions in Chafe (1985, 1988), we found that written samples indeed contained a larger ratio of extended ideas as compared to speech. Absence of other empirical investigations exploring differences between speech and writing in terms of extended ideas precludes the comparison of our findings to previous research. For this reason, we can only advance a theoretical explanation to our results.

To explain this finding, we again resort to language production models. Idea units are assumed to relate to the conceptualization stage (Ellis & Barkhuizen, 2005) and, we would like to suggest, also to the stage of linguistic formulation where the assembly of phrases and clauses takes place (Kormos, 2006). Time pressure conditions of oral production might induce speakers to organize the propositional content into syntactically simple clause-like units. In contrast, less time constraint in written production allows ideas to be formulated into more complex, integrated linguistic wholes, as (Chafe, 1985) was suggesting. Additionally, the inherent slower pace of writing causes thoughts to “constantly get ahead of our expression of them” (Chafe, 1982, p. 37). This may result in the accumulation of information in the pre-execution phase, which, as a consequence, may also induce writers to create more informationally packed ideas with the resulting higher ratio of extended idea units in writing.

9.2.2.3 *Similar length of idea units in speech and writing*

To explore differences between speech and writing, we also employed the metric of the length of idea units. Our results showed that the length of ideas was similar in speech and in writing. This finding runs counter to Chafe and Danielewicz (1987) who stated that “a principle way in which written language differs from spoken is in the greater length of written intonation [idea] unit” (p. 102). This conclusion is based on the results of the empirical study in which 20 L1 English adult participants, who were university teachers and students, provided samples of oral and written discourse, which were analyzed for idea units. The oral samples collected in this study represented casual conversations and lectures, while the written samples were informal letters and excerpts from academic articles. The authors segmented oral and written production into ideas and found that there was a gradation of the length of the ideas, with the shortest ideas

found in conversation (mean of 6.2 words per idea unit), followed by lectures (7.3 words), letters (8.4 words) and academic papers (9.3 words). On the basis of these findings, the authors conclude that the most distinctive feature of oral and written ideas is their length, which can be attributed to the different processing constraints in the two modes. In Chafe and Danielewicz's (1987) view, limitations in the working memory capacity coupled with the on-line pressures of oral production keep down the size of idea units in speech. On the other hand, writers produce language at the self-paced speed which lowers the pressures on the cognitive resources. Such more relaxed conditions allow writers to produce longer idea units.

Caution, however, is required when interpreting the finding in Chafe and Danielewicz (1987) and when comparing their study with our investigation. The first point to take into account is that the oral and written samples in this study were different in their register and purpose. As Chafe and Danielewicz (1987) highlight, while some of the differences in language production have much to do with the difference in how language is produced (e.g., under more or less level of pressure), other differences in performance are on the account of other factors. In this regard, Chafe and Danielewicz (1987) stress that "the context of language use, the purpose of the speaker or writer, the subject matter of what is being said or written – these are some of the factors which influence the form language takes" (p.84). The importance to control for register and the type of task employed to elicit discourse samples was also highlighted by Biber (2009) who posited that "communicative task is also an important predictor of linguistic variation; therefore, equivalent communicative tasks should be compared to isolate the existence of mode differences" (p.75). Given that Chafe and Danielewicz (1987) did not control for register, it is difficult to conclude that the variation in the length of ideas can be attributed to mode or to register. Another important limitation is that this study

provides only descriptive statistics, which makes it impossible to determine whether or not the difference between written and oral samples was statistically significant.

We should also be cautious when comparing the results of our study with the findings in Chafe and Danielewicz (1987) because of the different populations employed. While in our study the subjects were L2 learners of intermediate L2 English proficiency, the subjects in Chafe and Danielewicz's (1987) study were educated L1 English speakers. Moreover, we elicited discourse samples by means of an argumentative task, which is very different from the discourse samples obtained in Chafe and Danielewicz (1987). It is a matter of future empirical research to determine if the type or length of idea units in oral and written production is moderated by register or type of task. Another consideration concerns the way the discourse samples were segmented. Because Chafe and Danielewicz (1987) do not provide the detailed guidelines for the identification of idea units, it is hard to determine exactly if we segmented discourse into idea units in the similar or different ways. In the description of the procedure, Chafe and Danielewicz (1987) broadly indicate that an idea unit was a clause-like unit, and that intonation was used to identify the boundaries of idea units in speech, while in writing the boundaries were identified by relying largely on punctuation. Based on this definition, we could state that our methodological decisions were roughly comparable to those in Chafe and Danielewicz (1987). However, it is impossible to determine with precision if the idea units in our understanding represented exactly the same type of units in Chafe and Danielewicz (1987).

In sum, we found that the length of idea units was similar in oral and written production. This finding runs counter to the results in Chafe and Danielewicz (1987), who reported that the ideas were markedly longer in writing. Because of the number of methodological factors, we must be cautious when interpreting and comparing findings

in Chafe and Danielewicz (1987) with the results in our study. Undoubtedly, more research is required to determine if there are differences in the length of idea units in L2 oral and written production.

9.2.2.4 *Interpretation of the results for propositional complexity*

In sum, we found that the total number of idea units was higher in speech, the ratio of extended idea units was higher in writing, while the length of ideas was similar in the two modes. The question that arises now is whether speakers' or writers' propositional complexity appeared to be more complex in our study. To answer this question, we can resort to the ideas offered by different schools of linguistics, which we reviewed in the introductory part. Typological linguistics, for example, takes an objective and quantitative view and defines complexity in terms of the number of linguistic elements and their interrelationships (Bulté & Housen, 2012; 2014; Dahl, 2004; Miestamo, 2008; Pallotti, 2015). In this view, more units, longer units and greater embeddedness mean higher complexity. In this regard, Bulté and Housen (2012) stated that “the L2 performance of a speaker who encodes 55 idea units in narrative a story or in describing a picture will be propositionally more complex than that of a speaker who only encodes 25 idea units” (p.24). Following this interpretation, oral production in our study would be interpreted as more complex in terms of propositional complexity because the speakers produced more idea units overall than writers.

However, from the point of view of typological linguistics, higher embeddedness also represents higher complexity. In our study, the ratio of extended idea units (i.e., idea units with embedded clauses) was higher in writing, which can also be interpreted as indicating higher complexity in written production. We must note, however, that in our guidelines not all types of subordinate clauses form part of

extended idea units. For this reason, the ratio of extended idea units cannot be equaled to the measure of subordination. Syntactically, this metric rather constitutes a measure of embedded clauses (Holmes, 1995; Huddleston, 2002) which are defined as subordinate clauses representing high semantic, intonational and syntactic integration with the main clause.

Concerning the length of an idea unit, at the first sight, this metric might appear as similar to the length-based measures employed to assess linguistic complexity, such as the T-unit and AS-unit length (which are considered to be the measures of global syntactic complexity) or the length of clause (which Norris & Ortega [2009] consider to be the measure of phrasal elaboration). In our view, despite its sharing a superficial similarity with the length-based metrics of syntactic complexity, the mean length of idea unit is a distinct type of metric, because it measures the length of a semantic and not syntactic unit of discourse. For this reason, we would not venture to claim that the length of idea units could be interpreted as, for example, a measure of overall propositional complexity. However, to interpret this measure, we find it useful to resort to the objective and quantitative perspective from typological linguistics which treats longer linguistic units as more complex (see also, Bulté & Housen, 2014). In line with this view, a longer idea unit could be interpreted as indicating higher propositional complexity. Given that in this study oral and written ideas were of similar length, it could be stated that oral and written samples exhibited similar levels of complexity along this sub-dimension of propositional complexity.

Our findings for propositional complexity can also be explained by drawing on functional-typological linguistics. Specifically relevant to this dissertation is Givón's (1975, 1981, 1985) distinction between *pragmatic* and *syntactic* communicative modes, and his theorizing about the assessment of functional or message complexity, which, in

our view, is comparable to what we define as propositional complexity. The syntactic mode, is characterized by “large chunks under one intonation contour” and “tight coordination”, while pragmatic mode exhibits “small chunks under one intonation contour” and “loose coordination” (Givón,1985, p.140-141). According to Givón (1985):

The syntactic mode [...] then will turn out to be more complex than the pragmatic mode of speech processing by virtue of packing more information - presumable both asserted and presupposed- into the same unit of time. And some syntactic constructions are more complex than others in terms of information/message because they simply pack in more information per time as compared to other, less-syntacticized constructions. (p. 148-149)

In line with these ideas, we could suggest that, by virtue of packing more information, the extended idea units in our definition would be more aligned with the syntactic (i.e., more complex) mode of communication. Following this logic, it could be stated that writing production in our study was more propositionally complex, because it exhibited a higher rate of informationally dense idea units. However, we must be cautious with this interpretation, because the density of information might not be the sole parameter which determines the complexity of an idea unit. Additionally, Givón (1985) indicates that “the “same” unit of information is never really the same when construed in the context of a different organizational schema” (p.152), which means that the information conveyed in the clause-like ideas will not be exactly the same when encoded by means of extended idea units. This renders the interpretation of lower or higher levels of propositional complexity intricate and suggests the existence of different types of propositional complexity.

9.2.2.5. *More versus different propositional complexity*

Thus, rather than interpreting our findings in quantitative terms and deciding which mode exhibited more or less propositional complexity, we would like to suggest that speech and writing display different types of propositional complexity. The dynamic nature of spoken production may explain why our speakers conveyed the information through more numerous clause-like idea units, and our writers relied more on informationally dense ideas to express their message. This interpretation is in line with the ideas in functional linguistics, which views each mode of production as “more complex in its own way” (Halliday, 1987, p.66). As outlined in the introductory part, Halliday (1987) distinguishes between *synoptic* and *dynamic* views. In Halliday’s view, speech is associated with a dynamic mode, which encodes information primarily by means of verbs and verbal constructions, while writing represents a synoptic view, which relies on different means of meaning-making, such as nominalization. The relevant idea here is that writing and speech exploit different features of the linguistic system to achieve complexity. Accordingly, meaning-making (and, as we would like to suggest, information packaging) is highly idiosyncratic in oral and written modes of production. In line with these arguments, we consider that writing and speech construct propositional complexity in different ways, as it was evidenced by the findings in our study.

9.2.3 *Similar accuracy in speech and writing*

For the purposes of this study, we defined accuracy as the degree of conformity of L2 learners’ performance to a norm (i.e., Standard English), and we employed the general measure of the number of errors per 100 words to assess accuracy. We initially advanced a null hypothesis for the differences between writing and speech in terms of

accuracy. This null hypothesis was primarily motivated by mixed findings obtained in previous empirical studies. Thus, some studies reported higher accuracy in writing (e.g., Ellis & Yuan, 2005; Kormos, 2014), while others reported higher accuracy in speech (e.g., Ferrari & Nuzzo, 2009; Granfeldt, 2008). In our experiment, we found that the levels of accuracy were similar in writing and speech. Thus, our findings run counter to previous investigations. Caution, however, is required when comparing findings for accuracy in this study to previous research, as there are certain methodological differences between our experiment and the previous ones.

In our investigation, we compared oral and written production of Spanish/Catalan learners of English, with intermediate L2 proficiency, whose ages ranged from 18 to 40, with the majority being in their 20s. Our participants performed exactly the same argumentative task in speech and in writing. Ferrari and Nuzzo (2009), for example, who found higher accuracy in speech, employed young learners of L2 Italian in their experiment. Because of the differences in the ages of the participants, we can hardly compare findings in Ferrari and Nuzzo (2009) to the findings in this study. Similar to our study, Kormos (2014), Ellis (1987) and Ellis and Yuan (2005) also employed adult learners of English. Although the learners in these studies differed from our participants in terms of L1 (L1 Hungarian in Kormos [2014], mixed L1 backgrounds in Ellis [1987] and L1 Chinese in Ellis & Yuan [2005]), and their levels of proficiency ranged from early intermediate (Ellis, 1987) to upper intermediate (Ellis & Yuan, 2005; Kormos, 2014), we believe that the profile of the learners in these investigations is similar to the profile of the participants in our own study, as they all represent adult learners of English of intermediate L2 proficiency. Similarity in terms of the participants' profile sets the basis for the meaningful comparison of our results and findings in Ellis (1987), Ellis and Yuan (2005) and Kormos (2014).

As mentioned previously, we found similar levels of accuracy in the two modes. These findings run counter to Ellis (1987), Ellis and Yuan (2005) and Kormos (2014) who reported higher accuracy in written production. We would like to suggest that certain methodological differences in terms of the tasks and production measures can account for the discrepancy in findings. Thus, in our study, we employed a between-learner design in which one group of participants performed orally a reasoning task based on the visual cues, while the other group of learners, of a similar profile, performed exactly the same task in writing. Other investigations, in contrast, employed a within-learner design, in which the same participants performed oral and written tasks, which involved different visual cues. There are also differences in the procedure: Ellis (1987), for example, reports that “some assistance was given with spelling, punctuation, and vocabulary” to writers (p.6). In our study, however, the participants did not have access to any external resources. This could also have accounted for the discrepancies in findings in Ellis (1987) and in our experiment.

In terms of tasks, Ellis and Yuan (2005) and Ellis (1987) employed different picture tasks from Heaton (1975) to elicit oral and written performance. In Ellis and Yuan (2005), for example, speakers were required to tell a story about the boys who could not get on the bus because four bigger boys pushed in front of them, while writers had to tell a story about the boy who had lost a parcel and the man returned the parcel to him. Although the pictorial information in the two tasks was different, Ellis and Yuan (2005) argued that the tasks were comparable, as they contained a similar amount of information (6 pictures each), the outcomes of the narratives were equally difficult to predict, there was a similar number of elements, and the situations depicted in the pictures could be considered equally familiar to the learners. Similarly, Kormos (2014) employed different picture cues to elicit writing and speech (the detailed description of

these tasks was provided in the section in which we discussed lexical sophistication). Similar to Ellis and Yuan (2005), Kormos (2014) also argued that oral and written prompts are comparable.

In our opinion, however, oral and written samples in the above mentioned studies might not be comparable when it comes to the analysis of accuracy. Thus, we agree that the parallel task versions employed to elicit oral and written performance in Ellis (1987), Ellis and Yuan (2005) and Kormos (2014) were similar in terms of the overall structure and the cognitive load imposed by the task. If we analyze the task in Ellis and Yuan (2005) in terms of Robinson's (2001) criteria for task complexity, we can conclude that oral and written prompts were roughly equivalent in terms of the imposed cognitive load, as they involved the same number of elements, both tasks were narrations (which means that the imposed reasoning demands were the same), the participants had no prior knowledge of the contents of either story and had access to the pictures while performing both oral and written tasks. The same analysis could be applied to Kormos (2014), whose task versions included similar types of elements, and were designed to elicit similar type of discourse. Oral and written tasks were also performed under the same condition, as the visual prompt was available to the learners during both oral and written production.

We must stress, however, that the similarity of the task versions in terms of the cognitive load they impose does not necessarily mean that oral and written production, elicited by means of these tasks, can be straightforwardly compared in terms of accuracy. Our argument is that differences in the picture cues call for the use of different lexical items. It is plausible that the oral task versions demanded the use of the words that were recently acquired or even unknown to the learners, while the written tasks called for the use of more familiar lexical items, which resulted in higher accuracy

in writing. In contrast to previous investigations, our study avoided this limitation, as the visual cues were exactly the same for speakers and writers. This allows us to claim that the difference in performance in oral and written samples was attributable to the mode of production, and not to any other factor.

The discrepancy in findings for accuracy in our study and in previous investigations could also be attributed to different accuracy measures employed. For the purposes of our study, we used the measure of the number of errors per 100 words. We set the criteria for accuracy in accordance with Standard English, and defined an error as non-target use of a linguistic form or combination of forms. The error, in our definition, subsumed errors in lexis, grammar, pragmatics or any other type of non-target language use. In this way, we strived to ensure global, objective and comparable analysis of oral and written samples in terms of accuracy.

Previous studies, however, took a somewhat different approach to the choice of accuracy measures. Ellis (1987), for example, solely examined the accuracy of the past verb forms. Such a stringent approach to the analysis of accuracy in Ellis (1987), precludes meaningful comparison of Ellis's findings with our results. A more global approach in the analysis of accuracy was taken in Ellis and Yuan (2005) and Kormos (2014) who measured general accuracy via the ratio of error-free clauses, which was calculated relative to the total number of clauses. Additionally, these studies employed a number of specific measures of accuracy, such as the ratio of the error-free relative clauses, the ratio of the error-free verbs and past-tense verbs (Kormos, 2014) or the ratio of error-free verbs (Ellis and Yuan, 2005).

In our study, we did not employ any specific measure of accuracy in order to avoid multicollinearity. However, our study, as well as previous investigations, employed a global, albeit different, measure of accuracy (the number of errors per 100

words), which makes our findings comparable to the results in Ellis and Yuan (2005) and Kormos (2014). We must nevertheless highlight, that the measure of the ratio of the error-free clauses employed in these studies has been criticized for a number of reasons. One of the problematic issues concerns the difficulty in defining the clause and dividing the performance into clauses in a reliable manner. The biggest problem, however, is that a clause, which contains a single error, is treated in the same way as a clause, which contains more than one error. This issue is problematic, as it can obscure the interpretation and analysis of the results. In our study, however, we overcome this limitation, as we adopt a global measure of accuracy, which consists in counting the number of global errors per 100 words. In contrast to a clause, a word represents a smaller unit, which can be identified in discourse by employing word-count software. Although the identification of errors was manual, the use of the software in the identification of the units of analysis (i.e., words) enhanced the reliability of the metric we employed. Also, in order to ensure the reliability in the error analysis, we repeatedly analyzed 35% of data, obtaining an acceptable degree of intra-rater reliability. In sum, we consider that the methodological decisions we took in the procedure and in the analysis allow us to claim that our findings for accuracy are valid, notwithstanding the discrepancy of our results with previous studies.

Thus, in our experiment oral and written samples exhibited similar levels of accuracy. We consider that a theoretical analysis of the sources of errors in performance can help us explain this finding. Thus, from a theoretical point of view, written performance appears, at the first sight, as offering more favorable conditions for the production of more accurate discourse as compared to speech. As discussed in previous chapters, the physical characteristics of the act of production and output in writing favor planning, formulation and monitoring. Writing is slow and self-paced, and its output is

visible and permanent. This means that learners have more time for the careful planning of the message, and they can pay attention to forms during both formulation and monitoring. Visibility of written output can also make it more consequential in the perception of the learner (i.e., the learner perceives that his/her output will be available for scrutiny and evaluation). As a consequence, the learner may pay more attention to the accuracy of output in writing. Higher accuracy in written production can also be theoretically justified by drawing on experimental research in psychology. Thus, it has been well-established that for a person working on a particular task, accuracy will decrease as the person works faster (Luce, 1986). This phenomenon is known as the speed-accuracy trade-off, and it is regarded to be a within-person feature that can be expected for any task (Wickelgren, 1977). Following this logic, it could be suggested that, theoretically, a person's written production would be more accurate than oral production, which is inherently faster. Fast delivery of speech and the evanescence of oral output also mean that the opportunities for the attention to forms in speech are more stringent than in writing. Theoretically, these conditions should lead to higher accuracy in written production. In addition to enhanced focus on forms, writers also have more possibilities to employ explicit and metacognitive knowledge during production. The possibility to apply the learned rules may also favor more target-like production in writing.

However, the favorable conditions for the attention to forms and explicit knowledge use may not be sufficient to ensure accuracy of production. Thus, according to Ellis and Barkhuizen (2005) and Wolfe-Quintero et al. (1998), errors in performance can be attributed to different sources. First of all, the degree of accuracy depends on the state of the interlanguage itself. There may be gaps in the transitional interlanguage, which means that lack of linguistic knowledge (i.e., lack of L2 representations *per se*)

can result in errors. On the other hand, the L2 representations may exist, but they may be weak and not fully automatized. Such weak representations can lose a competition to strong, earlier-established L2 forms, which can be non-standard. This means that a learner may not cope with the pressures of the on-line production and fail to produce an L2 representation accurately, resorting, instead, to an earlier-established non-standard L2 form or even to an L1 representation. A related cause of errors is the insufficient degree of automatization of language production. In hard-pressed on-line production, L2 learners may experience difficulty in accessing forms that have not been fully mastered, which can also result in errors. It is generally accepted that once production is fully automatized, there tend to be less errors (MacKay, 1992). However, Wolfe-Quintero et al. (1998) point out that deviant representations can also be automatized, which may lead to production, which is error-full and, at the same time, impervious to change (i.e., *fossilization* can happen).

In line with these theoretical ideas about the sources of errors, we would like to suggest that writing conditions can help prevent some types of errors, but not others. In our view, the slow pace and the cyclical nature of writing can potentially prevent errors stemming from the weakness of the newly acquired L2 representations as well as errors stemming from the insufficient degree of automatization of language production. The on-line production pressures are of no consequence if written production is not timed, which allows learners to devote their attentional resources to the retrieval and monitoring of the L2 representations. The newly acquired or weak representations may require more time and resources in order to be produced in a target-like manner, as compared to the established representations. However, under untimed condition, writers can afford this expenditure, which enhances the possibility of the target-like production of the non-automatized representations. Favorable writing conditions cannot, however,

prevent fossilized errors or errors which are the result of the gaps in L2 knowledge. Gaps in the interlanguage knowledge mean that the L2 representations are faulty or do not exist. As a result, the possibility to devote more time and attentional resources in writing will be of no consequence for the target-like production of these representations, unless there is access to external sources (e.g., a dictionary). The same would apply to the fossilized errors, which represent well-established non-target L2 representations. Such errors will appear in production because, even with the availability of time, these errors will not be detected by the learner.

In line with these ideas, we would like to suggest that the similar levels of accuracy in speech and writing in our study were due to the fact that the participants in the oral and writing groups were of the similar level of L2 proficiency, which means similarity in the quality and quantity of L2 representations. Although writing conditions may have aided our writers in coping with the production of the non-automatized L2 representations and with the monitoring of their output, there was no remedy for the gaps in linguistic knowledge, which may be the cause of the similar levels of accuracy in the oral and written task performance. On the basis of these findings, we can conclude that, although writing conditions can theoretically aid the target-like production along certain lines, writing does not necessarily ensure higher accuracy, as the errors can also stem from the gaps in linguistic knowledge.

Similar ideas were also expressed in Kormos (2014). As mentioned previously, Kormos's overall conclusion was that written production was more accurate than speech. At the same time, Kormos indicates that the patterns of accuracy varied depending on the type of linguistic structure. Thus, while the levels of accuracy for the past-tense forms were straightforwardly higher in writing, the accuracy of production of the relative clauses showed considerable variation in both modes. Kormos suggests that this

variation is attributable to the fact that the learners have only partially acquired relative clauses (i.e., there were gaps in the knowledge of relative clauses or their use was not automatized). This led Kormos (2014) to conclude that the beneficial effects of writing conditions “may only apply to structures for which students have already developed high accuracy in the form of declarative knowledge” (p.208). In other words, Kormos suggests that the advantage of writing consists in the enhanced focus on form of the learned linguistic structures. Although we agree with this interpretation, we also believe that the benefits of writing for accuracy may not be limited to the enhanced monitoring of the acquired structures, as writing may offer additional advantages for the target-like production of the non-automatized L2 representations. All these benefits, however, can be overridden by the gaps in linguistic knowledge, which can nullify writing advantages for accuracy.

9.2.4 Higher time on task in writing

Finally, as expected, we found higher time on task scores in writing, with time on task operationalized as the time from task onset to task completion. Higher score in written performance was predictable, as the inherent labor-intensive nature and slow pace of writing were likely to result in longer engagement with the language task. Higher time-on-task is precisely one of the arguments used in support of the potential language-learning potential associated with writing tasks (Manchón, 2011; Manchón & Williams, 2016), a prediction for which our data offered empirical evidence. As reviewed in the introduction, the proponents of the language learning potential of writing attribute the L2 learning benefits of writing to a combination of a number of factors. Among the features, which are thought to be conducive to learning, are the visibility and permanence of written text, the problem-solving nature of academic writing tasks, the

metalinguistic reflection during writing and also enhanced time-on-task, which stems from the slowness and laboriousness of writing activity. However, what exactly can time buy for a language learner? According to Manchón (2011) and Manchón and Williams (2016), availability of time during task performance allows to engage in deeper planning and formulation activities, engage in careful monitoring, and also employ all cognitive and knowledge resources to a deeper extent, which might be unavailable during hard-pressed oral production. It is also believed that availability of time allows writers to access both implicit and explicit knowledge (Manchón & Williams, 2016). As an extension of this theoretical ideas, we would like to suggest that time-on-task in writing could be related to the deployment of controlled processing. We build this argument on findings from educational psychology, where time on task has been shown to relate positively to task success in problem-solving situations, which require controlled mental processes (Goldhammer et al., 2014). Undoubtedly, the theoretical supposition we advance needs empirical verification, and it is still an empirical question how exactly different types of knowledge (implicit vs. explicit) and type of processing (controlled vs. automatic) are deployed during written production.

9.2.5 Summary and evaluation of the mode effects on L2 production

Summarizing our findings, we obtained evidence that L2 learners can draw on their interlanguage variably and display different linguistic behaviors depending on the mode (oral or written) in which a task is performed. Our participants spent more time and used more complex syntax as well as more varied vocabulary when they performed the task in writing. Mode also influenced propositional complexity, as speakers produced more idea units overall, but writers employed extended idea units to a higher degree. These findings provided empirical evidence for the theoretical prediction that the oral

and written modes offer different opportunities for language practice and development, and, consequently, mode can constitute a task design feature that may contribute to the development of distinct L2 competencies (see also Tavakoli, 2014).

The inherent emphasis on fluency, for example, could be interpreted as pointing to oral tasks as unique vehicles to train L2 learners to mobilize their interlanguage system to communicate meanings in real time. Sustained practice with oral tasks could thus be assumed to lead to the improvement of learners' access to and control over their existing interlanguage over time and facilitate the proceduralization of their declarative knowledge (DeKeyser, 2007; Towell, Hawkins, & Bazergui, 1996). Practice along oral tasks can be related to the development of automaticity of processing. According to Gatbonton and Segalowitz (1988), automaticity is related to fluency. The construct of fluency is multicomponential and it minimally involves (1) the skills concerned with the selection of utterances (knowing what to say, to whom and when) and (2) skills concerned with the actual production of the utterances (smooth and rapid production without hesitation and pauses). The two types of fluency skills are very distinct: the learner can be fluent in the selection of utterances, but still unable to produce them smoothly. Gatbonton and Segalowitz (1988) relate this second type of fluency skills to automaticity, which they define as "the speed and ease of handling utterances" (p.474). Given that the definition of automaticity includes the component of velocity, we would like to suggest that the development of automaticity is directly related to the practice of oral tasks (see also DeBot, 1996). Automaticity is important because it signifies progress in L2 development. On the one hand, automaticity of performance entails reduction in the consumption of attentional resources, which means that greater amount of processing capacity will be available for other aspects of the activity (Segalowitz & Gatbonton, 1995). Moreover, when a given component becomes automatized it

becomes modularized (Fodor, 1963), which means that this aspect of performance will be carried out more accurately and efficiently, as there will be no interference from other processes (Stanovich, 1991). Segalowitz and Gatbonton (1995), however, propose that automatic processing is not restricted to faster or otherwise quantitatively improved processing of information, but it also entails qualitative changes in the underlying system: “when performance has become relatively more automatized, the manner by the which the underlying mechanisms operate changes, and these changes reflect restructuring” (p.138).

To develop automaticity, practice is necessary. As mentioned above, we consider that oral tasks provide a perfect vehicle for this purpose. Moreover, Segalowitz and Gatbonton (1988, 1995) and Gatbonton and Segalowitz (2005) posit that the context of genuine communication (which TBLT precisely provides) is the best condition to develop automaticity. In this argument, the conditions of real-life meaningful communication are ideal to promote *creative automatization*, which entails the mastering of utterances, rather than structures. In this regard, Segalowitz and Gatbonton (1988) state that “what is needed is an activity designed to enable learners to practice (repeat) many tokens of target sentences while they are engaged in real communication” (p.484). According to Segalowitz and Gatbonton (1988), to effectively promote creative automatization, communicative activities must not only be genuine, but they also have to possess psychological authenticity. This means that an activity, intended to promote automaticity, has to pose the same psychological pressures felt by people engaged in real communication.

In line with these arguments, we can conclude that TBLT tasks in the oral mode constitute a perfect means to produce quantitative (speeding-up of processing) and qualitative (restructuring) developmental changes in underlying L2 system. Although

we do not deny the possibility that written tasks can also contribute to the development of automaticity (see Gilabert et al., 2016), our position is that the tasks in the oral mode are specifically appropriate for this purpose, as they pose on-line pressures and demand speed in production. Future research is required to determine the role of writing tasks in the development of automaticity.

In relation to the language-learning potential of writing tasks in particular, the unique characteristics of individual writing, such as the inherent problem-solving nature of certain writing tasks, the availability of time, and the visibility of written output are variables that can potentially foster L2 development through deeper linguistic processing, enhanced focus on form, hypothesis testing, and self-assessment of one's own linguistic options (Manchón, 2011; Manchón & Williams, 2016; Williams, 2012). Therefore, it is worth reiterating that the results obtained in our study provided confirmatory evidence of the purported language-learning potential of writing, with the written mode emerging as providing more opportunities for exploiting certain facets of linguistic complexity (e.g., subordination, lexical diversity) and creating more complex, informationally dense idea units. According to Skehan and Foster (1999), complexity is of paramount importance in L2 development, as it is also considered "to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system" (p.97). As mentioned above, the process of restructuring entails qualitative changes in the underlying L2 system (Segalowitz & Gatbonton, 1995). Restructuring in SLA context is also defined as the process by which the interlanguage becomes more elaborate and structured (McLaughlin, 1990), more efficient in communication (Cheng, 1985), and more consistent with the input data, and more native-like (Cook, 1994). As mentioned above, Foster and Skehan (1999) and also Skehan (1996) directly relate restructuring with the complexity of production. This entails that the use of complex

linguistic structures, which may be at the limits of the learner interlanguage, can cause qualitative reorganization of the internal L2 system. Such qualitative reorganization would signify progress in the development of interlanguage. Skehan (1996) also highlights the importance of complex performance, as it is believed to be more suitable to communicate elaborated messages efficiently and with precision. In this regard, Skehan (1996) also highlights that because of the pressures during on-line communication, there might not be enough time for restructuring to occur, “since processing resources have to be excessively committed to achieve certain communicative outcomes” (p.48). In line with these ideas, we would also like to suggest that more relaxed language production conditions in writing would provide optimal conditions for the deployment of what Skehan (1998) terms a rule-based system. This system is posited as generative and restructurable; in other words, it is related to the qualitative changes in the underlying system. Conversely, on-line pressures in speech might induce learners to rely more on the exemplar-based system, which is less resource-demanding.

9.3 Research question 2: The effects of task complexity in speech and writing

Our second research question asked whether an increase in task demands would differentially influence L2 oral and written production. The relevance of this research question was accounted by the fact that the existing cognitive models of task-based language teaching, such as the Trade-off Hypothesis (Skehan, 1998, 2009) or the Cognition Hypothesis (Robinson, 2001, 2011a) do not account for the idiosyncrasy of language production in oral and written modes. The predictions that these models advance are, by default, intended for oral task performance and it is still an empirical question whether or not Skehan`s and Robinson`s theorizing can apply equally well to

written production. In this dissertation we compare the way in which the predictions of Robinson's Cognition Hypothesis apply to L2 oral and written performance. We have specifically focused on the factor of reasoning demands. According to Robinson, increase in task complexity along the resource-directing factor of reasoning demands is expected to be detrimental to fluency, leading, at the same time, to greater accuracy and complexity of L2 production. This tenet has been tested in both oral and written production with mixed results. Of specific relevance for this dissertation are those empirical studies which have compared the effects of task complexity within the same experimental design (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Tavakoli, 2014). The results of these studies are mixed, which precludes from making definitive conclusions about the effects of task complexity on different modes of production.

With an aim to contribute to the knowledge about the nature of task complexity effects in different modes of production, we formulated the following research question:

Research Question 2: Does task complexity affect lexical complexity, syntactic complexity, propositional complexity, accuracy, and time on task in L2 oral performance in the same way as in L2 written performance?

Although previous studies have produced mixed results, there seems to be an indication that increases in task complexity play out somewhat differently in writing and speech. This allowed us to formulate a non-directional hypothesis for our second research question, namely:

Hypothesis Research Question 2: We hypothesized that task complexity, operationalized as +/- reasoning demands, would produce different effects in L2 oral and written task performance.

Overall, our findings confirmed this prediction: We found that task complexity affected speech and writing differently on a number of production dimensions, but not on others. More precisely:

- (1) We obtained different effects of task complexity, especially regarding ratio of extended idea units, accuracy and time on task. We found that writers produced more extended idea units, spent more time on task and were more accurate in the complex task condition as compared to the simple condition. At the same time, there were no changes in these dimensions in oral performance.
- (2) We found similar effects of task complexity as both writer and speakers produced longer AS units, more sophisticated words, and more ideas in the complex task, compared to the simple condition. At the same time, task complexity did not produce detectable effects on subordination, phrasal complexity, lexical diversity and length of idea units neither in writing nor in speech.

Below, we discuss our findings for each production area separately.

9.3.1 Similar effects on linguistic complexity in speech and writing

According to Robinson (2001, 2011a), increases in task complexity along such variables as reasoning demands would increase the level of linguistic complexity of L2 production. This prediction is based on the recognized interactive relationship between

conceptual and linguistic representations. Thus, the need to express more complex thoughts is expected to call for more complex linguistic encoding, with the resulting increment in the indices of linguistic complexity. This assumption was only partially borne out in our study, as out of the five measures of linguistic complexity only two showed a higher value in the complex task. Thus, we found that both writers' and speakers' productions were more syntactically complex overall and exhibited more sophisticated vocabulary in the complex version of the task, compared to the simple version. At the same time, certain aspects of linguistic complexity (such as subordination, noun-phrase complexity, and lexical diversity) were not affected by task complexity either in writing or in speech. Notably, the pattern of results was identical for oral and written performance, demonstrating that the effects of cognitive task demands on linguistic complexity seem to be of rather similar nature in the two modes.

9.3.1.1 *Similar levels of lexical diversity, but higher lexical sophistication in the complex task for writers and speakers*

For lexical complexity, we found that increases in task demands affected the sub-dimension of lexical sophistication, but not lexical diversity, and this pattern was similar for both writers and speakers. Thus, we found that both writers and speakers employed more lexically sophisticated language in the complex task, but there were no changes in the diversity of vocabulary. Absence of task complexity effects on lexical diversity was also reported in Kuiken and Vedder (2011) for speech and in Kuiken and Vedder (2008, 2011), Vasylets and Gilabert (2014) for writing. Similar to our results, Vasylets and Gilabert (2014) also found higher lexical sophistication in the complex writing task.

The explanation of our findings might lie in the nature of the experimental task, which was a reasoning task that required to provide arguments and to justify actions. It

could be suggested that, in the perception of the learners, the successful achievement of the communicative goal of the task did not depend so much on the variety of the words employed, but on the quality of the words. Thus, it is plausible that the increased reasoning demands induced our participants to use the infrequent lexical items which could serve to reinforce the persuasiveness of their arguments, hence the higher level of lexical sophistication in the complex task in speech and writing.

9.3.1.2 *Higher overall structural complexity, but no changes in subordination and noun-phrase complexity in the two modes*

In terms of the comparison of our findings to previous research, higher values for overall linguistic complexity, as gauged by a length-based measure, was also reported for oral production in Tavakoli (2014) and for written production in Vasylets and Gilabert (2014). Higher indices for overall structural complexity are in line with Robinson's theoretical prediction. However, we also found that the level of subordination and noun-phrase complexity were not affected either in writing or in speech. Similar results for subordination were also found in other studies: thus, absence of task complexity effects on subordination was reported by Fukuta and Yamishita (2015) for speech and by Kuiken and Vedder (2008) for writing. Absence of the effects on subordination in our experiment was, at the first sight, surprising, as we expected that increases in task demands would induce learners to complexify their linguistic production through different means, including subordination. Various reasons, however, can account for the absence of task complexity effects on this sub-dimension of linguistic complexity. On the one hand, the subordination index we employed (S-nodes per AS unit) might not be sensitive enough to detect the subtle differences between

simple and complex task condition. Another reason could be that the levels of subordination did not change because the use of subordinate structures was specifically relevant in the experimental task. Although it might seem counterintuitive, this explanation has its logic: the experimental task we used was an argumentative task, which required to provide a solution to a problem and to justify a sequence of actions to take. This means that the task required the learners to employ their reasoning abilities, involving such operations as analysis and comparison of information, juxtaposition of ideas and establishment of the cause-and-effect links. Theoretically, the linguistic expression of all these operations would require the use of subordinate constructions (Matthiessen & Thompson, 1998). The specific relevance of the subordinate structures might have encouraged learners to employ subordination intensively in both simple and complex tasks, which could explain similar levels of subordination in the two conditions. We must mention, however, that the descriptive values for subordination were slightly higher in the complex condition for both writers and speakers. Thus, it is also plausible that a relatively small number of participants precluded from obtaining potential statistically significant results for subordination.

A similar line of reasoning about the use of subordination was also advanced by Kormos (2014) in a study in which she compared oral and written task performance. The author suggested that the amount of subordination in performance could be determined by the particular task characteristics, other than cognitive task complexity or the mode of performance. Although the expectation might be to find more complex syntax (i.e., higher subordination) in writing, Kormos found that there were no differences between writing and speech, and she argued that the explanation could lie in the characteristics of the experimental task, which belonged to a narrative genre. In Kormos's argument, the completion of a narrative task requires characterization of the

participants and event elaboration. As all these features are frequently expressed by means of subordinate clauses, both writers and speakers considered subordination as specifically relevant and employed it intensively in the production – hence, the similar levels of subordination in the two modes, in spite of the alleged proclivity of the written mode to display more complex syntax. Although highly tentative, these arguments might be instrumental to provide explanation for the absence of task complexity effect on production in some studies. It is plausible that the shape that L2 performance takes does not only depend on task complexity, but also on the interplay of task complexity with other factors such as genre of the task or the level of proficiency of L2 learners. Undoubtedly, more research with different genres of tasks and types of populations is required to elucidate this issue.

Concerning noun-phrase complexity, we found similar levels in the simple and complex task in the two modes. The same finding was obtained in Vasylets and Gilabert (2014) for written production. As mentioned in the previous section, the index of noun-phrase complexity did not distinguish between oral and written samples either. One of the possibilities could be that the measure we employed was not sensitive enough to detect subtle fluctuations in the levels of noun-phrase complexity in the simple and complex condition. To assess noun-phrase complexity, we used the mean number of modifiers per noun phrase which we obtained by means of the Coh-Metrix 3.0 program. It must be pointed out that some researchers have cautioned against the overreliance on the software in the assessment of complexity of production. In particular, Bulté and Housen (2014) stated that:

In spite of the considerable advances being made in recent years, the algorithms used by today's linguistic software tools (e.g., Coh-Metrix, L2 Syntactic Complexity Analyzer) may still be too rigid to accurately and fully identify,

segment, and parse the L2 learner productions (especially productions from the earlier stages of acquisition), which may create measurement noise. (p.48)

Another possible explanation to our findings could be related to our participants' level of proficiency in their L2. As mentioned in the previous sections, the ability to achieve complexity by means of phrasal elaboration is characteristic of the advanced stages of L2 development. It is plausible that both writers and speakers in our study had not mastered phrasal elaboration enough to draw on this resource when producing discourse, hence the similar levels of noun-phrase complexity in the simple and complex task.

Summarizing our results for linguistic complexity, we found that both writers' and speakers' productions were more syntactically complex overall and exhibited more sophisticated vocabulary in the complex version of the task, compared to the simple version. At the same time, certain aspects of linguistic complexity (such as subordination, noun-phrase complexity, and lexical diversity) were not affected by task complexity either in writing or in speech. These results provide evidence of the similar effects exerted by task complexity in speech and writing in the area of linguistic complexity. In contrast, other dimensions of production revealed differences between the two modes in terms of the susceptibility to the increases in cognitive task complexity.

9.3.2 More idea units in both complex oral and complex writing tasks, but higher ratio of extended ideas only in the complex writing task

As mentioned previously, propositional complexity refers to the encoding of the semantic content by means of idea units. In his theorizing, Robinson (2001, 2011a) does not advance any predictions concerning the effects of task complexity on idea units. However, we believed that idea units would be specifically revealing in the exploration of the effects of task complexity on L2 production. In his most recent publication on the Cognition Hypothesis, Robinson (2011a) explicitly relates the effects of task complexity to the conceptualization stage of production (Levelt, 1989). On the other hand, idea units represent a notion that is also believed to be linked to the conceptualization stage of production (Butterfield, 1975; Chafe, 1980, 1982). Along the similar line, Ellis and Barkhuizen (2005), in their book on the analysis of L2 performance, posited that the analysis of production for idea units provided the degree to which learner engaged into conceptualization during task performance. Based on the purported direct connections of both constructs to the stage of conceptualization, we believed that the exploration of the way in which task complexity might affect propositional complexity, operationalized as idea units, would be particularly relevant.

We mentioned previously that Robinson's Cognition Hypothesis does not advance any assumptions concerning the effects of task complexity on the quantity or quality of idea units. It would be logical to hypothesize, however, that the need to convey more complex concepts would entail the use of a higher number of ideas overall or would increase the ratio of the informationally dense ideas, or both. As we have argued in previous sections, both of such changes could be interpreted as an increase in propositional complexity. This prediction was fully borne out in the written production in this study. Thus, we found higher number of ideas overall and a higher ratio of the extended ideas in complex writing task as compared to the simple writing task. Task

complexity also affected speech, as speakers produced more ideas overall in the complex task. Scarcity of previous research on idea units in L2 limits the comparison of our findings to other studies. Some point of comparison is, however, provided in Choong (2011), who explored the effects of reasoning demands on syntactic and propositional complexity in oral production. Choong's operationalization of idea units is broadly comparable to our coding guidelines, although he does not make a distinction between extended and non-extended idea units. Similar to our study, Choong found that speakers produced more idea units in the complex task, which was explained by the need to convey more information in the complex task.

In sum, we found that increases in task complexity resulted in higher rates of propositional complexity in both modes, but the effects were more prominent in written production. This finding provides evidence that task complexity has an impact on the way content is conveyed in oral and written performance. A second observation that can be made is that the effect of increased task complexity is indeed intensification of propositional complexity, as we have previously suggested. We also obtained evidence that task complexity affected writing and speech in somewhat different ways, with propositional complexity in writing being more susceptible to the effects of task demands. A notable finding is that, in speech, task complexity affected the overall number of idea units, but not the ratio of the extended ideas. On the other hand, in writing, the effects were notable on both measures. First of all, we believe that, in a way, this finding provides indirect evidence to our supposition about differences in speech and writing in terms of the preferred means to construct propositional meaning. We argued that the dynamic nature of oral production might induce speakers to convey information primarily through clause-like ideas, while more relaxed writing conditions would allow writers to compose more informationally dense ideas. For this reason, we

can see the logic behind our finding that task complexity affected the preferred means to achieve propositional complexity in each mode: speakers increased the number of ideas in the complex task, while the ratio of the extended idea units increased only in writing. Additionally, writers increased the number of idea units in the complex task. This greater sensitivity of written production could be attributed to the lower level of pressure in writing as compared to speech. Because of the in-built planning opportunities that characterize writing (see Manchón, 2014), writers can fully attend to the induced need to enhance the efforts at conveying the propositional content of a complex task. In sum, we can conclude that the use of idea units did allow us to gain new insights into the effect of task complexity on L2 production. Two findings were specifically notable: (1) increases in task demands enhance propositional complexity of L2 production, and (2) propositional complexity in written production was affected to a greater extent as compared to speech.

Finally, we also found that the length of idea units was the same in the simple and complex condition for both writers and speakers. This finding also resonates with our results of a similar length of ideas in the two modes. This shows that in our study in particular, the length of idea units did not show up as a specifically sensitive measure to distinguish between modes of production or between simple and complex tasks. Undoubtedly, more research is required to explore if the length of idea units varies depending on the type of task or the characteristics of participants.

9.3.3 Higher accuracy in complex writing task, but no differences in speech

Concerning accuracy, the Cognition Hypothesis (Robinson, 2003, 2011a) predicts that a complex reasoning task, like any other task with enhanced resource-directing demands, would exhibit higher levels of accuracy than a counterpart simple task. Robinson

justifies this prediction by drawing on the ideas of Talmy (2000) about different ways a concept can be encoded grammatically. Thus, a language user can resort to the open-class lexical sub-system or to the closed-class grammatical sub-system. Languages, however, differ in the way conceptual domains are grammaticized (i.e., encoded). Because of the limited inventory of L2 form-meaning mappings, and often under the L1 influence, learners may grammaticize the concepts in an L2 in a way which is divergent from the target language. In Robinson's view, gradually increasing the cognitive and conceptual demands of L2 tasks has the potential of drawing learners' attention to the similar and divergent ways concepts are encoded in L1 and L2, leading, eventually, to gains in accurate grammaticization of meanings in L2. In our interpretation, these gains in accuracy might be expected in the long term after practicing numerous task sequences, which will gradually increase learners' awareness of the idiosyncrasy of the encoding concepts in an L2. Such heightened awareness coupled with the availability of necessary linguistic means would lead to higher accuracy in the encoding of meaning in L2.

We would like to suggest that, at the level of immediate performance, increased task demands will increase the levels of accuracy primarily due to the enhanced focus on form (i.e., enhanced monitoring). As mentioned in earlier sections, in his most recent reconceptualization of the Cognition Hypothesis, Robinson (2011a) connects task complexity effects to the stages of speech production: "in terms of Level's (1989) model of speech production, increasing the conceptual demands of tasks (naturally) leads to greater effort at conceptualization, and "macroplanning" at the stage of message preparation" (p.15). Complex message would subsequently call for more complex language to encode it, which would consequently affect the working of the stage of linguistic formulation. As an extension of these ideas, we would like to suggest that

increased effort at conceptualization and formulation would also transmit to monitoring, with the subsequent positive effects for the accuracy of production. Concerning the results in our own study, we found that writers were more accurate in the complex task, while there were no changes for speakers. These findings could be interpreted as suggesting that, in writing, increased reasoning demands induced a higher effort at monitoring with an observable reduction in the errors of production. In speech, however, increased task complexity did not produce observable effects, as measured by the number of production errors.

In terms of the comparison of findings with previous research, accuracy enhancement in complex writing tasks was also found in Kuiken and Vedder (2008, 2011). However, there were no effects for accuracy in written production in Vasylets and Gilabert (2014), Kormos and Trebits (2012), or Salimi and Dadashpour (2012), and Ruiz-Funez (2015) found lower accuracy in the complex task. Caution, however, is required when comparing the findings in some of these studies with our research. Ruiz-Funez (2015), for example, employed only descriptive statistics; thus, the results can only be considered as tentative. Differences between our findings and Vasylets and Gilabert (2014) could be explained by the potential differences in the level of proficiency of the participants in the two studies, and also by the different types of tasks employed: while in our experiment we employed a visual cue with short instructions in L2, Vasylets and Gilabert (2014) used a task which provided a greater amount of textual information in the L2. In Vasylets and Gilabert (2014), the textual prompts were practically identical in the simple and complex versions (except for a few phrases); during task performance, the learners had access to the prompts and, it is feasible, that this could affect their levels of accuracy in the simple and complex tasks. Similar to our experiment, Kormos and Trebits (2012) also used visual cues, but the visual prompts

were different in simple and complex conditions. Additionally, as discussed in previous sections, one must be cautious when comparing accuracy in discourse samples elicited by means of divergent visual prompts given that the differences in the pictorial images will call for different lexical items, which might be more or less familiar to the learners.

It is specifically interesting to compare our findings for higher accuracy in writing with those obtained in a study by Salimi and Dadashpour (2012) that also employed the “Fire – chief” task. The findings in both studies are different as Salimi and Dadashpour (2012) found no differences in accuracy between simple and complex writing tasks. This divergency in findings can be attributed to the differences in the characteristics of the participants in the two studies, and also to the differences in the measures of accuracy. Thus, while in our study we employed the general measure of the mean number of errors per 100 words, Salimi and Dadashpour (2012) used the percentage of the error-free T-units. As discussed previously, this measure might be problematic because it can obscure the real number of the errors in the discourse sample (e.g., a T-unit with one error is treated in the same way as a T-unit with two or more errors).

Concerning oral production, our findings coincide with those in Shiau and Adams (2011), who also reported no substantial differences between simple and complex conditions. This comparison is specifically relevant because Shiau and Adams (2011) also used the “Fire – chief” task in their experiment. Divergent findings were reported in the study by Fukuta and Yamashita (2015), who obtained higher accuracy in the complex oral task. This study, however, has certain methodological problems, namely, different picture cues were used in the simple and complex tasks, and the percentage of error-free AS units was used as a measure of accuracy. As discussed

earlier, these methodological decisions might not be the most appropriate in the exploration of accuracy in production.

Concerning the explanation of our findings of higher accuracy in the complex writing task, with no changes in speech, we would like to suggest that the differences in pressures in writing and speech could account for these results. As we mentioned in the introductory chapters, one of the most important differences between writing and speech is the level of the production pressures, which has a direct bearing on the way planning, formulation and monitoring are implemented. Writing is slower and self-paced, and its output is visible, while speech is fast and evanescent. In contrast to hard-pressed speakers, writers can afford to plan, formulate and monitor in a more careful and relaxed manner. It could be tentatively suggested that increased cognitive complexity of the task does create an impetus for enhanced focus on form, observable in the decrease in errors. However, under the on-line pressure of oral production, learners might not be sensitive to this impulse. Further evidence for this claim comes from Gilabert (2007), who analyzed the same oral corpus for self-repairs, which is a measure of accuracy that denotes both attention to form and an attempt at being accurate. Gilabert (2007) found that there was a similar number of errors and repairs in the simple and complex conditions, but there were two repair measures (all repairs per 100 words and corrected ratio repaired/unrepaired errors), which pointed in the direction of the advanced hypothesis, which predicted fewer errors and higher attempt at repairs in the complex task. In our view, these findings provide evidence that the stimulus for the focus of form in the complex condition does exist: however, it might not be strong enough to overcome the on-line pressure of oral production.

The conditions in writing, however, make this focus on form stimulus palpable and effective. In other words, enhanced complexity will prompt writers to pay more

attention to form, and writers will feel this incitement towards accuracy and will be able to attend to it. It could be tentatively suggested that a major increase in the accuracy of production under task complexity effects would be attributed to the reduction in errors stemming from the lack of automatization or weakness of L2 representations. As we argued previously, the errors stemming from the gaps in linguistic knowledge or the fossilized errors might not be amenable to correction because of the absence in the interlanguage of the target-like alternative. On the other hand, the weak and non-automatized L2 presentations (which might be perfectly target-like) are poor contenders as compared to the well-established linguistic representations (which are not necessarily target-like). An additional impetus to focus on form might reinforce the winning possibilities of the weak, but target-like, representations, with a resulting increment in the accuracy of production. These theoretical suppositions are tentative, and it is a matter of future research to determine which types of errors are reduced, if the complexity of task is incremented.

In sum, we believe that our study has provided evidence that increase in task complexity induces enhanced focus on form. While the on-line pressures in speech might nullify this towards-accuracy stimulus, writing conditions allow to attend to it, with a consequent reduction in errors.

9.3.4 Higher time on task in complex writing task, but no differences in speech

For the time-based measure of fluency, the Cognition Hypothesis (Robinson, 2001, 2011a) predicts lower values in the complex task as compared to the simple task. In other words, Robinson posits that, in the complex task condition, the procedural skill of fluency will be slowed down, while the conceptual and linguistic work will be boosted. In a way, this tenet can also be viewed as a kind of a trade-off between procedural-

executive and conceptual-linguistic mechanisms, although Robinson never explicitly posits a cause-and-effect relationship between increases in accuracy and complexity, on the one hand, and slowing down of fluency, on the other hand. Robinson's prediction for lower fluency in the complex task has been borne out both in oral production (Fukuta & Yamashita, 2015) and in writing (Ruiz-Funez, 2015). As mentioned previously, findings in Ruiz-Funez (2015) are tentative because only descriptive statistical values are reported. Salimi and Dadashpour (2012), on the other hand, reported higher fluency in the complex writing task. However, the authors employed the length of T-unit as a measure of fluency, and, according to Norris and Ortega (2009) this measure also represents a measure of linguistic complexity. For this reason, the findings for fluency in Salimi and Dadashpour (2012) should be interpreted with caution.

For the purposes of this dissertation, rather than using a measure of speed fluency (e.g., words per minute), we decided to employ another time-based metric, which is total time on task. In terms of the predictions for the effects of task complexity on time on task, it could be tentatively suggested that increased cognitive demands would cause an increase in the total time spent on task performance. The findings in our study confirm this prediction, but only in the written modality. Thus, we found that writers spent more time on the complex written task, whereas speakers spent the same amount of time when performing either the simple or complex task version. Higher time on task in complex written production could be interpreted as suggesting that the involvement in more intensive conceptual and linguistic activities in the complex condition induced writers to engage with the task for a longer period of time. On the other hand, similar time on tasks indices in the simple and complex tasks for speakers may also reflect lower intensity of task complexity effects in oral production. As

mentioned previously, our results showed that speech was less sensitive to the increases in task demands: except for three complexity dimensions, there were no changes from simple to complex task in the oral performance. This might be taken as an indication that, for speakers, the intensity of the conceptual and linguistic work in the two conditions was roughly the same, which was also reflected in the similar time on task values.

9.3.5 Global interpretation of the findings

In sum, we found that increases in task demands produced a number of effects on L2 performance, and these effects were largely in line with the predictions of the Cognition Hypothesis (Robinson, 2001, 2011a). We also found that, overall, the nature of task complexity effects were rather different in writing and speech, with the written mode showing a higher susceptibility to the enhancement in task demands. While task complexity affected the area of linguistic complexity in the similar way in the two modes, the effects were manifestly different in the realms of propositional complexity, accuracy and time on task. On the basis of these findings, two major conclusions can be made. First, we obtained evidence that task complexity constitutes an important task variable that can influence L2 performance and subsequent learning in both oral and written modes. Our central finding, however, was that task complexity does not operate in isolation. Rather, its effects on L2 production might be modified by the mode (oral or written) in which a task is performed. Crucially, changes in written production were more numerous than those in oral production and showed a better fit to the Cognition Hypothesis (cf. Robinson, 2011a) predictions, which is particularly intriguing given that the Cognition Hypothesis was implicitly intended to account for oral production.

Importantly, writers exhibited an increase in accuracy and complexity of production in the more demanding task, which is a central tenet of the Cognition Hypothesis.

It must be noted, however, that the studies which compared the effects of task complexity on oral and written production have arrived to somewhat different conclusions. Although previous research also reports that, overall, there seem to be an indication of the differential effects of task complexity in writing and speech (which is also corroborated in our study), none of the previous investigations reports higher sensitivity of writing to task complexity effects. Tavakoli (2014), for example, compared the effects of task complexity on syntactic complexity in the two modes and found higher values in the complex task for both writers and speakers. However, the results were statistically significant only for speakers, which provided evidence of a higher sensitivity of speech to the effects of task demands. Tavakoli (2014) tentatively attributed the lack of significant differences between simple and complex writing tasks to the fact that writers “take a “writerly” stance, working toward constructing a more elaborated narrative regardless of the nature of the task” (p.228). In other words, there is a tendency in written production to construct syntactically complex discourse irrespective of task demands. We must note, however, that Tavakoli only assessed overall structural complexity by means of the length of T-unit and subordination by means of the ratio of clauses to T-unit. Because of such a stringent approach, it is difficult to derive any definitive conclusions about the nature of the effects of task complexity based on Tavakoli’s data. In our study, for example, the effects of task complexity played out differently in the two modes depending on the area of production in question, and linguistic complexity was precisely the area where the effects of task complexity were similar in the two modes. This shows the importance of assessing L2

production multi-dimensionally, because a too narrow approach can preclude from seeing the global picture.

The conclusions in our study are also somewhat different from Kormos and Trebits (2012), who found that written production appeared to be more susceptible to the impact of cognitive task demands along the dimension of syntactic complexity, while speech was more sensitive in the areas of accuracy and lexical variety. It must be noted, however, that Kormos and Trebits (2012) employed different visual prompts to elicit oral and written performance. As we have repeatedly argued, the use of different picture cues might have induced participants to make different linguistic choices especially in terms of lexis, with the consequent effects for the accuracy and lexical complexity. In our study, however, we employed the same visual prompts to elicit oral and written production and, therefore, it might be problematic to make straightforward comparison of our findings and conclusions with those in Kormos and Trebits (2012).

Finally, Kuiken and Vedder (2011) reported that syntactic complexity was affected somewhat differently in writing and speech, with lower values for subordination in the complex oral task and absence of changes in writing. In spite of these differential effects, Kuiken and Vedder (2011) concluded that the influence of task complexity on linguistic performance was not significantly constrained by the mode of production, and, thus, “it does not seem necessary to include “mode” in Robinson’s Triadic Componential Framework” (p.103). We find this conclusion to be somewhat problematic because our study, in addition to previous research, has shown marked differences in the effects of task demands on writing and speech. We could tentatively attribute the discrepancy between Kuiken and Vedder’s (2011) findings and the results in this dissertation to the differences in the characteristics of the participants or the tasks employed. In the two studies, the participants differed in terms of L1 and L2: while our

participants were L1 Spanish/Catalan learners of L2 English, with an intermediate level of L2 proficiency, the participants in Kuiken and Vedder's (2011) study were L1 Dutch learners of L2 Italian with low to intermediate levels of proficiency. There were also differences in the tasks employed in both studies: whereas we used visual cues with minimal text, Kuiken and Vedder (2011) used a textual prompt. These divergent findings clearly show that more research is needed before the definitive conclusions concerning the effects of task complexity in the two modes can be reached.

Turning to the findings in this dissertation, the question that arises is why, in contrast to writing, our participants' oral performance appeared to be less sensitive to the increases in task demands. A plausible explanation might be that the beneficial effects of increased task complexity are weakened or neutralized if the task is very demanding along resource-dispersing dimensions (Robinson & Gilabert, 2007). One of the prominent resource-dispersing variables, whose importance in language production has been widely acknowledged, is the availability of planning time (Ellis, 2005). Speakers are rather constrained in terms of planning due to time pressure; thus, it is plausible that the inherent constraint imposed on the resource-dispersing variable of planning constitutes the cause (or one of the causes) that weakened the effects of increased task complexity on oral performance found in our study. By the same token, if high resource-dispersing variables can nullify or diminish task complexity effects, low resource-dispersion could enhance them. This facilitative effect was also pointed out by Robinson (2003), who stated that the impact of enhanced resource-directing features (e.g., reasoning demands) might be boosted when the task is simultaneously simpler along one or more resource-dispersing dimensions (e.g., availability of planning).

Built-in planning possibilities constitute a fundamental characteristic of written production (Manchón, 2014). It could therefore be suggested that inherently low

resource-dispersion conditions in terms of planning in the written mode, together with the possibility of implementing other production processes in a more self-determined and recursive manner, can make the written modality a perfect arena for the manifestation of task complexity effects. This, we would suggest, constitutes one of the most revealing findings of our study and one that has important theoretical implications for past and current theorizing on task complexity. More specifically, the validity of a fit-all-modes task complexity model might appear as limited. As a result, we support Tavakoli's (2014) claim that "it becomes difficult to assume that a single model of task complexity can uniformly account for the different cognitive demands of writing and speaking tasks" (p. 234). Following from this, we would also tentatively suggest that cognitive TBLT models (including those reviewed in the introductory section) should be problematized in an attempt to account in full for the specificity of production conditions in different modalities. Consequently, a certain expansion of current theorizing about task complexity in terms of its reformulation as a mode-dependent construct might be appropriate.

9.4 Reconsidering models of task complexity

In our view, the reconsideration of the cognitive TBLT models should, first of all, take into account the distinctive nature of the internal dynamism of the production processes, which are less constrained by time and more recursive in the written mode (Manchón, 2014). Given that task complexity is purported to affect primarily the conceptualization stage, it is crucial, in our view, to consider differences between speech and writing in terms of planning possibilities, which vary substantially across modalities. It must also be taken into account that cognitive resources might be used differently in the two modes given that the cognitive window is believed to be open somewhat wider and for a

longer period of time in the self-paced domain of writing compared to the hard-pressed conditions of speech (Williams, 2012). As a consequence, writers have the opportunity to monitor their production and experience less pressure than speakers to divide their attentional resources between message conceptualization and its linguistic realization during task performance (Kormos & Trebits, 2012). These differences in the management of cognitive resources can potentially mediate the effects of the cognitive load imposed by task design in the two modes. Additionally, we must note that learners' perceptions of the cognitive load of the task might differ in speech and in writing, meaning that the interplay between task complexity variables and learners' ability/affective variables (e.g., motivation, aptitude, etc.) could function differentially in the two modes. In this respect, TBLT theorizing should also account for a wider and somewhat different range of learner variables that may mediate task complexity effects in writing (Manchón, 2014). This expansion should at least include the L2 user's writing expertise, which plays a rather distinctive role in written production compared, for example, to the role of L2 proficiency (Cumming, 1989). The consideration of writing expertise is crucial, as this cognitive ability determines the central process of task conceptualization and, as a consequence, has a bearing on the quality of the writing processes. We believe that the consideration of these issues would help refine task complexity theorizing and would constitute a step forward in the quest for answers to the theoretical and empirical questions about task-based performance and learning.

CHAPTER 10. CONCLUSION, IMPLICATIONS, LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In this dissertation we have explored the effects of mode and task complexity on L2 performance. Our first finding was that L2 linguistic performance on the same task differed depending on the mode (oral vs. written) in which the task was performed. Thus, when performing the task in writing the learners spent more time on task, and used more varied vocabulary and more complex syntax. Mode also influenced the way learners encoded the propositional content of the task, with speakers using more idea units and writers employing more extended idea units. These findings provide evidence that mode has an influence on language use and, potentially, development, although further empirical evidence is needed before coming to firm conclusions in this respect.

We believe that our findings about the effects of mode specifically contribute to the disciplinary debates on (1) the distinctive nature of the language learning potential of L2 writing and oral tasks and (2) on the need for a more integrative TBLT research agenda. The findings of this dissertation provided evidence that oral and written modes contribute to L2 learning in unique but complementary ways. Therefore, we concur with Bygate et al. (2014) in their claim that “an adequate version of TBLT must enable its learners to engage with it through both oral and written media” (p. x). Undoubtedly, a more mode-balanced TBLT curriculum would create an environment in which learners can capitalize on the assets of both speech and writing, with the consequent benefits of more effective and balanced language-learning opportunities. This idea of balancing and combining oral and written modes also resonates with Skehan’s (1998) ideas of balance in the development of complexity, accuracy and fluency of production. As we outlined

in the introductory part, Skehan (1998) suggests that tasks in the language classroom should be administered in such a way, so as to ensure a balanced development in the three main areas of production. Consistent prioritization of one of the performance dimensions can have negative consequences. For example, overemphasizing fluency can induce learners to rely mainly on the exemplar-based memory system, which can also result in the fossilization of errors. On the other hand, prioritization of complexity, for example, might lead to the development of the interlanguage system, which may be characterized by “cutting-edge” language, but which may be full of inaccurate and weak L2 representations. In Skehan’s view, the goal of instruction is to ensure balance in the development of different performance areas. And this balance can be achieved through a judicious combination of different task types, which would timely prioritize one performance dimension or the other.

Combination and balancing of oral and written tasks will also help achieve balance in the deployment of the rule-based and exemplar-based systems, which we tentatively associated with writing and oral task performance respectively. In Skehan’s (1998) view, an optimal situation is when both systems are called upon depending on the communicative goal and condition. In other words, the use of both systems is necessary, because they have different and complementary effects on language use and development. Engagement of the exemplar-based system (i.e., focus on oral tasks), is expected to reinforce control of and access to the interlanguage. On the hand, the deployment of the rule-based system (i.e., focus on writing tasks) would allow restructuring to occur.

In sum, we would like to suggest that the combination of oral and written tasks could be seen as a viable way to attain developmental balance in TBLT-oriented classroom. This is also in line with the ideas in Gilabert et al. (2016), who posited that

L2 acquisition may proceed in the mingling and interweaving of modes, hence the importance of combining oral and written tasks in pedagogical practices. However, it is a matter of future empirical research to determine the principles of the combination of oral and written tasks, depending on the goal and context of instruction and the characteristics of the learners. It will be necessary to determine, *inter alia*, the sequence of the oral and written tasks, their types (e.g., collaborative vs. individual) or the level of their complexity. Another important issue to address is which tasks and which type of oral-written sequences or combinations would be more appropriate for L2 learners of different levels of proficiency or with different learning goals. In sum, we believe that mode can be fruitfully employed as an efficient task design feature, which can contribute to a balanced interlanguage development.

The findings from our study may also have relevant implications for both task design and language pedagogy. Our finding that mode may mediate the way in which L2 learners use their linguistic knowledge might be relevant for task design as it suggests that oral and written modes may promote different kinds of L2 production and, by extension, may advance L2 competences in different but complementary ways. Another pedagogically relevant finding is that the written mode appeared to be advantageous for learners to exploit in full the benefits induced by an increase in the cognitive load of the task. This finding is also in line with the positions that claim an important language-learning potential for L2 writing in terms of consolidation and expansion of L2 resources as a result of the deeper linguistic and cognitive problem-solving processes that characterize many forms of written communication (Manchón, 2011, 2014; Williams, 2012). The findings of this study may also be relevant for task sequencing. It must be noted that current models of sequencing do not address the issue of mode directly (Baralt et al., 2014). However, we believe that knowledge about the

idiosyncrasy of language use and learning in different modes would help syllabus designers and teachers with the difficult task of sequencing and balancing tasks in a task-based syllabus.

The second aim of this dissertation was to explore if the effects of task complexity were of the same nature in oral and written production. This research question was specifically linked to the disciplinary debates of putting TBLT tenets to the empirical test across modalities. As for task complexity effects, we found that increases in task demands played out differently in writing and speech in the areas of propositional complexity, accuracy and time on task. Notably, changes in writing were more numerous and showed a better fit to the predictions of Robinson's Cognition Hypothesis, despite the fact that the theory's implicit original concern was the oral mode. Our explanation for this finding lies in the less constrained production conditions that characterize the written mode. In contrast, hard-pressed conditions of real-time speaking may appear to be too challenging for learners to be able to fully engage in the more complex processing induced by the increases in task demands. This nullifying effect may also be responsible for the mixed nature of the results obtained in previous studies exploring task complexity influences on oral production. Therefore, we concur with previous suggestions along the same lines (cf. Byrnes & Manchón, 2014a, 2014b) and conclude that the general principles underlying task complexity theorizing ought to be refined to better account for the specificities of language production in the oral and written modes.

Despite this potential contribution to past and current disciplinary debates on the role of mode and task complexity in language learning outcomes, there are a number of limitations that we need to acknowledge. We had a fairly small number of participants, who were all university students. Use of only one type of population demands caution

when generalizing our findings. Undoubtedly, more research is required with other types of population and in different contexts. We must also highlight that only one type of task was used to compare speech and writing, and task complexity was operationalized as only one of the diverse task variables from Robinson's framework (i.e., +/- reasoning demands). This can also limit the generalization of our findings, as it has been shown that task type or genre can also have an influence on language products and processes (Gilabert, 2007; Révész et al., 2016).

Also, our methodological decisions for segmenting discourse into idea units need further refining and empirical validation with different types of tasks and populations. We would like to suggest that such methods as stimulated recall or eye tracking might be instrumental in confirming if idea units constitute units of discourse planning. In their recent study, for example, Vasylets and Gilabert (2015) made an attempt to validate idea units by exploring the correlation between idea units and visual behaviours (number and length of fixations) during L2 oral production. The study found high correlation between the number of idea units and intensity of visual behaviours. Definitely, more research in this line is required.

Another limitation concerns the assessment of L2 proficiency. As explained in the Method section, we employed a combination of two receptive vocabulary tests as a proxy of general L2 proficiency. Although this decision might be justifiable, the use of standardized L2 proficiency tests would be an additional asset. It is also necessary to highlight that this study is of cross-sectional nature. Longitudinal designs, without doubt, would offer a different perspective on the development of L2 oral and written competencies and their interrelationship. Another limitation is that in this study we did not account for the individual differences of the participants, as we did not assess the subjects' working memory capacity or the level of their language aptitude. Recent

research, however, has shown that individual differences should be taken into account when exploring L2 learners' performance in different modes (Gilabert, Muñoz, Artieda, forthcoming). We must also point out that this study has explored prototypical types of oral and written production, which limits the application of our findings to other types of oral and written performance, such as "chat", which is a hybrid type of production collapsing features of oral and written discourse. As the hybrid modes of discourse combine characteristics of both oral and written production, it might be suggested that language use and language learning opportunities in the hybrid modes would differ from more conventional oral and written task performance (see also González-Lloret & Ortega, 2014). Also, in this study we examined the dimensions of production separately. Future research, however, might consider examining the contribution of each performance dimension to oral or writing performance using such statistical techniques as PLS (partial least squares) or SEM (structural equation modeling).

Despite these limitations, the research reported in this dissertation can serve to (1) draw attention to the idiosyncrasy of performance and learning opportunities in the two modes; (2) provide further empirical evidence for those positions that view L2 writing as an opportunity for language learning; (3) question prevailing tenets about language learning that may derive from the manipulation of task complexity variables; and (4) open new research avenues regarding (1), (2), and (3). We must also point out the methodological contribution of this study. The analysis of speech and writing for propositional complexity allowed us to gain deeper insights into the effects of mode and task complexity on L2 production. We would like to conclude by highlighting the need for future research efforts in the realms of mode and task complexity, as the knowledge to be gained can help refine cognitivist TBLT theorizing and can also provide the basis

for informed decisions concerning the mode (oral or written) in which tasks should be designed, administered, or assessed.

References

Adams, R. (2006). L2 tasks and orientation to form: A role for modality? *ITL: International Journal of Applied Linguistics*, 152, 7–34.

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection* (pp. 243–266). Ann Arbor: University of Michigan Press.

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing: With commentaries by Ronald T. Kellogg & John R. Hayes* (Vol. 9). Springer Science & Business Media.

Allal, L., Chanquoy, L., & Largy, P. (2004). *Revision: Cognitive and instructional processes* (Vol. 8). Springer Science & Business Media.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, 18, 191–208. doi:10.1016/j.jslw.2009.05.003

Baddeley, A.D. (1986). *Working memory*. Oxford, U.K.: Clarendon.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and instructed second language learning*. Place of publication Bloomsbury Publishing.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(01), 17-34.

Beattie, G. (1980). Encoding units in spontaneous speech: Some implications for the dynamics of conversation. In H.W. Deschert & M. Raupach (Eds), *Temporal variables in speech: studies in honour of Frieda Goldman-Eisler* (pp. 131–143). The Hague: Mouton.

Beattie, G. (1983). *Talk: an analysis of speech and non-verbal behaviour in conversation*. Milton Keynes: Open University.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Berman, R., & Ravid, D. (2009). Becoming a literate language user: Oral and written text construction across adolescence. In D. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 92–112). Cambridge, UK: Cambridge University Press.

Bialystok, E. (1991). Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 113-140). London: Cambridge University Press.

Biber, D. (1985). Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. *Linguistics*, 23, 337–360. doi:10.1515/ling.1985.23.2.337.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384–414. doi:10.2307/414678.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, England: Cambridge University Press.

Biber, D. (2009). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In R. Olson, & N. Torrance (Eds.), *The Cambridge Handbook of Literacy* (pp.75-92). New York: Cambridge University Press.

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, 15(02), 223-250.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *TESOL Quarterly*, 5-35.

Bitchener, J. (2012). A reflection on “the language learning potential” of written CF. *Journal of Second Language Writing*, 21, 348–363.

Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics*, 31, 193–214.

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 Development* (Vol. 96). Multilingual Matters.

Börner, W. (1987). Schreiben im Fremdsprachenunterricht: U^o berlegungen zu einem Modell [Writing in the foreign language classroom: thoughts on a model]. In W. Lörcher, & R. Schulze, *Perspectives on language in performance, studies in linguistics, literary criticism and language teaching and learning* (Vol. II) (pp. 1336–1349). Tübingen: Narr.

Bulté, B., & Housen, A. (2009). The development of lexical proficiency in L2 speaking and writing tasks by Dutch-speaking learners of French in Brussels. Paper presented in the colloquium “Tasks across modalities”, Task Based Language Teaching Conference, Lancaster 2009.

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic research*, 4(1), 75-87.

Brown, G., A. Anderson, R. Shilcock, & Yule G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Bygate, M. (2000). Introduction to special issue: Tasks in language pedagogy. *Language Teaching Research*, 4(3), 185-192.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2000a). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman.

Bygate, M., Van den Branden, K., & Norris, J. (2014). Series editors' preface to Volume 7. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp. viii–xi). Amsterdam: John Benjamins.

Byrnes, H. (2014). Theorizing language development at the intersection of 'task' and L2 writing: Reconsidering complexity. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (Vol. 7; pp. 79-107). Amsterdam: John Benjamins.

Byrnes, H., & Manchón, R. M. (Eds.). (2014a). *Task-based language learning—Insights from and for L2 writing* (Vol. 7). Amsterdam: John Benjamins.

Byrnes, H., & Manchón, R. M. (2014b). Task-based language learning: Insights from and for L2 writing. An Introduction. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (Vol. 7; pp. 1–27). Amsterdam: John Benjamins.

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal*, i-235.

Carless, D. (2012). TBLT in EFL settings: Looking back and moving forward. *Task-based language teaching in foreign language contexts: Research and implementation*, 345-358.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19(4), 727-752.

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy*

(pp. 35–54). Norwood, NJ: Ablex.

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 105–123). Cambridge, UK: Cambridge University Press.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Eds.), *Coherence and grounding in discourse* (pp.21-51). Amsterdam: John Benjamins.

Chafe, W. (1988). Linking intonation units in spoken English. In J. Haiman and S. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 1-27). Amsterdam: John Benjamins.

Chafe, W. (1992). The flow of ideas in sample of written language. In S. Thompson and W.C. Mann (Eds.), *Discourse description: Diverse Analyses of a fund raising text*. Amsterdam John Benjamins.

Chafe, W. (1994). *Discourse, consciousness and time*. Chicago: University of Chicago Press.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R.Horowitz & S. Samuels (Eds.), *Comprehending oral and written language*. (pp 55–87). New York: Academic Press.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. New York: Cambridge University Press.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, 92, 414-423.

Choong, K. (2011). Task complexity and linguistic complexity: An exploratory study. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 11, 1–28.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English*, 24, 67–87.

Cook, V. J. (1994). *Linguistics and second language acquisition*. London: Macmillan.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.

Cromer, R. (1974). The development of language and cognition: The cognition hypothesis. In B. Foss (Ed.), *New perspectives in child development* (pp. 184-252). Harmondsworth, UK: Penguin.

Crookes, G. (1986). *Task classification. A cross-disciplinary review*. Center for Second Language Classroom Research, Social Science Research Institute, University of Hawai'i at Mānoa, USA.

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied linguistics*, 11(2), 183-199.

Crookes, G. V. & K. Rulon. (1985). *Incorporation of Corrective Feedback in Native Speaker/Non-native Speaker Conversation*. Technical Report No.3. Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii, Honolulu.

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41, 965–981.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561-580.

Crystal, D. (1969). *Prosodic Systems and intonation in English*. Cambridge, UK: Cambridge University Press.

Crystal D., P. Fletcher, & M. Garman. (1976). *The Grammatical analysis of language disability: A procedure for assessment and remediation*. London: Edward Arnold.

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81–135. doi:10.1111/j.1467-1770.1989.tb00592

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482-511.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222.

DeBot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1–24.

DeBot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46, 529–555

De Jong, N., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34, 5–34. doi:10.1017/S0272263111000489

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge, UK: Cambridge University Press.

Dewaele, J. M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics*, 19(4), 471-490.

Dipper, L., Black, M., & Bryan, K. (2005). Thinking for speaking and thinking for listening: The interaction of thought and language in typical and non-fluent comprehension and production. *Language and Cognitive Processes*, 20, 417–441.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in communication. *Studies in Second Language Acquisition*, 20, 349–385.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognitive and second language instruction* (pp. 206–257). Cambridge, UK: Cambridge University Press.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–262). New York: Cambridge University Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, 9, 12–20.

Ellis, R. (2000). Task based research and language pedagogy. *Language Teaching Research*, 4, 193–220. doi:10.1177/136216880000400302

Ellis, R. (2005). Planning and task performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3–37). Amsterdam: John Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford, UK: Oxford University Press.

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167–193). Amsterdam: John Benjamins.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language learning*, 44(3), 449–491.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139–155. doi:10.1016/1060-3743(95)90004-7

Erlam, R., Loewen, S., & Philp, J. (2009). The roles of output-based and input-based instruction in the acquisition of L2 implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit*

knowledge in second language learning, testing and teaching (pp. 241–261). Clevedon, UK: Multilingual Matters.

Evans, M., & Fisher, L. (2005). Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: The case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research*, 9(2), 173-192.

Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, 38(3), 247-273.

Fayol, M. (1991). From sentence production to text production: Investigating fundamental processes. *European Journal of Psychology of Education*, 6, 99-117.

Ferrari, S., & Nuzzo, E. (2009) Meeting the challenge of diversity with TBLT: Connecting speaking and writing in mainstream classrooms. Paper presented in the colloquium "Tasks across modalities", Task Based Language Teaching Conference, Lancaster 2009

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE.

Flower, L. (1990). Introduction: Studying cognition in context. In L. Flower, V. Stein, J. Ackerman, M. Kantz, K. McCormick & W. Peck (Eds.), *Reading-to-write: Exploring and cognitive and social process* (pp.3-28), Oxford: Oxford University Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition. Monographies*, 32(3-4), 259-296.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323. doi:10.1017/S0272263100015047

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency and lexical diversity. *Language Learning*, 59, 866–896. doi: 10.1111/j.1467-9922.2009.00528.x

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. doi:10.1093/applin/21.3.354

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, 36, 98-116.

Fukuta, J., & Yamashita, J. (2015). Effects of cognitive demands on attention orientation in L2 oral production. *System*, 53, 1–12. doi:10.1016/j.system.2015.06.010

García Mayo, M. P. & Azkarai, A. (2016). EFL task-based interaction: Does task modality impact on language-related episodes? In M. Sato & S. Ballinger (eds.) *Peer Interaction and Second Language Learning Pedagogical potential and research agenda* (pp. 241–266). Amsterdam: John Benjamins.

Garrett, M. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions*. London: Academic Press.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL quarterly*, 22(3), 473-492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, 61(3), 325-353.

Gilabert, R. (2005). Evaluating the use of multiple sources and methods in needs analysis: A case study of journalists in the autonomous community of Catalonia (Spain). In M. H. Long (Ed.), *Second language needs analysis* (pp. 182–99). Cambridge, UK: Cambridge University Press.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, 45, 215–240. doi:10.1515/iral.2007.010

Gilabert, R., Barón, J., & Llanes, A. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47, 367–395. doi:10.1515/iral.2009.016

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics*, 36, 117–135. doi:10.1017/S0267190515000112

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol 1*. (pp. 1008-1025). Hillsdale, NJ: Lawrence Erlbaum.

Givón, T. (1995). *Functionalism and grammar*. Amsterdam: Benjamins.

Givón, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam: John Benjamins.

Gui, L. (2010). A contrastive study of lexical proficiency between L1 and L2 compositions via computerized assessment. *Foreign Language Teaching and Research*, 42, 445–450.

Guiraud, P. (1959). *Problèmes et méthodes de la statistique linguistique*. Dordrecht: Reidel.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, 106(3), 608.

González-Lloret, M., & Ortega, L. (Eds.). (2014). *Technology-mediated TBLT: Researching technology and tasks* (Vol. 6). John Benjamins Publishing Company.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34, 445–474.

Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In M. Torrance, L. van Waes & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (Studies in Writing, Vol. 20, pp. 165–179). Amsterdam: Elsevier.

Granfeldt, J. (2008). Speaking and writing in L2 French: Exploring effects on fluency, accuracy and complexity. In A. Housen, S. van Daele, & M. Pierrard (Eds.), *Proceedings of the conference on complexity, accuracy and fluency in second language* (pp. 87–98). Brussels, Belgium: Koninklijke Vlaamse Academi.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the Phraseological Continuum Model: Does semantic transparency matter? *Language Learning*, 66, 296–323. doi:10.1111/lang.12143

Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Halliday, M. A. K. (1967) *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. Hodder Education.

Halliday, M. A. K. (1987). Spoken and written modes of meaning. In R. Horowitz & S. Samuels (Eds.), *Comprehending oral and written language* (pp. 55–82). New York: Academic Press.

Halliday, M. A. K. (1990). *Spoken and written language*. Oxford, UK: Oxford University Press.

Halliday, M. A. K. (2003). *On language and linguistics* (Vol. 3). A&C Black.

Halliday, M. A. K., & Mattiesen, C. (1999). *Construing experience through meaning: A language-based approach to cognition*. London: Cassell.

Han, Z. H., Park, E., & Combs, C. (2008). Textual input enhancement: Issues and possibilities. *Applied Linguistics*, 29(4), 597–618.

Hayes, J. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369–388. doi:10.1177/0741088312451260

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Randall (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Erlbaum.

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Hayes, J., & Nash, J. (1996). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 29–55). Mahwah, NJ: Erlbaum.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>

Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, 3(4), 323-361.

Holmes, V. M. (1995). A crosslinguistic comparison of the production of utterances in discourse. *Cognition*, 54(2), 169-207.

Horowitz, M., & Newman, J. (1964). Spoken and written expression: An experimental analysis. *The Journal of Abnormal and Social Psychology*, 68, 640–647. doi:10.1037/h0048589

Horowitz, S., & Samuels, J. (1987). *Comprehending oral and written language*. San Diego, CA: Academic Press.

- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Housen, A., Schoonjans, E., Janssens, S., Welcomme, A., Schoonheere, E., & Pierrard, M. (2011). Conceptualizing and measuring the impact of contextual factors in instructed SLA—the role of language prominence. *IRAL—International Review of Applied Linguistics in Language Teaching*, 49(2), 83–112.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). Amsterdam: John Benjamins Publishing.
- Huddleston, R. (1984). *Introduction to the grammar of English*. Cambridge, UK: Cambridge University Press.
- Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press.
- Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research* (Vol. 41). Amsterdam: John Benjamins.
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English* 43:732-39.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24–49. doi:10.1093/applin/amm017
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, 24, 168–196.
- Jackson, D., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63, 330–367. doi:10.1111/lang.12008

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(s1), 87-106.

Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 89–108). Amsterdam: John Benjamins.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57–71). Mahwah, NJ: Erlbaum.

Knoch, U. (2009). *Diagnostic writing assessment: The development and validation of a rating scale*. Frankfurt, Germany: Peter Lang.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kormos, J. (2011a). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 39-60). Amsterdam: John Benjamins.

Kormos, J. (2011b). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 148-161.

Kormos, J. (2014). Differences across modalities of performance: An investigation of linguistic and discourse complexity in narrative tasks. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp.193–217). Amsterdam: John Benjamins.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62, 439–472. doi:10.1111/j.1467-9922.2012.00695

Krings, H. P. (1992). Empirische Untersuchungen zu fremdsprachlichen schreibprozessen — ein forschungsu ¨berblick [Empirical studies of foreign language writing processes — a research survey]. In W. Börner, & K. Vogel, *Schreiben in der Fremdsprache. Prozeß und Text, Lehren und Lernen* (pp.47–77). Bochum: AKS-Verlag.

Kroll, B. (1977). Combining ideas in written and spoken English: a look at subordination and coordination. In E. Ochs Keenan and T. L. Bennett (Eds.), *Discourse Across Time and Space. Southern California Occasional Papers in Linguistics* 5. Los Angeles, Department of Linguistics, University of Southern California.

Kuhn, D. (1991). *The skills of argument*. Cambridge, UK: Cambridge University Press.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45(3), 261-284.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48–60. doi:10.1016/j.jslw.2007.08.003

Kuiken, F., & Vedder, I. (2011). Task performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). Amsterdam: John Benjamins.

Kuiken, F., & Vedder, I. (2012). Speaking and writing tasks and their effects on second language performance. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 364–379). New York: Routledge.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619. doi:10.1093/applin/aml029

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–322. <http://dx.doi.org/10.1093/applin/16.3.307>.

Lee, Y. (2002). Effects of task complexity on the complexity and accuracy of oral production in L2 Korean (Unpublished Ph.D. dissertation). University of Hawai'i at Manoa.

Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics*, 12, 180–196. doi:10.1093/applin/12.2.180

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York, NY: Routledge.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö, Sweden: Liber Förlag.

Loban, W. (1966). *Language ability: Grades even, eight, and nine*. Washington, DC: Government Printing Office.

Long, M. (2014). *Second language acquisition and task-based language teaching*. Oxford, UK: Wiley.

Long, M. H. (1998). Focus on form in task-based language teaching. *University of Hawaii Working Papers in ESL*, 16, 35-49.

Long, M. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hylstenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77-99). Clevedon: Multilingual Matters.

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26, 27-56.

Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design—The case for task. In G. Crookes & S. Gass (Eds.), *Tasks in a pedagogic context: Integrating theory and practice* (pp. 9-54). Clevedon: Multilingual Matters.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190-208.

Luce, R. D. (1986). *Response times: Their roles in inferring elementary mental organization*. Oxford, UK: Oxford University Press.

MacKay, D. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, 89, 483-506.

Malvern, D., Chipere, N., Richards, B., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills: Palgrave Macmillan.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19, 85–104. doi:10.1191/0265532202lt221

Manchon, R.M. (2009). Introduction. Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R.M. Manchon (Ed.), *Writing in foreign language contexts: Learning, teaching and research* (pp. 1–19). Bristol, UK: Multilingual Matters.

Manchón, R.M. (2011a). Writing to learn the language: Issues in theory and research. In R.M. Manchon (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). Amsterdam: John Benjamins. DOI: 10.1075/llt.31.07man

Manchón, R.M. (Ed.). (2011b). *Learning-to-write and writing-to-learn in an additional language*. Amsterdam: John Benjamins. DOI: 10.1075/llt.31

Manchón, R.M. (2013a). Teaching writing. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell. DOI: 10.1002/9781405198431.wbeal1178

Manchón, R.M. (2014a). Learning and teaching writing in the FL classroom: Fostering writing-to-learn approaches. In P. Driscoll, E. Macaro, & A. Swarbrick (Eds.), *Debates in modern language education* (pp. 96–107). London: Routledge.

Manchón, R. M. (2014b). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning – Insights from and for L2 writing* (Vol. 7; pp. 27–53). Amsterdam: John Benjamins.

Manchón, R.M., & Roca de Larios, J. (2007). Writing-to-learn in instructed language learning contexts. In E.A. Soler & M.P.S. Jorda (Eds.), *Intercultural language use and language learning* (pp. 101–121). Berlin: Springer. DOI: 10.1007/978-1-4020-5639-0_6

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes: Implications for theory. In R.M. Manchón (Ed), *Writing in foreign language contexts: Learning, teaching, and research*, (pp.102-129), Bristol, UK: Multilingual Matters.

Manchón, R. M., & Williams, J. (2016). L2 writing and SLA studies. In R. M. Manchón & P. K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 567–586). Boston: de Gruyter.

Matthiessen, C., & Thompson, S. A. (1988). The structure of discourse and 'subordination'. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (Vol. 18) (pp. 275– 329). Amsterdam: John Benjamins.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, 29, 3-15.

McCarthy, P., & Jarvis, S. (2010). MTL D, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381–392.

McLaughlin, B. (1990). Restructuring. *Applied linguistics*, 11(2), 113-128.

McNamara, D., Graesser, A., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-metrix*. Cambridge, UK: Cambridge University Press.

McNamara, D.S., Louwse, M.M., & Graesser, A.C. (2002). *Coh-Metrix: Automated Cohesion and Coherence Scores to Predict Text Readability and Facilitate*

Comprehension. Institute for Intelligent Systems, University of Memphis, Memphis, TN.

MacWhinney, Brian (1995). *The CHILDES project: Tools for analyzing talk* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, 24, 856–868.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15, 323–337.

Meara, P. M. (1992). *EFL vocabulary tests*. New York: ERIC Clearinghouse.

Meara, P. M., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). London: CILT.

Meara, P. M., & Milton, J. (2003). *X_Lex*. Swansea, UK: Lognostics.

Meara, P. M., & Miralpeix, I. (2006). *Y_Lex*. Swansea, UK: Lognostics.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20(1), 83-108.

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemaki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23–42). Amsterdam: John Benjamins.

Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Bristol, UK: Multilingual Matters.

Miralpeix, I. (2012, March). *X_Lex and Y_Lex: A validation study*. Paper presented at the 22nd Lexical Studies Conference, Newtown, UK.

Muñoz, C. (2010). Staying abroad with the family: A case study of two siblings' second language development during a year's immersion. *ITL International Journal of Applied Linguistics*, 160, 24–48.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 51–84). Cambridge, UK: Cambridge University Press.

Nas, G. (1975). *Determining the communicative value of written discourse produced by L2 learners*. Utrecht, The Netherlands: Institute of Applied Linguistics.

Norris, J. M., & Manchón, R. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 221–244). Berlin, Germany: Walter de Gruyter.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555–578. doi:10.1093/applin/amp044

Niu, R. (2009). Effect of task-inherent production modes on EFL learners' focus on form. *Language Awareness*, 18, 384–402.

Nystrand, M. (1987). The role of context in written communication. In Horowitz, S., & Samuels, J. (1987). *Comprehending oral and written language* (pp. 197-212). San Diego, CA: Academic Press.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Syntax and semantics* (Vol. 12): *Discourse and semantics* (pp. 51–81). New York: Academic Press.

Olson, D. R., Torrance, N., & Hildyard, A. (1985). *Literacy, language and learning: The nature and consequences of reading and writing*. CUP Archive.

Ong, W. (1977). *Interfaces of the Word*. Ithaca and London: Cornell University Press.

Ong, W. (1982). *Literacy and orality: The technologizing of the word*. London: Methuen.

- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109–148.
- Ortega, L. (2012a). Epilogue: Exploring L2 writing—SLA interfaces. *Journal of Second Language Writing*, 21, 404–415. doi:10.1016/j.jslw.2012.09.002
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Berlin:Mouton.
- Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82-94.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead.
- Pallotti, G. (2009). CAF: Defining, refining, and differentiating constructs. *Applied Linguistics*, 30(4), 590-601.
- Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.
- Perdue, C. (Ed.) (1993). *Adult language acquisition: Crosslinguistic perspectives Vols. 1: Field methods*. Cambridge, UK: Cambridge University Press.
- Piolat, A., Roussey, J. Y., Olive, T., & Amada, M. (2004). Processing time and cognitive effort in revision: effects of error type and of working memory capacity. In L. Allal, L. Chanquoy and P.Largy (Eds.), *Revision Cognitive and Instructional Processes* (pp. 21-38). Springer Netherlands.
- Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language learning*, 47(1), 101-143.
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, 21(4), 375–389.

Poullisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36–57.

Quesada, Jose, Walter Kintsch and Emilio Gómez-Millán (2005). Complex problem-solving: A field in search of a definition? *Theoretical Issues in Ergonomic Science*, 6 (1), 5–33.

Quirk, R., & Crystal, D. (1964). *Systems of prosodic and paralinguistic features in English*. Place of publication missing: Mouton

Quirk, R., S. Greenbaum, G. Leech, & J. Svartvik. (1985). *A Comprehensive grammar of the English language*. London: Longman.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, 30(1), 3-26.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of child language*, 29(02), 417-447.
doi.org/10.1017/S0305000902005111

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Reinders, H. (2012). Towards a definition of intake in second language acquisition. *Applied Research on English Language*, 1(2), 15–36.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments. *Studies in Second Language Acquisition*, 38, 703–737. doi:10.1017/S0272263115000339

Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes*, 15(2), 129-157.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.
doi:10.1093/applin/22.1.27

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21, 45–105.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Clevedon, UK: Multilingual Matters.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). Amsterdam: John Benjamins Publishing.

Robinson, P. (2011a). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 3–39). Amsterdam: John Benjamins.

Robinson, P. (2011b). Task- based language learning: A review of issues. *Language Learning*, 61(s1), 1-36.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics*, 30(4), 533-554.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, 45, 161–176. doi:10.1515/iral.2007.007

Robinson, P., & Gilabert, R. (2013). Task-based learning: Cognitive underpinnings. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell. DOI: 10.1002/9781405198431wbeal143

Robinson, P., Mackey, A., Gass, S., & Schmid, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–268). New York: Routledge.

Roca de Larios, J. R., Manchon, R. M., & Murphy, L. (2006). Generating text in native and foreign language writing: A temporal analysis of problem-solving formulation processes. *The Modern Language Journal*, 90, 100–114.

doi:10.1111/j.1540-4781.2006.00387

Roca de Larios, J. R., Manchón, R. M., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing, 17*, 30–47. doi:10.1016/j.jslw.2007.08.005

Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing, 28*, 1-19.

Salimi, A., & Dadashpour, S. (2012). Task complexity and language production dilemmas. *Procedia, 46*, 643–652. doi:10.1016/j.sbspro.2012.05.177

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. New York: Palgrave Macmillan.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111-141). Amsterdam: John Benjamins Publishing.

Schiffrin, D. (1985). Everyday argument: The organization of diversity in talk. In T. van Dijk (Ed.), *Handbook of discourse analysis* (Vol.3; pp. 35–46). London: Academic Press.

Schilperoord, J., & Verhagen, A. (1998). Conceptual dependency and the clausal structure of discourse. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 141–163). Stanford, CA: CSLI Publications.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1-33). New York: Cambridge University Press.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–385.

Schenck, A. D., & Daly, E. (2012). Building a better mousetrap: Replacing subjective writing rubrics with more empirically-sound alternatives for EFL learners. *Creative Education*, 3(8), 1320.

Schoonen, R., Gelderen, A. V., Glopper, K. D., Hulstijn, J., Simis, A., Snellings, P. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53, 165–202. doi:10.1111/1467-9922.00213

Schoonen, R., Snellings, P., Stevenson, M., & Van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R.M. Manchón (Ed), *Writing in foreign language contexts: Learning, teaching, and research*, (pp.77-101), Bristol, UK: Multilingual Matters.

Segalowitz, N., & Gatbonton, E. (1995). Automaticity and lexical skills in second language fluency: Implications for computer assisted language learning. *Computer Assisted Language Learning*, 8(2-3), 129-149.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232.

Shaw, S.D., & Weir, C.J. (2007). Studies in Language Testing. In Milanovic, M., & Weir, C.J. (Eds). *Examining writing: Research and practice in assessing second language writing* (Vol.26). Cambridge, England: Cambridge University Press.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255–283.

Shewan, C. M. (1988). The Shewan Spontaneous Language Analysis (SSLA) system foraphasic adults: description, reliability, and validity. *Journal of Communication Disorders* 21: 103-38.

Shiau, Y., & Adams, R. (2011). The effects of increasing reasoning demands on accuracy and complexity in L2 oral production. *University of Sydney Papers in TESOL*, 6, 121–146.

Skehan, P. (1992). Strategies in second language acquisition. *Thames Valley University Working Papers in English Language Teaching*, No.1.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532. doi:10.1093/applin/amp047

Skehan, P. (Ed.). (2014). *Processing perspectives on task performance* (Vol. 5). Amsterdam: John Benjamins.

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. *Complexity, accuracy, and fluency in second language use, learning, and teaching*, 207-226.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), pp. 93–120.

Skehan, P., Foster, P., & Mehnert, U. (1998). Assessing and using tasks. *ANTHOLOGY SERIES-SEAMEO REGIONAL LANGUAGE CENTRE*, 227-248.

Slobin, D. I. (1993). Adult language acquisition: A view from child language study. In C. Perdue (Ed.), *Adult language acquisition: Crosslinguistic perspectives vol 2: The results* (pp. 239-252). Cambridge, UK: Cambridge University Press.

Stanovich, K. (1991) Word recognition: changing perspectives. In R. Barr, M.L. Kamil, P. Mosenthal & P.D. Person (Eds.) *Handbook of reading research, Volume II*. New York, NY: Longman, (pp. 418-452)

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.), *For H. G. Widdowson: Principles and practice in the study of language. A Festschrift on the occasion of his 60th birthday* (pp. 125–144). Oxford, UK: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–82). Cambridge, UK: Cambridge University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–83). Mahwah, NJ: Erlbaum.

Tabachnick, B. Fidell (1996), *Using multivariate statistics*. New York, NY: HarperCollins College Publishers,

Talmy, L. (2000). *Toward a cognitive semantics, Vol. 1: Concept structuring systems*. Cambridge, MA: MIT Press.

Talmy, L. (2008). Attention phenomena in language and cognition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*, (pp. 27–38). New York, NY: Routledge.

Tannen, D. (1982). The oral/literate continuum in discourse. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 1–16). Norwood, NJ: Ablex Publishing.

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp. 217–236). Amsterdam: John Benjamins.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58, 439–473. DOI: 10.1111/j.1467-9922.2008.00446.x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Amsterdam: John Benjamins Publishing Company.

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*. Minneapolis: The University of Minnesota Press.

Tomlin, R. (1990). Functionalism in second language acquisition. *Studies in Second Language Acquisition*, 12, 155-177.

Tonkyn, A. (2012). Measuring and perceiving changes in oral complexity, accuracy and fluency: Examining instructed learners' short-term gains. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 221–246). Amsterdam: John Benjamins.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84–119. doi:10.1093/applin/17.1.84

Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language Learning*, 62(2), 506–540.

Van Daele, S. (2007). Linguïstische vlotheid, accuraatheid en complexiteit in de verwerving en productie van een tweede taal [Linguistic Fluency, Accuracy and Complexity in the Acquisition and Production of a Second Language]. Unpublished doctoral dissertation, Vrije Universiteit Brussel.

Van Daele, S., A. Housen, F. Kuiken, M. Pierrard, M. & I. Vedder. (Eds) (2007). *Complexity, accuracy and fluency in second language use, learning and teaching*. Brussels: Contactforum.

Van de Branden, K. (2006). *Task-based language education: from theory to practice*. Cambridge: Cambridge University Press.

Van den Branden, K., Bygate, M., & Norris, J. (Eds.). (2009). *Task-based language teaching: A reader*. Amsterdam: John Benjamins.

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287-301.

VanPatten, B. (2011). Input processing. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition*. New York, NY: Routledge.

VanPatten, B., Williams, J., & Rott, S. (2004). *Form-meaning connections in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

van der Pool, E.M.C. (1995). *Writing as a conceptual process: a text-analytical study of developmental aspects*. PhD dissertation Catholic University Brabant, the Netherlands.

van Wijk, C. (1999). Conceptual processes in argumentation: a developmental perspective. In G. Rijlaarsdam, & E. Esperet (Series Eds.), & M. Torrance, & D. Galbraith (Vol. Eds.), *Studies in Writing: Vol 4. Knowing what to write: Conceptual processes in text production* (pp. 31-50). Amsterdam: Amsterdam University Press.

van der Pool, E.M.C., & van Wijk, C. (1995). Process and strategy in a psycholinguistic model of writing and reading. *Tijdschrift voor Onderwijsresearch*, 20, 200-214.

Vasylets, O., & Gilabert, R. (2013). *Testing the Cognition Hypothesis in L2 writing: Critical review of the accumulated findings*. Poster presented at the EuroSLA Conference, Amsterdam, Netherlands.

Vasylets, O., & Gilabert, R. (2014). Second language writing as moderated by cognitive task complexity. In A. Díaz Galén, M. Fumero Pérez, M. Lojendio Quintero, S. Burgess, E. Sosa Acvedo, & A. Cano Ginés (Eds.), *Proceedings of AESLA (Spanish Association of Applied Linguistics): Communication, cognition & cybernetics* (pp. 185–193). Published online at <http://www.aesla.org.es/es/intranet>.

Vasylets, O., & Gilabert, R. (2014). Second language writing as moderated by cognitive task complexity. In A. Díaz Galán, M. Fumero Pérez, M. Lojendio Quintero, S. Burgess, E. Sosa Acvedo, & A. Cano Ginés (Eds.), *Proceedings of AESLA (Spanish Association of Applied Linguistics): Communication, cognition & cybernetics* (pp. 185–193). Published online at <http://www.aesla.org.es/es/intranet>.

Vasylets, O., Gilabert, R. (2015). “Exploring the visual-dynamic and linguistic conceptualisation traces in task-based performance”. Paper presented at the 6th TBLT conference, colloquium “*Eye-movements during task-based performance: What can we learn?*” Leuven, Belgium, September 16-18.

Vygotsky, L. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wells, G. (1987). *The meaning-makers: children learning language and using language to learn*. London: Hodder and Stoughton.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67–85.

Wickens, C. (1984). Processing resources in attention. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York, NY: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3(2), 159–177.

Wigglesworth, G., & Foster, P. (2008). Measuring accuracy in second language performance. In *Annual Conference of Teachers of English to Speakers of Other Languages (TESOL)*, New York.

Wilkins, D. (1976). *Notional syllabuses*. Oxford: Oxford University Press.

Williams, J. (2008). The speaking–writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral–literate connection* (pp. 10–25). Ann Arbor, MI: The University of Michigan Press.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21, 321–331. doi:10.1016/j.jslw.2012.09.007

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity*. Honolulu: University of Hawai'i.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31, 236–259. doi:10.1093/applin/amp024

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, 1–27. doi:10.1093/applin/24.1.1

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27, 567–595. doi:10.1017/S0272263105050254

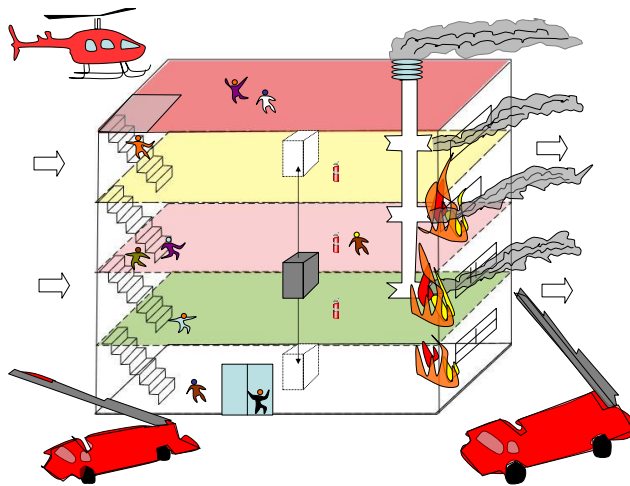
Zimmermann, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction*, 10(1), 73-99. doi.org/10.1016/S0959-4752(99)00019-5

Appendix

Instructions for the “Fire-chief” task

You volunteered for the university’s fire emergency team. As part of the training you are asked to handle the following situation: a fire has broken out in a building, and you must make sure everyone is saved. Say/write which actions you would take to save as many people as possible. Please explain the sequence in which you take those actions. Justify your choice of actions and their sequence.

Simple version of the “Fire-chief” task



Complex version of the “Fire-chief” task

