

GLOSSING FOR MEANING AND GLOSSING FOR FORM. A COMPUTERIZED
STUDY OF THE EFFECTS OF GLOSSING AND TYPE OF LINGUISTIC ITEM
ON READING COMPREHENSION, NOTICING, AND L2 LEARNING

A Dissertation
Submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirement for the
degree of
Doctor of Philosophy
in Spanish Linguistics

By

Claudia Guidi, M.S.

Washington, DC
May 29, 2009

GLOSSING FOR MEANING AND GLOSSING FOR FORM. A COMPUTERIZED
STUDY OF THE EFFECTS OF GLOSSING AND TYPE OF LINGUISTIC ITEM
ON READING COMPREHENSION, NOTICING, AND L2 LEARNING

Claudia Guidi, M.S.

Thesis Advisor: Ronald P. Leow, Ph.D.

ABSTRACT

The relative effectiveness of annotating L2 texts for improved comprehension has been generally supported by studies on glossing. The assumption has been that glosses are effective decoding devices that help L2 readers overcome their limited proficiency by making meanings more accessible to them. At the same time, glosses have been investigated in relation to noticing and learning in incidental contexts. In spite of its centrality, this question has not generally been addressed empirically in previous studies on glossing and L2 learning because qualitative measures of L2 processing, and concurrent data elicited via those measures, were generally not available in these studies. The present study addresses this important methodological problem in the glossing strand of research by utilizing a pretest-posttest-delayed posttest hybrid design that incorporates a combination of outcome and processing measures to investigate the issue of L1-translation glosses and type of linguistic item in relation to reading comprehension, noticing and learning. First-year college students learning Spanish were asked to read one of two versions of an L2 Spanish text on a computer. In the glossed version, exemplars of three types of linguistic items embedded in the text were glossed

with L1 English translations presented at the bottom of page. In the no-gloss conditions, participants were exposed to a different version of the same text without glosses. After exposure to the L2 reading task, participants completed a multiple-choice comprehension questionnaire previously announced to them, and two unannounced production and recognition immediate posttests, which were repeated three weeks later. The results of the study indicate that the presence of a gloss has a positive effect on a reading comprehension. However, the quantitative and qualitative evidence of the study does not support an effect of glosses on noticing and L2 learning of targeted items. Rather, the study provides evidence that glosses work in combination with other variables to impact different types of items differentially. Combined effects of glossing and type of linguistic item found in the study strongly suggest that glosses do not work in isolation but in combination with other factors to affect noticing and learning.

TABLE OF CONTENTS

Introduction.....	1
Statement of the Problem.....	1
Definition of Terms.....	7
Chapter 1: Review of Related Literature	9
Theoretical Foundation	10
Tomlin and Villa’s model of input processing	10
The noticing hypothesis	11
VanPatten’s model of input processing	13
Robinson’s model of attention and memory	15
Conclusion	17
Input Modification	19
Theoretical underpinnings	19
Empirical studies.....	22
Simplified written input	22
Textual enhancement	24
Glossing	36
Effects of glossing on L2 reading comprehension. Theoretical underpinnings.....	37
Effect of glossing on text comprehension: Empirical studies.....	41

Effects of glossing on L2 learning while interacting with an L2 text:	
Theoretical underpinnings	57
Effects of glossing on L2 learning while interacting with L2 texts:	
Empirical studies.....	59
Summary of findings of previous research on the effectiveness of glossing for improved L2 reading comprehension and learning	73
Operationalization and Measurement of Attention/Noticing.....	74
Other Variables not Considered in Previous Studies on Glossing.....	85
Type of linguistic item	85
Language difficulty	87
Research Questions	93
Chapter 2: Research Design and Methodology	95
Participants.....	95
Materials	96
The target of the study	96
The targeted types	99
The Present Perfect	99
The Impersonal <i>SE</i>	102
The lexical items.....	106
Operationalization of glossing	107
Selection of items targeted in the study	110

The experimental text	113
Procedure	115
Computerization of the materials of the study.....	116
The assessment tests	118
The Reading comprehension questionnaire	118
‘Learning’ measures.....	120
Overview of the Research Design.....	125
Chapter 3: Results.....	126
Part I. The think aloud protocols.....	126
Answer to RQ#(1) Reading patterns and styles as revealed through a think- aloud task	126
Reading with a gloss	127
Reading without a gloss	131
Summary of findings of the think-aloud protocol analyses	145
Part II. Quantitative analyses of the outcome measures	148
Introduction.....	148
Pretests	149
The reactivity of verbal reports.....	154
Quantitative analyses of the outcome measures	157
Chapter 4: Discussion and Conclusions.....	182
Discussion.....	182

Conclusions.....	211
Pedagogical Implications	222
Limitations and Future Research	223
Bibliography	227
Appendix A.....	245
Appendix B.....	247
Appendix C.....	249
Appendix D.....	252
Appendix E.....	258
Appendix F.....	264

LIST OF TABLES

Table 1. Summary Of Studies On Glossing And Reading Comprehension	42
Table 2. Summary Of Research Studies On The Effects Of Glossing On L2 Learning While Reading	59
Table 3. Difficulty Matrix For Two Dimensions	107
Table 4. Difficulty Matrix For Two Types Of Grammatical Items	107
Table 5. Phases Of The Experiment	116
Table 6. Mean, SD And SE Of Noticing For Two Gloss Groups.....	137
Table 7. Mean And SD For Noticing Of Three Types Of Items By Gloss Group	140
Table 8. Means, SD, And SE For Recognition Pretests Of Three Types Of Item By Think-Aloud Group.....	150
Table 9. Means, SD, And SE For Recognition Pretests Of Three Types Of Items By Gloss Group.....	151
Table 10. Mean And SD For Three Types Of Item At Pretest	151
Table 11. Means, SD, And SE For Production Pretests Of Three Types Of Item By Think-Aloud Group.....	152
Table 12. Means, SD, And SE For Production Pretests Of Three Types Of Items By Gloss Group.....	153
Table 13. Means And SD Of Production Pretests By Type Of Item	153

Table 14. Means, SD, And SE For Two Think-Aloud Groups On Text Comprehension	155
Table 15. Means, SD, And SE For Immediate Recognition Posttests Of Three Types Of Items By Think-Aloud Group.....	156
Table 16. Means, SD, And SE For Immediate Production Posttests Of Three Types Of Items By Think-Aloud Group.....	156
Table 17. Mean, SD, And SE Of Two Gloss Groups On Reading Comprehension	158
Table 18. Mean And SD For Recognition Of Three Types Of Item And Two Gloss Groups Over Time	160
Table 19. Recognition: ANOVA For Group By Time By Type Of Item.....	160
Table 20. Recognition Immediate Posttest: ANOVA For Group By Type Of Item	164
Table 21. Mean And SD Of Present Perfect Immediate Posttest Recognition For Two Gloss Groups	165
Table 22. Recognition Delayed Posttest: ANOVA For Group By Type Of Item	165
Table 23. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Lexical Item Versus Present Perfect)	166
Table 24. Mean And SD Of Delayed Recognition Of Lexical Item For Two Gloss Groups.....	167

Table 25. Delayed Posttest: ANOVA For Recognition Of Present Perfect Versus Lexical Item Within The Gloss Group.....	167
Table 26. Delayed Posttest: ANOVA For Recognition Of Present Perfect Versus Lexical Item Within The No-Gloss Group	167
Table 27. Mean And SD For Delayed Recognition Of Lexical Item And Present Perfect By Gloss Group	168
Table 28. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Lexical Item Versus Impersonal-SE).....	168
Table 29. Delayed Posttest: ANOVA For Recognition Of Impersonal-SE Versus Lexical Item Within The Gloss Group.....	170
Table 30. Delayed Posttest: ANOVA For Recognition Of Impersonal-SE Versus Lexical Item Within The No-Gloss Group	170
Table 31. Mean And SD For Group And Two Types Of Item On Delayed Posttest (Lexical Item Versus Impersonal-SE)	170
Table 32. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Present Perfect Versus Impersonal-SE).....	170
Table 33. Recognition Of Lexical Item: ANOVA For Group By Time	173
Table 34. Recognition Of Present Perfect: ANOVA For Gloss By Time	175
Table 35. Recognition Of Impersonal-SE: ANOVA For Gloss By Time	178
Table 36. Production: M And SD Of Three Items, Two Groups And Three Times	181

LIST OF FIGURES

Figure 1. Mean Scores On Recognition Of Three Types Of Item And Two Groups At Pretest (A), Immediate Posttest (B), And Delayed Posttest (C)	161
Figure 2. Mean On Delayed Recognition Of Two Types Of Item For Two Gloss Groups.....	166
Figure 3. Mean On Delayed Recognition Of Two Types Of Item For Two Gloss Groups.....	169
Figure 4. Lexical Item: Mean On Recognition For Two Groups Across Time..	173
Figure 5. Present Perfect: Mean On Recognition For Two Groups Across Time	176
Figure 6. Impersonal-SE: Mean Scores On Recognition Of Two Gloss Groups Across Time	179

INTRODUCTION

Statement of the Problem

Glossing an L2 reading passage for the L2 reader by way of adding L1-translations for 'difficult' or 'unfamiliar' vocabulary words and grammatical structures is a popular pedagogical practice in the L2 learning context. In spite of the popularity of these 'traditional' 'micro' level annotations, theoretical views of L2 reading and L2 learning make differing predictions about the purported benefits of these types of glosses for improved comprehension, and L2 learning.

The purported benefits of glossing linguistic (i.e., lexical or grammatical) information in L2 texts are supported by the premises of a bottom-up theoretical view of L2 reading. This view predicts that 'micro' level annotations (i.e., short definitions or explanations for linguistic elements of a text in the L1 or the L2) should help L2 readers overcome the decoding problems they face due to lack of lexical and grammatical knowledge. In this context, glosses are viewed as decoding aids. By providing the meanings of unknown or unfamiliar words or structures, and making them easily accessible to the L2 reader, glosses should make text contents more available to L2 readers. In contrast, exposing the learners to no glossed texts containing words and structures that are difficult to decode will force them to guess the meaning from context, or to use background knowledge to try to compensate for decoding difficulties. From a bottom-up point

of view, however, strategies such as these are not supported. Thus, bottom-up approaches to L2 reading, with their emphasis on the notion that L2 readers need to understand the linguistic aspects of the text before they are able to understand its textual and pragmatic elements, support the use of traditional 'micro' level glosses as decoding aids that may help bridge the gap between the L2 reader's characteristic lack of linguistic knowledge and the L2 text.

While the purported benefits of glossing linguistic (i.e., lexical or grammatical) information in L2 texts are supported by the premises of a bottom-up theoretical view of L2 reading, this type of glossing is put into question by top-down and bidirectional approaches to L2 reading, premised on the idea that decoding of every word and structure may not be necessary for text comprehension: Readers overcome decoding deficiencies by using strategies such as guessing from context or using background knowledge. From this point of view, glossing linguistic information may hinder, rather than facilitate, text comprehension.

Despite the theoretical debate in the reading literature about glossing linguistic aspect of the L2 text for the L2 reader (i.e., lexical or grammatical), the empirical research conducted from both top-down and bottom-up approaches appears to support the use of 'micro' level glosses for L2 reading comprehension based on evidence that (a) low proficiency L2 readers have shown a strong preference for L1 translation glosses, over other types of annotations, as a way to

access the meaning of unknown words and structures in their own language, and (b) glossing linguistic (i.e., lexical or grammatical) information in the L2 text has been found to positively impact performance on various outcome measures of text comprehension presented to the reader after exposure to the L2 reading text.

Research studies on glossing for improved comprehension appear to bring evidence of a positive effect of glossing (and L1-translation glosses in particular) on quantitative performance measures of L2 reading. The predominant use of post-exposure performance measures of L2 reading comprehension, with their implied overemphasis on L2 reading as a quantifiable outcome rather than as a qualitative process, has raised questions about the validity of the conclusion that glosses are effective *decoding* aids drawn from these studies. In other words, if *decoding* is defined in terms of a process, as the ability to access a semantic representation for the glossed word or phrase in an L2 text, then processing measures should be utilized that allow the researcher to elicit qualitative data on decoding processes while reading. Since this kind of evidence is scarce in previous studies on glossing for improved comprehension, any claims about the decoding advantages of L1-translation glosses is assumed rather than empirically established in previous studies. For this reason, the present study addresses the need to revisit the effectiveness of L1-translation glosses as decoding aids for improved comprehension.

While the first glossing studies were primarily concerned with glosses in relation to reading comprehension, later studies have addressed glossing in relation to the input-to-intake phenomenon, that is, whether external conditions of exposure to the input can be created through glossing with the purpose to impact the ‘salience’, ‘isolatability’, or ‘noticeability’ of linguistic form while interacting with the L2 text. In the L2 learning context, the question has been raised whether glossing can be defined as a ‘focusing’ or ‘noticing’ device. Raising this question has certainly produced a shift of interest in glossing studies, from an almost exclusive focus on glosses as devices for accessing meanings and processing texts for meaning, to an interest in glosses as ‘focusing devices’ that may help processing L2 texts for form (i.e., lexical or grammatical information embedded in the L2 text). In the context of L2 learning, the question has been whether glosses can help learners notice connections between meaning and form while they read an L2 text for meaning.

As in the L2 reading context, glossing studies focusing on glossing for L2 learning make different predictions based on differing theoretical assumptions. There is an ongoing debate in the glossing research about what is the best method to attract learners’ attention to unfamiliar forms, and their meanings: Would learners attend more to form-meaning connections when meanings are given to the L2 reader via glossing or when meanings are not given to the L2 reader? Which of these two strategies would be a better ‘noticing’ device?

The issues of noticing (i.e., processing form and form-meaning connections) and learning by reading is a central question in recent glossing studies. In spite of its centrality, this question has not generally been addressed empirically in previous studies on glossing and L2 learning because qualitative measures of L2 processing, and concurrent data elicited via those measures, were generally not available in these studies. The scarcity of direct empirical procedures to measure language processing has ultimately led to all sorts of unsupported claims about the internal attentional processes that take place during on-line processing of L2 texts. For example, some researchers have speculated that ‘purely incidental’ (e.g., control) conditions of exposure to the input should ‘force conscious’ attention to unknown forms (lexical or grammatical) while others have argued for conscious processing of form under less incidental conditions (e.g., L1-translation glosses). The issue of processing L2 form (i.e. lexical or grammatical) under various glossing conditions can be settled empirically by utilizing hybrid designs, that is, designs in which both outcome and processing measures are available, and qualitative and quantitative analyses conducted. The present study addresses this gap in the glossing strand of research by incorporating a hybrid design to investigate the issue of L1-translation glosses in relation to noticing and learning. Since Schmidt’s (1995) noticing hypothesis, SLA researchers have strived in search for the factors that, by affecting attention and noticing, promote learning. It is generally agreed in the SLA literature that

this search should not be restricted to external conditions of exposure to the input (e.g., input modification techniques such as glossing). Other important factors should be considered as well. One important factor in SLA research, which has not received attention in the glossing strand of research, is the type of linguistic element.

The type of item is important in light of the hypothesis, advanced by SLA researchers, that some types of linguistic elements seem to have more chances of being attended to and noticed than others. Among the various theoretical explanations for this phenomenon, one that is increasingly attracting the attention of researchers in SLA is the notion of inherent characteristics of different types of linguistic elements. It appears that what parts or features of the language will be noticed and learned depends on the ‘inherent difficulty’ of the item. In light of this, the ‘old’ research question about glossing effects *in general* proposed in the glossing strand of research should be replaced with a new research question with more rich theoretical implications for SLA: Whether glosses may impact different parts of language differentially. Unlike previous studies in the glossing strand of research, the present study isolates type of linguistic item as a variable, to investigate the combined effect of glossing and type of item on noticing, and learning.

In sum, the present study addresses important issues not taken into consideration by previous studies on glossing. In particular, the present study

includes a hybrid design, that is, a combination of quantitative and qualitative, process and outcome, measures and techniques to investigate empirically (a) the effectiveness of L1-translation glosses as decoding aids for improved comprehension, and (b) the effects of L1-translation glosses and type of linguistic item on noticing and L2 learning.

Definition of Terms

Attention: A cognitive mechanism of information processing and storage (Tomlin & Villa, 1994). In Tomlin and Villa's (1994) model, attention has three components: alertness, orientation, and detection. While alertness (i.e., a 'readiness' to deal with stimuli) and orientation (i.e., a mechanism of inhibition and facilitation) are important for the allocation of attentional resources to incoming stimuli, only detection is responsible for the registration of stimuli in memory. In this model, detection does not involve awareness.

Awareness: The "subjective experience of some cognitive content or external stimulus" (Tomlin and Villa, 1994, p. 193).

Gloss: An annotation to a reading text with pedagogical purposes, for example, to improve comprehension of text contents or to promote learning (Nation, 2001).

Although glosses have been predominantly annotations for 'difficult' vocabulary, the concept has been extended to include all sorts of annotations (e.g., audio, image, video) for various text elements (e.g., vocabulary, grammar, paragraphs;

see Roby, 1999, for overview) and presented to the learner with various modalities (e.g., computer versus paper and pencil).

Glossing: A pedagogical technique of input modification (Leow, 2009b) consisting of creating, via the use of glosses, a non-incident external condition of exposure to the input (in contrast to a control condition) with the purpose to promote the allocation of attentional resources to the modified input.

Incidental/Non-incident conditions of exposure to the input: The conditions by which the L2 learner is exposed to the input. The term incidental/non-incident refers to the amount of information about the language presented with the input. For example, in a purely incidental condition (i.e., control), no information about the language is provided with the input. Conditions of exposure to the input are external, and they should be carefully distinguished from any internal attentional mechanism of the learner (e.g., attention, awareness, noticing).

Inherent difficulty: The difficulty of the item depending on its intrinsic linguistic characteristics, such as salience, abstractness, and complexity of encoding (DeKeyser, 2005).

Input modification: A pedagogical technique consisting on modifying the input for the L2 learner (Leow, 2009b). It includes three strands of research: textual simplification, textual enhancement, and glossing.

Input: In the context of language learning, input is the language of exposure, which is a primary source for the learner to process linguistic information (Wong, 2005).

Intake: The portion of the input that has been attended to during input processing (Leow, 1993).

Noticing: The conscious registration of stimuli (Schmidt, 1994). In Schmidt's (1994) model, noticing is defined as attention plus a low level of awareness, and it plays a critical role in intake.

Textual enhancement: A specific input modification technique consisting on the use of various types of typographical manipulation of font, character size or style in a written text to promote the allocation of attentional resources to the input (Wong, 2005; Leow, 2009b).

Textual simplification: A specific technique of input modification consisting on the simplification of authentic texts by way of reducing text length, manipulating the structure of clauses, and avoiding lexical density. Although the primary purpose of textual simplification is to help L2 learners achieve L2 text comprehension (Nation, 2001; Hatch, 1983), the term has also been defined (e.g., Leow, 2009b) in relation to L2 learning.

CHAPTER 1: REVIEW OF RELATED LITERATURE

Theoretical Foundation

Several theoretical models have influenced SLA research in the past decades: Tomlin and Villa's (Tomlin & Villa, 1994) functional model of input processing in SLA, Schmidt's noticing hypothesis (Schmidt, 1990, 1993, 1994, 1995, 2001), VanPatten's model of input processing (VanPatten, 1994, 1996, 2004), and Robinson's (Robinson, 1995) memory-based model of attention. Since these models underpin the SLA research that will be presented in subsequent sections, they are briefly summarized below.

Tomlin and Villa's model of input processing

Tomlin and Villa (Tomlin & Villa, 1994) propose a model of input processing in which attention has three components: alertness, orientation, and detection. Alertness is defined as a general readiness to deal with incoming stimuli, and is not crucial for learning. Orientation refers to internal mechanisms of inhibition and facilitation. Although orientation is the part of attention that may facilitate the allocation of attentional resources, it is not required for the mechanism that, in this fine-grained analysis, is deemed responsible for the cognitive registration of stimuli: detection. Detection is the attentional process by which particular bits of information are encoded in short-term memory, and made

available to the internal system for further processing. In this model, detection is the only mechanism responsible for learning.

An important claim of this model is that there needs to be no awareness involved during detection. Awareness is defined as the “subjective experience of some cognitive content or external stimulus” (Tomlin & Villa, 1994, p. 193). According to Tomlin and Villa (1994), previous models of attention have failed to make critical distinctions between the cognitive processes involved in learning (i.e., attention as detection), and the subjective experience of what is being attended to during input processing (i.e., awareness). Based on their fine-grained analysis, they claim that while attention, as detection, is the crucial mechanism responsible for language learning, the conscious apprehension of stimuli (i.e., awareness) is a separate mechanism that is not required for learning to take place.

The noticing hypothesis

Schmidt (1995) postulates that attention and awareness can be distinguished theoretically and empirically, but he emphatically rejects the idea of learning without awareness. Schmidt claims that it is possible to distinguish conceptually between attending to stimuli and having a subjective experience of what is being attended to in the input. Empirically, studies on subliminal perception and blind sight abound in examples of subjects who are able to detect the stimulus (in the sense of Tomlin and Villa, 1994) without being aware of it.

Schmidt claims that these are not instances of learning in any sense of the word. Learning involves more than mere perception: Learning involves intake, and noticing, defined as attention plus a low level of awareness, plays a critical role in intake.

Crucial to Schmidt's (1995) model is the notion that the processing of information in the input is determined by awareness at the level of noticing: "What learners notice in input is what becomes intake for learning" (Schmidt, 1995, p. 20). Because of this, attention and awareness cannot be disassociated in the context of learning: they are isomorphic. Since awareness plays a critical role in learning, more noticing should lead to more learning.

The noticing hypothesis is a main contribution to SLA theory. Another important aspect of Schmidt's (1995, 2001) model is the need to distinguish awareness at different levels. The model has defined two levels of awareness: awareness at the level of noticing, and awareness at the level of understanding. As mentioned above, awareness at the level of noticing is the level of awareness necessary to move the input to intake, and make the input available for further processing. Awareness at the level of understanding is a level of awareness that involves higher order thinking, though it is not deemed necessary for processes that lead to L2 development.

There is agreement among SLA researchers that awareness at the level of noticing may be enough for learning single items and simple rules, and those

portions of language that are most salient to the learner (DeKeyser, 2005; Gass & Svetics, 2003; J. Hulstijn & de Graff, 1994; Leow, 1997a). However, other parts of language may require awareness at higher levels of processing. Having ruled out any possibility of unconscious abstraction or unconscious induction of patterns and rules from the input, and given the abstract and complex nature of natural languages, Schmidt (1995) hypothesizes that awareness must, at times, work at higher levels of processing (see DeKeyser, 2005, for a different interpretation). Schmidt (1995) postulates the notion of awareness at the level of understanding to explain the work of awareness at higher levels of processing involving processes of hypothesis testing, and an analytical plan to organize complex stimuli.

In sum, attention is isomorphic with awareness in the context of learning. Awareness at the level of noticing is necessary to convert input into intake. What information is processed depends on what is in awareness during online processing. For this reason, more studies are needed in SLA to investigate the role of awareness during input processing.

VanPatten's model of input processing

VanPatten (1994, 1996, and 2004) has postulated several stages of processing to explain language acquisition. The conversion of input into intake is only the first stage. A second stage is that of *accommodation* of intake, and

restructuring, while a third stage involving processes of *access* and *monitoring* has been postulated to explain aspects of outcome or language production.

The model of input processing is only concerned with the initial stage of processing, at which input becomes intake. Not all the information in the input becomes intake under a capacity constrained view of attention. If not all information in the input can be taken in at once, the question is then what mechanisms regulate what in the input gets attended to, and processed, and why. According to this model, what gets processed is largely determined by (1) an internal drive of the learner for communication, and (2) the varying degree of ‘meaningfulness’ of different parts of language (i.e., the notion that language is not all the same in terms of ‘communicative value’). These factors largely determine that lexical meanings will be processed before learners are able to attend to purely formal aspects of the language, that is, morph-syntactic features devoid of communicative value.

The claim that processing of form is primarily constrained by an internal drive of the learner to communicate, and by the communicative value of linguistic items, has caused some controversy in the SLA field. For example, DeKeyser (2005) has argued that other internal factors such as age and aptitude may play a more crucial role in how learners process the information in the input. Furthermore, among the external factors, inherent aspects of language not necessarily related to a notion of communicative value may need to be considered

in order to explain the seemingly fact that different aspects or parts of the language are not processed and stored in the same way. Examples of inherent aspects of language that may explain differential effects on attention are abstractness of meaning/reference, complexity of the form-meaning connection, and phonological salience (DeKeyser, 2005). These underlying characteristics of linguistic features may be important in explaining why different parts of language are processed differentially. Despite the controversy around the issue of what determines differences in processing, the hypothesis of a differential effect of the type of linguistic element on attention and noticing is becoming increasingly important in SLA research (e.g., Gass & Svetics, 2003).

Robinson's model of attention and memory

Robinson's model (1995) is complementary to the noticing hypothesis (Schmidt, 1995). Robinson agrees with this hypothesis that intake is the product of noticing, but he claims that the relationship between attention and memory has not been satisfactorily characterized in the literature. In particular, previous models of the role of attention and awareness in SLA (e.g., Tomlin & Villa, 1994; Schmidt, 1995) have not explained in detail how attentional mechanisms relate to encoding and retrieval of information from the memory system.

According to Robinson (1995), learning starts when information is detected (in the sense of Tomlin & Villa, 1994), and finishes when information is

encoded in long term memory. The process of detection is important because it is a first step in the learning process: Detection activates the content of short-term memory. However, information must remain active in short-term memory (i.e., through rehearsal) long enough to exceed a certain threshold before it can be accessed by awareness. In other words, detection is a first step in the process, but is not the mechanism responsible for learning. The mechanism responsible for learning, defined in this model as “detection plus rehearsal in short-term memory, prior to encoding in long term memory” (Robinson, 1995, p. 296) is noticing. The content of intake is “what is both detected and then further activated following the allocations of attentional resources from a central executive” (Robinson, 1995, p. 297). Awareness plays a crucial role in learning because what has been noticed (i.e., detected, rehearsed, and accessed by awareness) may be subsequently encoded in long-term memory. Therefore, there is no long-term storage of information (i.e., learning) without noticing.

The first question in Robinson’s (1995) model has been for the relationships between attentional mechanisms and the subsystems of memory. The second question asked by Robinson (1995) is that of what external factors may affect attentional mechanisms during rehearsal and elaboration. Based on a review of factors among which is how the input is presented to the learner (cf. Sharwood Smith, 1980; Sharwood Smith, 1991; Sharwood Smith, 1993), Robinson (1995) concludes that what the learner is asked to do with the input

(i.e., the task) is an important factor, due to the different demands of different tasks. In response to external demands, different processing mechanisms may be initiated that differentially affect the nature of encoding. In other words, encoding may be more or less elaborated, and storage more or less permanent depending on the design features of the task. For example, directing the learner to search for rules in the input (i.e., the task) may initiate conceptually-driven processing that requires the retrieval of higher-order schemata for plans of action from memory. In contrast, directing the learner to memorize a list of items may initiate data-driven processing, that is, a less elaborated form of learning according to this framework.

In sum, unlike Robinson (1995), previous models of attention and awareness in SLA have not made explicit the relationship between attentional mechanisms and the subsystems of memory. Furthermore, awareness may play a crucial role in learning, especially during processes of encoding and storage. Moreover, the nature of encoding and storage of the information depends largely on how the cognitive demands of the task differentially affect attentional mechanisms.

Conclusion

A decade or so ago, various theoretical models were developed in the SLA field that established an uncontroversial role for attention in language learning.

An important contribution to the field has been Schmidt's (1995) noticing hypothesis. Since then, SLA researchers have strived in search for an answer to a question with important implications for language pedagogy: what are the factors that, by affecting attention and noticing, promote learning.

The search for answers to this question has included the development of the ongoing hypothesis in SLA research that attentional mechanisms may be affected differentially by, for example, the type of linguistic element. This hypothesis stems from the observation that some types of linguistic elements seem to have more chances of being attended to and noticed than others. Explanations of why this is the case differ. For example, according to VanPatten (1996), it is the variation along the parameter of communicative value what makes some items more noticeable than others due to a natural drive of learners for communication, and the constraints of attention. Alternatively, DeKeyser (2005) has pointed to inherent aspects of language as important factors in predicting what parts or features of the language will be more difficult (or easier) for the learner to attend to, notice, process, and learn.

Other factors that might affect attention are the various modes of presentation of the input (e.g., Sharwood Smith, 1980, 1991, 1993) and the tasks the learner is asked to perform (e.g., Robinson, 1995). The notion that learning can be induced indirectly or incidentally (i.e., via exposure to modified input while processing the L2 input for meaning and communication) has been the main

motivation behind the broad area of research on input modification. However, the manipulation of features of the experimental task has received less attention within this area of research.

Input modification

Theoretical underpinnings

Since the development of the theoretical models summarized above, classroom SLA researchers have been especially concerned with the pedagogical implications of the hypothesis that attention and noticing play a critical role in language learning. Learning depends to the greatest extent on whether, and how, learners pay attention to and notice linguistic forms in the input. For this reason, it is a crucial question in the SLA field whether, and how, attentional mechanisms can be affected by, external, pedagogical choices.

Pedagogical choices seem unlimited because the input can be pedagogically manipulated in various ways to create an array of techniques varying (a) with respect to the amounts of information *about* the language offered to the learner (cf. Sharwood Smith, 1980, and elsewhere) and (b) with respect to the context of the task. The more information *about* the language is provided (e.g., a grammar lesson), the more explicit is the stimulus/instructional technique. At the same time, the more communicative the context of the task (e.g., during reading a novel in the foreign language), the more incidental or unobtrusive the

technique is. Of all the possibilities, this study will focus on input modification, that is, a pedagogical manipulation of the input usually considered toward the implicit/incidental end of the continuum (see below for discussion).

Despite the importance of distinguishing the pedagogical techniques that are employed from the mental processes these techniques are assumed to affect, the SLA literature has not always separated what is internal from what is external in language learning. For this reason, it is probably wise at this point to introduce some terminological clarifications. The term *explicit/implicit*, in the context of pedagogical techniques, does not refer to a type of processing (e.g., being aware of certain stimuli versus not being aware) or to a type of linguistic representation (e.g., declarative knowledge versus procedural knowledge). *Explicit/implicit*, used in the context of input modification, only refers to how much information *about* the language is presented with the input. This notion of *explicitness/implicitness of the stimulus* should be carefully distinguished from the notion of attention/noticing (Schmidt, 1995): attention/noticing are internal mechanisms that may or may not be affected by external stimuli.

Likewise, the notion of *intentionality/incidentalness* (Schmidt, 1995) does not refer to a type of processing (e.g., being aware of certain stimuli versus not being aware) or to a type of linguistic representation (e.g., declarative knowledge versus procedural knowledge). *Intentional/incidental* does not mean *attended/unattended*, or *noticed/unnoticed* (Schmidt, 1995). Rather,

intentional/incidental refer to the type of context, that is, more or less communicative, in which the language is processed. In the research on input modification, the term *indirect* is usually used to refer to the incidentality of the conditions of exposure to the input.

As mentioned above, this study focuses on techniques of input modification within the written mode. The written input has been modified in at least three ways for the L2 learner: (1) by way of simplifying the morphosyntax and lexicon of L2 authentic texts (i.e., textual simplification), (2) by typographically enhancing various linguistic forms in written texts (i.e., textual enhancement), and (3) by way of adding annotations (also called glosses) in the L1 or the L2 for individual vocabulary words, phrases, and grammatical structures (see Leow, 2009b, for overview).

The techniques of textual simplification and textual enhancement have been overtly defined in the literature as *implicit* and *unobtrusive* strategies which may have the potential to increase the salience of linguistic forms embedded in the input. They are *implicit* because there is no information about the language that is presented to the learner with the input (e.g., Alanen, 1995; Shook, 1994). They are unobtrusive because they are utilized in the context of communicative, as opposed to language focused, tasks. Like simplified and enhanced texts, annotated or glossed texts are considered unobtrusive for the learner, and for this reason, incidental. However, less clear in the research studies on annotated texts is

how explicit/implicit the technique is. The ambiguity in this case may be due to one of two reasons. For one reason, previous studies using annotations tend not to distinguish external manipulations from internal processes, that is, claims about the explicitness of the technique are usually derived as conclusions based on participant performance on learning measures. For another reason, foreign language texts can potentially be annotated or glossed in many different ways, not to exclude information about the language (Nagata, 1999). In spite of the differences, the centrality of the notion of attention is common to input modification studies. For this reason, models that emphasize the role of attention and noticing in L2 learning have greatly influenced the research on input modification. A summary of empirical studies on input modification is presented in the next section.

Empirical studies

Simplified written input

The primary purpose of textual simplification is to help L2 learners achieve L2 text comprehension (Nation, 2001). For example, an authentic text can be pedagogically modified to shorten sentences, simplify the structure of clauses, and avoid idioms and vocabulary of low frequency to make the text more readable (Hatch, 1983). While some researchers have argued that simplified texts lose coherence and authenticity, and they do not promote comprehension (Blau, 1982;

Doddis, 1985; Young, 1999) others have supported the effectiveness of simplification for the comprehensibility of L2 texts, and of learning, as a byproduct of comprehensibility (A. Davies, 1984; Nation, 2001). See also Long (2007), on the role of texts.

In debates of learning through exposure to simplified texts, some researchers have pointed out that simplification may not be effective in promoting L2 acquisition because it removes vocabulary and grammatical structures needed to feed the acquisition system (Long & Ross, 1993; Yano, Long, & Ross, 1994). A more elaborated form to achieve comprehensible input is needed, according to this view. Despite the debates, only a few studies have addressed the effects of textual simplification empirically, by investigating the intake of linguistic forms under conditions of textual simplification (Leow, 1993; Leow, 1995; Wong, 2003).

Leow's (1993) study with 137 college students enrolled in first and fourth semesters of Spanish investigated the effects of simplification (simplified vs. unsimplified), type of linguistic item (present perfect vs. subjunctive), and language experience (first vs. fourth semester) on intake, operationalized as recognition of targets in a multiple choice immediate posttest. Before addressing the independent variables, Leow first established that the simplified text was statistically more comprehensible than the unsimplified text via recall protocols of the content matter. The results indicated an effect for language experience:

learners with more language knowledge were able to intake more forms. With respect to type of item, it was observed that inflectional morphology was especially difficult to intake by the participants in this study. Results led the researcher to conclude, “The external manipulation of the input may not only be haphazard but also inadequate to address what may appropriately facilitate learners’ intake” (Leow, 1993, page 341). A replication of Leow (1993) in the aural mode (Leow, 1995) and a study by Wong (2003) both corroborated Leow’s (1993) finding not supporting a positive effect of textual simplification on intake.

In conclusion, studies on textual simplification have been more concerned with the notion of readability (cf. Nation, 2001), and simplification as a way to improve L2 reading comprehension. Despite the learning debate on the comprehensibility effects of simplification, only three studies have addressed textual simplification in relation to measures of learning. The results of empirical analysis strongly suggest that simplification of L2 written texts may not lead to improved L2 learning.

Textual Enhancement

Several studies in SLA have enhanced the input for the L2 learner by way of typographical manipulation of font, character size, or style in a written text (Alanen, 1995; Bowles, 2003; Izumi, 2000, and 2002; Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995; Leow, 1997b; Leow, 2001b; Leow, Egi, Nuevo, &

Tsai, 2003; Overstreet, 1998; Shook, 1994; White, 1998; Wong, 2003). These studies are summarized below.

A study by Jourdenais et al. (1995) with second semester students of Spanish investigated the effects of two conditions of exposure to the input (i.e., an enhanced group versus a non-enhanced group, or control) during a reading comprehension task on the written production of targeted linguistic forms on a post exposure written recall test. Additionally, the researchers asked the participants to think-aloud all their thoughts while completing the post exposure immediate production test. The targeted structure in this study was the Spanish preterit and imperfect forms, that is, a pair of alternates used in the language to mark aspectual choices in the past. The analysis of the verbal protocols revealed that, although the enhanced group mentioned past tense forms more often than the control group, an effect of enhancement of the functional aspectual alternation targeted in the study was not observed in this measure. The analysis of the written recall protocols replicated this result: While the enhanced group produced more past tense forms in general, enhancing the input had no effect on the accurate production of preterit versus imperfect as markers of aspectual choices. From these results, the researchers concluded that input enhancement had a “priming effect” that facilitated learning of past tense. From the results of the study, the researchers concluded that textual enhancement had a positive effect on L2 learning.

Alanen (1995) conducted a study with English speakers learning beginning Finnish. The purpose of the study was to compare the relative effectiveness of more explicit conditions of exposure to the input (i.e., two rule presentation with or without enhancement groups) versus a more implicit condition (i.e., enhanced only group) versus a purely 'incidental' condition (i.e., control) for learning of targeted forms during an online reading comprehension task. The researcher used a combination of online and offline measures of recognition and production of the targets. The targets in this study were the form-function mappings of three locative suffixes, and a contextually constrained rule of consonant alternation in Finnish. The results of the study indicated a positive effect of more explicit conditions of exposure on all measures. Furthermore, the study did not support the claim that typographically enhancing the input positively affects learning of targeted linguistic forms, especially in the case of output performance: the enhancement did not make a difference when learners had to produce the grammatical targets. These findings seem to corroborate the result often observed in incidental learning experiments (Hulstijn, 1992) indicating low rates of learning under incidental, indirect, or implicit conditions of exposure to the input, especially when learning is operationalized as the production of targeted forms in post exposure tests. The result of relatively little or no effectiveness of input enhancement (cf., Alanen, 1995; Jourdenais et al., 1995) was further supported by a study by White (1998) with ESL francophone

students. In this study, enhancing (versus not enhancing) English third person singular possessive pronouns in various reading passages did not have an effect on the production of targeted items on a post exposure oral picture description task.

Leow (1997) conducted a study with college students enrolled in a second semester Spanish. This study investigated the relative effects of two variables, that is, text length (long versus short), and input enhancement (i.e., enhanced versus non-enhanced input) on reading comprehension and intake, operationalized as recognition in a multiple-choice recognition assessment task, of the targeted forms: the Spanish formal imperative. The results of the study indicated a main effect for text length on reading comprehension, but no main effect for input enhancement, and no interaction. With respect to intake, no significant main effects for input enhancement or text length, and no interaction, were observed. Besides the finding of no effect for input enhancement, this study corroborated previous findings indicating that noticing of linguistic form may not be facilitated by text comprehensibility. (cf. the previous discussion in the section on input simplification).

In a study with college students in third-semester Spanish, Overstreet (1998) predicted that content familiarity, that is, a variable usually assumed to 'easify' reading (e.g., Davis, 1989), should make enhanced input more noticeable to the learner. In other words, input enhancement and content familiarity were

predicted to interact to produce the greatest effect on L2 recognition and production of targeted forms. The forms targeted in this study were the same as in the study by Jourdenais et al. (1995): the preterit/imperfect alternation. The dependent measures of the study were a comprehension test, a written narration task (replicated from Jourdenais et al., 1995), and a fill-in-the-blank task with preterit versus imperfect choice, which was also used as the pretest. No main effects were observed for enhancement or content familiarity on either the written recall task or the fill-in-the-blank test, and no interactions were indicated between the two independent variables of the study on these measures. However, a negative effect was indicated for comprehension. In conclusion, the study did not support the effectiveness of enhancing the input for L2 learner development.

In a study with intermediate ESL learners, Izumi (2002) investigated the relative effectiveness of output production versus input enhancement during an online reading comprehension for various measures of recognition and production of the targeted form: English relative pronouns restrictions (i.e., how to use when, who, which in relative clauses). Additionally, online and offline measures of noticing (think-aloud protocols, note taking) were included in the research design. The analyses of the dependent measures in this study pointed to the conclusion that output production seems to serve as an ‘internal priming device’ that proves to be more effective when compared with conditions of exposure that only included input enhancement, or no enhancement (i.e., control). Thus, input

enhancement seems to be too indirect or ‘incidental’ a strategy as to be effective in promoting the noticing necessary for learning.

As pointed out by Leow (Leow, 1999, and elsewhere), an internal validity problem of many studies on input modification is the “failure ... to operationalize the process of attention” (p. 65). In general, conclusions about learning effects are drawn without gathering any empirical data on the attentional processes that occurred during the experimental task. For this reason, it is not possible to investigate how the participants *process* form and content while interacting with the L2 text without using sound methodologies.

The first study that addressed this important methodological issue in the input enhancement strand of research was Leow (Leow, 2001b). This study employed a more robust research design that included concurrent online think aloud protocols to elicit the reported noticing of forms under an enhanced versus an unenhanced condition of exposure to the input. The data on noticing collected from these protocols were correlated then with the data on intake, operationalized as recognition of forms in a multiple-choice test. Leow’s (2001b) study was conducted with 74 college students learning beginning Spanish. The target structure of the study was the Spanish formal/polite imperative. Besides the measure of intake, the study included measures of comprehension (i.e., a content questionnaire), and written production of forms, by means of a fill-in-the-blank task. The results of the study revealed that noticing in the enhanced group was not

significantly different from noticing in the unenhanced group. Furthermore, in both the enhanced and unenhanced groups, noticing was positively correlated with intake. However, no significant difference was observed between these two relationships. This study is important in the input enhancement strand of research for several reasons. It disentangles input enhancement as an external condition of exposure from the internal mechanisms of attention and noticing (cf. Sharwood Smith, 1980, and elsewhere). Furthermore, it explains why previous studies generally found that input enhancement did not have a differential effect on learning. Finally, the study supports empirically Schmidt's (1995) noticing hypothesis. The result of no effect for input enhancement was further supported by a conceptual replication study by Bowles (2003) with intermediate Spanish learners.

Research studies on textual input enhancement do not always solve satisfactorily the issue of the research questions they ask in relation to the methodologies they use. More than often, what should be predictions are easily turned into assumptions. Of the studies summarized above, only two (Leow, 2001b; Bowles, 2003) satisfactorily solve this issue. These studies help explain why previous studies in textual input enhancement do not provide enough evidence to support the effectiveness of input enhancement on noticing and learning of targeted linguistic forms (see Leow, 2009a, for a discussion of studies in input enhancement).

Type of linguistic item

Several studies have found evidence that item type affects attention differentially during online processing. Among the first studies in SLA comparing effects of specific linguistic features is VanPatten (VanPatten, 1989; VanPatten, 1990). In a dual-task experiment with college students learning Spanish, the researcher compared the processing of lexical information (i.e., an individual vocabulary item), versus lexico-syntactic (i.e., a definite article) versus morphological (i.e., a verb ending) during a listening comprehension task. Explicit instructions to attend to only one of the three items were given to each group. The results of the experiment showed that attending to a piece of morphology, as opposed to lexical information, negatively affected reading comprehension (the dependent measure was a recall protocol). From these results, VanPatten concluded that morphology, likely due to its lack of communicative value, has the least chances among language parts to be attended to and be noticed by learners during normal listening behavior, suggesting that language features are not processed all alike.

The input simplification study by Leow (1993; see the section on text simplification) did not find an effect for type of linguistic item. The targeted items in this study were the Spanish present perfect and the present subjunctive. It was hypothesized that the participants would attend to and take in more forms of the

present perfect than forms of the subjunctive. This prediction was based on the assumption that a more meaningful and phonologically more salient form (i.e., the present perfect) would have more effects on attention and noticing than a redundant and phonologically non salient subjunctive ending. The results of the study indicated that, although the participants with more language experience did intake more forms in general, a type of item effect was not observed, probably due to the fact that both targeted forms in the study included morphological endings. The participants did not appear to have focused on the functional connections of these endings, but on the lexical meanings encoded in the verb stems, instead. However, whether this observation was true remains to be methodologically supported given that no concurrent data were gathered.

Following up on Leow (1993), and premised on VanPatten (VanPatten, 1990) notion of communicative value, Shook (1994) investigated whether exposing the learner to items varying in ‘meaningfulness’ would have an effect on performance in a series of post exposure tests of the recognition and production of the form and function of the items involved. The experiment included a comparison between a ‘meaningful’ form that carries meanings of tense and aspect (i.e., the Spanish *present perfect*) and a form (i.e., the purely syntactic alternation of Spanish relative pronouns *que/quien*, ‘*that/who*’) that was deemed ‘meaningless’. Additionally, type of item was investigated in relation to the condition of exposure. Three conditions of exposure varying in degree of

explicitness were created in this study: (1) input enhancement plus directions to search for rules (i.e., a more explicit condition of exposure), (2) input enhancement only (i.e., a less explicit or more implicit condition of exposure), and (3) control. The results indicated main effects for condition of exposure, favoring the more explicit conditions, and type of item, favoring performance on the more meaningful item, that is, the present perfect. According to Shook, the results of the study corroborated VanPatten's (1994) hypothesis, and suggested that more explicit conditions of exposure may be necessary to direct the learner's attention to 'meaningless' forms, as opposed to more meaningful forms, for improved learning of these forms. This study is important within the input modification area of research because it is one of the few (e.g., Alanen, 1995) that included a comparison between a less implicit condition of exposure (i.e., directions to search for rule) versus a more implicit one (i.e., input modification), versus the most implicit (i.e., control).

Following up on Leow (1993) but extending Leow's (2001b) more robust methodological approach to the investigation of item effects on noticing, intake, and text comprehension, Leow et al. (2003) included as targeted forms in this study the Spanish present perfect (e.g., *ha discutido* "has discussed") versus the subjunctive (*discuta*, "should discuss"). As in Leow (1993), the prediction was that the periphrastic form should be more salient to the learner than embedded and redundant subjunctive. The results of the study were similar to those found in

Leow (2001b). The amount of noticing by the two enhancement groups was statistically similar: different input enhancement conditions did not have a differential effect on noticing. However, the noticing was highly correlated with the intake, regardless of enhancement group: more noticing led to more intake (Schmidt, 1995). With respect to item effects, the study found an effect for item on noticing (present perfect noticed more substantially than subjunctive), but no effect on outcome measures (i.e., recognition and text comprehension).

Overall, the studies that isolated type of linguistic item as a variable have contributed empirical evidence to the hypothesis that different parts of language are not processed in the same way. While at the same time providing evidence for theoretically founded hypotheses, the results of an effect for type of linguistic item on L2 processing also serves to explain why there are some mixed results in the input enhancement strand of research. For example, one of the conclusions of the study by Jourdenais et al. (1995) was that typographical enhancement had had a ‘priming’ effect on the ‘intake’ and production of past tense in general. However, while the enhanced group produced more past tense forms in general, enhancing the input had no effect on the accurate production of preterit versus imperfect as markers of aspectual choice. Under the assumption of a type of linguistic item effect, this finding is easily explained by the fact that the form-to-meaning connections implicated in the expression of aspect are far more obscure to the learner than the mapping between a form and the meaning [+past tense] (cf.

DeKeyser, 2005). It should be noted at this point that the notion of communicative value could not be used to explain this finding because the rationale behind this notion would predict that the Spanish morphological marker of past tense should not be processed, and noticed, by beginning L2 learners given its redundancy. For this reason, the notion of inherent difficulty (DeKeyser, 2005; see also Gass and Svetics, 2003) of the item, based on parameters that make explicit the degree of transparency of the form-to-meaning connections involved is a more robust and powerful concept to explain differential processing effects of item type.

To summarize to this point, some but not all of the studies on textual input enhancement isolate type of linguistic item as a variable. Studies that do so provide evidence in support of the claim that language features are processed differentially by learners. In general, studies that have isolated the variable type of linguistic item tend to support the claim that explicit conditions of exposure are more effective for noticing and learning of linguistic forms, and especially for noticing and learning some forms instead of other forms. Furthermore, a notion of difficulty that makes explicit the different nature of different form-to-meaning relationships (e.g., DeKeyser, 2005) seems necessary to give an account of why some forms are more noticeable than other forms, or why different conditions of exposure seem to affect some connections but not others.

Glossing

Researching input modification by way of adding annotations for linguistic information, namely morphology, syntax, and the lexicon, to an L2 text is becoming increasingly important in the SLA field. However, studies within the glossing strand of research have had different scope and interests. For example, it is characteristic of this strand of research within input modification to treat text comprehension and L2 learning by reading as separate issues. The reason for this lies in the history of this research strand.

The first glossing studies in the 80's and 90's focused on glossing in relation to text comprehension and the implications of these effects for reading theory. The interest in investigating glossing effects in relation to the issue of learning the L2 while reading, that is, in the context of 'incidental learning', developed later. For this reason, the first studies do not include measures of L2 learning in their research designs. At the same time, these later studies on glossing in relation to learning the L2 while reading require that a reading comprehension task be included in the research design to insure that the focus of the processing is on meaning. However, these studies do not necessarily *measure* reading comprehension. In sum, while the studies primarily interested in glossing in relation to text comprehension do not include measures of L2 learning, the studies focused on L2 learning while reading do not always *measure* reading comprehension, even if they include a reading comprehension task in their

research design. Since reading comprehension and L2 learning tend to be treated as two separate issues in the glossing strand of research, the research studies investigating the effects of glossing on text comprehension are summarized and discussed first, before the research on glossing in relation to the issue of L2 learning through reading.

Effects of glossing on L2 reading comprehension. Theoretical underpinnings

Input modification by way of adding annotations for individual vocabulary words to L2 reading texts is a common pedagogical practice. The primary purpose of this practice is to bridge the gap between the L2 reader's characteristic lack of linguistic knowledge and the L2 text. Additionally, glosses for lexical information are usually viewed as unobtrusive devices that make text contents more accessible to L2 learners without diverting them from reading, that is to say, glosses may help avoid dictionary consultation.

The purported benefits of glossing lexical information in L2 texts are supported by the premises of a bottom-up theoretical view of L2 reading (e.g., Birch, 2002). This view predicts that these 'micro' level annotations (i.e., short definitions or explanations in the L1 or the L2) should help L2 readers overcome the decoding problems they face due to lack of lexical and grammatical knowledge. In this context, glosses are viewed as decoding aids. By providing the

meanings of unknown or unfamiliar words, glosses should make text contents more available to L2 readers. In contrast, exposing the learners to non glossed texts containing words that are difficult to decode will force them to guess the meaning from context, or to use background knowledge to try to compensate for decoding difficulties. From a bottom-up point of view, however, strategies such as these are not effective because (1) learners vary greatly in inferring ability, (2) inferences are not reliable, and (3) background knowledge may not be available (Birch, 2002; Hulstijn, 1993; Nation, 2001).

While the purported benefits of glossing lexical information in L2 texts are supported by the premises of a bottom-up theoretical view of L2 reading, the advantages of glossing vocabulary are put into question by other theoretical views of L2 reading, for various reasons. For example, the use of these annotations is not supported by top-down approaches to L2 reading (Goodman, 1973; Smith, 1973) which are premised on the idea that decoding of every word is not needed for text comprehension: Readers overcome decoding deficiencies by using strategies such as guessing from context or using background knowledge. From this point of view, glossing lexical information may hinder, rather than facilitate, text comprehension.

While the bottom-up perspective makes a strong emphasis on paying attention to the linguistic aspects of the text, the top-down view stresses the role of readers' knowledge, and de-emphasizes a focus on language. There is, yet, a

third view of the reading process that emphasizes the idea that the L2 reading process is bidirectional. This is the view that a focus is necessary at both the micro level of words, and the macro level of text (Bernhardt, 1991). It should be noted at this point that the glossing studies that are based on a bidirectional view of L2 reading are also based on theories of learning with multimedia in a computerized environment. In this context, the dual-coding theory (Paivio, 1969; Sadoski & Paivio, 2001) has been especially influential. This model maintains that exposure to dual-stimulus (i.e., exposing the learner to meanings encoded in both symbolic, as in a definition, and analog, as in a picture, format) strengthens memory representations. The idea is that dual-stimulus encourages short term processing and encoding, and long term memory storage of mental representations for word forms in *both* a symbolic and an analog format. This has been the main motivation for the inclusion of a comparison between glosses for vocabulary items in the traditional sense mentioned earlier (i.e., short definitions in the L1 or the L2) versus these same glosses with the added purported benefits of multimedia. In short, glossing studies based on the ‘bidirectional’ view of L2 reading (Chun & Plass, 1996; Lomicka, 1998) support annotations at the micro level of words. However, theories favoring dual (i.e., verbal plus visual) stimulus have greatly influenced glossing studies that have had a ‘bidirectional’ view of L2 reading. For this reason, these studies tend to *disfavor* the use of micro level glosses *in the traditional sense of short definitions in the L1 or L2*.

With respect to the use of multimedia at a macro level (e.g., pictures and video accompanying textual contents), glossing studies are based on Mayer's (Mayer, 1984) generative theory of learning with multimedia and hypertext. This model maintains that some mental representations activated during online reading have an analog format. From this perspective, understanding a text is like having a mental 'picture' of the text content and its subject matter. Thus, reading multimedia annotated texts should foster a "direct" construction of a mental model (Schnotz, 1993, p. 248). By contrast, reading texts not annotated with multimedia would require an extra step because a symbolic stage of representation must be achieved before the text information can be integrated with analog representations.


To summarize to this point, the practice of glossing individual vocabulary words in an L2 reading is popular among curriculum developers, the editors of textbooks, as well as among teachers and students. In spite of the popularity of traditional glosses, theoretical views of L2 reading make differing predictions about the purported benefits of these types of glosses. While bottom-up views support the use of vocabulary glosses for text comprehension, both top-down and bidirectional approaches do not, for different reasons. From a top-down perspective, the use of hypertext annotations, and advance organizers that would activate learners' background knowledge are preferred instead of traditional glosses. From a bidirectional approach, both the micro and the macro level are

important. However, since the glossing studies premised on a bidirectional view of the L2 reading are also premised on theories of learning with multimedia, micro level glosses are redefined in these studies to include multimedia (i.e., images, and sound). Based on these differing underpinnings, it is not surprising to find that vocabulary glosses in the traditional sense (i.e., short definitions in the L1 or L2) are compared with other types, at both the micro level (i.e., definitions accompanied by other media), and at the macro level (i.e., hypermedia annotations).

Effect of glossing on text comprehension: Empirical studies

Several studies have investigated the effectiveness of using traditional vocabulary glosses for improved comprehension. (Table 1 presents a summary of studies on glossing and reading comprehension).

Table 1. Summary Of Studies On Glossing And Reading Comprehension

Study	Control		Def. of Terms		Attention	Results
			Gloss	Comprehension		
Johnson (1982) N=72	No	Yes	Marginal definitions for individual words	Written recall protocols Cloze test Multiple choice questionnaires	No	No effect for gloss
Pak (1986) N=65	No	Yes	As in Johnson (1982)	Cloze	No	No effect for gloss
Davis (1989) N=71	No	Yes	Marginal, online, annotations (different types)	Written recall protocols	No	Effects for glossing (versus control)
Jacobs, et al. (1994) N= 85	No	Yes	L1 versus l2 word definitions	Written recall protocols	No	Interaction between glossing and language level
Jacobs (1994) N=116	No	Yes	L1 translations for words	Written recall protocols	No	Effect for glossing versus control
Bell & Le Blanc (2000) N= 40	Yes	No	As in Jacobs et al. (1994)	Multiple-choice questionnaire	Tracker	No effect for language of the gloss

Bowles (2004) N= 50	Yes	Yes	Computer versus paper	Multiple-choice questionnaire in the 11 & Think-aloud protocols	Tracker	Effect for glossing (versus control)
Davis and Lyman- Hager (1997) N= 42	Yes	No	Definitions for words	Written recall & multiple-choice	Tracker	No effect for verbal glossing
Lomicka (1998) N= 12	Yes	No	As in davis et al. (1997)	Think-aloud protocols	Tracker	No effect for verbal glossing
Chun and Plass (1996, 1997) N= 103 (study 2)	Yes	No	Same as in lomicka (1998) and davis et al. (1997)	Written recall protocols	Tracker	No effect for verbal glossing

While some researchers predict that this type of glosses should have a positive effect on comprehension (Bell & LeBlanc, 2000; Bowles, 2004; Davis, 1989; Hulstijn, 1993; Jacobs, 1994; Jacobs, Dufon, & Hong, 1994; Roby, 1991; Roby, 1999), others make the opposite predictions (Chun & Plass, 1996; Chun & Plass, 1997; Davis & Lyman-Hager, 1997; Davis & Lyman-Hager, 1997; Johnson, 1982; Lomicka, 1998; Lyman-Hager, Davis, Burnett, & Chennault, 1993). These studies are summarized below.

Johnson (Johnson, 1982) investigated the effects of exposure to traditional vocabulary glosses (operationalized as definitions on the margin of the text), content familiarity (operationalized as having a *schema* or prior knowledge representation of the text content in memory), and pre-reading vocabulary instruction on text comprehension by ESL learners. Comprehension was measured in this study by a written recall task, a content-based recognition test, and a 3-week delayed cloze test. Results indicated positive effects for content familiarity and for pre-reading vocabulary instruction (for familiar passages), but no effects for marginal glosses. Exposure to vocabulary glosses during reading might have encouraged the participants to approach the reading task from a bottom-up processing perspective, setting up favorable conditions for word-by-word reading. The conclusion of the study was that focusing on the linguistic aspects of the text at the micro level of words might have a detrimental effect on

text comprehension. In contrast, pedagogical strategies that encourage top-down processing, as for example the activation of background knowledge during reading, are more recommendable for L2 readers.

In a follow-up study, (Pak, 1986) investigated whether the use of vocabulary glosses, operationalized as ‘short explanations’, would facilitate or hinder reading comprehension. In this study, ESL learners of two aptitude levels, low and high reading ability, were randomly assigned to three conditions: (1) marginal glosses, (2) ‘elaboration’ glosses (i.e., word definitions embedded in the passage), and (3) control. The analysis of the text comprehension measure (i.e., a cloze test) indicated an effect for aptitude, no effect for treatment, and no interactions: High reading ability learners comprehended the text better regardless of gloss availability. Like the study by Johnson (1982), Pak (1986) does not support the use of vocabulary glosses for L2 reading comprehension.

Davis (1989) investigated the effects of background knowledge, operationalized as annotations provided before reading the text, and glosses (i.e., annotations accessible during reading) on reading comprehension. The study was conducted with US college students in third semester French. Davis asked the participants to read a text under three conditions: (1) with no aids (but with instructions to ‘read-write-then reread’), (2) with definitions of vocabulary items plus question and comments presented before reading, and (3) with the same questions and vocabulary definitions as in (2) but in a marginal gloss format, that

is, glosses presented during reading. Reading comprehension was measured via a written recall protocol in the L1 (i.e., English). Results indicated an effect for glossing (versus control), but no difference between pre-reading and online reading. Thus, participants exposed to two types of annotations (i.e., vocabulary glosses and informational glosses) read the text better, regardless of method of exposure (i.e., pre- versus on-line exposure). The conclusion in this study was that online glosses, including vocabulary glosses, have a positive impact on L2 reading comprehension.

With a focus on the micro or word level, Jacobs (1994) investigated the effect of vocabulary glosses on reading comprehension. The measure of comprehension was a written recall protocol. The participants in this study, that is, third-semester US college students of Spanish, were randomly assigned to one of two groups: participants either read a text with vocabulary glosses, operationalized as English translations for individual word items placed in the margin of the text, or read the same text without vocabulary glosses. Results indicated an effect for glossing (versus no-glossing), suggesting that vocabulary glosses have a positive impact on text comprehension, thus supporting the bottom-up view of the reading process. Furthermore, this study found no interaction effects between glossing and several learner's variables among which were aptitude, attitude, psychological type, and language level.

Vocabulary glosses were also operationalized as translation of words into the learners' L1 in two computerized studies with high school Dutch learners of English as a foreign language (N=44, for study 1; N= 38, for study 2) by Hulstijn (Hulstijn, 1993). Although they included glosses, these studies were not studies on the impact of glossing on text comprehension. Rather, the study investigated the internal and external factors that might influence the learners' look-up behavior (i.e., the learner's own decision to consult, or not to consult, the gloss). In spite of this, the study by Hulstijn (1993) has been interpreted in the glossing literature as a study that supports the use of vocabulary glosses for reading comprehension (cf., e.g., Bell & Le Blanc, 2000). This interpretation was based, perhaps, on the observation that all participants in this study "answered [comprehension] questions correctly" (Hulstijn, 1993, page 143), suggesting that giving the option to freely access a translation gloss did not impede, and it might have facilitated, comprehension. However, it should be noted that (1) no statistical analysis of the comprehension questionnaire was available, and (2) it was observed that using a 'maximal' strategy (i.e., looking-up more words) or using a 'minimal' strategy (looking-up less words) "were not related to qualitative differences in the product (answers to the comprehension questions)" (Hulstijn, 1993, p. 146).

The results of the analyses conducted by Hulstijn (1993) indicated that the look-up behavior of the participants was greatly influenced by the task goals, the

relevance of the words in the text, and the students' own aptitude and learning styles. Based on these findings, the researcher concluded that readers should be instructed in all available options for strategic reading comprehension, including but not limited to vocabulary glosses. This study is important because it was the first study in the computerized context that included a tracking mechanism to register and compile in log files the actual access to glosses by learners (i.e., look-up behavior). The use of this mechanism in glossing studies will be discussed later.

Given that individual vocabulary items can be glossed in two ways (i.e., the native or the target language), some researchers have included a comparison between L1 and L2 glosses in their research designs (e.g., Jacobs, Dufon, and Hong, 1994; Bell & LeBlanc, 2000). Jacobs, Dufon, and Hong (1994) set out to investigate the effects of language of the gloss, L1 versus L2, on reading comprehension, as measured by written recall protocols. The participants in this study, that is, fourth semester college students of Spanish, were randomly assigned to one of three groups: (1) no gloss, (2) English glosses, and (3) Spanish glosses. At the end of the experiments, participants also completed an exit questionnaire that probed their glossing preferences. The results of the study indicated no effect for glossing on recall, but a trend existed "favoring students who had glosses" (p. 26). Furthermore, an interaction effect was found between glossing and language level: participants with higher than average proficiency

recalled more of the text if they had read the text with glosses. With respect to glossing preferences, the participants preferred glossed to non glossed texts, and glosses in the L2, “if they can understand them” (p. 26). From these findings, the researcher concluded that using vocabulary glosses is beneficial to L2 readers, but that a threshold of L2 competence may be needed to be able to use the glosses effectively.

Following up on Jacobs, Dufon, and Hong (1994), Bell and Le Blanc (2000) investigated the effects of language of the gloss, L1 vs. L2, on comprehension. The participants in this study, that is, English speaking students in third semester Spanish courses, were assigned to one of two groups: L1 gloss or L2 gloss. A tracker mechanism was used to register the learners’ access to glosses. Results indicated that the difference between group means on the dependent measure (i.e., a multiple-choice comprehension questionnaire) was not significant. The conclusion was drawn from this study that language of the gloss did not have an impact on reading comprehension.

Besides the language of the gloss comparison, other parameters have been used by studies that focused on glossing at the level of words in relation to text comprehension. For example, studies have investigated the relative effectiveness of computerized vocabulary glosses vis-à-vis the paper-and-pen technology (Bowles, 2004; Lyman-Hager et al., 1993), or the effects of glosses in relation to dictionaries, or both (Aust, Kelley, & Roby, 1993; Roby, 1999).

Roby (1999) investigated the effectiveness of gloss presentation mode (i.e., computerized versus paper-and-pen glosses) and type of glossing support (dictionary versus dictionary plus vocabulary glosses) on text comprehension by US college students of Spanish. It was observed that, although the addition of glosses to dictionary entries helped speed up the reading task, no effects were found for glossing. Furthermore, the study indicated a strong preference of participants for computerized aids. The finding that learners look up more words when they are working on a computer replicated the findings of a prior study by Aust, Kelley, and Roby (1993). From these results, the researcher concluded that glosses and computerized glosses in particular, “would appear to lessen the disruption of the reading process caused by conventional dictionary look-ups” (Roby, 1999, p. 98), therefore supporting the use of vocabulary glosses for L2 reading comprehension.

While previous studies demonstrated a clear pattern in learner’s preferences for reading on computers (Roby, 1999; Aust et al., 1993), the relative effectiveness of computerized glosses versus paper-and-pen glosses for reading comprehension is less clear. Bowles (2004) recently revisited this issue in a study on glossing utilizing think aloud data elicitation technique in the methodology of the research design. Following up on a study by Lyman-Hager, Davis, Burnett, and Chennault (1993), Bowles (2004) set out to investigate whether exposure to vocabulary glosses (i.e., operationalized as translations into the learner’s native

language) on a computer would be more effective than exposure to the same L1 glosses presented via paper-and-pencil technology for measures of noticing, reading comprehension and vocabulary learning. The participants in this study were US college students of beginning Spanish. Comprehension was measured by means of a multiple-choice comprehension questionnaire in the L1 (i.e., a quantitative measure), and by think-aloud protocols (i.e., a qualitative measure). Learning was measured by recognition and productions posttests. The analyses of the dependent measures indicated a significant difference for glossing, with the gloss group outperforming the control, but no difference between glossing modes (i.e., reading computerized versus non computerized glosses). Thus, learners who were exposed to vocabulary glossed (regardless of technological mode) understood the text better and learned more targeted words than learner not exposed to these glosses did.

So far, this survey of previous research has included studies focusing on the effects of glosses versus no-gloss or control condition, and comparisons among different glossing techniques (e.g., L1 versus L2, computerized versus non computerized, dictionary versus glossing) at the level of glossing of vocabulary items. Beside these, other studies, theoretically grounded on a bidirectional view of the L2 reading process, have been especially concerned with the compounded effects of different types of glossing techniques, including but not limited to

traditional annotations for individual words, on processing of texts at different levels.

Three studies on glossing have addressed the predictions of theories of learning with multimedia at the micro and macro levels (Chun & Plass, 1996; Chun & Plass, 1997; Davis & Lyman-Hager, 1997; Lomicka, 1998). These studies are summarized below.

Premised on the idea that “L2 readers require much more informational support than word definitions” (p. 60), Davis and Lyman-Hager (1997) set out to investigate the relative effectiveness of ‘word definitions’ vis-à-vis different types of textual annotations on text comprehension. To this end, intermediate US college students of French read a computerized text that not only included word definitions in English (L1) and French (L2), that is, traditional vocabulary glosses, but also grammatical explanations, aids for intratextual reference relations, pictures and cultural notes. Text comprehension was measured by means of an immediate free-response written recall protocol, and a multiple choice test on the text content. As in previous computerized studies (e.g., Hulstijn, 1993; Bell et al., 2000), a tracking mechanism registered the look-up behavior of participants. Results indicated that, despite the variety of glosses available, 85 per cent of the consulted information was English translations of individual words. Participants seem to have the perception that accessing words meanings in their native language is a “key” factor (Davis et al., 1997, p. 62) in text comprehension.

Furthermore, it was observed that preference of gloss type (i.e., translation of words into English versus all other types) was not statistically correlated with text comprehension.

The researchers mentioned two factors to explain why glosses in this study did not seem to have facilitated text comprehension. One factor was that the participants did not access other glosses: “Comprehension might have improved had other program components also been accessed” (Davis et al., 1997, p. 58). Furthermore, the researchers speculated that there might have been a ceiling effect related to language level: the text was so difficult that glosses became useless. Based on these findings, the researcher concluded that more training in reading multimedia annotated texts is needed for students to be able to enjoy the benefits of reading in computerized environments.

Lomicka’s (1998) pilot study with English speaking learners of second semester French used online concurrent data elicitation techniques (i.e., think-aloud protocols) to study inferential processes during online reading. There were three conditions: (1) an ‘all gloss’ group with French definitions, images, references, questions, pronunciation, and English translation; (2) a ‘traditional gloss’ condition with French definitions and English translations, and (3) a ‘no gloss’ or control group. ‘Inferential process’ was operationalized as the number of explanations and other types of comprehension processes (e.g., paraphrases, predictions, associations) provided in the think-aloud protocols, which were the

only measure of reading comprehension in this study. A tracking mechanism was included to register 'look-up' behavior (i.e., number and type of glosses accessed, and gloss consultation time).

The results of the study indicated no differences among groups in percentages of explanations generated in the think-aloud protocols. The data on other types of comprehension processes, besides explanations, were excluded from the analysis due to the low number of exemplars available. With respect to the tracker data, it indicated a strong preference for translation glosses, even when multimedia glosses were available. Think-aloud and tracker analyses point to the conclusion that students "seem to be generally satisfied with achieving a minimal level of comprehension via quick and basic translation" (p. 49). Thus, with respect to the effects of different types of glosses on text comprehension, this study replicated previous findings by Lyman-Hager and Davis (1996), and pointed to similar conclusions about the need to instruct learners in the use of hypertext and multimedia annotations.

The predictions of multimedia learning at the micro and macro levels were investigated by Chun and Plass (1996) in three studies (N = 36 for study 1; N = 103 for study 2; and N= 21 for study 3) conducted with second-year university students studying German as a foreign language. In these three studies, participants read a story annotated with a multimedia reading software that included three conditions: (a) items for which access to visuals (i.e., either a

picture or a video) and access to traditional vocabulary glosses were available, (b) items for which only traditional glosses were available, and (c) no glosses available. Furthermore, in order to address the effects of annotations at the macro level, the researchers included a pre-reading activity that consisted in watching a video preview of the content of the text. This video was predicted to serve as an ‘advance organizer’, that is, as a way to give participants some knowledge of the content of the text that they could activate during online reading. All participants were exposed to this video prior to the online reading task. Then, a comparison was established in the dependent measure (i.e., written recall protocols) between frequency of propositional units that encoded information that appeared in the video preview and frequency of propositional units that did not appear in the video.

With respect to the glossing effects at the micro level, the results of the study indicated the biggest effect for exposure to dual stimulus, a smaller effect for no annotated items, while the effect for traditional glosses was the smallest. The effect for dual stimulus provides evidence for dual coding (Paivio, 1969). With respect to the glossing effects at the macro level, the recall protocols indicated that the participants mentioned information contained in the video preview more often than they mentioned information not contained in the video preview, suggesting - according to the researchers - that video previews may have an impact on text comprehension as a whole.

The researchers' general conclusion from these experiments was that while multimedia vocabulary glosses (i.e., presenting words via dual-stimulus) strengthen bottom-up processing, dynamic advance organizers seem to impact top-down processes of reading comprehension. In conclusion, different types of annotations are needed to affect processes of text comprehension at different processing levels.

This survey of previous research has presented the findings of studies investigating the effects of glossing for improved comprehension. Predictions vary across studies driven by different theoretical views of the L2 reading process. However, the empirical evidence points toward the following conclusions. First, there is a strong preference for traditional definitions in the language of the learner (e.g., Jacobs, 1994; Jacobs et al., 1994; Bell et al., 2000; Davis et al., 1997; Lomicka, 1998). Second, there is some evidence that these types of glosses are effective to help L2 readers comprehend L2 texts (e.g., Jacobs, 1994; Jacobs et al., 1994; Bowles, 2004). Finally, studies that predict no effects for this type of glosses do not provide enough evidence to support this prediction (e.g., Johnson, 1982; Davis et al., 1997; Chun & Plass, 1996; Lomicka, 1998).

Effects of glossing on L2 learning while interacting with an L2 text: Theoretical underpinnings

Like the studies on textual input enhancement, recent studies in the glossing strand of research are interested in input modification techniques in relation to the processing at the micro level of lexical and grammatical information while interacting with an L2 text, and in particular the issue of noticing and awareness of items in an incidental context¹. The issues of learning by reading, and processing linguistic form while reading for meaning are central questions in recent glossing studies. However, the lack of empirical procedures to measure language processing has ultimately led to all sorts of unsupported claims about the implicitness/explicitness of the internal processes that take place during on-line reading of L2 texts. For example, some studies have assumed that *vocabulary that has been learned is vocabulary that has been consciously attended to while reading* (Gettys, Imhof, & Kautz, 2001; Hulstijn, 1992; Hulstijn, Hollander, & Greidanus, 1996; Nagata, 1999; Watanabe, 1997) while others have assumed the opposite, that is, that new learned words were ‘picked up’ *unattended or unintentionally* during reading (see Coady & Huckin, 1997; N. Ellis, 1994; Schmidt, 1995, for discussion). Thus, an important underpinning of recent glossing studies is that they are theoretically premised on attention, although they do not always include adequate procedures to measure noticing and awareness during online processing of L2 texts.


Glossing studies not only differ in terms of the conclusions they draw but also in terms of the assumptions and predictions they make about the conditions of exposure included in their research designs. In other words, there is an ongoing debate in the glossing research about what is the best method to attract learners' attention to unfamiliar word forms, and their meanings: Would learners attend more to word forms-to-meaning connections when word meanings are given to the L2 reader via glossing or when word meanings are not given to the L2 reader? Which of these two strategies would be a better 'noticing' device? In the context of the debate on glossing, some researchers have argued that not giving the meanings of unknown words (i.e. a control or 'purely' incidental condition) to L2 readers is better because, under this condition, learning is like a problem solving task (see Hulstijn, 1992, for overview of this postulation). In order to solve this task, the L2 reader must try to guess the meaning of unfamiliar words from the context. In order to guess meaning, the L2 reader must make inferences. Inferential processes should "force" the L2 reader to pay conscious attention to words: While the word form is under focal attention, the L2 reader is searching for meanings to attach to that form. That is the postulation of the "mental effort hypothesis" (Hulstijn, 1992). The prediction is, then, that the holding in memory of the connection between the new word form and its inferred meaning should facilitate recall. Advocates of glossing emphatically reject this view by pointing out that learners vary greatly in inferring ability, and inferences are not reliable

(e.g., Watanabe, 1997; Nagata, 1999; Gettys et al., 2001). Therefore, when the glossing debate is presented as a dilemma between two options, many researchers prefer to gloss than not to gloss. Once the issue is settled in favor of glossing, the next question is how texts should be glossed. The research strand of glossing has moved from the broad question of ‘to gloss or not to gloss’ to the narrower question for the relative effectiveness of different types of glosses for the L2 reader. In this context, the effects of various types of glosses, including but not limited to traditional definitions or explanations in the L1 or L2 on the margin or at the bottom of the page, have been investigated in relation to various measures of L2 learning. These studies are summarized below.

Effects of glossing on L2 learning while interacting with L2 texts: Empirical studies

Several studies have investigated the effectiveness of glossing in relation to the issue of learning through reading L2 texts (Table 2 presents a summary of studies on glossing and L2 learning).

Table 2. Summary Of Research Studies On The Effects Of Glossing On L2 Learning While Reading

		Target	Op. def. 'incidental learning'	IV	Learning measures	Comprehen. measures
Hulstijn/92 Exp.1 N= 65 Turkish learners of Dutch	no	12 lexical words	Reading comprehension task + unannounced posttests.	4 gloss groups	Recog. and prod.	Six multiple choice comprehension questions (exp. Task)
Results. A. Recognition. Translation gloss diff. from the other groups. B. Production. No diff. observed C. Overall, the outcome means are very low (3.1/12 and 1.9/12). Low learning rates under all conditions. D. High percentage of wrong answers						
Exp. 3 N= 45 Sample=Dutch L2	no	Same	Same	Two gloss groups	Recogniti on word meanings put in their original context	Same
Results. A. Higher performance overall B. Effect for the MC condition C. high percentage of wrong answers						
Experiment 5 N= 35 Dutch learners	No	Same	comprehension task, unannounced recog. post-tests; and announced posttests.	Same	Recog. and prod. tests	Same
Result A. Higher performance overall with intentional tests. No diff. between gloss and MC in any of the intentional tests (rec. and prod.) B. Incidental tests: diff. for CMC.						
Jacobs et al., 1994 N= 85 Spanish	no	32 words	No	L1, L2 gloss plus control	Recog. (repeated 4-Week later)	Written recalls in L1
Results. Sig. effect for treatment on the imm voc test; no sig. diff. in the delayed posttest						

Hulstijn et al., 1996 N= 78 LF French	no	16 words	Reading task; Unannounced tests	L1-gloss versus dictionary; frequency of exposure	Recog. and prod. Posttests.	Comprehension questions
---	----	----------	------------------------------------	---	-----------------------------	-------------------------

Results *Rec*: frequency effect. No effect for gloss and no interaction. [low AVG reported dic. use]. *Recall (words in isolation)*: Effect for MG, frequency and frequency X. group *only for the full point score*. [high retention rate when words are looked up in a dic.] (Knowledge of words in context). [higher retention rates overall] . Effect for group (MG better), frequency effect, and interaction only for full point (i.e., exact response or fully correct response).

Chun & Plass, 1996 N= 103 (study 2) German, second year	yes	82 words with various annotations	Reading task. Learners are not told that they will be given a vocabulary test at the end.	L1-gloss vs. text+picture, vs. text+ video	Recog. and prod.	-Recall protocol in L1 English - Look up behavior
---	-----	-----------------------------------	---	--	------------------	--

Results. - higher rates of recall overall (in comparison to other studies). 24-26% [no control] - diff. favoring the text+picture, better than text only, and better than text+video. Variation in the types of glosses looked up appears in the look-up data. Maybe amount of exposure effect

Watanabe, 1997 N= 231 ESL	No	16 words	Reading task + translation task; Comprehension questions during reading; unexpected vocabulary posttest.	L2-gloss versus appositives versus MC versus control; Transl. task	Prod. posttest & 2 delayed posttests	1. Cloze test (to measure proficiency, and as distractor) 2. 5 Open-ended comprehension questions.
---------------------------------	----	----------	--	--	--------------------------------------	---

Results. -The gloss group scored sig better on the comprehension test than the original group - Low rates of learning overall.

Nagata, 1999 N= 26 Japanese	yes	20 words; 3 Gram. feature	Announced reading comprehension test; unannounced posttests.	L1-gloss versus MC+feedback; frequency (4 levels)	Pretest = translation task Post-test = pretest = delayed (1 month later)	no
-----------------------------------	-----	---------------------------	--	---	---	----

One month later: drop in recall. Sig. diff. between single and MC only for grammar. Feedback did not solve the problem

Gettys et al., 2001 N= 22 College students in a second-year Russian course	yes	43 lexical words	Reading task on computer. Were told that they would be asked 'about what you read', instructions on the test may direct students to believe that they will be word tested	L1-gloss versus dictionary form	Recog. posttest and time on task	global comprehensi on (recall protocols) ability to capture details (multiple choice questionnair e)
Results Effect for time: the dic form group spent more time reading the text than the exact translation group.						

Hulstijn (1992) conducted five studies on glossing within the broader context of incidental learning of vocabulary in L2 Dutch. Of these five studies, only the three studies that were conducted with L2 learners of Dutch are relevant here (i.e., studies I, III, and V). Experiments I and III were similar. In experiment I, the participants read a text in Dutch under four conditions: (1) marginal translation glosses, (2) a marginal sample sentence containing the target word, (3) the translation gloss embedded in a multiple choice task, and (4) no gloss. The experimental task included a reading comprehension task (i.e., six comprehension questions that the participants had to try to answer as they read). This task was included as a strategy to insure that the participants were processing the text for its content and meaning, instead of being focused on the targeted forms. Additionally, another strategy used in this experiment to insure the 'incidental' or 'non-intentionality' of all conditions in the experiment (i.e., all the gloss, and

the no-gloss conditions) was that the participants were not asked to memorize words, and they were not told in advance that they would be tested on the targeted items. The dependent measures in this study were a post exposure immediate recognition posttest of a list including the original targeted words from which the participants had to provide the meaning, and a production cloze test. The results of the study indicated an effect favoring the translation gloss on the recognition test. This difference was not observed on the production test. Hulstijn (1992) also reported the overall rates of learning under the incidental conditions of exposure included in the experiment. It was observed that learning under incidental conditions of exposure (including glosses) yields “very low” means (see Chart 1). However, learning under non glossed conditions resulted in high percentage of wrong answers. From these findings, the researcher concluded that although glosses may not be the most effective technique to help L2 learners acquire new words, it might help avoid making wrong inferences.

Experiment III was similar to Experiment I, except that it included glosses in Dutch, instead of glosses in the L1 (Turkish), which might have influenced the finding of an effect favoring the multiple choice condition (i.e., the L1 glosses in experiment I seem to be more effective than glosses in the L2). Unlike experiment I, experiment III included a recognition test in which the targets appeared in the original context. The result of higher rates of learning overall in this experiment need to be interpreted with caution because of the test effects caused by the

inclusion in the posttest of original contextual information from which the meaning of the target could have been inferred.

Unlike experiments I and III, experiment V had a unique feature: It included a comparison of intentional versus non-intentional (i.e., incidental) conditions of exposure to the input. The intentional condition consisted on manipulating the experimental task: The participants in this condition were told that they would be tested on their knowledge of all glossed words. The results of this experiment indicated higher performance on the intentional tests in all dependent measures (i.e., recognition and production). No measures of comprehension were reported.

Jacobs, Dufon, and Hong Fong (1994) conducted a study with fourth semester college students learning Spanish. The participants in this study were randomly assigned to one of three groups: (1) English gloss, (2) Spanish gloss, and (3) control. The dependent measures included recognition and a production tests that were repeated three weeks later. Additionally, the researchers also measured text comprehension by means of a written recall task. Results of the study indicated a statistically significant difference favoring the use of glosses (in the L1 or the L2), but these differences disappeared at delayed posttests.

Hulstijn, Hollander, and Greidanus (1996) conducted an incidental vocabulary learning experiment with Dutch advanced learners of French in order to investigate the effects of glossing and frequency on various measures of

recognition. The participants read the French text under three conditions: (1) L1 marginal glosses, (2) access to a bilingual dictionary, and (3) control. The frequency of occurrence of the targets varied from one to three occurrences. The results of the study indicated an effect for frequency, no effect for gloss, and no interaction on a test in which the participants had to indicate whether the word appeared in the text. However, the results of a recognition test in which the participants had to provide the meaning of the targets in the L1 indicated a main effect for gloss (favoring marginal glosses in the L1), a main effect for frequency, and a gloss by frequency interaction. Frequency has its biggest effect on the L1 marginal gloss condition.

Following up on the studies by Hulstijn (1992), Hulstijn et al. (1996), and Jacobs (1994), Watanabe (1997) conducted an incidental vocabulary learning experiment with ESL Japanese university students learning English in which the participants were given a reading task, and received a surprise unexpected vocabulary posttest afterwards. As in the studies by Hulstijn (e.g., 1992) comprehension question were given to students during the experimental task. There were two independent variables in this study: a translation task and glossing. Half of the participants in this study read the English text under one of five conditions of exposure: (1) L2 English glosses, (2) given meanings in embedded appositive clauses, (3) correct gloss embedded in multiple choices, (4) non glossed original text, (5) control. The other half were asked to translate the

unknown words into Japanese in writing as they read the texts under one of these five conditions. The researcher made two important predictions: (1) marginal glosses should have the largest enhancement effect because these glosses are expected to “provide explanations with clarity of the connection” between word forms and their meanings (p. 290). The dependent measures in this study were an immediate post exposure definition supply test that was repeated one week later. In addition, there was a word-in-its-original context delayed posttest (cf. Hulstijn, 1992). The results of the study indicated a main effect for gloss, no effect for the translation task, and no interaction. The single and multiple choice gloss groups differed statistically from the meaning-in-appositives and control groups, with both gloss groups outperforming the meaning-in-appositives, and the control. Furthermore, no differences were observed between the multiple choice condition and the marginal single gloss, and appositives behaved just as control. The analysis of scores in the comprehension test favored the glossing groups. Finally, the study replicated previous findings of low rates of learning under incidental learning conditions, whether they included a gloss or not.

Following up on studies by Hulstijn (1992) and Watanabe (1997), Nagata (1999) set out to investigate whether the multiple-choice gloss could be improved by adding a right/wrong indication every time the participants selected a choice. The participants in this study were college students in second semester Japanese. Unlike previous studies, this study included three grammatical structures: *hodo*,

that is, a comparative form (Eng. ‘*as much as*’) with syntactic restrictions, *kara* (Engl. ‘*because*’), and *to* (Engl. ‘*with*’). There were two independent variables in this study: frequency (with four levels), and gloss (single L1 glosses versus multiple choice plus feedback versus control). An L2-to-L1 translation task was the posttest, repeated one month later. The results of the study revealed that the multiple choice +feedback group outperformed the single gloss group on the immediate posttest. Furthermore, recall improves as a function of frequency. Moreover, there is an overall drop in recall in the delayed posttest, and a difference is observed between the single gloss and the multiple choice plus feedback group only for grammar. Finally, the study indicated a high percentage of wrong answers in the multiple choice condition, in spite of the feedback provided.

Gettys, Imhof, and Kautz (2001) conducted a study with second-year college students of Russian in which the effects of giving an exact translation to L1 English gloss (e.g., *went*), versus (2) dictionary form (e.g., *go*) were compared on various measures of comprehension, L2 learning, and time on task. The results of the study indicated a significant difference for time on task: the participants who read the L2 text with glosses as dictionary forms spent more time on the task than the participants who were exposed to the exact translation task. None of the other independent variables of the study was found to have a statistically significant effect on the learning and comprehension measures.

Several computerized studies have investigated the relative effectiveness of multimedia annotated glosses for the L2 learner (Al-Seghayer, 2001; Bowles, 2004; Chun & Plass, 1996; 1997; Lyman-Hager et al., 1993; Yoshii, 2006). Chun and Plass (1996) conducted a computerized incidental learning experiment with second year college students of German. The study included a within-subject design in which each participant had the choice to look-up various types of glosses. Premised on Paivio's (1969) hypothesis of dual-coding, the study aimed at comparing the effectiveness of dual-stimulus presentation (i.e., L1 verbal annotation plus picture, or plus video) versus single stimulus (i.e., L1 definitions only). The participants were exposed to these conditions while completing a reading comprehension task. The dependent measures of the study were two recognition tests: a supply definition test that measured receptive knowledge of the targets, and a multiple choice recognition test respectively. The results of the study indicated higher rates of recall overall, in comparison with prior studies (cf. Hulstijn, 1992). Furthermore, statistical analyses indicated a significant difference favoring dual-stimulus with still picture only, over text-plus-video or control. Finally, as it is usual in this type of studies, the look-up data, measured via a tracking mechanism, indicated great variability in the types of glosses looked up.

Following up on a study by Lyman-Hager, Davis, Burnett, and Chennault (1993), Bowles (2004) set out to investigate whether exposure to L1 vocabulary glosses on a computer would be more effective than exposure to the same L1

glosses presented via paper-and-pencil technology for measures of noticing, reading comprehension, and vocabulary learning. The participants in this study were college students of beginning Spanish. Comprehension was measured by means of a multiple-choice comprehension questionnaire in the L1 while intake was operationalized as the ability to select the correct meaning of a targeted form out of four choices in a multiple choice recognition posttest. Written production was measured by a written translation task from the L1 to the target language. Both recognition and production posttests were given immediately after exposure and three weeks later. The analyses of the dependent measures indicated a significant difference for glossing, with the gloss group outperforming the control, but no difference between glossing modes (i.e., reading computerized versus non computerized glosses). Thus, learners who were exposed to vocabulary glossed (regardless of technological mode) understood the text better and learned more targeted words than learner not exposed to these glosses did.

A special feature of Bowles' (2004) study was the inclusion of think-aloud protocols in her research design as a measure of noticing, and the various analyses conducted on the think-aloud data. The results of these quantitative and qualitative analyses, absent in previous studies, drew new light into the relationship between glossing, noticing, text comprehension, and L2 learning. With respect to the quantitative analysis of the think-aloud protocols, the results indicated that the participants exposed to glosses, regardless of glossing mode,

reported noticing of targeted vocabulary items significantly more than participants in the control groups. This result empirically supports the effectiveness of vocabulary glosses for “drawing learner’s attention to certain difficult words in context in order to explain them” (p. 549). Thus, glosses seem to affect noticing. In spite of this finding, the qualitative analysis of the think-aloud data revealed a low level of awareness of the targeted forms-to-meaning connections overall, explaining/supporting previous findings of a low rate of learning in these incidental learning conditions. The qualitative and quantitative analyses conducted by Bowles (2004) corroborated the finding of previous studies by Leow (e.g., Leow, 2001b) in the input enhancement strand of research (e.g., Leow 2001b, Leow et al., 2003) indicating that input modification by way of textual enhancement is not effective for L2 learning when compared to unenhanced or modified textual input.

Following up on a study by Chun and Plass (1996), Al-Seghayer (2001) investigated the effects of different types of multimedia annotations on vocabulary learning, in particular the issue whether accompanying textual annotations with a video clip would be more effective than accompanying textual annotations with a still picture. To test this hypothesis, thirty intermediate ESL learners of various language backgrounds read a computerized text including printed text annotations alone, printed text annotations accompanied with still pictures, and printed text annotations accompanied by video clips. In this within-

subject design, all the participants were exposed to all annotation modes.

Performance was assessed by using vocabulary recognition and production tests.

Additionally, a questionnaire and an interview were used to assess participant attitudes towards the different annotation modes.

With respect to performance on the vocabulary test, the results of the study indicated a statistically significant difference in favor of the printed text plus video annotation mode over the text plus picture, and the text only modes. No difference was found between the printed text plus picture and the text alone annotation modes. Furthermore, the results on the participants' ratings on the differential efficacy of modes (i.e., the questionnaire and the interview) paralleled those on performance on the vocabulary tests. In both questionnaire and interview participants found the text plus video mode to be more effective than the picture plus text and the text alone mode, with the text alone mode showing the least efficacy. From these results, the researchers drew the conclusion that video clips, by stating that "dynamic image and sound facilitate recall" (Al-Seghayer, 2001, p. 225).

Yoshii Makoto's study (2006) with ESL Japanese college students set out to investigate the effects of language of the gloss (i.e., L1 versus L2) and type of gloss (i.e., dual-stimulus versus single verbal stimulus) on recognition (i.e., a definition supply task, and a multiple choice test completed immediately after exposure and two weeks later). The results of the study indicated no main effect

for language of the gloss, a main effect for picture (where the definition-plus-picture performed significantly better than the definition only group) and no interaction. Furthermore, there was a language of the gloss by time interaction in both recognition tests. Apparently, the participants were able to remember L1 glosses better than L2 glosses over time. The researcher concluded that the study supports the use of L1 annotations, and that dual-stimulus may lead to dual-encoding, thus supporting the dual-coding theory.

Overall, the studies that have investigated the effectiveness of glossing on L2 learning, in the traditional sense of short definitions in the L1 or the L2, have yielded mixed results. Some studies support the use of these glosses (Hulstijn, 1992; Jacobs et al., 1994; Hulstijn et al., 1996; Lyman-Hager, 1997, Watanabe, 1997; Nagata, 1999; Bowles, 2004), while others do not (Chun & Plass, 1996; Al-Seghayer, 2001; Yoshii Makoto, 2006). Regardless if they support the use of these devices or not, the observation that learning overall is rather low under the incidental conditions of these experiments, that is, in the context of processing input with a primary focus on meaning and communication, is a generally accepted finding in the glossing strand of research. This overall result is similar to that observed in the context of textual input enhancement and textual simplification: The chances for input modification to influence L2 learning are rather low, especially in incidental conditions.

Summary of findings of previous research on the effectiveness of glossing for improved L2 reading comprehension and learning

This survey of research has presented and discussed the main findings of previous studies on the effects of glossing on comprehension and L2 learning. An overall conclusion from a survey of previous studies is that using traditional glosses for individual vocabulary items affects reading comprehension positively. This is not surprising in the context of incidental learning. In this context, learners are focused primarily on reading for meaning. In this context of language used for communication, *learners seem to be attending to the meanings presented in the gloss*. The issue whether the learners can also focus *on the forms* of the words glossed is a question that has not received a definite answer in those studies that have been especially concerned with the issue of learning while reading due to the pattern of mixed results found in this research strand.

Therefore, while glosses seem to be effective for focusing the learners on meaning, the input-to-intake phenomenon in the domain of the lexicon is not settled: Can learners also *pay attention to the forms of the glossed words*, and to the connections between the word forms and their meanings? Would glossing make the form-to-meaning mappings more “clear” and “transparent” for the learner, as predicted by previous studies (e.g., Watanabe, 1997)? This issue merits more attention in the research strand of glossing.

Mixed results may be due to several factors such as type of text (e.g., genre, length, and format), target language, and criteria for item selection, length of treatment and language background and knowledge. Furthermore, several methodological issues may be taken into consideration to explain the pattern of mixed results found in previous studies on glossing. Issues with the measures of reading comprehension and learning, issues with the techniques used to control for prior knowledge, or the lack thereof, are only a few of the many major methodological problems found across studies. Among the many factors that have the potential to impact the internal validity of glossing studies, the issue of finding an adequate definition and measurement of attention/noticing is perhaps one of the most important. This issue will be commented at length due to its especial relevance to the present study.

Operationalization and measurement of attention/noticing

Despite the centrality of attentional mechanisms to the studies on glossing, only one of the glossing studies included in this survey of research (i.e., Bowles, 2004) has included an operational definition of attention/noticing and a way to measure it during online processing via concurrent think-aloud protocols. This is an important issue for the internal validity of previous studies on glossing. It has been pointed out several times in the literature (Bowles & Leow, 2005; Bowles, 2008; Leow, 2000, and 1999; Leow & Morgan-Short, 2004; Rosa & Leow, 2004)

the internal validity issues of studies that lack operational definitions of attentional processes or a method to measure them directly, that is, empirically. Due to the lack of such empirical measures, conclusions about the amount of attention allocated to certain stimuli during task performance are often made based on the participant's scores in a posttest, for example, higher scores may suggest that more attention was paid to the input, or the type of stimulus presented to the learner, for example, the assumption that more attentional resources must be allocated to explicitly presented input. The problem for internal validity is that conclusions about attention and noticing patterns are never made based on empirical data. For this reason, the relationship between attentional mechanisms and measures of learning are ultimately based on mere speculation in these studies.

The point made in the previous paragraph should not be taken to mean that previous studies on glossing completely lack any measures of processing. The claim has been made that glossing studies have improved their internal validity considerably when they became computerized because computers allow the researcher to use an electronic tracking system that tracks the amount of information accessed by the participants while interacting with the L2 input (Hulstijn et al., 1996; Chun and Plass, 2001). Despite the merits they might have, these tracking systems are not effective methods to study cognitive processes because these processes are qualitative in nature while the tracking mechanisms

can only provide the researcher with quantitative data, for example, the number of times a participant accessed a gloss. As an insight into thought processes and noticing of L2 forms, these mechanisms are limited, and leave the researcher with a “blind” design. Added to the tracking mechanisms, concurrent data elicitation techniques, such as think-aloud protocols provide the researcher with a more adequate method to study the cognitive processes during online performance.

A line of research in SLA studies conducted within an attentional framework (Leow, 2000, 2001a & 2001b, 1997a, 1998a, 1998b, 1999; Rosa & O'Neill, 1999; Rosa & Leow, 2004) have begun to address empirically the role of attention (and awareness) in their studies by including in their research designs operational definitions of attention, noticing, and awareness, and employing concurrent verbal reports (e.g., think-aloud protocols) to study the learners' cognitive processes while they interact with the L2 input.

A study conducted by Leow (1997a) had found evidence in support of the claim that awareness is a sufficient condition for foreign language learning. Although awareness proved to have a ‘facilitative effect’ on the learning measures of the study, a claim for awareness as a necessary condition for learning could not be made because a definition of “unawareness” was ‘difficult to find’ (p. 494). In order to address this issue, Leow (2000) replicated Leow (1997a) with 32 beginning learners of Spanish and the same targeted items included in the original study, that is, the vowel shift undergone by stem-changing Spanish verbs in the

preterit. The study set out to investigate how being aware or being unaware during an exposure task would affect the ability to recognize and produce, in written, the targeted forms of the study. The original task was changed to a task that would allow the researcher to elicit data from unaware learners. Thus, instead of providing clues that would direct the participants' attention to the correct vowel shift, in the 2000 study the vowels were retained in the stems, and the participants' attention was diverted from the stems by asking them to fill in the morphological endings of the Spanish verbs. Think aloud protocols during completion of both the exposure and assessment tasks were utilized to measure noticing and awareness of the targeted items. The results of the study indicated that "participants who demonstrated awareness of the targeted morphological forms during the experimental exposure recognize and produce in writing significantly more of these forms when compared with the group that demonstrated a lack of such awareness" (p. 568). Thus, Leow found that there was no dissociation between learning and awareness in this study, and therefore that awareness plays a crucial role in language learning.

An important feature of this study was the use of think-aloud protocols to distinguish aware from unaware learners. It was the patterns of processing found in the protocols what provided the researcher with direct empirical evidence. For this reason, this study validated the use of concurrent think-aloud protocols as a

measure that does “provide richer source of information on learners’ attention and awareness than quantitative analyses” (p. 570).

A study by Rosa and O’Neill (1999) with 67 intermediate learners of Spanish used think-aloud protocols to investigate whether and how intake is affected by awareness and by two conditions of exposure to the input ([+/- formal instruction] and [+/- explicit directions to search for rules]). Intake in this study was measured via a multiple-choice recognition posttest. Results indicated statistical effects for both learning conditions and awareness. With respect to learning conditions, the analyses revealed a positive effect of explicit rule explanations on intake of grammatical forms, that is, formal instruction, regardless of rule search instructions, was significantly better than any of the implicit conditions. With respect to awareness, the analyses indicated that participants who showed ‘awareness at the level of understanding’ did statistically significant gains from pretest to posttest, as opposed to the ‘unaware’ learners. It was concluded from this study that “awareness at the level of understanding is related to a more sophisticated type of structural processing that may contribute to higher amounts of intake (...) the protocols from the [awareness at the level of understanding] group also support claims in the cognitive psychology and SLA fields that associate higher levels of awareness with a structural, conceptually driven way of processing input” (p. 543). With respect to the relation between awareness and learning conditions, the results led to the hypothesis that explicit

conditions of exposure to the input “may have contributed to raising the learners’ level of awareness” (p. 547). This study further validated the use of think-aloud protocols as an adequate measure of noticing and awareness during online completion of a problem-solving task.

Rosa and Leow (2004) set out to investigate the differential effects on intake and L2 development of a problem-solving task varying in degrees and types of explicitness. Participants in this study (N=100 advanced learners of Spanish) were randomly assigned to one of five experimental conditions and one control varying in explicitness. Multiple-choice recognition and controlled written production were utilized as performance measures. The results of the study can be summarized as follows. Performance in the recognition and production tests revealed that participants in all five experimental conditions statistically outperformed the control, and that this difference was maintained in the delayed posttest. Furthermore, all groups except the control improved significantly from pretest to posttest. Moreover, in the production test a difference was found between the most and the least explicit conditions in the experiment, in favor of more explicit conditions. The recognition and production tests including novel items indicated similar results to those found for the production of old items. However, it was observed that the effects of treatment on the ability to produce new items “tended to weaken over time” (p. 206).

The use of verbal reports as data in SLA studies premised on the role of attention is further validated because it has been widely used in empirical studies in psychology, cognitive science, and education (see Ericsson & Simon, 1993, for an overview). Within these disciplines, verbal reports are a tool for “elucidating a clearer picture of internal processes employed by adult learners” (Leow & Morgan-Short, 2004, p. 36). Unlike quantitative data, “qualitative analyses of verbal reports can provide unique insights into learners’ processing” (p. 37).

The studies on SLA summarized above have utilized a type of verbal report called online, concurrent or introspective. With concurrent verbal reports, the learner is asked to verbalize his thoughts while performing the experimental task. Since this type of report is “concurrent” with the task, it is believed not to be subject to memory decay. For this reason, concurrent verbal reports are considered to be more effective than non-concurrent or retrospective reports. Besides being concurrent to the experimental task, the type of verbal reports utilized in previous studies on SLA has also been non-metalinguistic. Non-metalinguistic reports are those in which the learners are not asked to reflect on their own thinking processes. In non-metalinguistic reports thinking aloud is secondary; the primary task being the experimental task. It has been argued in the literature that for verbalizations to reflect learners’ processes (e.g., Ericsson & Simon, 1993) introspective, non-metalinguistic verbalizations are most adequate.

Despite the relative advantages of concurrent non-metalinguistic verbal reports, some researchers (R. Ellis, 2001; Stratman & Hamp-Lyons, 1994) have pointed out that verbalization tasks have the potential to affect or impact internal processes of the learner. Reactivity is the term used by researchers to denote the potential for verbalization to modify the learners' processes.

A review of non-SLA studies conducted by Ericsson and Simon (1993) did not find reactivity effects, although there was some evidence that concurrent verbal reports may alter the time learners need to complete that task. While Ericsson and Simon reported no evidence for reactivity in the studies they reviewed, a review by Leow and Morgan-Short (2004) found an effect for concurrent verbalizations on accuracy on gambles and addition tasks, and for time on gambles and anagrams in a study by (Russo, Johnson, & Stephens, 1989). However, the non-metalinguistic nature of the concurrent verbalization task was not clear in this study, as participants who were doing the concurrent verbalizations "might have been metalinguistically primed by a previous retrospective verbalization task" (Leow & Morgan-Short, 2004, p. 39).

The issue of reactivity of concurrent non-metalinguistic verbal reports has motivated SLA studies to investigate empirically the potential of these reports to impact thought processes (Bowles & Leow, 2005; Bowles, 2008; Leow & Morgan-Short, 2004; Sachs & Polio, 2007; Sanz, Lin, Lado, Bowden, & Stafford, 2009). In a study with 77 beginning Spanish learners, Leow and Morgan-Short

(2004) investigated the effects of verbal reports on an L2 reading task that included enhanced and unenhanced items. A t-test comparison of the scores in the three dependent measures of the study (i.e., reading comprehension, recognition and written production) yielded no significant difference between the think-aloud and the non think-aloud groups. From these results, the researchers concluded that “thinking aloud while performing a reading task does not seem to have detrimental effects on learners’ comprehension, intake, or controlled written production when compared to a nonthink-aloud group performing the same task (...) thinking aloud was nonreactive in this study, and these results provide empirical support for the validity of the use of such procedures to gather concurrent data on learners’ cognitive processes” (p. 50).

Bowles and Leow (2005) study with 45 advanced learners of Spanish set out to investigate the effects of type of verbal report, that is, non-metalinguistic versus metalinguistic, on test performance, as measured by text comprehension and written production. Additionally, the present study investigated the effects of verbal reporting on system learning (the ability to apply recently learned grammatical information to new exemplars). The study also addressed the issue of latency: whether the type of verbalization had an effect on time on task.. The results of this study can be summarized as follows. An analysis of comprehension test scores revealed a significant main effect for group, with the non-metalinguistic group performing significantly better than the metalinguistic group.

However, neither experimental group performed significantly differently from the control group. From this finding, it was concluded that metalinguistic verbalizations caused a significant decrease in text comprehension when compared to non-metalinguistic verbalization, although “the presence of either type of verbal reports did not appear to have a detrimental effect on participants’ comprehension of the L2 text when compared to a silent control group” (p. 428). Furthermore, an analysis of production test scores revealed similar results of no effect of type of verbalization. Finally, there was an effect for time on task for both metalinguistic and non-metalinguistic groups.

Bowles (2008) study with 194 college students enrolled in beginning Spanish examined the effects of type of verbalization (non-metalinguistic versus metalinguistic versus silent group) and implicit versus explicit feedback using a L2 problem-solving task on the ability to produce in writing old, previously encountered, and novel exemplars of the psych verb *gustar* in Spanish. The analyses revealed a main effect for feedback, in favor of the more explicit mode, a main effect for verbalization, but no interaction, on the written production of ‘old’ exemplars. Post hoc analyses of type of verbalization indicated that the difference was between the metalinguistic and control groups, with the metalinguistic verbalization group performing worse than control. Furthermore, with respect to the ability to produce novel exemplars, there was a main effect for feedback, in favor of the more explicit mode, but no effect for type of verbalization and no

interaction. Finally, time on task was reactive for the metalinguistic verbalization group only. From these findings, the researcher concluded that metalinguistic verbalizations seemed to have a detrimental effect on participants' ability to produce 'old' exemplars. However, the type of verbalization did not interfere with system learning.

Despite the fact that reactivity studies in cognitive psychology and SLA have generally not found verbal reports to be reactive, in the sense of thought altering devices, the issue of reactivity is still an open question in these fields for several reasons. It has been pointed out, for example, that the reactivity of verbal reports might be affected by a number of factors including the type of task, text length and genre, text difficulty, targeted linguistic structure, and individual differences (Leow & Morgan-Short, 2004). For this reason, some researchers (e.g., Leow & Morgan-Short, 2004) have suggested that studies employing concurrent verbal reports include a silent group as a control for potential reactivity effects.

To summarize to this point, most previous glossing studies have not included measures of attention/noticing, or operational definitions of these constructs, in their research designs. This lack of empirical measures of cognitive processes has caused internal validity issues for previous studies on glossing. By contrast, several studies in SLA have utilized concurrent verbal reports, that is, think-aloud protocols, to gather qualitative data on language processing during

task completion as a way to address this internal validity issue. By including a verbalization technique that had been used in the cognitive psychology field for more than fifty years, SLA studies further validated the use of this technique in the SLA field. In spite of this, the potential of the technique for reactivity, that is, the potential for verbal reports to alter the cognitive processes under investigation has motivated SLA studies on reactivity. Due to the paucity of studies on reactivity in SLA, and the many factors that may potentially affect the use of verbal reports, researchers in the field have recommended that a silent group be included as a control in those research designs using verbal reports as data. In the present study that employs concurrent non-metalinguistic verbal reports a silent group will be added to the research design in order to control for reactivity effects.

Other variables not considered in previous studies on glossing

Type of Linguistic Item

Several studies in the textual enhancement strand of research have isolated type of linguistic item as an independent variable. These studies have generally contributed empirical evidence to the hypothesis that different parts of language are not processed in the same way (VanPatten, 1994). Researchers have advanced various hypotheses to explain the phenomenon that language is not processed all alike. While some researchers have proposed a notion of communicative value,

others have recently redefined the notion of difficulty to explain why some form-meaning connections are easier to establish than others (DeKeyser, 2005).

A type of linguistic item effect in relation to noticing, intake, and L2 development is an important issue that has received no attention in the glossing strand of research. Only one study in this strand has isolated some elements of an L2 text from others to measure their effects on reading comprehension (i.e., Hulstijn, 1993). In this study, word relevance was isolated as a variable to investigate how relevance of the item would affect the look-up behavior. This study has shown that glosses for more relevant words are consulted more often than glosses for less relevant words. Furthermore, it has shown that words that are difficult to infer from the context are not looked-up more often than those that are easy to infer. This was the only study in the glossing strand that has isolated some items of a text from others.

The fact that the variable type of linguistic item has not attracted the attention of researchers within the glossing strand of research is not surprising given that the interest for the effects of glossing on L2 learning stemmed from the prior interest in glosses in relation to reading comprehension. While the first studies on glossing have been interested in focusing the L2 readers on the meaning of words, later studies have been more interested in focusing the learner on word forms, word meaning, and their connections. In this context, only one study has investigated form-to-meaning connections within the domain of

grammar and syntax (i.e., Nagata, 1999). Although a type of linguistic item effect was suggested in this study, this study has not isolated type of linguistic item as a variable. Therefore, the research strand of glossing is lacking a study that isolates type of linguistic item as an independent variable, to investigate the effects of this variable on measures of noticing, intake, L2 development, and reading comprehension.

Language Difficulty

The notion of language difficulty encompasses a number of factors. For example, the setting in which the learning takes place (i.e., instructed versus non instructed settings), the amount of instruction, age, stage of interlanguage development are all important factors that influence language difficulty (DeKeyser, 2005). A factor of difficulty that has attracted the attention of researchers in instructed SLA is the notion of *inherent difficulty*. This is the notion that difficulty in language learning can be explained, in part, based on the linguistic characteristics of parts of language. If different parts of language present different intrinsic characteristics, it is plausible that this inherent variation of the types of linguistic items affects the way in which input is attended to and noticed by the L2 learner. For this reason, the notion of variation of type of linguistic item based on the inherent characteristics of the item is becoming increasingly important in the context of cognitive SLA.

According to DeKeyser (2005), the notion of *inherent difficulty* is not easy to define because of the many linguistic aspects involved. For example, some items may carry a ‘problems of form’ while others may present a problem for ‘form-meaning connections’. Besides these ‘problems’ of form, and form-meaning connections, there is the issue of how abstract is the encoded meaning. Among “problems of form” are, for example, the morphological marking of person by means of verbal inflections in Romance languages. Besides the number of different markings the learner needs to learn, there is the issue of the communicative value of the item (e.g., VanPatten, 1994). This is the notion that some items carry more semantic meaning than other items. From the point of view of their relative communicative value, items that carry more semantic meaning are deemed to be more ‘salient’ to the learner than items that carry a grammatical value only (see Leow, 1993 for discussion). For example, the Spanish Present Perfect may be relatively more salient than the Spanish subjunctive morphology because of the inherent redundancy of the subjunctive marking, that is, the fact that the semantic contents of the subjunctive mood (e.g., volition, emotion, doubt) are recoverable by context. In contrast, the Spanish Present Perfect carries specific tense-aspect semantic information that may not be recoverable by context.

Besides the ‘problems of form’, there are problems of ‘form-meaning connections’ that are relevant to defining *inherent difficulty*. The complexity of

form-meaning connections may depend on a number of factors including but not limited to the communicative value of the item (see DeKeyser, 2005, for discussion). Correct mappings of meaning and form may also depend on how ‘opaque/transparent’ and how ‘optional/obligatory’ form-meaning connections are *within* the language. For example, allomorphy (i.e., various grammatical meanings in one form as, for example, the clitic SE in Spanish) is a typical source of opacity. Complexity of the form-meaning connection may also be explained by the optionality of the item. For example, null subjects in Spanish and Italian are a source of difficulty to the L2 learner because null subjects are optional. Optional items may create an apparent irregularity in the input. Irregular or inconsistent patterns in the input may interfere with the process of hypothesis formation, making it difficult for the learner to test hypothesis about optional items. The picture of language difficulty is further complicated by the fact that a full description of inherent problems is not possible without taking into account L1-L2 contrasts (DeKeyser, 2005). For example, the challenge to learn the Spanish clitic SE for English native speakers is not only its inherent opacity but also the fact that an equivalent piece of morphology does not exist in English. Thus, ‘problems with form-meaning connections’ can be explained based on the inherent complexity of the connections to establish. Both the inherent characteristics of the L2 and the influence of the L1 are a source of complexity (DeKeyser, 2005). However, the intrinsic way in which meanings are encoded

within the language, and the complexity of encoding within the L2 is what causes the inherent 'opacity' of the item. The L1 influence may add to this complexity.

Within the language, items may differ in complexity for several reasons. For one reason, some encodings exist in some languages but not in others (e.g., classifiers in some African languages do not exist in Romance). Furthermore, grammatical encodings are highly arbitrary (e.g., the encoding of *Psy* verbs in Spanish; the variation of the semantic range for the Indirect Object construction in English and Spanish, the use of locative pronouns in French, the encoding of the 'possessive', etcetera).

The claim that complexity of the form-meaning connection is relevant to grammar should not be taken to imply that complexity of the form-meaning connection does not exist in the lexicon (Whitley, 1986). There is a general assumption in the SLA research literature that grammatical items are more difficult than lexical items because – it is assumed - lexical word form connect to lexical meaning in a one-to-one fashion. However, not all mappings in the lexicon are of the type one meaning to one form. Synonyms, and homonyms, are clear examples of *one-to-many* mappings between form and meaning that create problems within the lexical domain (Whitley, 1986). Despite the existence of complexity within the lexical domain, this study only deals with complexity within the grammatical domain.

Besides the notion of complexity of encoding, the abstractness of the meaning/referent is another source of inherent difficulty of the item. *Abstractness* of meaning/referent refers to the degree of grammaticality or functionality of the referent. This is the notion that some items are more abstract than other items. There is general agreement in the SLA literature that grammatical items are more difficult to process and learn than lexical items because grammatical meanings are abstractions that capture functional concepts and relationships, for example, tense, person, gender, number, case, to name just a few, that are strange to most language learners, even when these concepts and relationships are discussed in the context of their own language (DeKeyser, 2005). The systematic use of form to express such abstract concepts does not exist in the lexicon in principle, although some grammatical distinctions may become lexicalized. Thus, lexical abstraction is not generally considered a problem for language learning, while grammatical abstraction is (e.g., DeKeyser, 2005; Gass & Svetics, 2003).

It is probably worth noting at this point that this view of learning in the grammatical domain by no means implies denying that abstraction exists in the lexicon. In fact, many of the lexical concepts in our sophisticated societies are quite abstract (i.e., in the domains of science, technology, philosophy). However, most L2 teachers and researchers would probably agree that abstract lexical concepts are not *the* issue in L2 learning. Put another way, if abstract meanings/referents in the lexical domain is a learning issue, it is not an issue of

learning language, but rather of grasping concepts. The point of the argument is, then, that *abstractness in the lexical domain is different from abstractness in the domain of grammar.*

The notion of abstractness of meaning/reference not only distinguishes lexical from grammatical items, but it also welcomes further distinctions within the domain of grammar (DeKeyser, 2005). Within the grammatical domain, some forms are easier to process than other forms based on their relative abstractness. For example, “case” is a more abstract grammatical meaning than “gender” or “tense” because tense and gender marking are the grammaticalization of semantic notions. In contrast, “case” meanings are purely linguistic, based on syntactic functions. Furthermore, notions such as ergativity or accusativity are highly abstract. In sum, abstractness of meaning/referent may make lexical items, in general, more ‘isolatable’ than grammatical items (Gass & Svetics, 2003; VanPatten, 1990). Further distinctions of abstractness may also exist within the grammatical domain.

Based on the discussion above, linguistic items differ in difficulty based on their inherent salience, abstractness, and complexity of the form-meaning connection. Given the complicated picture of language difficulty, for the purpose of the present study the notion of inherent difficulty will be operationally defined (see chapter 2: Research Design and Methodology).

Research questions

Based on the methodological issues and limitations explained above, the research on glossing has left many questions unanswered. One of these questions deals with the effects of glossing on noticing and learning. In this study, this question will be adequately addressed by including concurrent verbal reports to elicit processes of attention/noticing. Another question left unanswered by previous studies are the effects of type of linguistic item on noticing and learning. Based on the discussion presented above, this study will address the following research questions:

RQ#(1) Does reading a text with L1-translation glosses versus reading a text without glosses have an effect on reading comprehension?

RQ#(2) Do readers exposed to glossed text report significantly more noticing of targeted items than readers exposed to no-gloss text?

RQ#(3)

(a) Does glossing have an effect on reported noticing?

(b) Does type of linguistic item have an effect on reported noticing?

RQ#(4) Is there a significant relationship between the reported noticing of glossed versus no-glossed targeted items and the recognition and production of these items? If so, are the two relationships significantly different?

RQ#(5)

(a) Does glossing have an effect on:

- (i) intake, as measured by immediate recognition of targeted items, and
- (ii) L2 development, as measured by immediate production of targeted items?

If so, will these effects be maintained three weeks later?

(b) Does type of linguistic item have an effect on:

- (i) intake, as measured by immediate recognition of targeted items, and
- (ii) L2 development, as measured by immediate production of targeted items?

If so, will these effects be maintained three weeks later?

CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY

Participants

Participants in this study were 65 college students enrolled in Beginning and Introductory Spanish courses (i.e., first year). From an original pool of 137 participants, 72 were eliminated from the study for one or other of the following reasons: (1) having produced inaudible think-aloud protocols, or not having followed the instructions to think-aloud, (2) having demonstrated prior knowledge of at least one of the three linguistic structures targeted in the study (i.e., Lexical Item, Present Perfect, and Impersonal-SE), (3) having received formal instruction in any other Romance language, (4) not having completed all phases of the study, and (5) not possessing any prior knowledge of simple past tense in Spanish. Additionally, all participants included in the study had received previous formal instruction in the Spanish simple past tense.

Pre-tests including both recognition and production tasks were utilized to assess prior knowledge of the targeted items of the study. The participants were given the recognition and production pre-tests three weeks before the treatment. Immediately before completing the treatment, participants completed a background questionnaire that aimed to elicit any prior language background, that

is, (a) whether they had had any formal instruction in other languages, (b) what language(s), and (c) for how long.

To ensure homogeneity of the sample, all the Spanish classes from which participants were recruited had a notional-functional syllabus, with a strong emphasis on teaching the four skills (i.e., listening, speaking, reading, and writing) based on communicative functions and the use of the target language in a variety of contexts.

Materials

The target of the study

The target of the study was a comparison between linguistic items of various degrees of ‘inherent difficulty’. The following contrasts were included in the study. Firstly, the study aimed to compare the lexicon (e.g., vocabulary items such as Sp. [un] *espiguelo*, Eng. [a] *search*) with grammar. Secondly, the study included a comparison between two types of grammatical items, i.e., the Present Perfect (e.g., Sp. *ha trabajado*, ‘he has worked’), versus (3) the Impersonal-SE (e.g., Sp. *se trabaja mucho aquí*, ‘one works hard here’).

The notion of inherent difficulty used in this study was based on DeKeyser (2005) definition. However, given the ‘complicated picture’ one finds in the search for a definition of inherent difficulty, a formal and explicit analysis of difficulty should include operational definitions of the feature-components of the

term. Thus, for the purpose of the present study, the notion of inherent difficulty was operationally defined based on an overtly defined matrix of difficulty with three operationalized features: (1) *salience*, (2) *abstractness of meaning/referent*, and (3) *complexity of the form-meaning connection*. An operational definition of these terms is presented below.

Salience. Salience was operationalized as *the prominence of an item* based on criteria of form and form-meaning mappings. With respect to form, items such as clitics, pronouns, articles, etcetera, are considered less salient than lexical words because lexical items carry stress. With respect to form-meaning mappings, grammatical encoding in natural languages tends to be less salient than lexical encoding because grammar tends to use formally discontinuous elements, such as periphrastic forms (e.g., Sp. *hacer*+*[time expression]*+*que* meaning “ago”) or verbal inflections in morphologically ‘rich’ languages, while lexical items use more consistent mappings. The distinction between lexical items and grammatical form usually accepted in the SLA literature is based partly on this criterion of salience (DeKeyser, 2005; Gass & Svetics, 2003; VanPatten, 1990; Whitley, 1986).

Abstractness of meaning/referent. The feature of *abstractness* was operationalized in this study in terms of the degree of grammaticality or functionality of the

referent. Grammatical meanings are abstractions that capture functional concepts and relationships. For example, the grammatical notions of tense, person, gender, number, case, to name just a few, tend to be strange to most language learners, even when these concepts and relationships are discussed in the context of their own language (DeKeyser, 2005). The systematic use of form to express such abstract concepts does not exist in the lexicon. Thus, lexical abstraction is not generally considered a problem for language learning, while grammatical abstraction is (e.g., DeKeyser, 2005; Gass & Svetics, 2003). From this point of view of abstractness, those linguistic items that carry an abstract function will be considered abstract.

Complexity of the form-meaning connection. This term was operationally defined as the *type of meaning encoding*. Type of encoding refers to the way in which a form encodes meaning, and it can refer to L1-L2 contrasting encodings (e.g., aspects, causation, definiteness both exist in English and Spanish but they are encoded differently) or to single versus multiple encodings of meanings within the L2. This notion of complexity differs from the previous notion of abstractness in that abstractness does not refer to how the language encodes meanings but rather to how functional (i.e., non semantic) the meaning encoded is. Complexity, on the other hand, is defined in this study not to include how abstract the referent is but rather how the meaning, regardless of abstraction, is encoded in the L2. For

example, learning case marking in German for a native speaker of English is *difficult* not only because case is inherently *abstract* (see discussion above) but also because the meaning-form mappings are *complex*, as English nouns and articles do not encode grammatical case. Thus, for the notion of complexity, as operationalized in this study, L1-L2 contrasts are more relevant to the ‘complexity’ of the item than its functionality. A description of the items included in the study, their definition in the matrix configuration, and a rationale for their feature specification are discussed below.

The targeted types

As stated above, each item type included in the study was operationally defined based on the difficulty matrix [+/-salient, +/-abstract, +/-complex]. Selection of the item types was based on their potential to instantiate a specific configuration of the features included in the matrix. Below is a description of items and their feature specification.

The Present Perfect

The Sp. Present Perfect (e.g., Sp. *ha trabajado*, Eng. *(he) has worked*) was specified in this study as [-salient, +abstract, -complex]. The Sp. Present Perfect consists of the AUX *haber* (Eng. *have*), conjugated in the present + past participle. This item received a negative feature value for salience because it is a

compound tense, i.e., “compounded” from a “helping” or auxiliary verb and a participle (Whitley, 1986). Thus, when the lexical part of the Present Perfect form is factored out, what is left is a discontinuous string of AUX plus bound morphology (e.g., Sp. *ha –ado*). Due to this discontinuity of form, the Sp. Present Perfect was predicted to be less ‘isolatable’ by the learner, and therefore, less salient.

In this study, the Sp. Present Perfect also received the value [+ abstract]. The Sp. Present Perfect is abstract because it conveys tense, which is an abstract functional meaning. With respect to its meaning/referent, the meaning of the Sp. Present Perfect has been explained by linguists as expressing ‘relevant anteriority’ to a present point (Whitley, 1986). That is, the Sp. Present Perfect emphasizes an event in the ‘relatively recent past’ and considers its ‘relevance’ in relation to the speaker’s present point. The nature of that relationship between the recent past event and the present point event depends on the event structure of the lexical verb. Thus, for example, applied to state verbs, the Present Perfect expresses continuity of the relevant event into the present point, as in (1):

(1)

Sp. Hemos estado muy contentos en esta casa.

Engl. We have been very happy in this house.

When used with resultative verbs (e.g., accomplishments or achievement verbs), the Present Perfect typically emphasizes the ‘abiding repercussions’ of the recent past event in the present point (see Whitley, 1986), as in (2):

(2)

Sp. El huracán ha destruido su casa.

Engl. The hurricane has destroyed his house.

The expression of a to-be-defined relationship between a recent past point and the speaker’s present point justifies the feature specification [+ abstract] for the Sp. Present Perfect.

In this study, the Sp. Present Perfect received the value [-complex]. The rationale for this is the similarity of form-meaning mappings between English and Spanish in this area. With respect to form, it should be noted that (a) both English and Spanish have an AUX (have/haber) + past participle construction, (b) in both languages the regular past participle is formed by adding a morpheme to a the verb stem (Sp. *-(a)do*, Engl. *-ed*). With respect to meaning, it should be noted that the nature of the relationship between past and present events is defined similarly in both languages. The semantics of the present perfect in English has been extensively described by linguists (see Bardovi-Harlig, 1997, for an overview). In these descriptions, the present perfect has been characterized as

carrying positive values for the features [anteriority] and [current relevance]. The notion of [anteriority] refers to “past” and serves to establish the common features between simple past and present perfect (i.e., the fact that both tenses share the feature [anterior]). The notion of [current relevance] was originally introduced by Comrie in 1976 to refer to the “continuing relevance of a past situation” (quoted by Bardovi-Harlig, 1997, p. 4), and it does not differ substantially from the notion of relevance introduced above to explain the meaning of the present perfect in Spanish. Since the present perfect in Spanish is similar in form (or cognate) to the present perfect in English, and since the semantics of this tense can be described in similar fashion in both languages, by way of the features of [anteriority] and [current relevance], the present perfect tense in Spanish and English presents similarity of encoding. In this study, similarity of L1-L2 encoding has been interpreted as presenting less complexity than cases with dissimilar encodings. For this reason, the feature [-less complex] is proposed for the Sp. Present Perfect.

The Impersonal SE

The Sp. Impersonal SE (e.g., Sp. *se caminó todo el día*, Eng. People walked all day) was specified in this study as [-salient, +abstract, +complex]. In Spanish, the combination of the morpheme SE (a clitic) with a conjugated lexical verb can be used to express an impersonal meaning. For example, contrast (1), an impersonal meaning, with (2) a non impersonal one:

(1)

Sp. se vive bien aquí.

Engl. one/you can live well here.

(2)

Sp. vive bien aquí.

Engl. he/she lives well here.

The impersonal SE+Verb construction in Spanish was specified as [-salient, +abstract, + complex]. This item received a negative feature value for salience because it involves a clitic, i.e., an unstressed form. In this study, the impersonal SE+Verb also received the value [+ abstract]. This form was categorized as [+abstract] because it conveys a meaning, i.e., the impersonal, which is an abstract functional meaning. With respect to the abstractness of the meaning/referent, some linguists (e.g., Whitley, 1986) have pointed out that in English impersonal subjects are not easy to disentangle from the personal ones because the language uses the same forms to convey both an impersonal and a personal meaning. Thus, for example, in spoken English *you*, *people*, or *they* can be used ambiguously to express impersonal and personal subjects (e.g., Eng. *How do you get good grades?*; personal interpretation: *I study a lot*; impersonal

interpretation: *You study a lot*). The picture gets complicated by the fact that the same ambiguity is found in passive sentences (e.g., Engl. *the tires were checked by someone/ by him*).

Besides the difficulty of conceptualization of the meaning ‘impersonal’, the functionality of the impersonal further contributes to its abstractness. The term impersonal refers to a grammatical category: in the context of grammar, ‘impersonal’ is used to indicate that a certain linguistic item in a language gives an impersonal meaning to the sentence, or that it refers to an impersonal grammatical subject. In this sense, the referent ‘impersonal’ should be considered highly abstract. The same grammatical meaning of ‘impersonal subject’ also exists in the Spanish grammar, and it may be conveyed in Spanish by a piece of morphology, i.e., the clitic SE. For this reason, the impersonal SE+Verb construction targeted in this study was defined as [+ abstract].

At this point, an important difference between English and Spanish should be pointed out: The notion of impersonal subject may or may not be conveyed in similar ways in both languages. Consider the following examples, which show the contrasts that exist between the L1 and the L2 in the expression of the impersonal:

(1)

Sp. Uno debe trabajar duro para ganar buenas notas / Se debe trabajar duro para ganar buenas notas.

Engl. One must work hard to get good grades.

An important difference here is that Spanish possesses a piece of morphology with no English equivalent to express the impersonal meaning. This makes the Spanish impersonal item more complex. However, it is not only the Spanish-English contrast what generates complexity but also the fact that the construction in Spanish is optional: the speaker has the option between *uno* (Engl. one) or SE (see DeKeyser, 2005, for a discussion of opacity as a source of complexity; see Mendikoetxea, 1999, for a discussion of 'impersonal' in relation to other grammatical meanings).

Besides the complex encoding of the form in Spanish, and the complexity of the L1-L2 relationship, one must also take into account the fact that the Sp. SE is prototypically a form to convey the reflexive. As it has been pointed out (e.g., Mendikoetxea, 1999; Whitley, 1986) formal instruction in Spanish makes a strong pedagogical emphasis on the reflexive SE (e.g., *se peina*, 'he combs his hair'). This emphasis on the reflexive has partially contributed to the delayed acquisition of the passive and other uses of SE morphology (Whitley, 1986). The multiplicity of encodings of the form SE in Spanish contributes greatly to its opaqueness, which is a feature of the complexity of the item. For these reasons, the impersonal SE+Verb construction received the value [+complex] in this study.

The lexical items

The lexical items included as targets in this study received the value feature specification [+salient, -abstract, -complex]. The rationale for categorizing the nouns and adjectives included in the study as salient but not abstract should be obvious at this point: by our definition, lexical items should be more ‘isolatable’ by the learners because of their salience and the ‘concreteness’ of their referents.

Beside these features, the nouns and adjectives included as targets of the study were defined as [-complex]. The claim in the SLA literature that complexity of the form-meaning connection is relevant to grammar should not be taken to imply that complexity of the form-meaning connection does not exist in the lexicon (Whitley, 1986). There is a general assumption in the SLA research literature that grammatical items are more “difficult” than lexical items because - it is assumed - lexical word forms connect to lexical meanings in a one-to-one fashion. However, not all mappings in the lexicon are of the type one meaning to one form. Synonyms, and homonyms, are clear examples of *one-to-many* mappings between form and meaning that create problems within the lexical domain (Whitley, 1986).

Despite the existence of complexity within the lexical domain, this study only deals with complexity within the grammatical domain. Besides the theoretical distinctions often mentioned in the literature, there is a practical reason for this. The reason for this decision is the study’s requirement to include

vocabulary targets that could be glossed with a translation in the L1. Spanish vocabulary words that did not have a clear equivalent in English were not included in the study. For this reason, all vocabulary words included in the Lexicon were ‘simple’ in our sense of the term, i.e., L1-L2 one-to-one mappings. Therefore, lexical items were defined as [-complex].

To summarize to this point, this study aims to compare different types of items based on a prediction of their inherent difficulty. Three parameters are relevant to define inherent difficulty of the linguistic item: salience, abstractness of the meaning-referent, and complexity of the form-meaning connection. Based on these features, the study aims to draw the following comparisons: (1) lexical items versus grammar, and (2) the Present Perfect versus the Impersonal-SE. (See Tables 3 and 4).

Table 3. Difficulty Matrix For Two Dimensions

Domains compared	Salience	Abstractness
Lexicon	+	-
Grammar	-	+

Table 4. Difficulty Matrix For Two Types Of Grammatical Items

Domains compared	Item	Abstractness	Complexity
Morphology	Present Perfect	+	-
Morphosyntax	Impersonal-SE	+	+

Operationalization of Glossing

This study aimed to investigate the effects of glossing on comprehension and learning within an attentional framework. From a theoretical standpoint, glosses are viewed as a kind of input modification: A pedagogical technique

aimed to direct learners' attention to linguistic form during input processing. The term [-incidental] is used in this study to refer to a context in which a participant is exposed to a glossed text. The rationale for this decision is that – under the assumption that glosses are an input modification technique - glosses are predicted to attract learner's attention to form. This external condition of exposure to the input contrasts with a purely incidental one, in which no glosses are provided, i.e., the control.

It should be noted at this point that in this study incidentality is conceptualized as an external condition of exposure to the input. In other words, no implications are made about the type of learning that takes place under such conditions. This point is worth clarifying because the term incidental (as well as its related term implicit) are often used ambiguously in the literature, to refer to (a) the way in which certain mental representations are formed, and (b) the external conditions of exposure to the input (e.g., directions to memorize, search for rules, etcetera; see (Robinson, 1996). In this study, the term [+/- incidental] is only used in the sense of an external condition of exposure to the input: No assumptions are made about how the participants learned under different external conditions. This is the issue that the study aimed to investigate empirically, instead.

In this study, glosses were L1-English translations of the targeted items embedded in a reading passage. The use of L1-translation glosses was justified

based on the finding in previous research of a strong preference of L2 learners for L1 glosses, and the claim supported by the evidence in previous studies that this type of gloss is relatively effective in comparison to L2 glosses or other types of annotations (see literature review section). Below are three examples of how the experimental text was glossed.

1. Lexical item:

...escucho opiniones pasmosas¹ sobre los argentinos.

1. *astounding*

2. Impersonal-SE:

... en Argentina se abultan² todas las cosas.

2. *people exaggerate*

3. Present Perfect:

... millones de personas que han surcado³ la Argentina no piensan eso.

3. *have traveled*

This researcher did the translations for the items glossed in this study. To insure accuracy of translation, the researcher asked two bilingual speakers to rate each item translated by her on a Likert-type scale of accuracy from 1 (accurate) to 5 (not accurate). Furthermore, raters were asked to give options every time the rate was not 1. The researcher made the necessary adjustments until all translations were rated 1.

Selection of items targeted in the study

The following sections discuss methodological issues that were addressed in the study during the selection of the targeted items. Issues with selecting lexical targets are discussed first, followed by the criteria utilized in the selection of items for the two grammatical structures included in the study.

Lexical items

This study included ten targeted lexical items (i.e., four adjectives and six nouns). Three methodological issues were addressed during the selection of these lexical targets: word frequency, word cognateness, and word inferability. These issues are discussed below.

How frequent a word is relative to the absolute lexicon of a language, that is, how likely it is that a language learner encounters a word in authentic texts and conversations in the L2 is an important issue for the internal validity of empirical studies because word frequency may potentially interfere with an experimental treatment: Even if a participant is not able to recognize or tell the meaning of a word, it is still possible that the learner might have encountered the word form before. Previous studies on glossing have utilized a variety of strategies to control for frequency effects (e.g., piloting, selection of moderate to low frequency words, or the use of words with no frequency, i.e., pseudo-words). In order to

address this issue, the vocabulary items included as targets of the present study were ten low frequency lexical words (i.e., six adjectives and four nouns) selected by the researcher upon extensive piloting. None of the vocabulary targets included in the present study belonged to the 5,000 most frequent Spanish words, as *lema* or as word form (Davies, 2006). Davies (2006) frequency dictionary of Spanish was selected because it is the most representative and up-to-date frequency dictionary for Spanish currently available: Its lexicon is based on a 20 thousand million word corpus, and genre and register were taken into account for frequency counts.

Besides the issue of frequency, similarity of form, i.e., cognateness is a methodological issue in empirical studies of Spanish vocabulary acquisition by English speakers because English and Spanish are highly cognate languages (e.g., Whitley, 1986). Since learners in a post-exposure test may perform differently based on their ability to guess or recognize the targets from L1-L2 word similarity alone, cognateness must be addressed as a methodological issue. In order to control for this potential source of variability, this study included non cognate lexical words. Non-cognateness of the targets included in the study was assessed by asking ten native speaker of English with no prior exposure to Spanish to rate, on a scale from 0 ('I can't think of any word in English similar to this word') to 5 ('this word reminds me of the English ...') the degree of cognateness of the form.

The ability to guess or recognize a L2 word based solely on its superficial similarity with an L1 word is one factor that should be controlled, but even if an L2 word is unrecognizable merely by its form, e.g., is a non-cognate, learners may still be able to identify the meaning of an unknown word in a reading passage based on contextual semantic and pragmatic cues. The term inferability is used among researchers to refer to the ability to guess words from context (Nation, 2001). In the glossing strand of research, inferability effects were found, for example, in a study by Hulstijn (1993). In order to control for this potential intervening factor, only 'not easy to infer' words were used in the present study.

Inferability of a lexical target was operationalized as the native speaker (in)ability to produce in writing the target or its synonyms. In order to identify non inferable targets, ten native speakers of Spanish of different dialects were asked to complete a fill-in-the-blank test based on the original text containing original targets. Prior to the task, the researcher had manipulated the text so that no contextual clues were provided. All targets whose meanings (or forms) were inferred by a native speaker during the fill-in-the-blank task were eliminated from the study, and an antonym was used instead. An online language dictionary (Kellogg) was utilized to generate the ten lexical targets of the study.

In sum, the lexical items included as target of the study shared the following characteristics: (1) they were low frequency words, (2) they were non cognates, and (3) they were not easy to infer.

Items in the Present Perfect and the Impersonal-SE

This study included ten items in the Present Perfect and ten items in the impersonal-SE. All twenty items were embedded in the reading text. With respect to the Present Perfect, the event structure of the lexical base was taken into account. In order to control for the ‘kind of relevance’ of the ‘recent past event’ to the present point, the study only included achievement/accomplishment verbs. With respect to the Impersonal-SE, only agentive bases were used, to insure an unambiguous impersonal meaning every time the se construction appeared in the text.

The experimental text

The text for this study was a 17-paragraph non-literary article written by the researcher describing common traditions in Argentina (Appendix A). The text was 667-word long and was designed to have the format of a magazine article for young travelers that could be found on a Web page. There were several criteria for using this text: (a) text length, (b) text genre/format, (c) readability, (d) topic. With respect to length, L2 texts included in previous studies have ranged from 300 to 800 words. The text included in this study was within this range, with 667 words, to ensure comparability with previous studies. Furthermore, previous studies used descriptive and narrative text, and literature excerpts were only used

with advanced students. Taking this into account, the text included in this study had both descriptive and narrative sequences, and it was not fictional or literary. With respect to text readability, the text met the requirements of those usually used in first and second year textbooks. In other words, contents were clearly organized around topical paragraphs that included short and syntactically simple sentences, with clear transitions, explicit pronouns and no implicit pragmatic references. The text's readability index (Fernandez Huerta, 1959; Krantz) was 76. Texts in the 70-80 range are labeled "fairly easy" according to the Fernández Huerta's scale. Finally, the researcher selected a culturally bound topic, that is, local customs and traditions in Argentina, to control for the familiarity effect (e.g., Davies, 1989).

The experimental text had originally a dozen drafts that were modified to include various possible targets. After several trials and extensive piloting of targets over the summer of 2008, a final version was produced containing all final targeted items.

The final version of the experimental text was subsequently modified to create a version that would be utilized in the glossed condition, which included bottom of page L1-translation glosses for each of the 30 items targeted in the study (i.e., ten vocabulary words, ten items in the Present Perfect, and ten items in the Impersonal-SE).

Procedure

In order to recruit participants for this experiment, this researcher visited 12 Spanish classes (Introductory and Beginning) and invited students to participate by giving a short presentation of the goals of the study, that is, to investigate how computers and technology can enhance foreign language learning. Additionally, students received a packet containing a written invitation to participate, a consent form and a document explaining the different phases of the experiment and its time commitment.

The experiment had three phases: (1) pretest, (2) treatment + immediate posttest, and (3) delayed posttest (see Figure 2). The pretests were administered three weeks before the treatment session. On the day of the treatment, participants reported to the language laboratory, where they were randomly assigned to one of four treatment groups. The treatment had been previously designed in the computer by using the Blackboard system. Four different instruction sheets were created for each one of the four groups of the study, which were formed based on crossing the variables [+/- think-aloud] and [+/-gloss] (Appendix B). Each participant was given a packet containing general instructions (i.e., how to record your voice in the computer, instructions to log in to Blackboard,) as well as a background questionnaire.

After a short think-aloud training and practice session, participants in the think-aloud groups were prompted to use the headset and start the experiment,

following each prompt on the computer. Those participants who did not think-aloud were prompted to start following the computer instructions. Upon reading the group instructions and experimental text, participants went on to complete the reading comprehension test, and the immediate posttests. Participants completed the production test before the recognition test. This test order was adopted in order to minimize test effects. The delayed posttest (i.e., phase three) was completed in the laboratory three weeks after the experiment (see Table 5).

Table 5. Phases Of The Experiment

Phase 1	Week 1	Pretest			
		↓	↓	↓	↓
Phase 2	Week 4	+TH+GL	+TH-GL	-TH+GL	-TH-GL
			Reading		
			Production Postest		
			Recognition Postest		
			Production Delayed Postest		
Phase 3	Week 8		Recognition Delayed Postest		

Computerization of the materials of the study

As mentioned above, all materials were computerized using Blackboard (see Appendix C for illustration). This system was utilized for several reasons.

First, as the course management system used by the university community, it was assumed that the participants would feel comfortable interacting with a reading passage within a well known and familiar platform. Second, the Blackboard system offers several features that make it especially suited for this research. The system is linked to a database that records each answer into a ‘grade book’ easy to access by the researcher. The Blackboard design features make it easy to incorporate a reading passage with glosses at the bottom of the page, which can be easily tracked, viewed, and clicked by the participant. The test manager includes a backtracking control system that allows the researcher to control how participants navigate the pages. For example, a parameter can be set to prevent users from coming back to a previous answer or page. The system also includes a controlled release feature to monitor how the different instructions, texts, and tests appear (or do not appear) on the screen as the participants navigate through the system. For example, when participants log in to the system, they cannot see all the tasks of the experiment (i.e., instructions, text, tests). What they can see, instead, is only one of the assignments that they have to complete. They can see the next assignment only after they have completed the first one.

Previous computerized studies on glossing have utilized a tracking system to track the number of times a participant accessed a glossed item in an experiment. A problem for internal validity appears when one act of clicking on an icon is counted by the researcher as one deployment of attention. In other

words, these systems, valid as they might be for other purposes, are problematic when they are used to operationalize the process of attention. For this reason, in this study computer tracking system were not included in the research design. Instead, participants in the glossed conditions were asked to click, every time they saw an L1-translation, on a box that appeared next to the gloss. This is the way in which attention was operationalized in this study. Additionally, the researcher included a measure of noticing and awareness by means of concurrent think-aloud protocols.

The assessment tests

The Reading comprehension questionnaire

Previous studies have used the following measures of reading comprehension, (a) cloze test, (b), recall protocols, (c), think-aloud procedure, and (d) reading comprehension questionnaires. This study utilizes a reading comprehension questionnaire (Appendix D).

The cloze (Oller, 1983) is a procedure that consists on random, or not random, deletion of words in a text. The reader then must fill-out the blanks as she reads the text. The use of this test as an outcome measure of L2 text comprehension is highly controversial. Researchers debate what the cloze really measures, that is, whether it is a reliable measure of vocabulary/grammar knowledge, the readability of a text, the global reading skill of the reader, or other

abilities and/or knowledge. Furthermore, there are many potential problems with construct validity, due to the possibility of text and scoring procedure effects (e.g., allowing or not allowing synonyms), and deletion/restorability issues. Perhaps for this reason, this procedure is rarely used in the glossing strand of research.

Although more popular than the cloze, the use of recall protocols to measure L2 reading comprehension also has issues for construct validity, and reliability (Bernhardt, 1983; Lee, 1986; Wells, 1986). The recall procedure consists of asking the reader to recall as much as she can immediately after exposure to the written text. Recalls are generally written, but they can also be oral. Researchers analyze recalls based on a count of idea units. Within the research on the effects on glossing for reading comprehension, the recall protocol is the most used measure. In spite of its popularity, this procedure has many problems. One of the main problems for this procedure is the unavailability of valid procedures to score idea units in the protocol (e.g., Wells, 1986), the average percentage of idea units typically recalled (approximately, 33%), and the potential role of memory.

Within the glossing strand of research, only one study used think-aloud protocols as the only measure of reading comprehension (i.e., Lomicka, 1998). Although a valid measure of the processes that occur during interaction with the L2 text, the question is raised whether this is a valid outcome measure of reading comprehension (Roby, 1999).

In this study, reading comprehension is operationalized as the ability to decode, or understand, the meaning of the targeted forms and words while reading the L2 text. This is a bottom-up perspective, or a micro-level focus, of L2 reading. In order to measure this ability, this study utilized an L1 comprehension questionnaire. This procedure was selected for several reasons. For one reason, none of the alternative measures mentioned above seems better suited as a reliable measure. Second, questionnaires in the L1 are the most common testing procedures as outcome measures of text comprehension in the classroom, at least at the college level. Third, if designed carefully, questionnaires can elicit relevant information on whether the participant was able to understand the L2 text at the level in which this study is interested, that is, the micro level of word recognition (Hughes, 2003; Hulstijn, 1993). Since this study focused on reading at the micro level of linguistic information, a questionnaire in the L1 seems the measurement best suited for the purpose of this study.

'Learning' measures

The term *learning* is rarely defined in the research strand of glossing. Measures of learning vary greatly in this strand too. More often than not, there is confusion in the operationalization of the terms *recognition* and *production*. Often, it is not clear what information about the learners' abilities and/or knowledge the researcher is hoping to elicit with the test. For example, decisions

to present target forms ‘in their original context’ in the recognition post-test are not justified (e.g., Hulstijn, 1992; Watanabe, 1997).

In this study, learning is operationalized as the ability to recognize and produce the targets immediately after exposure, and three weeks later. Taking into account the processing stages of learning (e.g., intake, L2 development), intake is operationalized as the ability to recognize targets, while L2 development is operationalized as the ability to produce the targets.

Recognition (pre- and post-) test. The purpose of this multiple choice test is to elicit whether the participant are able to recognize the target forms immediately after exposure, and three weeks later. Intake is operationalized in the study as the ability to recognize the targets in this test. This test elicits receptive knowledge of targets (Appendix E).

Production (pre- and post-) test. In this study, L2 development is operationalized as the ability to produce the targets in writing immediately after exposure and three weeks later. The purpose of this translation task is to elicit whether the participants are able to produce in writing the target forms and vocabulary words immediately after exposure, and three weeks later. This test elicits productive knowledge of targets (Appendix F).

Think-aloud protocols

This study included think-aloud protocols as a measure of an attention and noticing. The think-aloud protocols collected in this study served to elucidate the processing aspects of reading comprehension and language learning. The reading patterns found in the protocols were useful to address the issue whether the presence, versus absence, of a translation gloss would affect reading comprehension from the point of view of the reading processes involved while reading an L2 text on-line. Furthermore, protocol analyses provided the researcher with useful information about patterns of noticing, and awareness, within two glossing conditions, and how these patterns of noticing under two conditions were related, or not, to the learning measures of the study. At the same time, it was possible to establish in the protocols that participants were paying attention to the input during task performance. This was established before submitting posttests scores to statistical analyses.

In order to address the question of glossing and reading comprehension, on the one hand, and the question of the relationship between reported noticing of targets under two glossing conditions and learning, qualitative and quantitative analyses were conducted in two stages. In the first stage, think-aloud protocols were coded for instances of reading patterns in both the gloss and the no-gloss groups. In this context, the leading question for the analysis was on how the participants processed the text for its contents and ideas, and whether the

participants in the gloss group used the translation glosses, and if so, how. For the purpose of this analysis, the L2-text original clauses, as read aloud in the think-aloud protocols, were distinguished from ‘novel’ clauses, that is, any comments about the text contents produced by the participants throughout the reading task. Thus, any paraphrases, inferences, explanations, comments or questions the participants made while interacting with the L2 text were coded as *interpreting units*, in order to differentiate them from read-aloud clauses.

In the second stage of the analysis, the think-aloud protocols were coded for instances of noticing and awareness. In this context, the questions were if, and how, the participants processed linguistic form. The researcher followed standard procedures in previous SLA studies utilizing concurrent elicitation techniques to investigate noticing and awareness (e.g., Bowles, 2004). Thus, for the purposes of the analysis, any mention of a targeted form, lexical or grammatical, accompanied by a pause or a mention of its translation (gloss group) or a mention of not knowing the meaning of the word (no-gloss group) was coded as an instance of noticing. Any comment about the targeted items that indicated an explicit focus on the items as language *per se*, for example, linguistic analyses of forms and their meanings, connections between meaning and forms made explicit, overt comments about meaning-form relationships, was coded as an instance of awareness at a higher level than noticing.

All the protocols included for the analyses were transcribed and coded by the researcher. Additionally, in order to calculate the inter-rater reliability of the coding procedure, the researcher randomly selected ten per cent of the recordings and asked a graduate student to code them (Kappa coefficient was high: .98).

Since the purpose of the study was to investigate processing during a reading task, the participants were asked to think-aloud only during the experimental task, i.e., while reading the L2 text, and not while completing the post-exposure performance tests.

Seventy-two participants, out of an original pool of 137, were asked to think aloud while reading the experimental text. Of the 72 participants who thought aloud, 36 were not considered for the analysis because of one of the following reasons: (1) the protocols were inaudible, (2) the participants did not follow the directions to think-aloud, (3) the participants scored two or more points in the recognition and production pretests, and/or (4) the participants did not complete all phases of the experiment.

Both the translation gloss and the no-gloss experimental groups included participants who thought aloud. The distribution of protocols between the experimental groups was almost even: Nineteen think-aloud protocols were produced by participants in the gloss group, while seventeen protocols were produced by participants in the no-gloss group.

Overview of the research design

This study has a pretest-posttest design with four independent groups and a within-group factor. T-tests and ANOVA techniques were used for the quantitative analyses. Separate analyses were conducted for each dependent measure of the study: reading comprehension, recognition, and production immediate and delayed posttests. For the dependent measures of the study, reliability coefficients were reported. In addition to the previous analyses, the study included qualitative and quantitative analyses of the relationship between noticing/awareness and learning of items.

CHAPTER 3: RESULTS

For ease of exposition, this chapter is divided into two parts. The first part reports the qualitative and quantitative results of the think-aloud protocol analyses included in the study, while the second part reports the quantitative results of the outcome measures of the study.

Part I. The think aloud protocols

Answer to RQ#(1) Reading patterns and styles as revealed through a think-aloud task

Overall, the participants in both the gloss and the no-gloss groups approached the reading task in a linear and bottom-up fashion. Participants in both groups tended to read the text from beginning to end without making connections between paragraphs or recapitulating ideas. Reading aloud in Spanish or reading aloud by translating the text into English, or a combination of these, were the predominant reading behaviors in both groups (cf. also Leow, 2001 for similar findings). In general, participants in both groups did not show elaboration or integration of text contexts with their own schemata and knowledge in order to construct a coherent mental model for the text. In general, participants were not making inferences, connecting ideas to make causal or temporal relations, solving

contradictions or drawing conclusions. Only a few participants stopped to think or paraphrased, recapitulated, reorganized and summarized ideas, hypothesized, or asked questions, but even in these cases, the interpretation did go beyond the sentence level. Despite these commonalities between the groups, the presence of a translation gloss affected reading comprehension processes, as the text proved difficult to understand for the participants in the no-gloss group. The following paragraphs present a summary of the reading behavior observed in the gloss and no-gloss group.

Reading with a gloss

The participants in the gloss group used a combination of word by word reading aloud in Spanish and English translation to gain a minimal understanding of the text at the sentence level. Some participants read more in Spanish, using English mainly to translate the targets, while others preferred to read the text by translating into English, without necessarily reading aloud the targeted item, and some used a combination of these strategies, by reading some parts in Spanish and translating others.

Whether the participants were reading in Spanish, English or a combination of both English and Spanish, there was evidence that glosses were used as an aid to gain understanding of the text contents at the level of sentences. The following seven excerpts are taken from protocols of participants in the gloss

group who scored high reading comprehension questionnaire (in the 80th percentile or higher; fake names are used and targeted items are underlined):

(i) John

River y Boca .. River y Boca son los equipos más importantes. Muchos tienen encono porque River y Boca son los más populares. They are very important players. I'm assuming there is some resentment 'cause they're so popular (...) Muchos esta costumbre típicamente argentina el tango no es fácil encontrar el origen de esta danza. So many admirers después de un espiguelo... they did a search... is a search... algunos creen que el origen del tango es la danza nacido en los barrios portuarios de la ciudad So it is... it comes from the end of the 19th century in the barrios areas of the city. Al principio el tango era considerado la danza de furcias... oh! the tango is considered dance of prostitutes... oh I didn't know that! (Laughs)

(ii) Peter

Debo decirte que algunos visitantes [keeps reading in Spanish].... La lengua que usan es zafe? Sometimes they think ... oh! People use vulgar terms in Argentina... El estilo de hablar se platica.... spoken in los bares y clubs. El deporte is muy importante entre las costumbres [continues reading in Spanish]... siempre se agregan programas de fútbol... So sports are very important in Argentina. Football is the most favorite sport it is very popular and Sundays on TV sometimes are added programs.... Oh! Football programs are added to regular programing on TV.

(iii) Ian

Muchos me preguntas [reads in Spanish] una bombilla de plata... So el contenedor es el mate and la herba se embute.... is packet en el mate luego la bombilla se liba sucked... Talks about how is prepared.... Metal and straw metal and a straw.... Ummm... it's an infusion.... long pause...No se herve people do not boil it.... You prepare this with your family or

friends ... they have sipped this so they definitely... is something they drink.... Debes hacerlo con un amigo fetén But only do this with genuine friends.

(iv) Ethan

Hola soy Mariana Clavero y soy de Argentina. A menudo, escucho opiniones pasmosas PAUSE astonishing sobre los argentinos. Algunas personas some people dicen que los argentinos son macanudos PAUSE cool. Otras personas dicen que en Argentina se abultan todas las cosas PAUSE so people exaggerate!

(v) Laura

You can drink the mate by itself or in the company of acquaintances and friends. Many people have sipped el mate in the company... but to drink mate you must take it with a genuine friend, or do it with a genuine friend.

(vi) David

This appears to be talking about the language kind of the slang.... You should say that these people visiting think that the Argentineans speak a language that is quite vulgar but she doesn't agree with this idea. Without question I think that the style of speaking ... spoken in only the bars and clubs... like the kids... nothing really intense.

... Nuestro equipo nacional de fútbol ha descollado internacionalmente. So this one talks about futbol; doesn't talk about rugby; it's not a good sport... but ummmm talking about how sports... is a very important custom to Argentineans. Soccer is very popular on Sundays, kind of similar to football in America with the ... it's not bizarre finding people watching fútbol on TV in restaurants. And then their national team is internationally very good.

(vii) July

Other customs... one example is el mate.... to prepare el mate ... debes en container straw they're stuffing something with a straw ... la yerba is

packet la yerba is put and then the straw. . . I'm not sure it's sounds most like a piñata I guess.

They have water... hot water in container... The water for mate is not boiled. Puedes beber they drink el mate together with family and friends... En compañía de familiares y amigos so they drink it with a straw, you're supposed to drink it with like a genuine friend.

The above examples show how translation glosses were utilized by the participants to understand the text, at least at the level of sentences. The participants in the gloss group read and understood the glosses and how these fitted in the context. Furthermore, the think-aloud protocol analyses indicated that the participants in this group were reading the text and clicking on the glosses, that is, participants were generally not skipping the translations, which suggest that they needed them to understand the text. Moreover, evidence of how glossing affected comprehension processes was revealed by the fact that the interpreting units produced by participants in the gloss group were generally accurate: Glosses helped them understand the sentences in which the unknown words were included. That tendency is apparent in examples (i-vii) above. However, the sample lacks examples in which the participants show that they are integrating textual information at the paragraph or textual levels, or using their own schemata. Case (ii) is a clear example of a participant who did not integrate ideas at the paragraph level and for this reason his interpretation was incomplete and ultimately incorrect (e.g., *people use vulgar terms in Argentina*). Case (vii), July, is a case in which the lack of integration with the text information led to the

wrong inference (e.g., she thinks that mate is a piñata). Lack of integration of contents is apparent because, even in the face of further evidence that mate is a drink, July never corrected the inaccuracy in her thought process.

In sum, the participants in the gloss group used a combination of bottom-up strategies to gain a minimal understanding of the text at the sentence level. There was evidence that glosses were used as an aid to gain understanding of the text contents, but this understanding was only at the level of sentences. The participants were focused on meaning and on gaining a minimal level of understanding, but they were not using high order strategies (e.g., making inferences, connecting parts of the text, hypothesizing) to construct a coherent model for the text. Similar reading strategies were also reported in previous studies that employed concurrent data elicitation procedures during the reading process (e.g., Leow, 2001b; Leow et al., 2003; Bowles, 2004).

Reading without a gloss

The reading behavior of participants in the no-gloss group showed similar reading patterns to those in the gloss group. The participants in the no-gloss group read the text in a bottom-up and linear fashion, without engaging higher order strategies of elaboration and integration of the text contents. In the absence of translations for unknown words, some participants preferred to read aloud all the

text in Spanish, from beginning to end, without pausing to guess words or interpret sentences, while other participants used a combination of Spanish reading with English translations in their attempt to understand the text contents. When comments were made, they were mostly in reference to words, phrases or sentences, and not whole paragraphs.

In the gloss group, the protocol analysis indicated that accessing of the gloss produced several breaks in the reading process, in comparison to the no-gloss group. The gloss group spent an average of 12.58 (SD = 3.6) minutes on the experimental task, while the no-gloss group spent an average of 11.06 (SD = 3.7) minutes. However, one-way ANOVA revealed no statistically significant difference between two glossing groups on time on task, $F(1, 34) = 1.541, p = .223$.

The analysis of the 17 think-aloud protocols in this group indicated that reaching a minimal understanding of the text was difficult for the participants who read the text without glosses. Below are some excerpts from the think-aloud protocols of participants in this group who scored high (in the 80th percentile or higher) on the reading comprehension outcome measure:

(i) Tom

River y Boca are the most important teams muchos tienen encono porque River y Boca son los más populares... They have a lot of fans? A lot of fame? Because are very popular... I don't know...

... Ummm something about fire in metal and other things. Something about yerba and bombillas yeah!!!

(ii) Lucy

Mariana Clavero is from Argentina. mmmmm She is listening to opinions about Argentines. For example, some people say that in Argentina todos son macanudos. I don't know what macanudos mean. Other people say that in Argentina they do something with everything. LONG PAUSE. ... Some that have already read my page on the Internet. I tell them... I tell my readers that the culture and traditions in Argentina are special. They don't understand maybe PAUSE I don't understand se desdennan. Language I should discern PAUSE I should tell you that when two Argentines speak, the language that they use is zafia PAUSE I don't understand zafia. This includes some Argentines that have sometime... with this idea. I think that the style of speaking takes place in bars and clubs of the city. Fútbol or soccer Fútbol is our favorite. Fútbol is very popular on Sundays on TV. We always they always play sport ummm soccer programs for regular programming It's not estrambótico... I don't understand estrambótico to meet people watching TV in a nice restaurant. That's interesting. Our national fútbol team has already been known internationally. River and Boca are the most important teams. Many have encono. I don't know what encono means PAUSE because River y Boca are more popular. Oh! Maybe they have things in common, encono. Al principio, tango is considered una furcia de dance... una danza de furcias... I don't know what furcias means. For this reason the themes of the tango are sad themes PAUSE oh! It's a dance of frustration! Furcia means frustration. For example, tangos are canciones Oh! songs, los tangos are the songs of tango. So they are songs that speak of human beings that nothing paliado I don't know what paliado means where love is fragile and disappears...

(iii) Rebecca

Macanudos probably means the same as everyone else... Some lectures that have something my page in the Internet have asked me why Argentines are special. I respond to them that the culture and traditions

are special. Unfortunately sometimes desden the culture and... Algunos lectores que ha elogiado...ok yo les respondo que respond to them that the culture are special unfortunately sometimes they have lack of understanding.

Debo to tell you that... Some tourists think that when two Argentinians speak the language that they use is... This I include translated with this idea. Nevertheless I think that the style of speaking takes place in the bares.

zafia... .qué es zafia...The sport is very important among the customs... Soccer is our favorite sport... Soccer is more popular on Sundays on TV... always they get together programs of football on TV.... the players of the famous teams River and Boca han renido have duro... They lasted?

(iv)

This is pretty much how other people see Argentinians they think that they are macanudos, I don't know what that means.... Think that they... I think that's laugh all the time but other people don't think that at all. Overall Argentinians they are unique in many aspects like language, sports, and their customs.

Some lectors ask me why the Argentinians are special, and I respond to these people that they are special... Unfortunately there are many that do not know the culture and the traditions. The language. Should tell you that the following of what I think.... The language that they use is zafa.... I include ... changes with this idea....if I don't I think that this is from talking in the bars and clubs around the country. Fútbol. The sport that is most important fútbol is more popular ... el fútbol is the favorite sport it is so popular that the Sundays on the TV there are programs, so there is always regular programming on TV. Not sure what strambótico means ummm. You'll not be able to encounter a person watching part of a fútbol so you won't see people watching tv in a restaurants usually.

Many persons drink the mate with amigos.... But to drink the mate we should do with un amigo at a party? So it's kind of... a traditional beverage for Argentinian people mate and they drink it only with family and friends you shouldn't drink alone, maybe it has alcohol in it.

(v) Marian

Hola soy Mariana Clavero y soy de Argentina. A menudo, I listen to the opinions about the Argentinians. For example, some people talk that Argentina everything about something. Other people talk about Argentina something. Nevertheless, millions of people that have been to Argentina or something like that don't understand us... People think that Argentinians are unique and very respected how sports and something don't know what that is.

Other lectures groups that something my page on the Internet ask me questions about Argentinians specially. I respond that they are special. Unfortunately sometimes desdenan I don't know what that is the culture and the traditions. Nice.

I think talking to you that some visiting think that when two Argentinians talk the language they use is I don't know what zafia is... maybe different?

... they talk to us en something in the city. I don't quite understand that. No... I don't know what that is maybe it's more popular than American fútbol

No strambótico other people watch futbol on TV in a restaurant. Some watch the national fútbol cheer for international soccer maybe.

(vi) Brad

Hola soy mariana clavero y soy de argentina. A menudo escucho opiniones pasmosas sobre los argentinos. Por ejemplo, algunas personas dicen que en Argentina todo...

Mariana is from Argentina. She listens to different opinions. For example some persons say that Argentinians are macanudos ... they're... otras personas say that in Argentina se abultan all the las cosas. Sin embargo millones de personas que han surcado la Argentina no piensan eso. Ellos piensan que los argentinos son únicos en muchos aspectos, como la lengua los deportes y las costumbres. They think that argentinos are one-sided, are one way in a lot of aspects, like la lengua y las costumbres

These excerpts illustrate how the participants in the no-gloss group were reading the text without glosses. The general tendency was to read the text word by word in Spanish or English without thinking, for many, on the meanings of the

unknown words. When a targeted item was encountered, participants tend to avoid it by skipping it or replacing it with the word ‘something’ (e.g., v and vi), said that they did not know the meaning of the target (i-vi) or guess its meaning from context (iv and v). Furthermore, the majority of interpreting units in the no-gloss group led to wrong inferences.

In sum, a comparison of the reading behavior of both the gloss and the no-gloss groups shows commonalities and differences. A commonality was that the participants in both groups used similar approaches to the reading task, i.e., a tendency to read the text in a bottom-up and linear fashion without indications of higher order cognitive engagement. Furthermore, in both groups there were participants who used the target or the native language, or a combination of both. Despite the commonalities, some differences were observed. For example, the participants in the gloss group paid attention to, clicked and used the translation glosses as comprehension aids. Most revealing, the presence of a gloss helped the participants produce more accurate paraphrases of and comments on the text than the participants in the no-gloss group. In the no-gloss group, the lack of glosses sometimes led the readers to infer the wrong meanings. These findings strongly suggest that the presence of a translation gloss helped learners understand more of the text, in comparison to the control group.

Research questions (2-3) of this study asked about the relationship between glossing and noticing, and type of linguistic item and noticing. In order to answer these questions, think-aloud protocols were analyzed for instances of noticing and awareness, of the targeted items overall, and by item types. The results of these analyses are summarized in the next sections.

RQ#(2) Do readers exposed to glossed text report significantly more noticing of targeted items (overall) than readers exposed to no-gloss text?

An analysis of noticing of targeted items revealed that the participants in the gloss group noticed the targeted items, overall, more than participants in the no-gloss group. However, one way ANOVA revealed that this difference was not statistically significant, $F(1, 35) = 2.061, p = .160$. (See Table 6 for Mean, SD, SE for two gloss groups on noticing scores).

Table 6. Mean, SD And SE Of Noticing For Two Gloss Groups

Group	Mean	SD	SE
Gloss (N=19)	13.58	8.3	1.9
No-Gloss (N=17)	9.82	7.2	1.7

Note. Maximum score on noticing = 30.

When noticing scores are broken down into three type of items, ANOVA comparisons of two groups on the noticing for three types of item revealed no statistically significant differences: for noticing of Lexical Item, $F(1, 35) = 1.141, p = .293$; for noticing of Present Perfect, $F(1, 35) = 1.394, p = .246$; for noticing

of Impersonal-SE, $F(1, 35) = 2.502, p = .123$. It appeared that having or not having a gloss did not have an impact on the noticing of three types of items. Qualitative analyses of the think-aloud protocols did not support the claim that participants in one group or the other demonstrated higher levels of awareness of the targets, i.e., beyond noticing. In fact, there were only a few cases of participants who demonstrated awareness of the items, beyond mere noticing, and these few cases were found in both the gloss and the no-gloss groups.

Only a few of the participants in both groups demonstrated awareness of items beyond the noticing level. For example, the participant who demonstrated the highest level of awareness in the gloss group, according to our coding system, was Vivian. As it is shown in the example below, Vivian made explicit the connections between word meanings and their forms by introducing the copular verb *is* between the target and its English translation, and by repeating the glossed items (underlined) at the end of paragraph:

(i) Vivian (Gloss group)

Nuestro equipo nacional de futbol ha descollado internationally... So has excelled so the argentinian futbol has excelled internationally... So descollado is excelled, agregan is added, and estrambótico is bizarre.
Fútbol no se veda... So I think it's basically saying that there is a big rivalry between River y Boca... Reñido es quarrelled.... vilipendiado is instulted.... Veda is banned and encono is resentment.
Sin duda el espiritu del tango is lóbrego El espíritu del tango is lóbrego.... So is gloomy.... Lóbrego is gloomy.

Tango wasn't famous... Carlos Gardel lo inmortalizó.... No ha espichado so he's died sin antes catapultar el tango a la fama.... Hoy el tango es tan popular que todos quieren bailar cuando se vista la banda. So it's so popular... and vista is people see, and espichado is died ... and lóbrego is gloomy.

Besides Vivian, no participant in the gloss group made explicit form-meaning connections in this or other ways.

In the no-gloss group, participants were aware that they did not know the meaning of the targeted lexical words, i.e., they *noticed* the items, in our coding view. In this group, there were only a few participants showing higher levels of awareness by, for example, pausing to reflect on the item and trying to infer its meaning from context. Below is an example.

(i) Paula

zafia... qué es zafia? (seven seconds pause)

The rivalry between the two teams is very strong... han vilipendiado..... vilify each other? (six seconds pause) Maybe?

Besides these examples that were coded as *awareness* in this study in order to distinguish them from the *noticing*, it should be noted that participants in both groups did not engage in language analysis. For example, they did not established explicit relations with previously known words in the target language (e.g., 'Sp. furcia is the same as Sp. Prostituta', 'fetén is like amigo bueno'; 'estrambótico is like raro'; 'lóbrego is the same as triste'), made word

associations or repeated a target more than twice. In sum, the data corpus showed patterns of noticing, with the gloss groups noticing more items overall than the no-gloss groups. However, the participants in both groups did not demonstrate higher levels of awareness of the targeted items overall.

RQ#(3a-b) Do glossing and type of item (i.e., Lexical Item versus Present Perfect versus Impersonal-SE), have an effect on noticing?

In order to answer research question (3) of the study, mixed ANOVA analyses were conducted on the noticing scores of each one of three types of items comparing two gloss groups. These analyses revealed a main effect for type of item, $F(2, 34) = 6.607$, $p < .05$, eta squared = .16, no main effect for gloss, $F(2, 34) = 1.975$, $p = .169$, and no interaction, $F(2, 34) = .126$, $p = .882$. (See Table 7 for Mean and SD for noticing of type of item by gloss group).

Table 7. Mean And SD For Noticing Of Three Types Of Items By Gloss Group

Type of Item	Group	Mean	SD
Lexical Item	Gloss (N=19)	5.11	2.8
	No-Gloss (N=17)	4.06	3.0
Present Perfect	Gloss (N=19)	4.53	3.3
	No-Gloss (N=17)	3.29	2.8
Impersonal-SE	Gloss (N=19)	3.95	2.7
	No-Gloss (N=17)	2.53	2.5

Paired Samples T-tests on the noticing scores within each group were conducted in order to establish whether there were differences between types of item. This analyses revealed statistically significant differences between noticing of Lexical Item and noticing of Impersonal-SE within the gloss group, $t = 3.139$,

df = 18, $p < .05$, on the one hand, and noticing of Lexical Item and Impersonal-SE within the no-gloss group, $t = 2.227$, $df = 16$, $p < .05$, on the other hand. No significant differences were observed between Lexical Item and Present Perfect in any of the groups (for the gloss group, $t = 1.398$, $df = 18$, $p = .179$; for the no-gloss group, $t = 1.307$, $df = 16$, $p = .210$, or between Present Perfect and Impersonal-SE in any of the groups (for gloss, $t = 1.037$, $df = 18$, $p = .314$; for no-gloss, $t = 1.540$, $df = 16$, $p = .143$).

In sum, results revealed a type of item effect on noticing. Lexical Items were noticing significantly more than Impersonal-SE items in both the gloss and no-gloss groups. Present Perfect did not differ statistically from Lexical Item or the Impersonal-SE with respect to noticing in both the gloss and the no-gloss groups. Furthermore, a result of no observed effect for gloss was confirmed, that is, participants in both the gloss and the no-gloss groups did not appear to differ statistically in amount of noticing of targets during the on line reading task.

RQ#(4) Is there a significant relationship between the reported noticing of glossed versus no-glossed targeted items and the recognition and production of these items? If so, are the two relationships significantly different?

In order to answer research question (4) the scores on reported noticing and the recognition tests² were submitted to Pearson product-moment correlation analyses. The results of these analyses revealed significant correlations for the reported noticing of targeted items and the immediate recognition of items for

both the gloss ($r = .66$; shared variance = .43) and the no-gloss group ($r = .68$; shared variance = .46). With respect to the delayed recognition posttest, results revealed a similar pattern of significant correlations between noticing and delayed recognition of items for both the gloss group ($r = .58$; shared variance = .33), and for the no-gloss group ($r = .49$; shared variance = .24).

Subsequently, a Fisher's z Transformation and Comparisons between Independent r s was conducted on the two correlation coefficients, for immediate and delayed recognition separately, in order to determine whether the relationships differed statistically. With respect to the immediate recognition posttest, the results indicated that there was no significant difference between the two relationships ($z = .09$; $p = \text{n.s.}$). Similarly, no significant difference was found between the correlation coefficients corresponding to the relationships between reported noticing by group and delayed recognition posttest ($z = .34$; $p = \text{n.s.}$). Similar analyses were conducted on each type of item separately. With respect to Lexical Item, the results of these analyses revealed significant correlations for the reported noticing of targeted items and the immediate recognition of items for the gloss ($r = .53$; shared variance = .28) but no significant correlation for the no-gloss group ($r = .21$). With respect to the delayed recognition posttest, results revealed a similar pattern: the correlation between amount of reported noticing and delayed recognition of Lexical Item was significant for the gloss group ($r = .53$; shared variance = .28), but not for the no-gloss group ($r = .42$).

Subsequently, a Fisher's z Transformation and Comparisons between Independent r s was conducted on the two correlation coefficients, for immediate and delayed recognition of Lexical Item separately, in order to determine whether the relationships differed statistically. With respect to the immediate recognition posttest, the results indicated that there was no significant difference between the two relationships ($z = 1.03$; $p = \text{n.s.}$). Similarly, no significant difference was found between the correlation coefficients corresponding to the relationships between reported noticing of Lexical Item by group and delayed recognition posttest ($z = .38$; $p = \text{n.s.}$).

With respect to Present Perfect, the results of these analyses revealed significant correlations for the reported noticing of targeted items and the immediate recognition of items for the gloss ($r = .60$; shared variance = $.36$) and no-gloss groups ($r = .56$; shared variance = $.31$). With respect to the delayed recognition of Present Perfect, no significant correlations were observed between amount of reported noticing and delayed recognition of Present Perfect, for both gloss group ($r = .24$), and no-gloss group ($r = .13$).

Subsequently, a Fisher's z Transformation and Comparisons between Independent r s was conducted on the two correlation coefficients, for immediate and delayed recognition of Present Perfect separately, in order to determine whether the relationships differed statistically. With respect to the immediate recognition posttest, the results indicated that there was no significant difference

between the two relationships ($z = .16$; $p = \text{n.s.}$). Similarly, no significant difference was found between the correlation coefficients corresponding to the relationships between reported noticing of Present Perfect by group and delayed recognition posttest ($z = .31$; $p = \text{n.s.}$).

With respect to Impersonal-SE, the results of these analyses revealed no significant correlation for the reported noticing of targeted items and the immediate recognition of items for the gloss ($r = .29$). However, the correlation was significant for the no-gloss group ($r = .67$; shared variance = .44). With respect to the delayed recognition of Impersonal-SE, a significant correlation was indicated between amount of reported noticing and delayed recognition of Impersonal-SE, for gloss ($r = .47$; shared variance = .22), but no significant correlation was found for the no-gloss group ($r = .22$).

Subsequently, a Fisher's z Transformation and Comparisons between Independent r s was conducted on the two correlation coefficients, for immediate and delayed recognition of Impersonal-SE separately, in order to determine whether the relationships differed statistically. With respect to the immediate recognition posttest, the results indicated that there was no significant difference between the two relationships ($z = 1.39$; $p = \text{n.s.}$). Similarly, no significant difference was found between the correlation coefficients corresponding to the relationships between reported noticing of Present Perfect by group and delayed recognition posttest ($z = .78$; $p = \text{n.s.}$).

In sum, results overall revealed that the amount of reported noticing was positively correlated with the ability to recognize those items immediately after exposure to the reading text and three weeks later. However, these relationships were statistically similar for the gloss and the no-gloss group. At the level of each type of item, noticing of Lexical Item was positively correlated with item recognition for the gloss group but not for the no-gloss group. Noticing of Present Perfect was positively correlated with the immediate retrieval of these items for both the gloss and the no-gloss group, but not with item retention. Finally, noticing of Impersonal-SE was positively correlated with immediate recognition of items for the no-gloss group, and with retention of item in the gloss group.

For all types of items, there was no statistically significant difference between the relationships for the two groups.

Summary of findings of the think-aloud protocol analyses

The think-aloud protocols collected in this study served to elucidate the processing aspects of reading comprehension. The think-aloud protocols showed similar patterns of reading behavior in both gloss and no-gloss groups, and also pointed out some important differences.

A comparison of the reading behavior of both the gloss and the no-gloss groups shows commonalities and differences. A commonality was that the participants in both groups used similar approaches to the reading task, i.e., a

tendency to read the text in a bottom-up and linear fashion without indications of higher order cognitive engagement. Furthermore, in both groups there were participants who used the target or the native language, or a combination of both. Despite the commonalities, some differences were observed. For example, the participants in the gloss group paid attention to, clicked and used the translation glosses as comprehension aids. Most revealing, the presence of a gloss helped the participants produce more accurate paraphrases of and comments on the text than the participants in the no-gloss group. In the no-gloss group, the lack of glosses sometimes led the readers to infer the wrong meanings. These findings strongly suggest that the presence of a translation gloss helped learners understand more of the text, in comparison to the control group.

The quantitative analyses of the think-aloud protocols conducted in this study revealed patterns of relationships between glossing, noticing, and a measure of intake of targeted items. With respect to glossing and noticing, the results did not support a claim for a glossing effect on noticing. Furthermore, a type of item effect found in the study indicated differences in processing between Lexical Item and the Impersonal-SE, while no differences were observed between the noticing of Lexical Item and the Present Perfect. The analyses also revealed that the amount of reported noticing was positively correlated with a measure of intake taken immediately after exposure to the reading text and three weeks later. It appeared that the more the participants were able to notice the items, the more

they were able to intake, and vice versa, the less the participants noticed, the less they were able to intake the targeted items embedded in the input. Perhaps more revealing, the relationship between reported noticing and intake of items did not depend on the condition of exposure. In light of this evidence, glossing condition does not appear to be a good predictor of noticing.

Besides these quantitative analyses, the study also included qualitative analyses of noticing and awareness patterns, as revealed in the think-aloud protocols. These analyses did not support the claim that participants in one group or the other demonstrated higher levels of awareness of the targets, i.e., beyond noticing. In fact, there were only a few cases of participants who demonstrated awareness of the items, beyond mere noticing, and these few cases were found in both the gloss and the no-gloss groups. Furthermore, the few cases in which there was evidence of a higher level of awareness, e.g., via explicit form-meaning connection verbalizations, corresponded to Lexical Items. Virtually no awareness, beyond the noticing level, was observed for the form-meaning connections that define the grammatical items included in the study. Based on this evidence, there were reasons to believe that any ‘focus on form’ was a focus on *what was lexical in the text* and not on what was grammatical, and this was true regardless of group. When considered together, the quantitative and qualitative analyses of the study strongly suggest that Lexical Item is more ‘isolatable’ than grammatical

items. While the contribution of noticing to this ‘isolatability’ is clear, the contribution of glossing remained uncertain.

Part II. Quantitative analyses of the outcome measures

Introduction

Besides investigating empirically effects of glossing on amount of reported noticing, and the relationship between reported noticing in two gloss conditions and learning measures, this study also investigated the effects of glossing and type of linguistic item on the learning measures of the study, that is, recognition and production immediate posttests and delayed posttests. The qualitative and quantitative results for glossing and noticing have been presented in the previous section. In this section, the results for glossing and type of item in relation to recognition/production of targeted items are reported. But before doing so, we present a summary of results that pertain to two important methodological issues: establishing no prior knowledge of the target items (i.e., analyses of pretests scores), and reactivity (i.e., whether a think-aloud task *per se* would impact thought processes).

Pretests

Statistical analyses were conducted in order to determine whether the participants had similar levels of abilities at the outset of the study, regardless of the groups to which they were randomly assigned (i.e., +think-aloud or -think-aloud, gloss versus no-gloss) and regardless of the item type. The results of these analyses are presented below.

Recognition Pretests

Independent sample *t*-tests of the [+think-aloud] versus [-think-aloud] mean scores on three recognition pretests included in the study indicated no significant difference between groups. Thus, there was no significant difference between think-aloud groups on the vocabulary recognition pretest ($t = 1.146$, $df = 63$, $p = .256$), the Present Perfect pretest ($t = .153$, $df = 63$, $p = .879$), or the Impersonal-SE pretest ($t = .712$, $df = 63$, $p = .479$) (See Table 8 for Mean, SD, and SE of recognition pretests of three types of item by think-aloud group).

Table 8. Means, SD, And SE For Recognition Pretests Of Three Types Of Item By Think-Aloud Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
+ TK (n=36)	.11	.31	.05	.03	.16	.02	.06	.23	.03
- TK (n=29)	.03	.18	.03	.03	.18	.03	.10	.31	.05

Note. Maximum score = 10.

* $p < .05$.

In order to investigate whether the gloss versus no-gloss groups started out at statistically similar ability levels, the mean scores for glossing on the recognition pretests were submitted to independent samples t-tests. These analyses indicated no statistically significant differences between groups on any of the three recognition pretests. Thus, there was no significant difference between the groups, i.e., L1-gloss versus control, in their ability to recognize the lexical items ($t = 1.433$, $df = 63$, $p = .157$), the Present Perfect ($t = .022$ $df= 63$, $p = .983$) or the Impersonal-SE ($t = 1.433$, $df = 63$, $p= .157$) on the respective pretests. (See Table 9 for Mean, SD, and SE for recognition pretests of three types of item by gloss group)³

Table 9. Means, SD, And SE For Recognition Pretests Of Three Types Of Items By Gloss Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
+ Gloss (n= 33)	.03	.17	.03	.03	.17	.03	.03	.17	.03
- Gloss (n= 32)	.13	.33	.05	.03	.17	.03	.13	.33	.05

Note. Maximum score = 10.

* $p < .05$.

Finally, since type of linguistic item was an independent variable in the study, it was important to address the issue whether, at the outset of the study, each participant did or did not possess differential amounts of knowledge or ability level across item types. In order to address this issue, ANOVA on the recognition pretests was conducted, in which type of item (i.e., vocabulary versus Present Perfect versus Impersonal-SE) was entered as a within subject factor. The results of these analyses indicated no statistically significant differences of each participant across different types of items at the pretest stage ($F = .899$, $df = 2$, $p = .410$). (See Table 10 for means and SD on pretests by three item types).

Table 10. Mean And SD For Three Types Of Item At Pretest

	Mean	SD	N
Vocabulary	.08	.26	65
PP	.03	.17	65
Impersonal-SE	.08	.26	65

Note. Maximum score =10.

* $p < .05$.

Production Pretests

Independent sample *t*-tests of the [+think-aloud] versus [-think-aloud] mean scores on three production pretests included in the study indicated no significant difference between two think-aloud groups on the vocabulary production pretest ($t = 1.116, df = 63, p = .269$). The *t* value could not be computed for the Present Perfect and Impersonal-SE because the standard deviations of both groups were equal to zero. (See Table 11 for means, SD, and SE for production pretests of three types of item by think-aloud group).

Table 11. Means, SD, And SE For Production Pretests Of Three Types Of Item By Think-Aloud Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
-TK (n= 29)	.03	.18	.03	.00	.00	.00	.00	.00	.00
+TK (n=36)	.00	.00	.00	.00	.00	.00	.00	.00	.00

Note. Maximum score = 10.

* $p < .05$.

In order to investigate whether the gloss versus no-gloss groups started out at statistically similar ability levels, the mean scores for gloss on the production pretests were submitted to independent samples *t*-tests. These analyses indicated no statistically significant differences between groups on any of the three production pretests. Thus, there was no significant difference between the gloss groups, i.e., L1-gloss versus control, on the vocabulary production pretest ($t = .984, df = 63, p = .329$). *T* values could not be computed for the Present Perfect and the Impersonal-SE because the standard deviations of both groups were equal to

zero. Descriptive statistics (means, SD, and SE for production pre-tests of three types of item by gloss group) are displayed in

Table 12. Means, SD, And SE For Production Pretests Of Three Types Of Items By Gloss Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
+ Gloss (n= 33)	.03	.17	.03	.00	.00	.00	.00	.00	.00
- (n= 32)	.00	.00	.00	.00	.00	.00	.00	.00	.00

Note. Maximum score = 10.

* $p < .05$.

Finally, in order to address differential knowledge across items at the outset of the study, an ANOVA test on the production pretests was conducted, in which type of item (i.e., vocabulary versus Present Perfect versus Impersonal-SE) was entered as a within-subject factor. The results of these analyses indicated no statistically significant differences of each participant across different types of item at pretest, $F(2, 158) = 1.000, p = .371$. Descriptive statistics (means and SD of production pretests by type of item) are displayed in Table 13.

Table 13. Means And SD Of Production Pretests By Type Of Item

	Mean	SD	N
Vocabulary	.02	.12	65
PP	.00	.00	65
Impersonal-SE	.00	.00	65

Note. Maximum score =10.

* $p < .05$.

To summarize the results of the pretests, the quantitative analyses performed on the pretests indicated that all participants regardless of group were

at a similar level of abilities at the outset of the study regarding their recognition, production, and prior knowledge of the target items.

The reactivity of verbal reports

Prior to addressing the research questions of the study, the reactivity of concurrent verbal reports on reading comprehension, recognition and production tests was investigated⁴. The results of these analyses are summarized below.

Reactivity of think-aloud protocols on reading comprehension

In order to investigate whether non-metalinguistic concurrent verbal reports were reactive, in the sense of ‘thought altering’ devices, the mean scores for the plus versus minus think-aloud groups on the reading comprehension post-exposure test were submitted to an independent samples t-test. No significant difference was found between groups on the comprehension test ($t = .763$, $df = 63$, $p = .448$). In other words, thinking aloud during task performance did not appear to be reactive on participants’ subsequent reading comprehension. Descriptive statistics (means, SD, and SE for two think-aloud groups on text comprehension) are displayed in Table 14.

Table 14. Means, SD, And SE For Two Think-Aloud Groups On Text Comprehension

Group	M	SD	SE
+TK (n= 36)	13.06	4.50	.76
- TK (n=29)	13.90	4.18	.77

Note. Maximum score in the reading comprehension test= 30.

* $p < .05$.

Reactivity of think-aloud protocols on recognition posttest

To investigate the reactivity effect of verbal reports on the three recognition posttests, the raw scores on these tests were submitted to a t-test for independent samples that included think-aloud as the group factor. The results of these analyses indicated no statistically significant differences between groups, i.e., [+/-] think aloud on the ability to recognize the targeted items (Vocabulary: $t = .040$, $df = 63$, $p = .968$; Present Perfect: $t = .176$, $df = 63$, $p = .861$; Impersonal-SE: $t = .028$, $df = 63$, $p = .977$). In other words, thinking aloud during task performance did not appear to be reactive on participants' subsequent recognition of the targeted items. Descriptive statistics (means, SD, and SE for recognition immediate posttests of three types of item by think-aloud group) are displayed in Table 15.

Table 15. Means, SD, And SE For Immediate Recognition Posttests Of Three Types Of Items By Think-Aloud Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
- think (n= 29)	3.62	1.78	.33	3.86	2.90	.53	3.62	2.61	.48
+ think (n= 36)	3.64	1.88	.31	4.00	3.32	.55	3.64	2.54	.42

Note. Maximum score = 10.

* $p < .05$.

Reactivity of think-aloud protocols on production posttests

The reactivity effect of verbal reports on the three immediate production posttests was investigated by submitting the raw scores on these tests to a t-test for independent samples that included think-aloud as the group factor. The results of these analyses indicated no statistically significant differences between groups, i.e., plus versus minus think-aloud on the ability to produce lexical items ($t = .274, df = 63, p = .785$). T values could not be computed for the Present Perfect and the Impersonal-SE because the standard deviations for both groups were equal to zero). Thinking aloud during task performance did not appear to be reactive on participants' subsequent production of the targeted items. Descriptive statistics (mean, SD, and SE for production immediate posttests of three types of item by think-aloud group) are displayed in Table 16.

Table 16. Means, SD, And SE For Immediate Production Posttests Of Three Types Of Items By Think-Aloud Group

Group	Vocabulary			PP			Impersonal-SE		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
-Think (n= 29)	.10	.31	.05	.00	.00	.00	.00	.00	.00
+Think (n= 36)	.08	.28	.04	.00	.00	.00	.00	.00	.00

Note. Maximum score in these tests= 10.

* $p < .05$.

In sum, the analyses of the reading comprehension, recognition, and production posttests in these study indicated that thinking aloud versus not thinking aloud did not have an effect on any of our dependent measures. Therefore, there is no reason to believe that thinking aloud *per se* affected the cognitive processing of the target items during the reading task. In this sense, the use of think-aloud protocols was not reactive in this study.

The analyses conducted in the previous sections allowed the researcher to establish (a) that the participants in this study started out at similar levels of ability, and (b) that thinking or not-thinking aloud did not affect performance. Upon establishing these facts, this researcher proceeded with the analyses of the outcome measures of the study, in relation to glossing and the outcomes of reading comprehension (i.e., RQ#1-b), and glossing and type of item in relation to recognition and production of items at posttest and delayed posttest (i.e., RQ#5, a-b). These questions are restated below⁵.

Quantitative analyses of the outcome measures

RQ# 1-b. Does reading a text with L1-translation glosses versus reading a text without glosses have an effect on reading comprehension, as measured by a multiple-choice reading comprehension questionnaire?

In order to answer research question (1-b) on the effects of glossing on reading comprehension, a univariate analysis was conducted for glossing (i.e., L1-gloss versus no-gloss) on the raw scores of the multiple-choice reading comprehension test. This analysis revealed a statistically significant difference between the gloss and the no-gloss groups, with the gloss group outperforming the control on reading comprehension ($F(1, 63) = 45.286, p < .05; \eta^2 = .41$). Cronbach's Alpha reliability coefficient of the reading comprehension test was .69, an acceptable level of reliability of the measurement utilized. Descriptive statistics (mean, SD, SE of two gloss groups on reading comprehension) are displayed in Table 17.

Table 17. Mean, SD, And SE Of Two Gloss Groups On Reading Comprehension

Group	Mean	SD	SE
+Gloss (n=33)	16.21*	4.01	.69
- Gloss (n=32)	10.56	2.57	.45

Note. Maximum score =30.

* $p < .05$.

In other words, the results indicated that the L1-translation glosses appeared to have helped the participants understand the text better, in comparison to the no gloss group. The effect size was large, indicating a big difference between the groups.

RQ#(5a-b) Do glossing (i.e., L1-gloss versus no-gloss) and type of linguistic item (i.e., Lexical Item, Present Perfect, and Impersonal-SE) have an effect on:

- (a) intake, as measured by immediate recognition of targeted items, and
- (b) L2 development, as measured by immediate production of targeted items?

If so, will these effects be maintained three weeks later?

Before submitting recognition test scores to statistical analyses to answer these questions, test consistency was calculated for the item recognition test.

Cronbach's Alpha coefficient of reliability was high ($\alpha = .87$).

In order to investigate the effects of gloss and type of linguistic item on the immediate and delayed recognition of the targeted items of the study, a 2 X 3 X 3 factorial ANOVA for mixed designs was performed on the raw scores of the recognition posttests, in which glossing (L1-gloss versus control) was entered as the between-subjects factor, while time with three levels (pretest versus posttest versus delayed posttest) and type of linguistic item, with three levels (lexical item versus Present Perfect versus Impersonal-SE) were entered as within-subjects factors.

Descriptive statistics (means and SD for recognition of three types of item and two gloss group over time) are displayed in Table 18. Results of the ANOVA indicated a significant main effect for time, $F(2,126) = 169,105, p < .05$, eta squared = .72, a significant interaction between gloss and time, $F(2, 126) = 3.798, p < .05$, eta squared = .05, and a significant triple interaction between gloss, time, and item $F(4, 252) = 5.180, p < .05$, eta squared= 0.07. No main effects were found for gloss, $F(1, 63) = 1.879, p = 0.175$, type of item $F(2, 126) = .717, p = .49$,

and no interaction between gloss and type of item, $F(2,126) = 0.526$, $p = 0.592$

(see Table 19, Figure 1).

Table 18. Mean And SD For Recognition Of Three Types Of Item And Two Gloss Groups Over Time

Item	Group	Pretest		Imm. Posttest		Group	Delayed Posttest	
		Mean	SD	Mean	SD		Mean	SD
Vocab.	L1-gloss (N=33)	.03 ^a	.17	3.94 ^b	1.85	L1-gloss	4.61 ^b	1.83
	No-gloss (N=32)	.13 ^a	.33	3.31 ^b	1.76	No-gloss	3.44 ^c	1.93
Present Perfect	L1-gloss (N=33)	.03 ^a	.17	4.91 ^b	3.07	L1-gloss	3.61 ^c	3.08
	No-gloss (N=32)	.03 ^a	.17	2.94 ^b	2.87	No-gloss	4.50 ^c	2.97
Imp-SE	L1-gloss (N=33)	.03 ^a	.17	4.00 ^b	2.70	L1-gloss	3.58 ^b	2.31
	no-gloss (N=32)	.13 ^a	.33	3.25 ^b	2.36	No-gloss	3.69 ^b	2.72

Note. Maximum score in recognition tests= 10.

(^{a, b, c}) are used to indicated significant differences (at $p < .05$) between times.

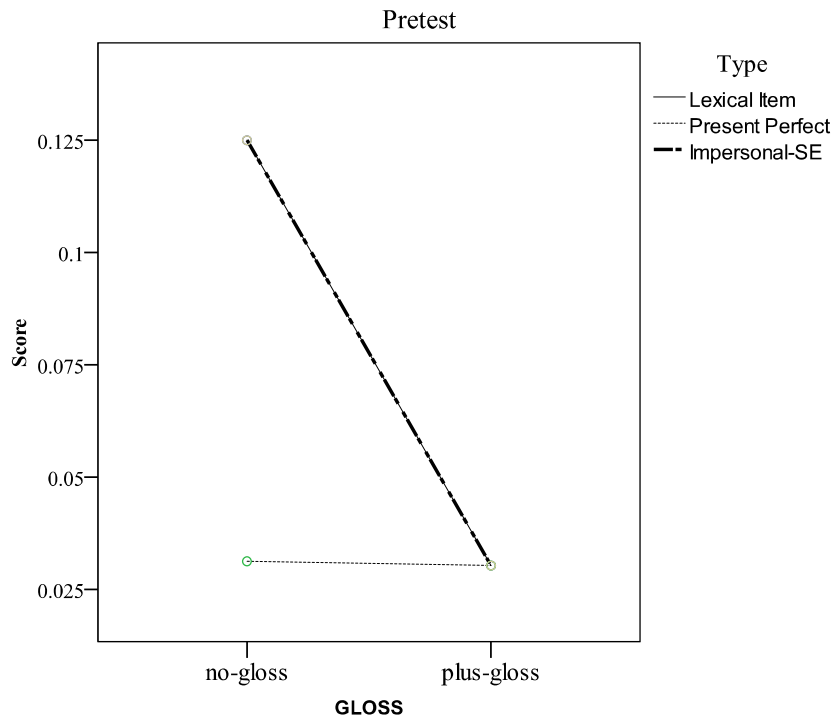
Table 19. Recognition: ANOVA For Group By Time By Type Of Item

Source	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>Eta S</i>
Gloss	1	2.212	2.212	1.879	.175	
Time	2	1832.145	916.072	169.105	.000*	.729
Time × Gloss	2	41.150	20.575	3.798	.025*	.057
Residual	126	682.566	5.417			
Type of Item	2	4.946	2.473	.717	.490	
Type of Item × Gloss	2	3.633	1.817	.526	.592	
Residual	126	434.903	3.452			
Time × Type of Item	4	6.157	1.539	.642	.633	
Time × Type of Item × Gloss	4	49.644	12.411	5.180	.001*	.076
Residual	252	603.747	2.396			

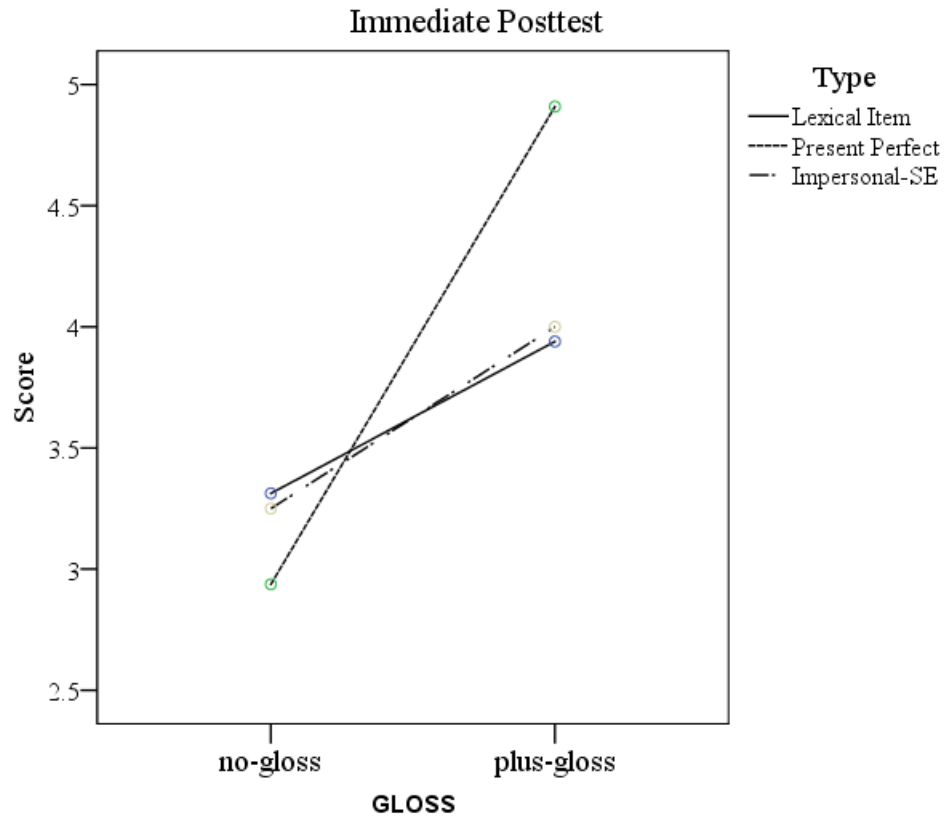
* $p < .05$.

Figure 1. Mean Scores On Recognition Of Three Types Of Item And Two Groups At Pretest (A), Immediate Posttest (B), And Delayed Posttest (C)

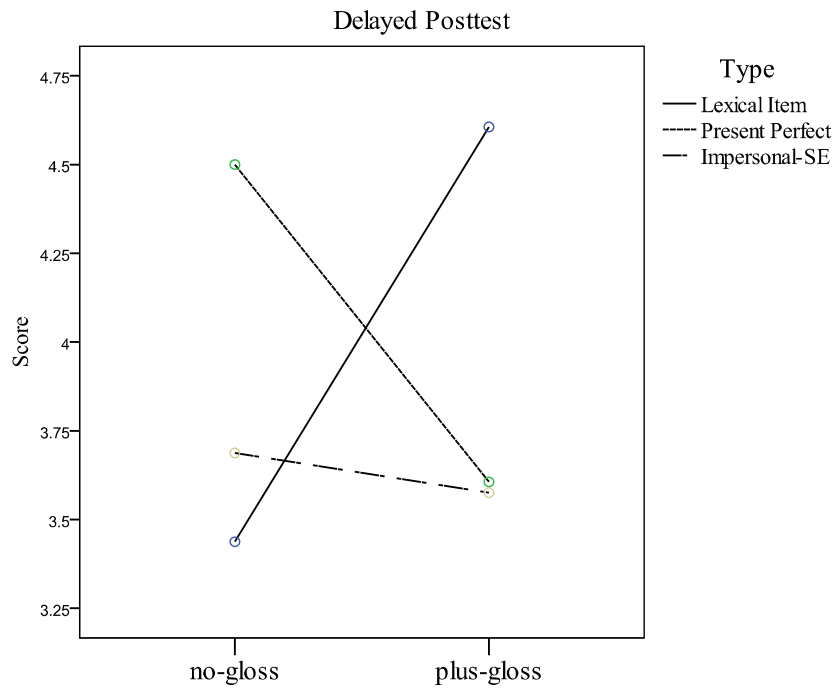
(a)



(b)



(c)



Further analyses were conducted in order to elucidate the patterns of the three-way interaction found in the study. In order to investigate whether there were interaction effects between type of item and gloss at each point in time analyses were conducted on the posttests and delayed posttest scores separately. In order to investigate whether gloss effects varied across time, recognition scores were broken down into types of item and individual analyses on each subset of scores were conducted. The results of these analyses are summarized in the following sections.

Interaction effects between type of item and gloss on recognition

Immediate Posttest

A 2 (gloss) by 3 (item type) mixed ANOVA on the immediate recognition posttest scores revealed a significant main effect for gloss, $F(1, 63) = 5.737$, $p < .05$, eta squared = .08, no main effect for type of item, $F(2, 148) = .126$, $p = .88$, and no interaction, $F(2, 148) = 2.183$, $p = .11$ (Table 20).

Table 20. Recognition Immediate Posttest: ANOVA For Group By Type Of Item

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>Eta</i> ²
Gloss	1	20.240	20.240	5.737	.020*	.083
Type of Item	2	3.843	1.921	.478	.621	
Type of Item × Gloss	2	17.956	8.978	2.233	.111	
Residual	126	506.609	4.021			

* $p > .05$.

A gloss effect was observed at posttest. However, there was no main effect for type of item, and no interaction between the two independent variables. According to these results, glosses appeared to have a positive effect on item recognition at immediate posttest regardless of the item type.

Separate analyses of each type of item will be presented later in this paper. However, it is worth noting at this point that, when types of items are considered separately, a significant difference was observed between the gloss and the no-gloss groups on immediate recognition of Present Perfect. No significant differences were observed for Lexical Item or Impersonal-SE. Mean and SD of Present Perfect immediate posttest recognition for two groups are presented in Table 21.

Table 21. Mean And SD Of Present Perfect Immediate Posttest Recognition For Two Gloss Groups

Group	Mean	SD
L1-gloss (N=33)	4.91*	3.07
No-gloss (N=32)	2.94	2.87

* $p > .05$.

Delayed Posttest

A 2 (gloss) by 3 (item type) mixed ANOVA on the delayed recognition scores revealed a type of item by gloss interaction, $F(2,126) = 4.223$, $p < .05$, $\eta^2 = .06$, no main effect for type of item, $F(2,126) = .859$, $p = .426$, and no main effect for gloss, $F(1, 63) = .013$, $p = .908$ (Table 22). (see Figure 1-c).

Table 22. Recognition Delayed Posttest: ANOVA For Group By Type Of Item

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>Eta S</i>
Gloss	1	.048	.048	.013	.908	
Type of Item	2	7.165	3.583	.859	.426	
Type of Item × Gloss	2	35.227	17.613	4.223	.017*	.063
Residual	126	525.563	4.171			

* $p < .05$.

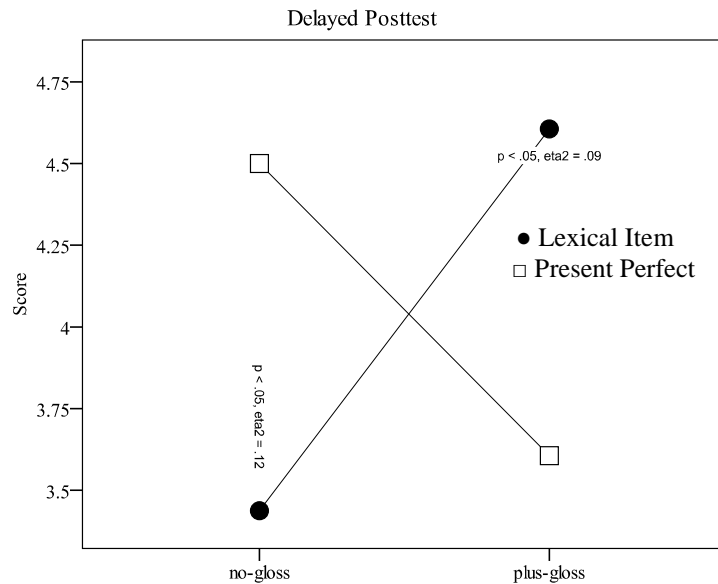
Further analyses were conducted in order to investigate the gloss by type of item interaction found at delayed posttest. A 2 (gloss) by 2 (Vocabulary versus Present Perfect) mixed ANOVA on the recognition delayed scores indicated a statistically significant type by gloss interaction, $F(1, 63) = 7.406$, $p < .05$, $\eta^2 = .10$, but no main effect for type of item, $F(1, 63) = .007$, $p = .935$, and no effect for gloss, $F(1, 63) = .076$, $p = .784$ (Table 23, Figure 2).

Table 23. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Lexical Item Versus Present Perfect)

Source	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>Eta S</i>
Gloss	1	.306	.306	.076	.784	
Type of Item	1	.032	.032	.007	.935	
Type of Item × Gloss	1	34.555	34.555	7.406	.008*	.105
Residual	63	293.938	4.666			

* $p < .05$.

Figure 2. Mean On Delayed Recognition Of Two Types Of Item For Two Gloss Groups



Group comparison analyses within each type of item on the delayed posttest indicated a statistically significant difference between the gloss and the no-gloss groups on Lexical Item, $F(1, 63) = 6.246$, $p < .05$, eta squared = .09 (see Table 24 for Mean and SD of delayed recognition of Lexical Item for two gloss groups), but no statistically significant difference between groups for the Present

Perfect at delayed posttest ($F(1,63) = 1.415, p = .239$). Furthermore, a comparison between two types of item within each gloss group indicated no statistically significant difference between the ability to recognize Present Perfect versus Lexical Item within the gloss group at delayed posttest, $F(1, 32) = 3.239, p = .081$ (Table 25). However, a statistically significant difference in the ability to recognize Present Perfect versus Lexical Item within the no-gloss group at delayed posttest was indicated, $F(1, 31) = 4.276, p < .05, \eta^2 = .12$ (Table 26). Descriptive statistics (Mean and SD for Delayed Recognition of Lexical Item and Present Perfect by Gloss group) are presented in Table 27.

Table 24. Mean And SD Of Delayed Recognition Of Lexical Item For Two Gloss Groups

Group	Mean	SD
L1-gloss	4.61*	1.83
No-gloss	3.44	1.93

* $p < .05$.

Table 25. Delayed Posttest: ANOVA For Recognition Of Present Perfect Versus Lexical Item Within The Gloss Group

Source	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig.</i>
Type of Item	1	16.50	16.500	3.239	.081
Residual	32	163.000	5.094		

* $p < .05$.

Table 26. Delayed Posttest: ANOVA For Recognition Of Present Perfect Versus Lexical Item Within The No-Gloss Group

<i>Source</i>	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig.</i>	<i>Eta2</i>
Type of Item	1	18.063	18.063	4.276	.047*	.12
Residual	31	130.938	4.224			

* $p < .05$.

Table 27. Mean And SD For Delayed Recognition Of Lexical Item And Present Perfect By Gloss Group

Item	Group	Mean	SD
Vocabulary	L1-gloss	4.61	1.83
	No-gloss	3.44	1.93
Present Perfect	L1-gloss	3.61	3.08
	No-gloss	4.50 †	2.97

† indicates $p < .05$.

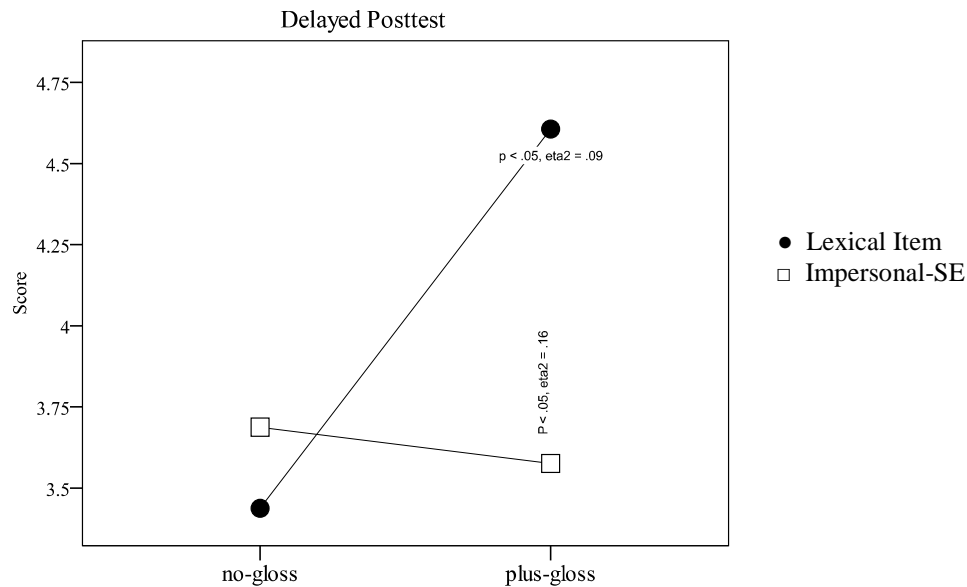
A 2 (gloss) by 2 (Vocabulary versus Impersonal-SE) indicated a statistically significant gloss by type of item interaction, $F(1, 63) = 4.291$, $p < .05$, eta squared = .064, no main effect for gloss, $F(1, 63) = 1.325$, $p = .254$, and no effect for type of item, $F(1, 63) = 1.594$, $p = .211$ (Table 28 and Figure 3).

Table 28. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Lexical Item Versus Impersonal-SE)

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>Eta2</i>
Gloss	1	4.536	4.536	1.325	.254	
Type	1	4.946	4.946	1.594	.211	
Type × Gloss	1	13.315	13.315	4.291	.042	.064
					*	
Residual	63	195.485	3.103			

* $p < .05$.

Figure 3. Mean On Delayed Recognition Of Two Types Of Item For Two Gloss Groups



Unlike the gloss groups comparison of the scores on Lexical Item, a group comparison on the Impersonal-SE delayed posttest scores indicated no statistically significant difference between the gloss and the no gloss group for this type of item ($F(1,63) = .032, p = .859$).

Furthermore, a comparison between two types of item, i.e., Lexical Item versus the Impersonal-SE, within each gloss group indicated a statistically significant difference between the ability to recognize Impersonal-SE versus Lexical Item within the gloss group at delayed posttest, $F(1, 32) = 6.481, p < .05, \eta^2 = .16$ (Table 29). However, a statistically significant difference in the ability to recognize Impersonal-SE versus Lexical Item within the no-gloss group at delayed posttest was not indicated, $F(1, 31) = 1.000, p = .284$ (Table 30). Mean

and SD for gloss group and two types of item (Lexical Item versus Impersonal-SE) on delayed posttest are displayed in Table 31.

Finally, a 2 (gloss) by 2 (Present Perfect versus Impersonal-SE) mixed ANOVA on the delayed recognition posttest indicated no main effect for gloss, $F(1, 63) = .759$, $p = .387$, no effect for type of item, $F(1, 63) = 1.216$, $p = .274$, and no interaction, $F(1, 63) = 1.047$, $p = .310$ (Table 32).

Table 29. Delayed Posttest: ANOVA For Recognition Of Impersonal-SE Versus Lexical Item Within The Gloss Group

Source	Df	SS	MS	F	Sig.	Eta2
Type of Item	1	17.515	17.515	6.481	.016*	.16
Residual	32	86.485	2.703			

* $p < .05$.

Table 30. Delayed Posttest: ANOVA For Recognition Of Impersonal-SE Versus Lexical Item Within The No-Gloss Group

Source	Df	SS	MS	F	Sig.
Type of Item	1	1.000	1.000	.284	.598
Residual	31	109.000	3.516		

* $p < .05$.

Table 31. Mean And SD For Group And Two Types Of Item On Delayed Posttest (Lexical Item Versus Impersonal-SE)

Item	Group	Mean	SD
Lexical Item	L1-gloss	4.61 †	1.83
	No-gloss	3.44	1.93
Impersonal-SE	L1-gloss	3.58	2.31
	No-gloss	3.69	2.72

† $p < .05$.

Table 32. Recognition Delayed Posttest: ANOVA For Group By Two Types Of Items (Present Perfect Versus Impersonal-SE)

Source	SS	Df	MS	F	p
Gloss	4.108	1	4.108	.759	.387
Type of Item	5.770	1	5.770	1.216	.274
Type of Item × Gloss	4.970	1	4.970	1.047	.310
Residual	298.922	63	4.745		

* $p < .05$.

To summarize to this point, a combined effect of gloss and type of item was observed at delayed posttest. It appears that the presence or absence of a gloss impacted the way in which the participants were able to recognize, three weeks after exposure to the experimental text, lexical items as opposed to grammatical items.

With respect to the variation of glossing across two levels of the variable type of item, i.e., Lexical Item versus Present Perfect, at delayed posttest, it appeared that not having a gloss was beneficial for the recall of Present Perfect, compared to Lexical Item. In fact, the participants in the no gloss performed statistically better on the delayed Present Perfect test items than on the delayed Lexical test items (see Figure 3). At the same time, not having a gloss, as opposed to having a gloss, was detrimental for the recognition of Lexical Item at delayed posttest, as the gloss group outperformed statistically the no-gloss group on Lexical Item. In other words, glosses positively impacted the recognition of Lexical Item at delayed posttest.

With respect to the variation of glossing across two levels of type of item, i.e., Lexical Item versus Impersonal-SE, having a gloss, as opposed to not having one, proved beneficial for the recognition of Lexical Item, and there was statistically more recognition of this type of item in comparison to the Impersonal-SE. It appeared, then, that glosses had no effect on the delayed recognition of Impersonal-SE.

When the results for the delayed posttest analyses are compared with the immediate posttest, a different picture can be observed. At posttest, the type of item did not appear to make a difference, and a positive effect for glossing was found.

Glossing effects across time

In order to investigate interaction effects across three times, the recognition scores were broken down into types of item, and separate analyses were conducted on each set of type of item scores. The results of these analyses are summarized below.

Glossing and Lexical Item

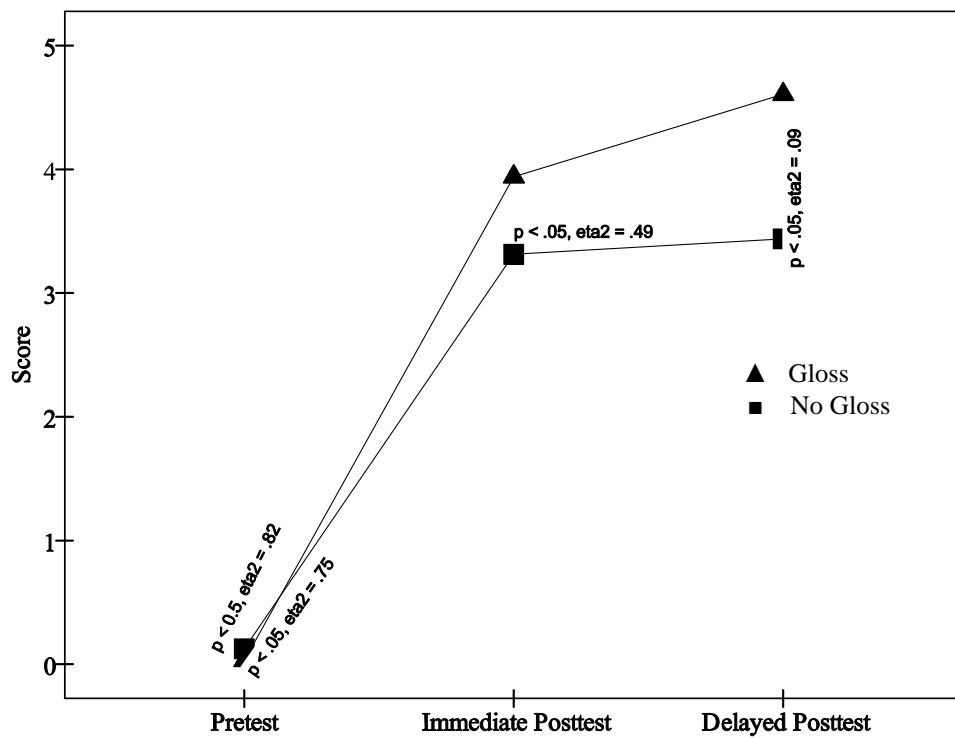
The recognition scores on vocabulary were submitted to a 2 (glossing) X 3 (pretest versus posttest versus delayed posttest) ANOVA for mixed designs. This analysis revealed a main effect for time, $F(2,126) = 164.741$, $p < .05$, eta squared = .72, a main effect for gloss, $F(1, 63) = 4.921$, $p < .05$, eta squared = .072, , and a significant time by gloss interaction, $F(2, 126) = 3.507$, $p < .05$, eta squared = .05. These results are summarized in Table 33 (Figure 4).

Table 33. Recognition Of Lexical Item: ANOVA For Group By Time

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>Eta S</i>
Gloss	1	15.664	15.664	4.921	.030*	.072
Time	2	613.092	306.546	164.741	.000*	.723
Time × Gloss	2	13.051	6.525	3.507	.033*	.053
Residual	126	234.457	1.861			

* $p < .05$.

Figure 4. Lexical Item: Mean On Recognition For Two Groups Across Time



At posttest, one-way ANOVA comparing the vocabulary posttest scores between groups indicated no statistically significant difference between groups, $F(1,63) = 1.945, p = .168$. At delayed posttest, the group difference was

statistically significant, $F(1, 63) = 6.246$, $p < .05$, eta squared = .09 (See Table 12 for mean comparisons).

Repeated measured ANOVA were conducted in order to test the significance of gains over time within each group. Within the gloss group, time differences were statistically significant, $F(2,64) = 108.716$, $p < .05$, eta squared = .77. Contrast analyses revealed a statistically significant difference between pretest and posttest, $F(1,32) = 145.734$, $p < .05$, eta squared = .82 (see Table 12), and a trend toward significance in the difference between posttest and delayed posttest, $F(1,32) = 3.321$, $p = .078$. Thus, the participants in the gloss group made significant gains from pretest to posttest, with a large effect size.

Within the no-gloss group, repeated measures ANOVA test revealed a statistically significant difference among the three times, $F(2, 62) = 60.415$, $p < .05$, eta squared = .66 (see Table 12). Contrast analyses indicated statistically significant gains from pretest to posttest, $F(1, 31) = 96.104$, $p < .05$, eta squared = .75, and significant gain from posttest to delayed posttest, $F(1, 31) = 30.857$, $p < .0$, eta squared = .49. Thus, the participants in the no-gloss group made significant gain from pretest to immediate posttest, and from immediate posttests to delayed posttests. Effect sizes indicated large differences.

In sum, the analyses on the vocabulary scores indicated that both the gloss and the no-gloss groups made significant gain from pretest to posttest. The no-gloss group performance improved statistically over time, i.e., from immediate

posttest to delayed posttest, while the gloss group performance did not. However, the gloss group at delayed posttest outperformed the no-gloss group statistically.

Glossing and the Present Perfect

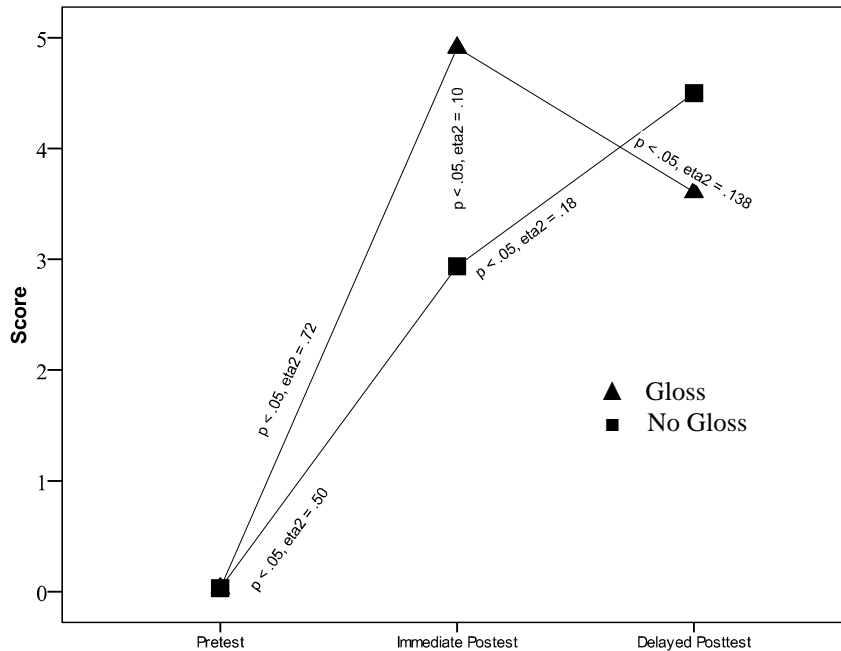
The recognition scores on the Present Perfect were submitted to a 2 (glossing) X 3 (pretest versus posttest versus delayed posttest) ANOVA for mixed designs. This analysis revealed a main effect for time, $F(2,126)= 69.555$, $p < .05$, eta squared = .52, a time by gloss interaction, $F(2,126)= 7.156$, $p < .05$, eta squared = .10, and no effect for gloss, $F(1, 63)= .756$, $p = .388$ (see Table 34 and Figure 5).

Table 34. Recognition Of Present Perfect: ANOVA For Gloss By Time

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>Eta S</i>
Gloss	1	2.093	2.093	.756	.388	
Time	2	679.026	339.513	69.555	.000*	.525
Time × Gloss	2	69.856	34.928	7.156	.001*	.102
Residual	126	615.036	4.881			

* $p < .05$.

Figure 5. Present Perfect: Mean On Recognition For Two Groups Across Time



One-way ANOVA comparisons between groups at posttest and delayed posttest, conducted separately, revealed that the group difference was statistically significant at posttest, $F(1, 63) = 7.122$, $p < .05$, eta squared = .10, but not at delayed posttest, $F(1,63) = 1.415$, $p = .239$. Thus, it seems that a difference that was observed at posttest was not maintained three weeks later. Results support a claim for a gloss effect at posttest on the ability to recognize the Present Perfect, with a moderate effect size. However, this difference was lost three weeks after exposure to the reading text.

In order to investigate the statistical significance of both the gloss and the no-gloss group gains over time, repeated measures ANOVAS were conducted on

the scores of each group individually. Results for the gloss group indicated a statistically significant difference for time, $F(2,64)= 41.917$, $p < .05$, eta squared = .56. Contrast analyses indicated statistically significant differences between pretest versus posttest, $F(1,32)= 82.273$, $p < .05$, eta squared = .72., and between posttest and delayed posttest, $F(1,32)= 5.138$, $p < .05$, eta squared = .138. Thus, the participants in the gloss group made significant gain from pretest to posttest in their ability to recognize the Present Perfect, with a large effect size. This performance dropped significantly from the immediate posttest to the delayed posttest.

Within the no-gloss group, similar analyses indicated an effect for time, $F(2, 62)= 34.763$, $p < .05$, eta squared = .52. Contrast analyses indicated that the difference between pretest to posttest was statistically significant, $F(1, 31)= 31.892$, $p < .05$, eta squared = .50, as well as the difference between the immediate posttest and delayed posttest, $F(1, 31)= 7.211$, $p < .05$, eta squared = .18. Like the gloss group, the no-gloss group made significant gain from pretest to posttest, with large effect size.

In sum, the statistical analyses revealed that both groups made significant gain from pretest to posttest, with large effect sizes. A glossing effect in favor of the gloss group was confirmed at posttest. Furthermore, while the gloss group performance dropped statistically three weeks later, the no-gloss group improved statistically over the same period of time, with a large effect size. Nevertheless,

between groups differences at delayed posttest were not statistically significant. It appears that the advantageous effect of having a gloss for the immediate recall of the Present Perfect was lost over time. By contrast, reading the L2 text without glosses was more effective for the recall of the Present Perfect over time.

Glossing and the Impersonal-SE

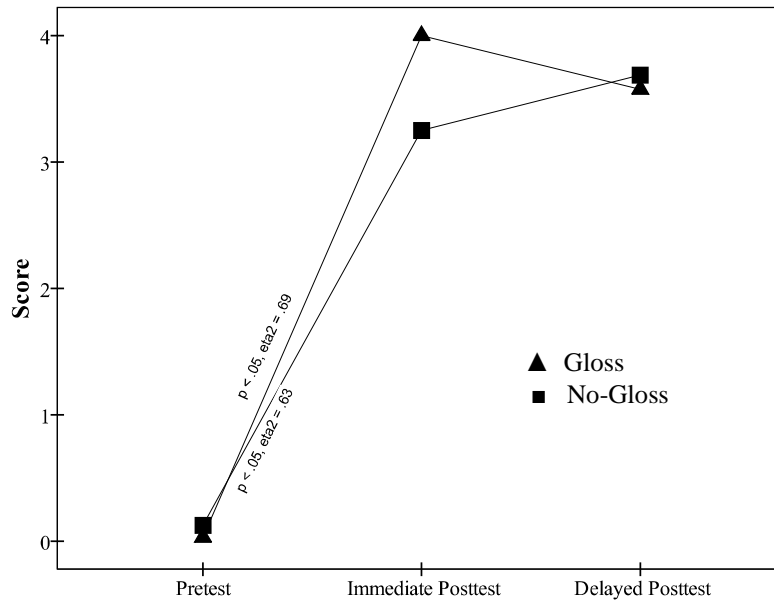
The recognition scores on the Impersonal-SE were submitted to a 2 (glossing) X 3 (pretest versus posttest versus delayed posttest) ANOVA for mixed designs. This analysis revealed a main effect for time, $F(2,126) = 78.773$, $p < .05$, eta squared= .55, no main effect for gloss, $F(1,63) = .266$, $p = .608$, and no interaction, $F(2, 126) = 1.137$, $p = .324$. These results are summarized in Table 35 (See Figure 6).

Table 35. Recognition Of Impersonal-SE: ANOVA For Gloss By Time

Source	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>Eta S</i>
Gloss	1	.533	.533	.266	.608	
Time	2	546.184	273.092	78.773	.000*	.556
Time × Gloss	2	7.887	3.943	1.137	.324	
Residual	126	436.821	3.467			

* $p < .05$.

Figure 6. Impersonal-SE: Mean Scores On Recognition Of Two Gloss Groups Across Time



Group comparison analyses revealed that the difference between groups was not statistically significant at posttest, $F(1, 63) = 1.411$, $p = .239$, or at delayed posttest, $F(1,63) = .032$, $p = .859$ (see Table 12).

In order to test for significant gains within each group over time, repeated measures ANOVAs were conducted on both the gloss and no-gloss sets of scores separately.

Within the gloss group, the results of the analysis indicated a main effect for time, $F(2, 64) = 41.398$, $p < .05$, eta squared = .564. Contrast analyses indicated that the difference between the pre and posttest was statistically significant, $F(1,32) = 71.430$, $p < .05$, eta squared = .69, while the difference

between the posttest and the delayed posttest was not, $F(1,32) = .576, p = .454$ (see Table 12). Thus, within the gloss group, the participants made significant gain from pretest to posttest, with a big effect size.

Within the no-gloss group, the results indicated a main effect for time, $F(2, 62) = 38.521, p < .05, \eta^2 = .554$. Contrast analyses indicated that the difference between the pre and posttest was statistically significant, $F(1,31) = 53.969, p < .05, \eta^2 = .635$, while the difference between the posttest and the delayed posttest was not, $F(1,31) = 1.145, p = .293$ (see Table 12). Thus, within the no-gloss group, the participants made significant gains from pre to posttest, with big effect sizes. Performance within the no-gloss group did not improve statistically from posttest to delayed posttest.

The statistical analyses revealed that both the gloss and the no-gloss groups made significant gains from pretest to posttest, with large effect sizes. However, their performance did not differ statistically from posttest to delayed posttest. Furthermore, the group comparisons revealed no glossing effect on the ability to recognize the Impersonal-SE. Therefore, an effect of L1-translation glosses on the ability to recognize the Impersonal-SE was not confirmed in this study.

Production of targets

Before submitting test scores to statistical analyses, Cronbach's Alpha coefficient of reliability were calculated for the production tests. Test consistency was low (.56).

Table 36 shows the means and SD for the gloss and no-gloss group by type of item in the production immediate posttest and delayed posttest. These descriptive statistics indicated that both the gloss and the no-gloss groups obtained means near zero in the posttest and delayed production posttest. Since the three production tests had a maximum score of ten points, and neither the gloss nor the no-gloss groups obtained a mean score higher than 1, thus indicating a floor effect, any comparison between groups would be meaningless. Therefore, no further analyses were conducted on the production tests.

Table 36. Production: M And SD Of Three Items, Two Groups And Three Times

Item	Group	Pretest		Immediate Posttest		Delayed Posttest	
		Mean	SD	Mean	SD	Mean	SD
Vocabulary	L1-gloss (N=32)	.03	.17	.12	0.33	.03	.17
	no-gloss (N=30)	.00	.00	.06	0.24	.16	.44
Present Perfect	L1-gloss (N=32)	.00	.00	.00	.00	.00	.00
	no-gloss (N=30)	.00	.00	.00	.00	.09	.53
Impersonal-SE	L1-gloss (N=32)	.00	.00	.00	.00	.00	.00
	no-gloss (N=30)	.00	.00	.00	.00	.00	.00

Maximum score= 10.

CHAPTER 4: DISCUSSION AND CONCLUSIONS

Discussion

In this study, qualitative and quantitative analyses were conducted to address the issue of glossing in relation to reading comprehension, noticing, and learning. This chapter discusses the research questions of the study in light of previous findings in SLA studies. With respect to the first research question, on the effectiveness of glossing for reading comprehension, this question received a positive answer. Both the outcome and process measures of the study (i.e., multiple-choice comprehension questionnaire and think-aloud protocol analysis, respectively) indicated that the participants in the gloss group were able to understand the L2 text better than the participants in the no-gloss group. Results indicated a statistically significant difference between the gloss and the no-gloss group on a multiple-choice test assessing reading ability at the 'local' level of the targeted items included in the study. Thus, according to these results, L1-translation glosses appeared to have helped the participants in the gloss group understand the basic level of the L2 text better, in comparison to the no-gloss group.

From the point of view of reading as a process, the think-aloud protocol analyses conducted in this study further confirmed the quantitative finding of a positive effect of using glosses to achieve basic understanding of the L2 text. This analysis indicated that L1-translation glosses helped the participants in the gloss

group understand the meaning of sentences in which the unknown words and phrases were embedded, and produce generally accurate inferences and comments. In contrast, the participants in the no-gloss group demonstrated little understanding of text content, as they were generally not able to understand the sentences in which the unknown words were embedded. In sum, the results of the analyses of both outcome and process measures of L2 reading, when L2 reading is viewed from the perspective of a bottom-up approach, strongly suggest that glosses are effective decoding aids.

At the same time, when reading comprehension is viewed in terms of top level processes, for example, at the level of interpretation of textual information and integration with previous knowledge, data from the process measure appear to indicate that both the gloss and the no-gloss groups might not have gained an understanding of the L2 text at higher levels of reading comprehension.

A qualitative analysis of the reading patterns in the think aloud protocols included in this study indicated that the participants in both the gloss and the no-gloss groups approached the reading task in a linear and bottom-up fashion. Participants in both groups tended to read the text from beginning to end without making connections between paragraphs or recapitulating ideas. Reading aloud in Spanish or reading aloud by translating the text into English, or a combination of these, were the predominant reading behaviors in both groups (cf. Leow, 2001b for similar findings). In general, participants in both groups did not appear to

show elaboration or integration of text context with their own schemata and knowledge in order to construct a coherent mental model for the text. In general, participants were not making inferences, connecting ideas to make causal or temporal relations, solving contradictions or drawing conclusions.

To summarize to this point, the answer to research question (1), on the effects of glossing an L2 text for improved comprehension, may depend on the approach to L2 reading. When reading is viewed in terms of decoding a text at the “bottom” level of its language features, the answer to research question (1) was positive: Glosses served as effective decoding aids. However, the process measure appear to indicate that when L2 reading is viewed from the point of view of top-down approaches to L2 reading, the answer to research questions (1) may not be the same: Both the gloss and the no-gloss group did not appear to have integrated the text contents enough, as to be able to construct a coherent mental model for the text.

The result that ‘micro’ level glosses, such as the L1-translations for lexical and linguistic elements utilized in this study, are effective decoding aids is not surprising given the fact that, in theory, the primary purpose of this pedagogical practice is to bridge the gap between the L2 reader’s characteristic lack of lexical and/or linguistic knowledge and the L2 text (Birch, 2002). In other words, from a bottom-up perspective of L2 reading, glosses for lexical or linguistic information are usually viewed as unobtrusive devices that make text contents more accessible

to L2 learners. This study empirically supported this theoretical claim. By making meanings easily accessible, the participants exposed to the glossed version of the L2 text were able to decode the sentences better, and gain more understanding, than the participants exposed to a non-glossed version of the same text. For these participants decoding the text at the sentence level without glosses proved especially difficult, as inferences proved not to be reliable.

The empirical studies on glossing for improved comprehension conducted from a bottom-up perspective of the reading process support the theoretical claim that glosses are effective reading comprehension aids (e.g., Davis, 1989; Jacobs, 1994; Jacobs, Dufon, and Hong, 1994; Bell & Le Blanc, 2000; Roby, 1991, 1999; Hulstijn, 1993; Bowles, 2004). However, only two studies measuring comprehension (Lomicka, 1998; Bowles, 2004) supported their quantitative results with qualitative analyses of concurrent data, via think-aloud protocols. In Bowles (2004), glosses for vocabulary words had a positive effect on reading comprehension, as measured by a multiple-choice comprehension questionnaire. Think-aloud protocol analyses further confirmed this finding. The participants in the gloss group noticed more targeted words than those in the no-gloss group, and were able to utilize the translations as reading comprehension aids. However, it was observed that the participants in this study “consulted the glosses in the passage to glean essential meaning from the text” (p. 550), that is, text comprehension was at the ‘minimal’ level of words and sentences. In-depth

processing of text content that would have led to integration and the construction of a coherent mental representation for the text was not observed in this study.

The present study corroborates these findings.

The present study also partially corroborates the findings in Lomicka's study (1998). In this study, it was observed that the use of L1 translation glosses was effective to gain a minimal level of understanding of the L2 reading text. This study is important because its participants had the option to choose from various types of annotations. It was observed that, of all the annotation types included (i.e., Lomicka), participants overwhelmingly preferred L1 translations for vocabulary words. As it was pointed out in a previous study by (Lyman-Hager & Davis, 1996), accessing word meanings via L1 translations was the "key factor in understanding the passage" (p. 62). It appeared that the L1 translation glosses for the L2 text in the present study also played an important role in helping readers understand the L2 text.

Although L1 translation glosses were the most preferred type of annotations in previous studies, Lomicka (1998) pointed out that annotating the linguistic elements of a text may not be beneficial for learners in the long term because this type of annotations tend to over-emphasize the linguistic information of the text, and ultimately force the readers to a bottom-up reading strategy. This finding was not corroborated in the present study. In the present study, the participants in both the gloss and the no-gloss groups utilized similar reading

strategies. They read the text in a bottom-up fashion without making connections or integrating text contents. Based on the findings in the present study, there was no reason to believe that a preference for a bottom-up orientation to L2 reading was due to the presence or absence of a gloss in the text.

The question that should be answered at this point is why, then, were the participants in the present study reading the text in a ‘bottom-up’ fashion, regardless of condition of exposure. One plausible explanation is lack of reading proficiency. It appears that the participants in this study had a ‘linear’ view of the reading process by which one starts reading from the first word and follows the words one by one until the end without stop. By contrast, proficient or “good” readers are equipped with an array of strategies that allow them to scan and skim through the text, stop to check for comprehension, make hypotheses, to ask questions, in sum to interact with the text in a highly active way (e.g., Grabe, 1991) Perhaps, the participants in the present study did not have a strategic reading ability, or perhaps they knew how to use these strategies in their native language but were not able to transfer them to the L2 reading context. This explanation is plausible given the general agreement in the reading literature that strategic reading requires training (e.g., Kern, 1989).

Another possible explanation for why the participants in this study did not reach higher levels of understanding of the L2 text was text difficulty. It was apparent from the think-aloud protocols the lack of a homogeneous and constant

level of difficulty throughout the nine paragraphs of the experimental text, that is, some parts of the text were more difficult than others. In fact, the mean for comprehension, in the outcome comprehension measure, was rather low for both groups (total $M= 13.38$ out of a total of 30 points). It is plausible that the variation in text difficulty might have been a factor in the amount of effort, cognitive but also in terms of motivation, invested by the participants in both groups. In particular with respect to the issue of motivation, the fact that the participants were taking part in a study with no grade assigned might account for both the reading style and the cognitive effort observed in the study.

In conclusion, this study corroborates the finding in previous studies that glosses are beneficial for reading comprehension, especially when a bottom-up approach to L2 reading is emphasized. Furthermore, this study does not support the claim that glossing a text at the level of linguistic information has an impact, positive or negative, on higher levels of reading comprehension. What may be the factors that affect L2 reading at higher levels of reading comprehension is a question that remains open.

Research question (2) in this study, on the effect of glosses on amount of reported noticing, did not receive a positive answer. The amount of reported noticing by participants in the gloss group ($M= 13.52$, $SD = 8.3$) did not differ statistically from the amount of reported noticing by participants in the no-gloss group ($M= 9.82$, $SD = 7.2$). There was no evidence in this study to support the

claim that having a gloss, versus not having a gloss, may impact the allocation of attentional resources to the targeted items embedded in the input. For this reason, the null hypothesis must be retained.

Besides the quantitative analyses, qualitative analyses of the think-aloud protocols were also conducted in the study. The analyses of noticing patterns conducted in this study shed light on the attentional processes as they occurred during the experimental task. Furthermore, these analyses were important in light of the small sample size (i.e., N= 19 in the gloss group and N=17 in the no-gloss group).

The think-aloud protocol analyses revealed that participants in both the gloss and the no-gloss groups did not appear to process targeted items beyond mere noticing. In other words, these analyses did not support the claim that participants in one group or the other demonstrated higher levels of awareness of the targeted items. In fact, there were only a few cases of participants who demonstrated awareness of the items, beyond mere noticing, and these few cases were found in both the gloss and the no-gloss groups. The lack of awareness of linguistic form-meaning connections was evidenced by the fact that participants did not make explicit the connections between meaning and form, did not establish relations with previously known words in the target language (e.g., ‘Sp. *furcia* is the same as Sp. *Prostituta*’, ‘*fetén* is like *amigo bueno*’; ‘*estrambótico* is like *raro*’; ‘*lóbrego* is the same as *triste*’), made word associations, or consciously

parsed grammatical information (e.g., ‘the *se* in *se embute* must mean...’, “the *han* of *han paliado* is *have*...”) (I will come back to this issue later).

The finding of no effect for condition of exposure on the amount of reported noticing should be discussed in light of the findings in previous studies on input modification (i.e., text simplification, textual enhancement, and glossing) that have measured directly, empirically, the effects of input modification techniques on noticing (i.e., Leow, 2001b, Leow, Egi, Nuevo, Tsai, 2003; Bowles 2003, and 2004).

Employing a more robust research design that included think-aloud protocols to elicit the reported noticing of forms under enhanced versus non-enhanced conditions, Leow (2001b) found that noticing in the enhanced group was not significantly different from noticing in the non-enhanced group. Thus, this study did not support the theoretical view of input modification as an effective way to direct learners’ attention to linguistic form while interacting with an L2 text. The finding of no effect for input modification techniques on noticing of targeted forms embedded in the input was further corroborated by two other studies conducted within an attentional framework (Leow et al., 2003; Bowles, 2003).

Within the glossing strand of research, Bowles (2004) submitted to ANOVA test the scores on reported noticing for four groups in her study (computerized and no computerized +/- L1 translation gloss). The results of the

analyses revealed that the amount of reported noticing differed statistically by glossing condition. However, the qualitative analyses of the think aloud protocols revealed patterns of noticing in both the gloss and the no-gloss groups, thus leading the researcher to conclude that “a crucial element of language learning may lie not in what external manipulations are made to input but rather in what learners *do* with the input as they interact with it” (p. 549), echoing similar observations put forth by Leow (1997b). Based on this, Bowles’ study (2004) did not support the theoretical postulation that input modification techniques are effective ways to direct learners’ attention to linguistic form while interacting with the L2 text (e.g., in the glossing strand, see Watanabe, 1997).

The finding that participants in both the gloss and the no gloss groups did not appear to consciously parse the targeted items in terms of meaning, form, form-meaning connections, as revealed by their verbalizations in the think aloud protocols, strongly suggest that there was no ‘focus on form’ in this study, in the sense that participants did not appear to be focused on the linguistic aspects of the targeted items. This finding corroborates the findings in previous SLA (e.g., Bowles, 2004; Leow, Hsieh, & Moreno, 2008). Bowles (2004) observed that participants “did not seem to be engaged in higher-level processes” (p. 549). Lack of “deeper-level” processing in the context of a L2-reading task was further corroborated in Leow et al. (2008). In this study, the researchers explained the finding of no effect for attentional condition (i.e., requesting L2 readers to attend

to specific forms in the input while processing a written text for meaning) on reading comprehension based on the notion of ‘depth of processing’ (Craik & Lockhart, 1972; see Leow et al., 2008). Leow and his colleagues found “overall a relatively minimal level of processing of the targeted forms in the input” across conditions (p. 685). Thus, it appeared that processing of linguistic form was not “deep” enough as to ‘hinder’ comprehension of the text by ‘taxing’ attentional resources. The present study further corroborated the absence of deep levels of processing of linguistic form in the context of a reading task, under two glossing conditions.

The finding of lack of deep processing of form-meaning connections under various conditions of exposure in the context of an L2 reading task is consistent across studies. Thus, task modality might provide an explanation for the lack of effect for input modification in these studies. Perhaps, utilizing tasks other than reading may be a more effective way to promote higher levels of processing of linguistic form (e.g., puzzles and other problem solving tasks utilized in previous SLA studies conducted within an attentional framework (e.g., Bowles, 2003; Bowles, 2004; Leow, 2000, 2001a & 2001b, 1997a, 1998a; Leow et al., 2003; Leow & Bowles, 2006; Rosa & O'Neill, 1999; Rosa & Leow, 2004). In sum, both quantitative and qualitative analyses conducted in this study did not provide enough evidence to support the claim that L1-translation glosses are effective ‘noticing’ devices.

Research question 3 of this study, on the combined effect of glossing and type of item on noticing, did not receive a positive answer. The quantitative analyses revealed that glossing and type of item did not have a combined effect on noticing scores. However, there was a type of item effect on noticing. Lexical Items were noticed significantly more than Impersonal-SE items in both the gloss and no-gloss groups. Present Perfect did not differ statistically from Lexical Item or Impersonal-SE with respect to noticing in both the gloss and the no-gloss groups.

At first sight, the finding in this study of a type of item effect on noticing seems to bring evidence in support of the claim that ‘language is not processed all alike’ (DeKeyser, Salaberry, Robinson, & Harrington, 2002; DeKeyser, 2005; VanPatten, 1989, and 1990). It appeared from the results of the present study that Lexical Item is more ‘isolatable’ (Gass & Svetics, 2003) than the Impersonal-SE, as participants noticed statistically more Lexical Item than Impersonal-SE. However, the finding of no difference in noticing between Lexical Item and the Present Perfect is not compatible with the claim, generally accepted in the SLA field, of a priority for processing lexical information over grammatical information. Furthermore, the finding of no difference between Present Perfect and Impersonal-SE was unexpected, under the assumption that the Present Perfect should be more ‘isolatable’ than the Impersonal-SE based on its relative ‘inherent difficulty’.

The finding, corroborated in this study, that participants reported noticing vocabulary words more than items carrying the impersonal-SE can be explained based on the theoretical postulation of relative difficulty of types of items (e.g., DeKeyser, 2005). In this study, vocabulary items were predicted to be more noticeable than the impersonal-SE because of the inherent characteristics that differentiate these two items. The comparison between less and more difficult items was based in a difficulty matrix with three defining features: (a) salience (b) abstractness of the meaning-referent, and (c) complexity of encoding. With respect to the salience of the item, the syllabic structure of the lexical targets of the study ranges from two to five syllables. Any of these stressed polysyllabic constructs was predicted to be more salient to the learner than the unstressed clitic *se*. This prediction was confirmed in the study. With respect to the abstractness of the meaning/referent, grammarians agree that the grammatical notion of ‘impersonality’ is ‘fuzzy’, in the sense that its semantic values are easily confounded, even by native speakers, with similar such as ‘de-causative’, ‘median’ and passive (Mendikoetxea, 1999; Whitley, 1986). With respect to the third feature of difficulty, English does not grammaticalize the “impersonal” by means of a clitic. Thus, the L1-L2 encodings of the grammatical ‘impersonal’ are different. Furthermore, from the point of view of encoding within the language (i.e., intra-linguistically speaking), the impersonal-SE is highly complex because the clitic *se* may carry many different values, not only grammatical but also

lexical, depending on the context. Thus, for example, for the foreign language learner *se* is the prototypical reflexive, but it can also be used as a de-causative and passivizing device (see Whitley, 1986 for discussion). Thus, *se* is highly obscure even for native speakers of Spanish. For these reasons, it is not surprising to find that the participants in this study showed very little noticing of the form-meaning connection that defines the impersonal-SE, in comparison to the more salient and semantically stable connections between the meaning and form of the vocabulary words included in this study.

Another finding of the study was that differences in processing between Lexical Item and Impersonal-SE were not influenced by the presence or absence of an L1-translation gloss. There are various explanations for this finding. It is plausible that L1-translation glosses are not effective ways to direct learners' attention to linguistic form, or form-meaning connections. Under this hypothesis, glosses may be viewed as effective reading aids that help readers to access the meanings of unknown words in their native language, without further implications for language processing. However, there is another plausible explanation. The think-aloud protocol analyses indicated that participants were focused on what was lexical in the text, not on what was grammatical. In fact, there were virtually no verbalizations showing awareness of grammatical connections. At the same time, it should be noted that the grammatical items presented both a grammatical and a lexical part. Furthermore, in order to insure comparability to Lexical Item,

all the lexical parts of the grammatical items selected as targets for the study were unfamiliar. Thus, it is plausible that the Impersonal-se turned out to be too difficult. In other words, the difficulty inherent in the unfamiliar lexical part of the grammatical item might have been added to the inherent difficulty of the grammatical form-meaning connection per se. It was evident from the think-aloud protocols that, when they encountered an item in the impersonal-se, most participants in both groups were focused on understanding the semantic contribution of the whole (e.g., Sp. *se avista*, people see) to the text. This was because the lexical part of that ‘whole’ (e.g., Sp. *avista* in *se avista*) was unfamiliar to them. It is not surprising, then, that the typical think-aloud behavior in this case was the verbalization of the lexical part only (e.g., ‘I don’t know what *avista* means”). In this context, the semantic contribution of Sp. *se* to *se avista* (i.e., *people see*) had very little chance to be noticed.

Previous SLA studies have compared processing of lexical versus grammatical information during an L2 reading task (Gass & Svetics, 2003; Greenslade, Bouden, & Sanz, 1999; Leow et al., 2008; see VanPatten, 1990 for original postulation of processing differences in the oral mode; Wong, 2001). Of these studies, only Leow et al. (2008) have directly measured how attention/noticing processes would differ across various types of items. In this study, Leow (2008) and colleagues have found that vocabulary items (i.e., the monosyllabic Sp. *sol*) was processed more often and deeper than the grammatical

items included as targets of the study (Sp. feminine, singular definite article *la*; the objective pronoun *lo*, and the morphological ending for the plural of verbs *-n*). Although no statistical analyses of noticing scores by type of item were conducted, a qualitative observation of the study was that “the difference in reporting processing the targeted forms a little deeper was not as clear for the rest of the attentional conditions as for the *sol* group” (p. 679). In their discussion of findings, Leow & colleagues (2008) speculated that “type of targeted form might have also played a partial role in these findings. *Sol*, being a content word, carried more meaning and clearly attracted readers’ attention more easily to it, whereas the bound verbal morpheme *-n*, carrying less semantic weight, did not” (p. 685). This observation appears to corroborate the observation made in the present study that L2 readers tend to focus on the semantic aspects of the L2 text they are reading, and not on the grammatical ones. In this study, as in Leow et al. (2008), readers made virtually no comments of the form-meaning connections governing grammatical items.

Although the present study did not include an operational definition of ‘depth’ of processing, two levels of awareness were identified. For example, a comment about meaning-form connections made explicit was coded as an instance of awareness at a level higher than noticing. As in level ‘three’ of depth of processing identified in Leow et al. (2008), in the present study there were very few instances of higher levels of awareness, and there was virtually no indication

that the form-meaning connections governing the grammatical targets of the study were been consciously parsed.

An SLA study that has directly measured the effects of type of linguistic item on reported noticing was Leow et al. (2003). Although a lexical item was not included as a target of the study, this study did compare a ‘more salient’ (i.e., the Sp. Present Perfect) versus a ‘less salient’ (i.e., the Sp. subjunctive marker) form. The researchers found that “learners exposed to the present perfect forms reported more noticing than those exposed to the present subjunctive forms” (p. 102). In the present study, the quantitative results for noticing of the Sp. present perfect were not consistent with the predictions of the study. Noticing of present perfect items did not differ statistically from the impersonal-se, arguably a less salient form, though it should also be pointed out that while the Spanish subjunctive marker was bound, the Spanish impersonal-se is a free standing word.

The finding that participants in this study did not notice significantly more Lexical Items than Present Perfect items was unexpected. The Sp. present perfect was coded as a less ‘isolatable’ item in comparison to lexical items, and therefore it was predicted that participants should notice lexical items more than items in the present perfect. One possible explanation for this unexpected result is that the statistical comparison conducted with a small sample size is misleading the researcher to conclude that *there was no real* difference from the statement that *a difference was not found*. In other words, this result might be due to a type II

error. However, an explanation should be provided, in the case there was not, and result of no difference is the correct one.

Based on considerations of inherent difficulty, a prediction was made in this study that lexical item should be noticed more than present perfect. Lexical items were defined as more salient, less abstract, and less complex than the present perfect. At the same time, the present perfect was less salient (than lexical item), more abstract, and more complex. However, the hypothesis that lexical item is more difficult than present perfect was not supported by the evidence found in this study. Therefore, this unexpected finding merits an explanation.

A plausible explanation why noticing of vocabulary items was similar to noticing of the Sp. present perfect might be that (a) lexical item in this study was more difficult than predicted, and (2) present perfect was easier than predicted. A variable that is seldom studied in the literature on glossing and that may be a key to finding a plausible explanation for the results of the study is the notion of frequency.

Frequency can explain both why present perfect was easier, and why vocabulary was more difficult than predicted. In this study, ten targeted items were in the present perfect, and ten targeted items were lexical items. However, there was a crucial difference in how the participants were exposed to one type of item or the other: While there were ten different lexical targets, with no repetition of these items in the L2 text, the grammatical part of the present perfect (i.e., the

discontinuous form of auxiliary *ha(n)* + part participle *-(a)do*) appeared with a frequency of ten in the experimental text.

It is highly plausible that by being exposed to the same form with a constant meaning ten times, participants were able to notice the present perfect more than expected, as noticing the same element several times during exposure must be cognitively less effortful than noticing many new different things only once. By contrast, noticing lexical items in this study must have been for the participants exactly like noticing ‘many new different things only once’. In other words, it is highly plausible that participants might have been overloaded with the high number of unfamiliar vocabulary words included in the text. Being exposed to too many new vocabulary words in a text that did not result as easy to read as expected might have caused a loss in motivation. In fact, the think-aloud protocols abound in ‘affective’ verbalizations reflecting a feeling of frustration with the amount of new words participants encountered in the text. Thus, while high frequency can explain the noticing of present perfect in this study, high number of unknown items presented with low frequency may explain the results for noticing of vocabulary items. Another plausible explanation for why vocabulary items and the present perfect were noticed equally statistically in this study may be that both items are equally salient, against our predictions.

The empirical evidence in previous studies (Hulstijn et al., 1996; Nagata, 1999; S. Rott, 1999, and 2007) seems to support an explanation based on

frequency effects. A study by Nagata (1999) comparing three grammatical items found an effect for frequency on intake of forms. It appeared from this study that the target had more chances to be learned when it appeared four times in the text, as opposed to appearing only once. Furthermore, a combined (lexical) word frequency and gloss effect was found in Hulstijn et al. (1996). In this study, a condition with a frequency of 3 was compared with a condition with a frequency of 1. Results indicated that when a glossed vocabulary words appeared three times in the text, there was more chances that the participants scored higher in an immediate posttest recognition test than when the glossed words appeared only once. A study by Rott (1999) found further corroborated a frequency effect on the retrieval of lexical words, with that data showing a difference between two and six encounters with a targeted word during exposure to an L2 text. Finally a study by Rott (2007) further confirmed that frequency is an important factor in word learning. An important conclusion of this study combining three types of word interventions with a frequency of four encounters with the targeted words was that “the increased frequency of obtrusive semantic word interventions resulted in the increase of word retention” (page 190). In other words, a frequency of four, in combination with glossing, had an impact on word retention. Unfortunately, none of these studies included think-aloud protocols to measure the amount of reported noticing under each condition of exposure, or tested statistically the impact of frequency on amount of reported noticing. However, they provide evidence of

intake of targeted items suggesting that frequency may be an important factor to consider in explaining differences in processing different types of items.

Besides the effects of glossing and type of item on noticing, this study also investigated the relationships between the reported noticing of gloss versus no-gloss targeted items and the recognition and production of these items (i.e., research question 4). If a significant correlation between noticing and 'learning' was indicated, the question was, then, whether the two relationships would be significantly different.

With respect to item recognition overall, the answer to this research question was positive. Results revealed that the amount of reported noticing was positively correlated with the ability to recognize items immediately after exposure to the reading text and three weeks later. Furthermore, these relationships were statistically similar for the gloss and the no-gloss group.

Similar analyses were conducted on each type of item separately. With respect to Lexical Item, results revealed significant correlations between the reported noticing of targeted items and the immediate, and delayed, recognition of items for the gloss group only. For the participants who read the text without glosses, there was no evidence that noticing was significantly correlated with immediate or delayed recognition of item.

With respect to Present Perfect, the results for immediate recognition contrasted with those for delayed recognition. Significant correlations between

amount of reported noticing and immediate recognition were found for both groups. By contrast, no significant correlations were indicated between amount of reported noticing and delayed recognition of Present Perfect, for both groups.

With respect to Impersonal-SE, the finding that reported noticing did not correlate with immediate (for the no-gloss group) and delayed recognition (for the gloss group) can be explained on the same lines of a frequency effect. In other words, the patterns observed for the grammatical items of the study raised the interesting question whether reported noticing of grammatical form was sensitive to the frequency of the form, and how strong that frequency is for grammatical items hypothetically varying in degree of difficulty.

Overall results revealed that the amount of reported noticing was (1) positively correlated with recognition of Lexical items immediately after exposure to the reading text and three weeks later for only the glossed group, (2) positively correlated with recognition of the Present Perfect only for the immediate test for both groups, and (3) not correlated with recognition of Impersonal-se for either the immediate test (for the no-gloss group) or delayed test (for the gloss group). Furthermore, these relationships were statistically similar for the gloss and the no-gloss group. The finding that noticing was generally positively correlated with learning, independently of the conditions of exposure was not surprising. This finding replicated the findings in Leow et al. (2003), and corroborated those in Leow (2001b). Leow et al. (2003) found that “the amount of reported noticing of

targeted forms in the input was statistically similar for both enhanced and unenhanced groups” (102). Similarly, in this study the amount of reported noticing was statistically similar for both the gloss and the no-gloss groups.

The case of Lexical Item merits further comments because the gloss and no-gloss groups differed in terms of noticing-intake correlation. The result of no significant correlation between reported noticing and recognition of lexical item within the no-gloss group (versus the correlation indicated within the gloss group) is not surprising considering the coding system for noticing utilized in this study. This result is explained by the fact that verbalizations in which the participant made explicit being aware of ‘not knowing the word’ were coded as instances of noticing. In fact, think-aloud protocol analyses also revealed higher amount of accurate verbalizations in the gloss as compared to the no-gloss group (see answer to research question 1), and a tendency to no- or wrong guessing in the no-gloss group. Given the difficulty that participants had in inferring the correct meanings of the unknown lexical words *they noticed* in their attempt to attain a minimal level of text comprehension, the result of no evidence for correlation between amount of reported noticing and immediate and delayed recognition of lexical items within the no-gloss group was not surprising. Perhaps, as suggested by the results of the gloss group, more ‘accurate’ noticing might have been positively correlated with accuracy in retrieval.

The answer to research question (4) should be discussed in the context of the finding of no effect for glossing on noticing in the present study (cf. answer to research question 2 above). With the exception of Bowles (2004), no previous study on glossing has measured the contribution of a gloss to the noticing of targeted forms directly, empirically, via think-aloud protocols. Bowles (2004) found a positive effect of glossing on noticing. However, her qualitative results did not match the quantitative (see answer to research question 2). As in the study by Leow et al. (2003), Bowles (2004) concluded that noticing is a more important factor in explaining learning than the external conditions of exposure crucially because noticing may occur *independently* of condition and still have a positive impact on learning. The present study confirms that noticing of items overall is positively correlated with learning, thus providing empirical evidence for Schmidt's noticing hypothesis (1990), and corroboration for previous SLA empirical studies on noticing and learning (Alanen, 1995; Leow, 1997, 1998, 2000, 2001; Rosa, 2004; Rosa & O'Neill, 1999).

The analyses at the level of each type of item revealed an interesting complex picture. With respect to Lexical Item, positive correlations were corroborated for the gloss group but not for the no-gloss group. This was explained by disentangling the notion of noticing of form from the notion of accurate noticing of form-meaning mappings. With respect to the grammatical items of the study, the hypothesis of a frequency effect was postulated to account

for the finding that reported noticing in both gloss and no-gloss group did not necessarily always correlate with retrieval or retention. Under the hypothesis of a threshold of four for a frequency effect, correlations between reported noticing of grammatical mappings and retrieval/retention do not need to be maintained.

Up to this point, the results on the effects of glossing on noticing, and the relationships between reported noticing in gloss or no-gloss groups and two 'learning' measures have been discussed. The last research question of the study (i.e., research question 5) comes back to the issue of glossing in relation to learning, and asks whether glossing and type of linguistic item have an effect on (a) recognition and (b) production of targeted items, both immediately after exposure and three weeks later. Research question (5-a), on the combined effect of glossing and type of item on recognition did not receive a positive answer. The analyses conducted in this study on item recognition revealed different patterns for immediate intake of items, and delayed retention. At immediate posttest, results indicated a main effect for gloss, no effect for type of item, and no interaction. Although a main effect for gloss was not revealed when items were considered overall, results indicated a positive effect for condition on the immediate recognition of Present Perfect. Given the fact that L1-L2 encoding are similar for present perfect, and given that the item appeared in the text ten times with a constant meaning, it is plausible that glosses might have interacted with the relative simplicity (in comparison with the Impersonal-SE) and high frequency of

the item to produce a difference on immediate recognition. In contrast, an effect for gloss was not indicated for immediate recognition of vocabulary item or for the Impersonal-SE. It should be noted that (a) being exposed to Lexical Item was like being exposed to ‘many different things only once’, and that (b) the Impersonal-se was predicted to be more complex, than Present Perfect. It is plausible, than, that glosses might not have had a combined effect in the case of Lexical Item and Impersonal-SE. These observations may explain the findings with respect to gloss and type of item on the immediate posttests.

With respect to delayed posttests, results indicated two type of item by gloss interactions. Gloss effects varied across two levels of type of item: Lexical versus Present Perfect. The participants who read the text without glosses performed statistically better on the present perfect than on the vocabulary test. This finding strongly suggests that the hypothesis of a frequency effect (i.e., high frequency available for Present Perfect versus low frequency available for lexical item) in conjunction with the observation of generally inaccurate noticing of lexical item in the no-gloss group may be a plausible explanation for these results. Again, we have a case in which Present Perfect appears to be easier than predicted, and vocabulary items appear to be more difficult than predicted, for those participants who read the text without the aid of glosses. At the same time, results indicated that participants who read the text with the aid of glosses outperformed those who read the text without glosses on delayed recognition of

vocabulary items. If differences are due to inaccurate response in the no-gloss group, and not to less noticing within this group (cf. answer to research question 3), the hypothesis that retention of lexical item, in the presence of ‘too many new items’, added a considerable amount of cognitive effort to the task seems to be grounded empirically.

The results of the delayed posttest also revealed a significant interaction between gloss and type of item. Glosses appeared to vary across levels of two types of item: Lexical Item versus Impersonal-SE. Participants who read the text with the aid of glosses were able to retain significantly more lexical items than items in the Impersonal-SE. This result replicated the result of a significant difference between the amounts of reported noticing of lexical item versus the impersonal-SE. Regardless of group; participants reported noticing more vocabulary items than items in the impersonal-SE. This result strongly suggests that impersonal-SE was highly difficult for the participants in this study given the fact that the item appeared in the input with constant meaning with a frequency of ten. This finding was not unpredicted.

Finally, with respect to a gloss effect across time for each individual type of item separately, results revealed a combined effect of gloss and time only for Present Perfect. Results indicated that the positive gloss effect observed for intake was lost over time. Glosses did not have a lasting effect on Present Perfect. In fact, the performance of participants who read the text with glosses dropped

statistically from immediate to delayed posttest. This finding merits further comments.

The finding of a significant drop in Present Perfect performance for the gloss group was unpredicted vis-à-vis the finding, documented in previous research of combined effects of frequency and various word interventions (e.g., Hulstijn et al., 1996; Rott, 2007). If gloss interacts with frequency, the participants in the gloss group should be able to show better retention of Present Perfect than it was observed in this study. Retention was not better, it was worse. A plausible explanation for the apparent ‘weakness’ of a purported delayed frequency effect might be lack of ‘deep’ level of processing of the form-meaning connections during interaction with L2 input in the gloss group. Retention over time is cognitively more demanding than immediate retrieval. More demanding tasks require deeper processing during exposure (Leow, 2001b; Leow et al., 2003). There is evidence from the think-aloud protocol analyses in this study that grammatical items were not processed beyond the level of mere noticing. While this low level of awareness might have been enough to promote immediate intake of Present Perfect, a higher level of awareness might have been required for retention of the Present Perfect pattern. It should be noted at this point that, although the participants in the no-gloss group improved significantly from immediate posttest to delayed posttest, performance was not statistically different

than the participants in the gloss group. Therefore, retention of Present Perfect was similarly low, regardless of condition of exposure.

To summarize the findings related to research question (5-a), the results for both analyses of immediate and delayed recognition did not provide enough evidence to support a claim for a glossing effect. The scores on the recognition posttests indicated a rather low performance across both condition and type of item (i.e., no mean score was higher than 5, over a total score of 10). Low performance overall was explained by the observed lack of ‘deep’ processing during exposure to the input. Furthermore, effects for gloss were only indicated for the immediate retrieval of Present Perfect, and the retention of vocabulary items. With respect to the former, it was postulated that exposure to gloss might have interacted with other factors affecting the Present Perfect, such as its high frequency of appearance in the text and its relative ‘transparency’ (i.e. similarity of L1-L2 grammatical encoding). With respect to the latter, the inaccuracy of responses in the no-gloss group, rather than a ‘noticing’ effect of gloss, might have explained the result of better retention of lexical items. Moreover, gloss did not have an effect on the immediate intake of vocabulary items, immediate intake of Impersonal-se, delayed retention of Present Perfect, or delayed retention of Impersonal-se.

Research question (5-b) of the study, on the effect of gloss and type of item on production posttest and delayed posttest could not be addressed in this

study given that the standard variation of scores on the production posttests was equal to zero. The low mean scores on all production posttest strongly suggests that production of targeted forms was too demanding a task in the context of learning while reading glossed or no-glossed texts with the targeted items employed in this study. The observed lack of processing of form beyond mere noticing might explain why the participants in this study did not show the ability to produce the targeted L2 lexicon and grammar.

This study corroborates the findings in previous glossing studies indicating that ‘incidental’ learning while reading is rather low (cf. Hulstijn, 1992; see Rott, 2007 for further discussion). The key to learning in this context might be investigated in more explicit interventions that promote deeper levels of language processing (see Laufer & Hulstijn, 2001, for discussion; S. Rott, 2007).

Conclusions

The purpose of the present study was to revisit the issue of the relative effectiveness of ‘micro’ level glosses for improved comprehension and L2 learning. The need to explore this issue stemmed from (a) the inconclusive findings for L2 learning of previous studies that have included ‘micro’ level glosses in their statistical comparisons, (b) the research gap in this area produced by an overemphasis on the lexicon in detriment of grammar, and (c) the

unavailability in previous studies of concurrent data to support research claims empirically.

The evidence provided in the study led the researcher to draw different conclusions with respect to the use of L1-translation glosses in relation to reading comprehension, on the one hand, and L2 learning, on the other hand. Conclusions with respect to glossing in relation to reading comprehension will be presented first.

The quantitative results indicated a significant positive effect for gloss on reading comprehension. Given that the quantitative measure of the study had a bottom-up, ‘micro’ or ‘local’ focus, the quantitative result of a positive effect of glosses on comprehension was interpreted *stricto sensu*, that is, as an effect on the text comprehension at the local level of word and sentences. Thus, when reading comprehension is viewed from a bottom up perspective, that is, the idea that text comprehension is primarily comprehension of the ‘surface’ aspects or text at the level of word and sentences, the study supported the claim that L1-translation glosses are effective comprehension aids. This quantitative result was further confirmed by the concurrent data gathered in the study.

The finding of a positive effect of ‘micro’ level glosses for ‘micro’ level understanding is not surprising given the fact that, in theory, the primary purpose of this pedagogical practice is to bridge the gap between the L2 reader’s characteristic lack of lexical and linguistic knowledge and the L2 text (e.g., Birch,

2002). In other words, from a bottom-up perspective of L2 reading, glosses for linguistic information are usually viewed as unobtrusive devices that make text contents more accessible to L2 learners. This study empirically supported this theoretical claim. Having access to L1-translations in their native language gave participants in the gloss group an attested advantage over their un-glossed peers. For these participants understanding the text at the sentence level without glosses proved especially difficult, as inferences proved not to be reliable.

The empirical studies on glossing for improved comprehension conducted from a bottom-up perspective of the reading process support the theoretical claim that glosses are effective reading comprehension aids (e.g., Davis, 1989; Jacobs, 1994; Jacobs, Dufon, and Hong, 1994; Bell & Le Blanc, 2000; Roby, 1991, 1999; Hulstijn, 1993; Bowles, 2004; Lomicka, 1996; Lyman-Hager and Davis, 1996). This study corroborated this result.

From the point of view of reading as a process, the think-aloud protocol analyses conducted in this study corroborated the quantitative finding of a positive effect of using glosses to achieve basic understanding of the L2 text. This analysis indicated that L1-translation glosses helped the participants in the gloss group understand the meaning of sentences in which the unknown words and phrases were embedded, and produce generally accurate inferences and comments. In contrast, the participants in the no-gloss group demonstrated little understanding of text content, as they were generally not able to understand the

sentences in which the unknown words were embedded. Thus, the qualitative analyses conducted in this study further confirmed the finding in support of the claim that glosses are effective comprehension aids.

The qualitative analyses of the think-aloud protocols also revealed that higher levels of text comprehension (e.g., at the level of interpretation of textual information and integration with previous knowledge) were not achieved by the participants in this study. Participants in both the gloss and the no-gloss groups approached the reading task in a linear and bottom-up fashion. Participants in both groups tended to read the text from beginning to end without making connections between paragraphs or recapitulating ideas. Reading aloud in Spanish or reading aloud by translating the text into English, or a combination of these, were the predominant reading behaviors in both groups. In general, participants in both groups did not show elaboration or integration of text context with their own schemata and knowledge in order to construct a coherent mental model for the text. In general, participants were not making inferences, connecting ideas to make causal or temporal relations, solving contradictions or drawing conclusions.

In conclusion, this study corroborates the finding in previous studies that accessing the meaning of unknown items in the native language is an effective way to achieve minimal understanding of L2 texts. However, it cannot be concluded from this study (as the advocates of top-down approaches to L2 reading might desire) that L1-translation glosses are detrimental to gain higher

levels of text comprehension, as an emphasis on reading in a bottom-up fashion was attested in both the gloss and the no-gloss groups. Therefore, how to help low proficiency learners acquire better reading habits is a question that remains open.

In this study, qualitative and quantitative analyses were conducted to address the issue of the effects of gloss and type of linguistic item on noticing, immediate retrieval via immediate recognition of items (i.e., a measure of intake), retention via a three week delayed recognition posttest, and immediate and delayed production posttest (i.e., a measure of L2-development).

The results of these analyses did not provide enough evidence to support the claim that L1-translation gloss has a positive and independent effect on noticing and learning. With respect to noticing, the amount of reported noticing by participants in the gloss group did not differ statistically from the amount of reported noticing by participants in the no-gloss group. Therefore, there was no evidence in this study to support the claim that having a gloss, versus not having a gloss, impacted the allocation of attentional resources to the targeted items embedded in the input. The qualitative analyses of the think-aloud protocols conducted in the study corroborated this finding. Think-aloud protocol analyses revealed that participants in both the gloss and the no-gloss groups did not appear to process targeted items beyond mere noticing. In other words, these analyses did not support the claim that participants in one group or the other demonstrated higher levels of awareness of the targeted items.

The finding of no effect for condition of exposure on the amount of reported noticing corroborated the findings in previous studies on input modification (i.e., text simplification, textual enhancement, and glossing) that have measured directly, empirically, the effects of input modification techniques on noticing (i.e., Leow, 2001b; Leow, Egi, Nuevo, Tsai, 2003; Bowles 2003, and 2004). In line with previous studies, the present study did not support the theoretical view of input modification as an effective way to direct learners' attention to linguistic form while interacting with an L2 text.

The finding of lack of deep processing of form-meaning connections under various conditions of exposure in the context of an L2 reading task is consistent across previous studies conducted within an attentional framework. In this study, it was hypothesized that task modality might provide an explanation for the lack of effect for input modification in these studies. Perhaps, utilizing tasks other than reading may be a more effective way to promote higher levels of processing of linguistic form (e.g., puzzles and other problem solving tasks utilized in previous SLA studies conducted within an attentional framework (e.g., Leow, 1997, 1998, 2000; Rosa & Leow, 2004; Rosa & O'Neill, 1999). In sum, both quantitative and qualitative analyses conducted in this study did not provide enough evidence to support the claim that L1-translation glosses are effective 'noticing' devices.

Like previous SLA studies (Leow et al., 2003), this study found type of item effects on noticing. Participants reported noticing vocabulary words more than items carrying the impersonal-SE. As expected, this finding was explained based on the theoretical postulation of relative difficulty of types of items (e.g., DeKeyser, 2005). In other words, the prediction that vocabulary items should be easier to notice and learn in comparison to Impersonal-se, by virtue of the differences in inherent difficulty of the item, was confirmed in the study.

In the present study, reported noticing of Present Perfect items did not differ statistically from noticing of lexical items. This unexpected finding was explained in this study by postulating that (a) lexical item turned out to be more difficult than predicted, and (b) Present Perfect was easier than predicted. Two factors were provided to account for why these types of items turned out to be different than predicted: frequency and cognitive load. The argument was made that the form-meaning connection governing the present perfect had high frequency in the input (i.e., it appeared ten times with constant meaning). In contrast, a high number of new different vocabulary words presented only once might have produced cognitive overload. The difference between being exposed to the same thing many times (i.e., present perfect) versus being exposed to many different things only once (i.e., vocabulary items) might make present perfect more noticeable than predicted, and lexical item less noticeable than predicted.

The results of the analyses of the recognition posttests were not incompatible with the findings for noticing: While an independent effect for gloss on item recognition was not supported by the evidence, glosses appeared to work in combination with other factors. For example, the finding of an effect for gloss on immediate recognition of present perfect, but no effect for vocabulary and no effect for impersonal-se, strongly suggests that glosses do not act independently but they may act in combination with factors such as frequency or inherent simplicity of the item.

Another example that the results of the analyses of the recognition posttests were not incompatible with those on noticing comes from the two interactions found between gloss and type of item at delayed posttest. Results indicated that gloss effects varied across two levels of type of item: Lexical versus Present Perfect. The participants who read the text without glosses performed statistically better on the present perfect than on the vocabulary test, thus corroborating the hypothesis of cognitive overload and frequency effects that explained differences in noticing. This finding strongly suggests that present perfect was easier to notice and retrieve than vocabulary words when meanings for unknown words are not available *through glossing or accurate differencing*. Again, L1-translation glosses do not seem to be acting independently of other factors.

The results of the delayed posttest also revealed a significant interaction between gloss and type of item. Glosses appeared to vary across levels of two types of item: Lexical Item versus Impersonal-SE. Participants who read the text with the aid of glosses were able to retain significantly more lexical items than items in the Impersonal-SE.

This result replicated the result of a significant difference between reported noticing of lexical item versus the impersonal-SE. Regardless of group participants reported noticing more vocabulary items than items in the impersonal-SE. Furthermore, the qualitative analyses of the concurrent data indicated that when participants paid attention to items in the Impersonal-SE, their focus of attention was the lexical part of the item, and not the grammatical part. This result strongly suggests that impersonal-SE was highly difficult for the participants in this study given the fact that the item appeared in the input with constant meaning with a frequency of ten. This finding was not unpredicted.

In sum, the analyses of gloss and noticing, on the one hand, and gloss and learning, on the other hand, yield similar results. These analyses did not provide enough evidence to support a claim for a glossing effect. Nevertheless, there was evidence coming from both quantitative and qualitative analyses supporting a claim for a combined effect of gloss with other factors. The interaction effects of gloss with type of items revealed in this study strongly suggest that glosses do not work in isolation but they do so in combination with factors relative to the

inherent characteristics of different types of items (e.g., salience, abstractness of their meaning, L1-L2 grammaticalization patterns, etc.). Beside these, other factors were hypothesized to play important roles within the context of glossing L2 text for improved L2 learning such as frequency of exposure, and number of targeted items embedded in the input.

Unlike most of previous studies on glossing, this study incorporated non-metalinguistic concurrent think-aloud protocols to study the processing of targeted forms during an L2 reading task. By incorporating concurrent data, the researcher was able to disentangle those concepts related to the internal processes of the learner (i.e., processing, attention, noticing, and awareness) from those related to the external conditions of exposure to the input. The distinction of one from the other is important because, once these concepts have been distinguished, the empirical contribution of external conditions of exposure to learning measures can be measured independently of the contributions of internal conditions (i.e., noticing/awareness) to learning measures. Along this line of reasoning, an important finding of the present study was that reported noticing of items overall was positively correlated with immediate and delayed recognition. However, these relationships (i.e., noticing-to-recognition) were statistically similar for the gloss and the no-gloss group. Thus, while previous studies on glosses have generally confounded attention/noticing with glossing, the present study produced an empirically grounded explanation for previous confounded results: Glosses do

not appear to have an isolated independent effect on noticing, but they may act in combination with other factors to enhance learning.

The results revealed, then, that while L1-translation glosses proved to be effective reading aids for improved comprehension, the claim for an independent effect of this type of glosses on noticing and learning was not supported by evidence. However, there was evidence that glosses act in combination with other factors, such as the type of linguistic item. Furthermore, the predictions of this study regarding relative difficulty of three types of item were only partially confirmed. The impersonal-se proved to be not easy to isolate in the context of L2 reading with glosses (or without them). This item proved highly obscure to the learner given the high frequency of appearance in the L2 experimental text. The hypothesis that the impersonal-se is not salient, abstract, and complex in terms of its grammatical encoding (between and within the language) was confirmed by the evidence in this study. Nevertheless, the predictions regarding the present perfect and the lexical items were not confirmed. The present perfect proved to be less difficult than expected. This item was hypothesized to behave differently than lexical items. However, the Present Perfect proved to be (a) not more difficult to notice than lexical items, and (b) easier to recognize than lexical items when lexical meanings were not available.

In conclusion, this study supported the finding in previous glossing studies with a bottom-up approach to L2 reading that L1-translation glosses are effective

reading aids. Furthermore, the study corroborated the findings in previous SLA studies supporting Schmidt's (1990) noticing hypothesis. Noticing is highly correlated with learning measures. However, the study did not support the notion that glosses are effective 'noticing' devices. Rather, the evidence in the study led to a new hypothesis: Glosses do not act in isolation but they do so in combination with other factors. Finally, the findings of the study seem to point in the direction of previous SLA studies conducted within an attentional framework which have found (a) that deeper levels of processing lead to more learning, and (b) more explicit interventions within the context of reading might have stronger effects on noticing, thus leading to higher and long lasting leaning gain.

Pedagogical Implications

This study has some important pedagogical implications. Perhaps, the most salient one is the use of L1-translations as reading aids with the caveat that its use may not promote higher levels of the reading process.

From the findings in this study, the effectiveness of L1-translation glosses for learning is uncertain. For this reason, teachers may be encouraged to use, prior to the reading task, proved methods of vocabulary learning and learning of difficult structures that might appear in the reading text.

An important feature of this study was the connection, finally made, between L2 reading and grammar. L2 texts abound not only in lexical but also grammatical complexity. The present study has revealed that glossed linguistic items may attract different amount of noticing and the more salient ones appear to be more feasible for glossing.

Limitations and Future Research

The evidence in this study led the researcher to formulate the hypothesis that glosses do not act independently but in combination with other variables. What factors interact with the use of L1-translation glosses of grammatical versus lexical items is a ripe area for research. The findings in this study strongly suggest that frequency of exposure plays an important role in how glossed information is processed and learned. However, frequency was not properly isolated in this study, as it conflated with other factors, such as the inherent characteristics of various types of items. Future research is needed to clarify the combined contribution of an L1-translation gloss and the frequency of exposure, especially in the realm of glossing grammatical items.

For the reason that frequency of exposure was not controlled for in this study, a comparison of three types purely in terms of their purported degrees of inherent difficulty was not straightforward. Another limitation for the construction of the variable type of item in this study was that L1-translations were used to

gloss items carrying both lexical and grammatical information. In other words, the glossing of grammatical information was not distinct and direct. Consider, for example, glossing objective pronouns (e.g., English *her*) to make case, gender, and number information available to L2 readers. This seems a more direct way of glossing for grammar. This was another reason why a comparison between types of items was difficult to establish.

Another important issue that should be considered in future studies is the operationalization of partially overlapping terms such as amount of noticing, ‘depth’ of processing, processing of form, processing of form-meaning connections, and so forth (see Leow et al., 2008, for discussion). Under the hypothesis, corroborated in the present study, that noticing is highly correlated with learning (cf. Schmidt, 1990), the finding of discrepancies between reported noticing and learning at the level of specific items strongly suggests that noticing might not be a single unitary construct. Perhaps, there are various dimensions to noticing. An argument on these lines has been recently developed in a study by Leow et al. (2008). In this study, the researchers identified three levels of processing that accounted for a differential ‘depth’ of processing of the targeted forms in the input. Thus, circling a form may not indicate the depth of processing indicated by translation of a form, commenting on it, rehearsing it, etcetera. Perhaps, more fine grained analyses of processing, in the lines of Leow and

colleagues (2008), may shed light on the relationship between noticing and learning for different types of linguistic items.

With respect to the notion of learning, this notion was operationally defined as the ability to both recognize and produce the targeted items immediately after exposure to the reading text, and three weeks later. However, no conclusions were drawn in this study from the production test data due to several limitations. For one reason, statistical analyses could not be conducted on the production tests scores because the test had a mode of scores equal to zero, which strongly suggest that many participants did not complete the test, or gave wrong answers. Furthermore, the low reliability coefficient seems to indicate that this test was too difficult for the participants. Besides the issue of low reliability, there is the issue of scoring and coding procedure. Due to the automatization of the scoring and coding procedure utilized in this study, all the potential production options could not be covered. For these reasons, no conclusions could be safely drawn from the production tests included in the study.

In this study, there were many constraints for inclusion of participants due to the several phases of the experiment, and the need to control for prior knowledge of three types of linguistic items. Furthermore, the need to control for reactivity effects, in the sense of the think-aloud task having an impact on the participants' thought processes, further increased the need for high recruitment of participants. Although these constraints were a methodological improvement in

comparison with prior studies that did not control for prior knowledge or reactivity effects, future research can increase the number of participants in order to gain statistical power.

Finally, the use of concurrent data elicitation technique (i.e., think-aloud protocols) provided rich qualitative data on reading processes. The analyses of the protocols indicated that the reading process had a clear bottom-up emphasis for both the gloss and the no-gloss groups. These qualitative data corroborated the quantitative results of the outcome measure assessing reading comprehension at a 'local' level of words in context indicating that glosses are effective reading comprehension aids. Nevertheless, the issue of levels of comprehension was not addressed in this study due to the absence of an outcome measure of reading comprehension at a global level. Therefore, whether, and how, L2 reading may be affected by external conditions of exposure at different levels of text comprehension is an important empirical issue that warrants future research.

BIBLIOGRAPHY

- Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259-302). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232.
- Aust, R., Kelley, M. J., & Roby, W. B. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research & Development*, 41, 63-73.
- Bardovi-Harlig, K. (1997). Another piece of the puzzle: The emergence of the present perfect. *Language Learning*, 47(3), 375-422.
- Bell, F., & LeBlanc, L. (2000). The language of glosses in L2 reading on computer: Learners' preferences. *Hispania*, 83, 274-285.

- Bernhardt, E. B. (1983). Testing foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis / Teaching German*, 16(1), 27-33.
- Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical, and classroom perspectives*. Norwood, NJ: Ablex Publishing.
- Birch, B. (2002). *English L2 reading. getting to the bottom*. Mahwah, NJ: Lawrence Erlbaum.
- Blau, E. K. (1982). The effect of syntax on readability for ESL students in puerto rico. *TESOL QUARTERLY*, 16(4), 517-528.
- Bowles, M., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*, 27, 415-440.
- Bowles, M. A. (2003). The effect of textual input enhancement on language laerning: An online/offline study of fourth semester spanish students. In P. Kempchinsky, & C. E. Pineros (Eds.), *Theory, practice, and acquisition: Papers from the 6th hispanic linguistic symposium and 5th conference on the acquisition of spanish & portuguese*. Summerville, MA: Cascadilla Press.

- Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Applied Linguistics*, 87(3), 541-552.
- Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA. A first look at a L2 task other than readin. *Studies in Second Language Acquisition*, 30, 359-387.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, 80, 183-198.
- Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, 1, 61-80.
- Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.
- Davies, M. (2006). *A frequency dictionary of spanish. core vocabulary for learners*. New York: Routledge.
- Davies, A. (1984). Simple, simplified and simplification: What is authentic? In J. Alderson, & A. Urquhart (Eds.), *Reading in a foreign language*, New York: Longman.

- Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal*, 73(1), 41-48.
- Davis, J. N., & Lyman-Hager, M. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, 30(1), 58-72.
- DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55(1), 1-25.
- DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "processing instruction: An update". *Language Learning*, 52(4), 805-823.
- Doddis, A. (1985). La cohesión textual en un discurso expositivo auténtico y simplificado y en sus correspondientes recreaciones. *Lenguas Modernas*, 12(136-148)
- Ellis, N. (1994). Implicit and explicit language learning. an overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages*, San Diego, CA: Academic Press.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, 51, 1-46.

- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Fernandez Huerta, J. (1959). *Medidas sencillas de lecturabilidad*. Buenos Aires: Consigna.
- Gass, S. M., & Svetics, I. (2003). Differential effects of attention. *Language Learning*, 53(3), 497-545.
- Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals*, 34, 91-106.
- Goodman, K. (1973). The psycholinguistic nature of the reading process. In K. Goodman (Ed.), *The psycholinguistic nature of the reading process*, Detroit, MI: Wayne State University Press.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL QUARTERLY*, 25(3), 375-406.
- Greenslade, T., Bouden, L., & Sanz, C. (1999). Attending to form and content in processing L2 reading texts. *Spanish Applied Linguistics*, 3, 65-90.

- Hatch, E. (1983). Simplified input and second language acquisition. In R. W. Anderson (Ed.), *Pidginization and creolization as language acquisition*, Rowley, MA: Newbury House.
- Hughes. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hulstijn, J., & de Graff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, 11, 97-112.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. A. a. H. Béjoint (Ed.), *Vocabulary and applied linguistics* (pp. 113-125). London: McMillan.
- Hulstijn, J. H. (1993). When do foreign language readers look up the meaning of unfamiliar words? the influence of task and learner variables. *The Modern Language Journal*, 77
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, 80(3), 327-339.

- Izumi, S. (2000). Promoting noticing and SLA: An empirical study of the effect of output and input enhancement on ESL relativization. (Diss., Georgetown University).
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. an experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541-577.
- Jacobs, G. (1994). What lurks in the margin: Use of vocabulary glosses as a strategy in second language reading. *Issues in Applied Linguistics*, 5(1), 115-137.
- Jacobs, G., Dufon, P., & Hong, F. C. (1994). L1 and L2 glosses in reading passages, their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, 17(1), 19-28.
- Johnson, P. L. (1982). Effects on reading comprehension of building background knowledge. *TESOL QUARTERLY*, 16(14), 503-516.
- Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). *Does textual enhancement promote noticing?: A think aloud protocol analysis*. Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Kellogg, M. *WordReference.com*. Retrieved May 2009, 1999, from

<http://www.wordreference.com/>

Kern, R. (1989). Second language reading strategy instruction: Its effects on comprehension and word inference ability. *The Modern Language Journal*, 73(2), 135-149.

Krantz, P. *Standards schmandards*. <http://www.standards-schmandards.com/exhibits/rix/>

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1-26.

Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *SSLA*, 8, 201-212.

Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *SSLA*, 15, 333-355.

Leow, R. P. (1995). Modality and intake in second language acquisition. *Language Learning*, 47, 467-505.

- Leow, R. P. (1997a). Attention, awareness, and foreign language learning. *Language Learning*, 47, 467-505.
- Leow, R. P. (1997b). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning*, 8(2), 151-182.
- Leow, R. P. (1998a). The effects of amount and type of exposure on adult learners' L2 development in SLA. *Modern Language Journal*, 82, 49-68.
- Leow, R. P. (1998b). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics*, 19, 133-159.
- Leow, R. P. (1999). The role of attention in second/foreign language classroom research: Methodological issues. In F. Martínez-Gil, & J. Gutiérrez-Rexach (Eds.), *Advances in hispanic linguistics: Papers from the 2nd. hispanic linguistics symposium* (pp. 60-71). Somerville, MA: Cascadilla Press.
- Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, 22, 557-584.

- Leow, R. P. (2001a). Attention, awareness, and foreign language behavior. *Language Learning, 51*, 113-155.
- Leow, R. P. (2001b). Do learners notice enhanced forms while interacting with the L2?: An online and offline study of the role of written input enhancement in L2 reading. *Hispania, 84*(3), 496-509.
- Leow, R. P., Egi, T., Nuevo, A. M., & Tsai, Y. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning, 13*(2), 93-108.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud. *Studies in Second Language Acquisition, 26*, 35-57.
- Leow, R. P., & Bowles, M. A. (2006). Attention and awareness in SLA. In C. Sanz (Ed.), *Adult second language acquisition. Methods, theory, and practice*, Washington, DC: Georgetown University Press.
- Leow, R. P., Hsieh, H. C., & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning, 58*(3), 665-695.
- Leow, R. P. (2009a). Input enhancement and L2 grammatical development: What the research reveals. In J. Watzinger-Tharp, & S. Katz (Eds.), *Conceptions of*

L2 grammar: Theoretical approaches and their application in the L2 classroom, Boston, MA: Heinle Publishers.

Leow, R. P. (2009b). Modifying the L2 reading text for improved comprehension and acquisition: Does it work? In ZhaoHong Han, & N. J. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries*, Ann Arbor, MI: The University of Michigan Press.

Lomicka, L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning and Technology*, 1, 41-50.

Long, M., & Ross, S. (1993). Modifications that preserve language and content. In M. L. Tickoo (Ed.), *Simplification: Theory and application*, RELC Anthology Series No 31.

Long, M. H. (2007). *Problems in SLA* (Nahwah ed.). NJ: Erlbaum.

Lyman-Hager, M., & Davis, J. N. (1996). The case for computer-mediated reading: Une vie the boy. *The French Review*, 69(5), 775-790.

Lyman-Hager, M. A., Davis, N., Burnett, J., & Chennault, R. (1993). Une vie the boy: Interactive reading in french. In F. B. & E. M. T. Johnson (Ed.), *Proceedings of the CALICO 1993 annual symposium on assessment* (pp. 93-97). Durham: NC: Duke University.

- Mayer, R. E. (1984). Aids to text comprehension. *Educational Psychologist*, 19, 30-42.
- Mendikoetxea, A. (1999). Construcciones con se: Medias, pasivas e impersonales. In I. Bosque, V. Demonte & F. L. Carreter (Eds.), *Gramatica descriptiva de la lengua espanola* (pp. 1631-1778). Madrid: Espasa.
- Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32, 469-479.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Oller, J. W. (Ed.). (1983). *Issues in language testing research*. Rowley: MA: Newbury House.
- Overstreet, M. (1998). Text enhancement and context familiarity: The focus of learner attention. *Spanish Applied Linguistics*, 2, 229-58.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, 241-263.
- Pak, J. (1986). *The effect of vocabulary glossing on ESL reading comprehension*. Unpublished manuscript.

- Robinson, P. (1995). Review article. attention, memory, and the "noticing" hypothesis. *Language Learning*, 45(2), 283-331.
- Robinson, P. (1996). *Consciousness, rules and instructed second language acquisition*. New York: Peter Lang.
- Roby, W. B. (1991). Glosses and dictionaries in paper and computer formats as adjunct aids to the reading of spanish texts by university students.
- Roby, W. B. (1999). What's in a gloss? A commentary on lara L. lomicka's "to gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 94-101.
- Rosa, E., & O'Neill, M. (1999). Explicitness, intake, and the issue of awareness. another piece of the puzzle. *SSLA*, 21, 511-556.
- Rosa, E. E., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, (25), 269-292.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21, 45-74.

- Rott, S. (2007). The effect of frequency of input-enhancement on word learning and text comprehension. *Language Learning, 57*(2), 165-199.
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*, 759-769.
- Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29*, 67-100.
- Sadoski, M., & Paivio, A. (2001). *Imagery and text. A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum.
- Sanz, C., Lin, H. -, Lado, B., Bowden, H. W., & Stafford, C. A. (2009). Concurrent verbalizations, pedagogical conditions, and reactivity: Two CALL studies. *Language Learning, 59*(1), 33-71.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2)
- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.

- Schmidt, R. (1994). Implicit learning and the cognitive unconscious. In N. E. (Ed.) (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). London: Academic Press.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning*, Honolulu, HI: University of Hawai'i Press.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Schnotz, W. (1993). On the relation between dual coding and mental models in graphics comprehension. *Learning and Instruction*, 3, 247-249.
- Sharwood Smith, M. (1980). Consciousness-raising and the second language learner. *Applied Linguistics*, 11(2), 160-168.
- Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7(2), 118-132.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *SSLA*, 15, 165-179.

- Shook, D. J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning, 5*, 57-93.
- Smith, F. (1973). *Psycholinguistics and reading*. Toronto: R&W.
- Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). London: Sage.
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183-203.
- VanPatten, B. (1989). Can learners attend to form and content while processing input? *Hispania, 72*, 409-417.
- 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. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review, 11*, 27-36.
- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex Publishing Corporation.

- VanPatten, B. (2004). *Processing instruction. theory, research, and commentary*.
Nahwah, NJ: Lawrence Erlbaum Associates.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing
on incidental learning of foreign language vocabulary. *Studies in Second
Language Acquisition, 19*, 287-307.
- Wells, D. R. (1986). The assessment of foreign language reading comprehension:
Refining the task. *Die Unterrichtspraxis / Teaching German, 19*(2), 178-184.
- White, J. (1998). *Getting the learners' attention: A typographical input
enhancement study*. Cambridge & London: Cambridge University Press.
- Whitley, M. S. (1986). *Spanish/English contrasts*. Washington, DC: Georgetown
University Press.
- Wong, W. (2001). Modality and attention to meaning and form in the input.
Studies in Second Language Acquisition, 23, 345-368.
- Wong, W. (2003). Textual enhancement and simplified input: Effects on L2
comprehension and acquisition of non-meaningful grammatical form.
Applied Language Learning, 13(2), 17-45.

Wong, W. (2005), *Input Enhancement: From theory and research to the classroom*, New York, NY: McGraw-Hill.

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language comprehension. *Language Learning*, 44, 189-219.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology*, 10, 85-101.

Young, D. J. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *Modern Language Journal*, 83(3), 348-366.

APPENDIX A

The experimental text (targeted items have been underlined and numerated)

LOS ARGENTINOS, SU CULTURA Y TRADICIONES LA PÁGINA WEB DE MARIANA

¡Hola! Soy Mariana Clavero y soy de Argentina. A menudo escucho opiniones pasmosas¹ sobre los argentinos. Por ejemplo, algunas personas dicen que en Argentina, todos son macanudos². Otras personas dicen que en Argentina se abultan³ todas las cosas.

Sin embargo, millones de personas que han surcado⁴ la Argentina no piensan eso. Ellos piensan que los argentinos son únicos en muchos aspectos, como la lengua, los deportes y las costumbres.

Algunos lectores que han elogiado⁵ mi página en la Internet me preguntan por qué los argentinos son especiales. Yo les respondo a mis lectores que la cultura y las tradiciones argentinas son especiales. Desafortunadamente, a veces se desdeñan⁶ la cultura y las tradiciones.

LA LENGUA

Debo decirte que algunos visitantes piensan que cuando dos argentinos hablan, la lengua que usan es zafia⁷. Incluso algunos argentinos han transigido⁸ con esa idea. Sin embargo, yo pienso que este estilo de hablar se platica⁹ en los bares y clubes de la ciudad.

EL FÚTBOL

El deporte es muy importante entre las costumbres argentinas. El fútbol es nuestro deporte favorito. El fútbol es tan popular que los domingos, en la TV, siempre se agregan¹⁰ programas de fútbol a la programación regular. No es estrambótico¹¹ encontrar personas mirando un partido de fútbol por TV en un restaurante costoso.

Nuestro equipo nacional de fútbol ha descollado¹² internacionalmente. River y Boca son los equipos más importantes. Muchos tienen encono¹³ porque River y Boca son los más populares. En los últimos partidos, los jugadores de los dos

equipos famosos, River y Boca, han reñido¹⁴ duro. La rivalidad entre estos equipos es tan fuerte que alguna vez los hinchas han vilipendiado¹⁵ a los jugadores.

El fútbol no se veda¹⁶ a las mujeres. Sin embargo, las mujeres prefieren ver el fútbol desde la casa, por la televisión.

EL MATE

Hay otras costumbres argentinas. Un ejemplo es el mate. El mate es una infusión que las personas preparan con una planta especial: la yerba. Los argentinos han cotejado¹⁷ el mate con las más antiguas tradiciones americanas.

Muchos me preguntan cómo preparas el mate. Para preparar el mate, debes colocar la yerba en un contenedor de metal o madera que tiene una bombilla¹⁸ de plata.

Este contenedor es el mate. La yerba se embute¹⁹ en el mate. Luego, la bombilla se liba²⁰.

Para beber, debes añadir agua caliente en el contenedor. El agua de mate no se hierve²¹.

Puedes beber el mate solo o en compañía de familiares o amigos. Muchas personas han sorbido²² el mate en compañía de familiares y amigos. Pero para beber mate, debes hacerlo con un amigo fetén²³.

EL TANGO

Muchos extranjeros admiran esta costumbre típicamente argentina. El tango es nuestra danza. No es fácil encontrar el origen de esta hermosa danza. Después de un espiguelo²⁴, algunos creen que el origen del tango es a fines del siglo XIX, en los barrios portuarios de la ciudad.

Al principio, el tango era considerado una danza de furcias²⁵. Por esta razón, los temas del tango son temas tristes. Por ejemplo, los tangos son canciones que hablan de seres humanos que nunca han paliado²⁶ sus penas, donde el amor es frágil y se desbarata²⁷. Sin duda el espíritu del tango es lóbrego²⁸.

Al principio, el tango no era famoso. Carlos Gardel, nuestro cantante nacional, lo immortalizó. Carlos Gardel no ha espichado²⁹ sin antes catapultar el tango a la fama.

Hoy, el tango es tan popular que todos quieren bailar cuando se avista³⁰ la banda.

FIN

APPENDIX B

Instructions for four experimental groups (a, b, c, and d)

(a) Instructions for LeKtor (i.e., [+think-aloud], [-gloss]) group.

INSTRUCTIONS

In this task, you will read an excerpt talking about Argentines, and their culture and traditions.

Read at your own pace. As you read, **think aloud all your thoughts to the recorder. This is very important: Do not stop thinking aloud at any moment. If you stop, I will ask you to keep thinking aloud.**

Immediately after this reading task, you will complete a comprehension questionnaire. So, as you read, try to understand the content of the text (i.e., the what, who, when, where, etc.).

END OF INSTRUCTIONS

(b) Instructions for Lektor-EM (i.e., [-think-aloud] [-gloss]) group

INSTRUCTIONS

In this task, you will read an excerpt talking about Argentines, and their culture and traditions. Read at your own pace.

Immediately after this reading task, you will complete a comprehension questionnaire. So, as you read, try to understand the content of the text (i.e., the what, who, when, where, etc.).

END OF INSTRUCTIONS

(c) Instructions for LeKtor-G (i.e. [+think-aloud], [+gloss]) group

INSTRUCTIONS

In this task, you will read an excerpt talking about Argentines, and their culture and traditions.

On the upper section of the screen, you will see the reading passage. Some words and phrases of the text have been underlined, and translated for you.

The English translation of these words and phrases appear at the bottom of the screen.

Each time you find an underlined word or phrase in the text, please find the referenced translation at the bottom of the screen.

When you see the English translation, please click on the box that appears to the left of the translation. In this way, I will know that you saw the translation, which is very important!!!!

Read at your own pace. As you read, **think aloud all your thoughts to the recorder. This is very important: Do not stop thinking aloud at any moment. If you stop, I will ask you to keep thinking aloud.**

Immediately after this reading task, you will complete a comprehension questionnaire. So, as you read, try to understand the content of the text (i.e., the what, who, when, where, etc.).

END OF INSTRUCTIONS

(d) Instructions for LeKtor-GEM (i.e., [-think-aloud], [+gloss]) group

INSTRUCTIONS

In this task, you will read an excerpt talking about Argentineans, and their culture and traditions.

On the upper section of the screen, you will see the reading passage. Some words and phrases of the text have been underlined, and translated for you.

The English translation of these words and phrases appear at the bottom of the screen.

Each time you find an underlined word or phrase in the text, please find the referenced translation at the bottom of the screen.

When you see the English translation, please click on the box that appears to the left of the translation. In this way, I will know that you saw the translation which is very important!!!!

Read at your own pace. Immediately after this reading task, you will complete a comprehension questionnaire. So, as you read, try to understand the content of the text (i.e., the what, who, when, where, etc.).

END OF INSTRUCTIONS

APPENDIX C

Illustrations of the computerized version of the experimental text for two experimental groups (a, b) and three tests (c, d, and e).

(a) Screen shot of the LeKtor G (i.e., [plus gloss] [no think-aloud] group

LOS ARGENTINOS, SU CULTURA Y TRADICIONES
LA PÁGINA WEB DE MARIANA

¡Hola! Soy Mariana Clavero y soy de Argentina. A menudo escucho opiniones pasmosas¹ sobre los argentinos. Por ejemplo, algunas personas dicen que en Argentina, todos son macanudos². Otras personas dicen que en Argentina se abultan³ todas las cosas. Sin embargo, millones de personas que han surcado⁴ la Argentina no piensan eso. Ellos piensan que los argentinos son únicos en muchos aspectos, como la lengua, los deportes y las costumbres.

1. astonishing 2. cool 3. people exaggerate 4. have traveled across

Moving to the next question prevents changes to this answer. Question 1 of 3 | ▶

(b) Screen shot of the LeKtor (i.e., control) group

LOS ARGENTINOS, SU CULTURA Y TRADICIONES
LA PÁGINA WEB DE MARIANA

¡Hola! Soy Mariana Clavero y soy de Argentina. A menudo escucho opiniones pasmosas sobre los argentinos. Por ejemplo, algunas personas dicen que en Argentina, todos son macanudos.

Otras personas dicen que en Argentina se abultan todas las cosas.

Sin embargo, millones de personas que han surcado la Argentina no piensan eso. Ellos piensan que los argentinos son únicos en muchos aspectos, como la lengua, los deportes y las costumbres.

When finished, click on this box, and click "submit" to go to the next section. Do not click on this box.

Moving to the next question prevents changes to this answer. Question 1 of 9 | ▶

(c) Screen shot of Comprehension Questions (i.e. Reading Comprehension Test)

▼ **Question Completion Status:**

Question 1 **1 points** [Save](#)

In the first paragraph of the article, how does the author react to the opinions about Argentinians?

with surprise.

with anger.

with anger and surprise.

None of the above.

Moving to the next question prevents changes to this answer. Question 1 of 30 | ▶

(d) Screen shot of Test B (i.e., Production test).

▼ Question Completion Status:

Question 2 **1 points** [Save](#)

Juan, mi amigo argentino, es (cool).

Moving to the next question prevents changes to this answer. Question 2 of 40 | ▶

(e) Screen shot of Test C (i.e., Recognition test)

▼ Question Completion Status:

Question 8 **1 points** [Save](#)

se platica

- it knows
- it speaks
- is spoken
- None of the above.

Moving to the next question prevents changes to this answer. Question 8 of 40 | ▶

APPENDIX D

Multiple-Choice Comprehension Questionnaire

Note. Correct answers are in bold.

1. In the first paragraph of the article, how does the author react to the opinions about Argentineans?
 - a. **with surprise.**
 - b. with anger,
 - c. with anger and surprise.
 - d. none of the above.

2. Some people say that Argentineans are:
 - a. bigheaded.
 - b. **great.**
 - c. OK people.
 - d. none of the above.

3. What does the author say that happens in Argentina?
 - a. people exaggerate about Argentineans, but it is true that everything is big.
 - b. some say that people exaggerate about Argentineans.
 - c. some say that everything is big.
 - d. **none of the above.**

4. Millions of people...
 - a. **have traveled across Argentina.**
 - b. were passing through Argentina.
 - c. had traveled in Argentina.
 - d. none of the above.

5. The readers...
 - a. had admired the page.
 - b. **have admired the page.**
 - c. were reading the page.
 - d. none of the above.

6. What does the author say about culture and traditions?
- a. what matters is to be aware.
 - b. important things do matter.
 - c. people do not pay attention to these important things.**
 - d. none of the above.
7. What do some visitors think about the way Argentines speak?
- a. It is crude.**
 - b. It is different.
 - c. It is elegant.
 - d. none of the above.
8. Some Argentines...
- a. have accepted the ideas held by the visitors.**
 - b. had formed an opinion about the Argentinean style.
 - c. were tolerating the ideas about the Argentinean style.
 - d. none of the above.
9. What does the author think about the Argentinean style?
- a. Argentines tend to be formal but they can be informal sometimes.
 - b. Argentines enjoy speaking with visitors in bars and clubs.
 - c. people speak that way in certain social contexts but not in others.**
 - d. none of the above.
10. What happens on Sundays on TV?
- a. there are more soccer programs than regular programs.
 - b. there are less programs because of the games time extensions.
 - c. some soccer programs get extended.
 - d. none of the above.**
11. According to the information in the text, watching a soccer game in a fancy restaurant is...
- a. not eccentric.**

- b. not out of taste.
- c. vulgar but not eccentric.
- d. none of the above.

12. The Argentinean national soccer team:

- a. had never had better times in the past.
- b. had won championships worldwide.
- c. has performed tremendously well abroad.**
- d. none of the above.

13. What are the feelings toward River and Boca (for many)?

- a. not grudge but pure envy.
- b. grudge.**
- c. jealousy.
- d. none of the above.

14. About women in Argentina the article says that

- a. women do not like soccer, but in general they watch the game at home.
- b. women prefer watching the game at home because they do not care for soccer.
- c. soccer is boring, so women prefer to watch it at home once in a while.
- d. none of the above.**

15. In past games

- a. the players had chatted.
- b. the players have quarreled.**
- c. the players were arguing.
- d. none of the above.

16. According to the text, fans

- a. have insulted the players.**
- b. had insulted the players.
- c. took offense.
- d. none of the above.

17. Argentineans

- a. were studying the ancient tradition of *mate*.
- b. had compared the *mate* to the most ancient traditions.
- c. have compared the *mate* to older traditions.**
- d. none of the above.

18. The *mate* has a

- a. silver container.
- b. silver straw.**
- c. silver straw and silver container.
- d. none of the above.

19. In the preparation of the *mate*

- a. you must pack the yerba in the *mate*.**
- b. the yerba settles in the *mate*.
- c. you must stir the yerba in the *mate*.
- d. none of the above.

20. What does the author say about the straw?

- a. the straw is light sometimes but not always.
- b. the straw is light.
- c. you must suck the *mate* through the straw.**
- d. none of the above.

21. In the *mate* preparation

- a. you must not let the water boil.**
- b. the water does not become cloudy.
- c. the water does not evaporate.
- d. none of the above.

22. Many people

- a. had sipped the *mate* before the water boiled.
- b. have sipped the *mate* in company of their friends.**
- c. had drunk the *mate* in the company of friends.
- d. none of the above.

23. The mate is to be shared with a friend who is...
- a. **sincere.**
 - b. fabulous.
 - c. fiercest.
 - d. none of the above.
24. Some people now believe that the origin of tango was in the XIX century because
- a. they found important clues in the city.
 - b. they found some important documents.
 - c. **they conducted a search.**
 - d. none of the above.
25. Who used to dance tango at the very beginning?
- a. very well dressed women.
 - b. people in general.
 - c. **prostitutes.**
 - d. none of the above.
26. According to the text, tango songs are...
- a. **about men who have not coped with their grief.**
 - b. had never been acquainted with their grief.
 - c. were trying to cope with their grief.
 - d. none of the above.
27. According to tango songs,
- a. love is everything.
 - b. love is cheap.
 - c. love is fragile but not cheap.
 - d. **none of the above.**
28. How does the author describe the origins of tango?
- a. coarse but not dark.
 - b. refined but sad.
 - c. **dark and sad.**

d. none of the above.

29. Carlos Gardel

- a. **has made tango famous before he died.**
- b. had died before tango became famous.
- c. was popularizing the tango when he became sick.
- d. none of the above.

30. When do people like to dance the tango?

- a. when both the band and the dancers are ready.
- b. when the band is ready.
- c. when they are ready.
- d. **none of the above.**

APPENDIX E

Recognition Test

Note. Correct answers are marked in bold.

Directions. This is a multiple-choice test of the text you just read. What is the meaning of the following Spanish words and phrases? Choose the correct answer from (a), (b), (c), (d).

1. pasmosas
 - a. stupid but not false.
 - b. stupid.
 - c. false.
 - d. none of the above.**

2. macanudo
 - a. arrogant.
 - b. gorgeous.
 - c. just OK.
 - d. none of the above.**

3. zafia
 - a. vulgar.**
 - b. formal.
 - c. senseless.
 - d. none of the above.

4. estrambótica
 - a. exaggerated.
 - b. strange.**
 - c. amusing.
 - d. none of the above.

5. encono

- a. **resentment.**
- b. scorn.
- c. jealousy.
- d. none of the above.

6. bombilla

- a. type of silver.
- b. type of *mate*.
- c. type of spoon.
- d. **none of the above.**

7. fetén

- a. **honest**
- b. familiar
- c. rich
- d. none of the above

8. espiguelo

- a. type of plant.
- b. type of old document.
- c. **type of search.**
- d. none of the above.

9. se agregan

- a. are adding.
- b. **are added.**
- c. some of them add.
- d. none of the above.

10. furcia

- a. **prostitute**
- b. woman
- c. flower
- d. none of the above

11. lóbrego

- a. with no light
- b. depressively dark**
- c. not completely dark
- d. none of the above

12. se abulta

- a. people exaggerate**
- b. things are big.
- c. things abound
- d. none of the above

13. se desdeña

- a. is destitute.
- b. is disregarded.**
- c. is disrupted
- d. none of the above

14. se platica

- a. it knows
- b. it speaks
- c. is spoken**
- d. none of the above

15. se veda

- a. is prohibited.**
- b. it prohibits.
- c. he prohibits.
- d. none of the above

16. se embute.

- a. is cramming.
- b. he packs
- c. it packs
- d. none of the above**

17. se liba

- a. **is sucked.**
- b. he sucks
- c. is sucking
- d. none of the above

18. se hierva.

- a. is boiled.
- b. is boiling
- c. **is boiled**
- d. none of the above.

19. se desbaratan

- a. **are destroyed.**
- b. are destroying.
- c. they destroy.
- d. none of the above.

20. se avista

- a. is seeing.
- b. **is seen**
- c. he sees
- d. none of the above.

21. han surcado

- a. were crossing
- b. had crossed
- c. **have traveled across**
- d. none of the above

22. han elogiado

- a. had admired.
- b. are admiring.
- c. **have admired.**
- d. none of the above.

23. han transigido

- a. had had concurred with
- b. had compromised
- c. were concurring
- d. none of the above**

24. han descollado

- a. have excelled**
- b. had excelled
- c. were excellent
- d. none of the above

25. han reñido

- a. had quarreled
- b. have quarreled**
- c. were quarrelling
- d. none of the above

26. han vilipendiado

- a. had insulted.
- b. have insulted.**
- c. Were insulted.
- d. None of the above.

27. han cotejado

- a. had compared.
- b. compared.
- c. were compared.
- d. none of the above**

28. han sorbido

- a. have sipped.**
- b. were sipping.
- c. had sipped

d. none of the above

29. han paliado

a. were alleviated

b. had alleviated

c. have alleviated

d. none of the above

30. ha espichado

a. has died.

b. had died.

c. was dying

d. none of the above

APPENDIX F

Production Test

Directions. This is a fill-in-the-blank test of the text you have read. For each sentence, fill in the blank with the Spanish translation of the English word or phrase that appears in parenthesis.

1. Ese libro es _____ (astonishing).
2. Juan, mi amigo argentino, es _____ (cool).
3. Ese discurso es _____ (vulgar).
4. Ese deporte no es _____ (bizarre).
5. Boca y River son los favoritos y por eso muchas personas tienen _____ (resentment).
6. Para tomar *mate* usas una _____ (straw).
7. El *mate* es como un amigo _____ (genuine).
8. Los investigadores hacen un _____ (search) para un trabajo.
9. En esta calle hay una _____ (prostitute).
10. El tango es _____ (gloomy).
11. En Argentina _____ (people exaggerate).
12. A veces _____ (people ignore) la verdad.
13. En los bares y clubes _____ (is spoken) el tema social.
14. Para hacer un plato famoso _____ (are added) huevos.
15. Fumar _____ (is prohibited).
16. En el *mate*, la yerba _____ (is packed).
17. El *mate* _____ (is sucked) por la bombilla.
18. El agua _____ (is boiled).
19. Muchas teorías malas _____ (are destroyed) con facilidad.
20. Cuando la banda _____ (is seen), todos salen a bailar.
21. Muchos visitantes _____ (have traveled across) la Argentina.
22. Muchas personas _____ (have admired) nuestro territorio.
23. Algunas personas _____ (have concurred) con esa idea.
24. Nuestro equipo _____ (has excelled) en el extranjero.
25. Ellos _____ (han reñido).
26. Los hinchas nunca _____ (have insulted) a un jugador.
27. Ellos _____ (have compared) las lenguas.

28. Muchos argentinos _____ (have sipped) el *mate* con amigos.
29. Tus palabras _____ (have alleviated) mi pena.
30. Carlos Gardel _____ (has died).

¹ Unlike the research strand of textual input enhancement, the issue whether learners can process linguistic forms while reading for meaning, and especially whether they would pay attention to, notice and become aware of linguistic form during an L2 reading task, has not been a main research focus from the very beginning. The focus of the first studies in the glossing strand of research has been on reading comprehension. In these first studies, the issue of learning through reading glossed texts is rarely mentioned because learning, in this context, is often viewed as a by product of text comprehensibility. In other words, the first studies on glossing tend to assume that learning took place under certain treatment conditions. With time, as researchers become more interested in L2 learning through glossing L2 texts, glossing studies take an empirical turn to incorporate learning measures in their research designs. Although this is considered an improvement, there is still the question whether the outcome measures utilized in these studies are appropriate in relation to the research plans and goals (see J. H. Hulstijn, 1993, for discussion). This methodological concern stems from the observation that while the majority of the studies on glossing for L2 learning view learning as a process, the performance tests utilized are more appropriate to measure learning as an outcome (see Leow, 1999, for further discussion of this methodological issue in SLA research).

² The analyses on the production scores were not conducted because these scores had a standard deviation equal to zero.

³ The final sample of the study was N=65, with a distribution across cells as shown below:

Cell Distribution - Final Sample

Group	+Gloss	-Gloss	Total
+TK	19	17	36
- TK	14	15	29
Total	33	32	65

The distribution of participants across groups in the final sample was not even due to attrition, i.e., death of participants. This explains why there are different numbers in the cells in the final sample.

⁴ Following standard procedures in previous studies on reactivity in SLA, only the immediate posttests were included for this analysis.

⁵ For the purpose of these analyses, the variable Think-aloud was collapsed and the original four groups were recoded into two groups: gloss and no-gloss.