

The Effects of Hypertext Glosses on L2 Vocabulary

Acquisition: A Meta-Analysis

By

Jee Hwan Yun

Submitted to the Department of Curriculum and Teaching and the Faculty of
the Graduate School of the University of Kansas in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Dissertation Committee

Dr. Manuela Gonzalez-Bueno

Dr. Bruce Frey

Dr. Paul Markham

Dr. Phil McKnight

Dr. Young-jin Lee

ABSTRACT

In the field of second language acquisition (SLA), “comprehensible input” (Krashen, 1985) has been considered a critical factor to help learners acquire foreign and second languages (L2). From this perspective, the notion of extensive or free voluntary reading (Day & Bamford, 1998; Krashen, 1993) has emerged that L2 learners should be given more pleasure reading by minimizing a burden look-up behavior. At the same time, technology innovation has made it possible for extensive reading to occur through technology over the past decades. In particular with hypertext glosses or multimedia annotations, a number of studies have indicated that hypertext glossed input is comprehensible input and has made it possible for L2 readers to benefit all from extensive reading.

This study examines (1) effects of hypertext gloss use on L2 vocabulary acquisition in computerized reading contexts, and (2) which specific combination of either text-only (single) or text + visual (multiple) hypertext glosses is more effective on L2 vocabulary acquisition and 3) What potential moderators to systematically account for between study variation are. In addition, it aims to synthesize characteristics of studies, technology use and research methods from empirical research studies for a comprehensible and insightful review of the effect of hypertext glosses on L2 vocabulary acquisition. Meta-analysis as a quantitative method was conducted to synthesize overall findings of empirical studies by calculating a standardized mean difference effect size. From 300 papers considered, 10 met the *Criteria for Inclusion* through a final filtering process, and were finally meta-analyzed to extract effect sizes in the present study. On the basis of 35 weighted mean effect size, 0.46 (Cohen, 1988: medium), the magnitude of text + visual (multiple) hypertext gloss

combination was moderately effective on L2 vocabulary acquisition when L2 learners were given two conditions: a text-only or a text + visual hypertext glosses. The results revealed that various L2 learners, including English as a second or foreign language (ESL/EFL), Spanish as a foreign language (SFL), Japanese as a foreign language (JFL), and German as a foreign language (GFL), benefit from multiple hypertext glosses while reading computerized texts. In terms of research design, hypertext gloss studies have been almost always conducted in settings of class session-based quasi-experiment design with a researcher-developed program at a university or college level. More implications are discussed for future research.

TABLE OF CONTENTS

ABSTRACT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES, GRAPHS AND FORMULARS

I. INTRODUCTION

1.1. Background of the Study

1.2. Statement of problem

1.3. Significance of the Study

1.4. Purpose of the Study

1.5. Research Questions

Chapter Summary

II. REVIEW OF THE LITERATURE

2.1. Meta-analysis

2.2. Meta-analysis and CALL

2.3. Theoretical Background

2.4. Defining a Gloss and a Hypertext Gloss

2.5. Hypertext Gloss Examples

2.6. Hypertext Glosses and SLA

2.7. Hypertext and Extensive Reading

2.8. The Studies of Hypertext Glosses

Chapter Summary

III. METHOD

3.1. Meta-Analysis Statistics

3.2. A Random-effects Model

3.3. Location and Selection of Publication

3.4. Criteria for Inclusion

3.5. Descriptions of Coding Characteristics

Chapter Summary

IV. RESULTS

4.1. Three Filtering Processes

4.2. Research Findings

4.3. Descriptive Data of the Present Meta-Analysis

Chapter Summary

V. DISCUSSION, CONCLUSION, RESEARCH LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

5.1. Discussion

5.2. Implications for Future Research

5.3. Research Limitations

REFERENCES

APPENDIXES

APPENDIX A: Keywords Used for Searches

APPENDIX B: Academic and Educational Databases and Journals

APPENDIX C: Studies included for the present meta-analysis

APPENDIX D: Meta-Analysis Coding Manual for the Effects of

Hypertext Annotations on L2 Vocabulary Acquisition

APPENDIX E: Statistics for Effect Sizes & Characteristics in the

Analysis

APPENDIX F: Technology Characteristics

APPENDIX G: Effect Size Statistics

APPENDIX H: Forest Plot for 37 Effect Sizes

LIST OF TABLES

Table 1	Taxonomy of Glosses -----	24
Table 2	Summary of Hypertext Gloss Studies-----	42
Table 3	Coding Characteristics -----	51
Table 4	10 Studies Selected for the 37 Effect Sizes-----	56
Table 5	Descriptive Effect Size Statistics -----	59
Table 6	Descriptive Effect Size Statistics (without the # 2 and # 4 outliers) --	61
Table 7	Summary of Moderator Variables-----	73

LIST OF FIGURES GRAPHS AND FORMULARS

FIGURE 1	Mayer's Generative Theory of Multimedia Learning -----	22
FIGURE 2	Screen Shot of Electronic Book with Hyper-Reference -----	27
FIGURE 3	Screen Shot of An Electronic Gloss with Picture -----	28
FIGURE 4	Screen Shot of Gloss Types -----	29
FIGURE 5	Screen Shot of <i>CyberBuch</i> -----	29
FIGURE 6	Screen Shot of Screen Layout from a Card of Hypertext-based Courseware -----	31
FIGURE 7	Screen Shot of Example of a Hypertext Structure -----	31
FIGURE 8	Screen Shot of Reactive Areas in a Reading Passage -----	32
FIGURE 9	Interactive Hypertext Example 1-----	33
FIGURE 10	Interactive Hypertext Example 2-----	34
FIGURE 11	Screen Shot of Main Course Window -----	35
GRAPH 1	Scatter Plot of 37 Effect Sizes-----	61
GRAPH 2	Funnel Plot -----	61
FORMULAR 1	Cohen's D -----	46
FORMULAR 2	Hedge's G -----	47

CHAPTER I

INTRODUCTION

1.1. Background of the Study

Technology has been playing such a crucial role in the field of language education that the United States of Department of Education (USDE) announced that the integration of technology use in English as a second language (ESL) teaching should be mandated for ESL teachers in the United States (Beatty, 2003; Chapelle, 2001; Levy, 1997). The Standards for Foreign Language Learning (SFL) in the 21st Century (NSFEP, 1999) included technology as one of the elements in the weave of foreign language learning. The International Society for Technology Education (ISTE) provided standards for technologically literate students in 1999. That is, a variety of disciplines in the field of education have required technology-enhanced language learning (TELL) in school settings. In addition, language education has recognized the needs of pedagogical TELL integration that can broaden the scope of language teaching and learning. In the field of major second and foreign languages (L2), such as English, Spanish, Korean, German, French, Japanese and Chinese, for example, a number of researchers and educators have made efforts in order to integrate technology-embedded language learning into classroom teaching and learning as efficiently as possible.

Computer-mediated communication (CMC) – (1) Asynchronous such as Email, Net pals, WebCT, Blackboard and ICON, and (2) Synchronous such as Instant Messengers, MOO and Internet Relay Chat – has become daily base language teaching resources across

classes (Beatty, 2003; Bush & Terry, 1996; Warschauer & Kern, 2000). Especially through the Internet or the Web, it became more plausible for L2 learners to contact native speakers of the target language and culture; the borders between countries in the world have even disappeared. As a result, English, as a second and a foreign language (SL/FL), has finally become an “*Internet lingua franca*” and its power has extended toward the “outer circle” people, L2 speakers of English (Kachru, 1985) faster than ever.

Various L2 readers, in particular, 750 million English as-a-second-or-foreign-language (ESL/EFL) learners have been frequently exposed to authentic reading materials on a computer monitor so that more self-instructional devices for authentic reading have been raised (The British council, 2000). For this demand, hypertext glosses have been developed and extensively integrated into authentic reading material necessary for L2 vocabulary acquisition and reading comprehension. A hypertext gloss has served particularly as a key component of vocabulary acquisition in self-instructional technology and web-based reading (Dunkel, Brill & Kohl, 2002).

A hypermedia or hypertext gloss refers to short definitions or explanations with nonlinearly linked data associated with text, graphics, audios, and videos in computerized text (Kommers, Grabinger and Dunlap, 1996). Its nonlinearity makes it possible to distinguish linear paper-based reading from online or electronic reading. It gives L2 readers more freedom to choose texts and references on their own. This electronic L2 reading device also allows readers to read more texts. Especially for low L2 ability learners, it is evident that this self-instructional assistive device enhanced with pictorial input facilitates L2 readers’ cognitive involvement in reading and helps them comprehend what they read.

Hudson (1982) supports that pictorial input increases comprehension of a reading passage, in particular with low proficient learners (Plass, Chun, Mayer and Leutner, 2003).

Accordingly, it well corresponds to the contemporary L2 educational trend, “communicative language teaching (CLT) with authentic material,” emphasizing language learners’ communicative competence and autonomy through technology (Canale and Swain, 1980; Hymes, 1971; Larsen-Freeman, 1986; Nunan, 1991).

Mayer has developed Generative Theory of Multimedia Learning (1997, 2001) based on Paivio’s (1971, 1991) Dual Coding theory (DCT), which has also led to the blossom of hypertext or hypermedia gloss research in CALL. Mayer (2001) suggests there exist two coding systems in our brain: visual and verbal. When comprehensible and high quality input enhanced with both visual and verbal representations is provided for L2 readers simultaneously, the readers are more likely to remember and retain the input better than text-only input in the brain. In other words, in order to help L2 learners better understand reading texts and vocabulary, and retain more information in the brain, comprehensible written (verbal) and pictorial (visual) input should be given to assist readers’ brain activity which consists of two separate but interrelated codes for information processing.

A number of reading researchers have emphasized that vocabulary learning is a key factor to develop reading comprehension; vocabulary learning has been actively researched over the past decades (Chun, 2006; Grabe, 1991; Leloup and Ponterio, 2003; Nikolova, 2004). Grabe (1991) argued that “virtually all L2 reading researchers agree that vocabulary development is a critical component of reading comprehension.” (p. 392) Vocabulary as a single factor appears to be one of the strongest predictors of L2 reading especially at the

lower grade levels (Schooner, Hulstijn & Bosser, 1998, reprinted from Grabe & Stroller, 2002). Chun (2006) also stated in his article that one critical topic that has been emerged and actively researched most in vocabulary acquisition via technology is the use of hypermedia or multimedia glosses. In other words, no matter how much vocabulary is needed for “reading threshold,” research supports that vocabulary acquisition is a basic but fundamental part for L2 learners to handle in order to become proficient readers.

Second language acquisition (SLA) studies have examined the value of modified interaction, “which refers to the learners’ interrupting their reading to receive help with vocabulary by clicking on unknown words in the written input.” (Chapelle, 2001, p. 71) This modified interaction allows L2 readers to access the definitions of difficult or unknown vocabulary with the help of hypertext glosses just by clicking a mouse, which leads them to have extra freedom to focus more on texts. It appears that modified interaction via technology-embedded hypertext glosses helps L2 readers read more with less look-up behaviors. With modified interaction, it is possible that L2 readers are more likely to have autonomy and access communicative and authentic reading material, depending on their interests and L2 levels.

1.2. Statement of the Problem

In order to make better comprehensible input (Krashen, 1994) within modified interaction, the integration of a hypermedia or hypertext gloss into L2 reading material has long appealed to CALL, L2 reading and SLA, especially to vocabulary acquisition because of its authenticity, salience and nonlinearity. In particular, the characteristics of nonlinearity

or “the networking of information units” associated with a variety of multimedia-embedded comprehensible input, have led to flourish more recent vocabulary acquisition studies than ever (Rouet, Levonen, Dillon, & Spiro, 1996). Nonlinearity, which is a distinct dimension, compared to linearity found in conventional paper-based texts, is multidimensional and allows readers to surf information relevant to their interests and needs, depending on their language proficiency and learning strategies. Accordingly, L2 readers are more likely to have had opportunities to read interactive texts on the computer monitor.

For this reason, there is a large body of literature that has reported the use of hypermedia or hypertext glosses in L2 computerized reading over the recent decades (Akbulut, 2007a; Al-Seghayer, 2001; Ariew, 2006; Aust, Kelly, and Roby, 1993; Khan; 1997; Koyama, & Takeuchi, 2004; Chun & Plass, 1996; Lomicka, 1998; Martinez-Lage, 1997; Nagata, 1999; Nikolova, 2004; Plass, Chun, Mayer & Leutner, 1998; Robin, 2007; Rogers, 1995; Rogers, 2000; Salem, 2006; Salem, & Aust, 2007; Smidt, & Hegelheimer, 2004; Yoshii, 2006; Yoshii, & Flaitz, 2002). In this research, one finding is that major foreign and second languages such as French, Spanish, English, German, and Japanese have been extensively involved and employed for hypertext gloss studies at a university level; however, little is known at a K-12 level.

Despite the fact that there are increased interests and an emerging literature in hypertext glosses, it is surprising that research has revealed inconsistent results, showing wide variation in the effectiveness of hypertext gloss use on L2 vocabulary acquisition. Chun (2006) argued that most results of hypermedia annotation research were quite mixed,

not providing decisive evidence, even though the research suggests promising combinations such as text, text + picture or text + audio.

For example, some studies of hypertext glosses showed that hypertext glosses have an overall effect on vocabulary acquisition and reading comprehension by increasing L2 learners' retention time and lessening their look-up behavior time (Abuseileek, 2008; Lomicka, 1998; Miyasako, 2006; Nagata, 1999; Nikolova, 2002). On the other hand, others indicated that the use of hypertext glosses might not directly affect or minimally influence L2 learners' vocabulary acquisition, even though reading with the help of a variety of hypertext glosses appears to have a positive impact on the L2 learners' perceptions, motivation and attitudes toward hyperlinked reading (Aust, Kelley & Roby, 1993; Gettys, Imhof, & Kautz, 2001; Levine, Bejarano, Carrell, & Vered, 2004; Lim, & Shen, 2006; Plass, Chun, Mayer, and Leutner, 1998, 2003; Sakar, & Ercetin, 2004).

With regard to research methods and technology integration, it is suspicious that some of the previous research study results were found inconclusive due to research reliability and validity issues. Others might have been due to the fact that researcher-created programs have their own limitations. Accordingly, there is a lack of systematic reviews to examine whether overall results of hypertext gloss studies are effective on vocabulary acquisition (Pearson, Ferdig, Blomeyer, & Moran, 2005; Taylor, 2006). From an L2 and CALL research perspective, it is now imperative that systematic meta-analysis from empirical evidence of previous studies should be conducted.

What remains to be explored for better research in hypertext glosses now is whether or not the use of hypertext glosses/annotations has a conclusive effect on L2 vocabulary

acquisition, and if positive, what specific types of hypertext glosses such as visual (picture, video and image), audio and text are effective in which environments with which types of technology use and research designs. Lomicka (1998) suggests that three variables should be clarified and further researched such as: 1) text type, 2) learner level and 3) outcome measures. Accordingly, more decisive results might be claimed when a generalizable research design has been made with robust outcome measures and applicable technology designs such as text types, interface design, and display of hypertext glosses. However, Chappelle (2001) also explained complexity in hypertext research, arguing that integrating individual studies into hypertext glosses might be complicated as follows:

...the summary of this growing body of research is difficult because of the variety of issues investigated, including preferences for various types of glosses (e.g., L1, L2, text, audio, image), influences on reading comprehension, and vocabulary acquisition, and the variety of research methods employed, including experimental and within-group designs as well as interaction analysis and think-aloud procedures.

1.3. Significance of the Study

The use of hypertext glosses in computerized reading has been, on the one hand, recognized as a key component across fields as different as L2 reading and vocabulary acquisition, SLA and CALL. A number of hypertext studies have been conducted over the past two decades. The study results have been, to some extent, evolutionary for hypertext

gloss users as well as researchers in terms of 1) no-gloss vs. gloss and 2) CALL L1 glosses vs. traditional L1 glosses groups in experimental designs. According to Taylor's meta-analysis (2006), there was a statistically significant difference between two conditions: a CALL L1 gloss group and a traditional L1 gloss group with a large effect size ($g = 1.09$) of the CALL L1 gloss group. That is, the CALL L1 gloss group outperformed the traditional L1 gloss group on a reading comprehension test.

On the other hand, the majority of the empirical research studies that have tested of which gloss features are most beneficial to L2 learners revealed fairly mixed results. For example, it is argued that too many hypertext features with video, audio and sound are not likely to enhance L2 learners' vocabulary acquisition and reading comprehension (Plass, Chun, Mayer and Leutner, 2003). L2 learners' learning styles, affective, language proficiency and technology preferences should be further taken into consideration to boost the potential effects of hypertext glosses. The reasons might be first found from research methodological issues: reliability and validity of researcher-created programs and research designs, test validity, and construct validity of vocabulary acquisition and reading comprehension. Moreover, very few studies have attempted to utilize a longitudinal research procedure to explore the long-term effectiveness of hypertext glosses on L2 vocabulary acquisition. In other words, the empirical evidence of the short-term quantitative studies might be more often than not misleading by inferential statistics of the multidimensionality: the relationship between the construct of vocabulary acquisition and research design.

The obvious question is how the multiple features of different types of hypertext glosses can be combined to optimize the effectiveness on L2 vocabulary acquisition. In order to answer this question more comprehensively, it appears better to synthesize the overall results of hypertext gloss studies.

1.4. Purpose of the Study

The primary goals of this meta-analysis study were to (1) examine effects of hypertext glosses on L2 vocabulary acquisition, (2) synthesize which features of effective technology use and research design have been employed, and, more importantly, (3) find out which particular combination of hypertext glosses (text-only vs. text + visual) is more effective and beneficial on L2 vocabulary acquisition. A dependent variable was outcome measure of vocabulary test scores and an independent variable was types of hypertext glosses. As a research methodology, meta-analysis was conducted to triangulate all data across empirical studies and synthesize a weighted standardized mean effect size (Lipsey, & Wilson, 2001) that calculates a corrected standardized mean difference effect size of between a control (text-only) group and an experiment (text + visual) group. Pedagogical implications were discussed for future research.

1.5. Research Questions:

1. Does a group with access to multiple glosses (text + visual) perform significantly better than a group with access to a single gloss (text-only) on a post vocabulary test?

2. What are the features of meta-analyzed studies regarding characteristics of studies, research methodologies and technology programs?
3. What are some potential moderators to systematically account for the between study variation of these meta-analyzed studies?

Chapter Summary

Chapter I began with background of the present meta-analysis study and the statement of problem. The purpose of the study with research questions was followed to give clear ideas of why the present study should be conducted.

Chapter II will provide the nature of meta-analysis and the pertinent literature review of hypertext gloss studies with current examples of hypertext glosses illustrated in empirical studies.

CHAPTER II

LITERATURE REVIEW

2.1. Meta-analysis

Meta-analysis is a quantitative method to synthesize empirical studies conducted for selected domains, compared to narrative literature reviews, which heavily rely on the results of statistical significance for evaluating and comparing studies. It is also comprehensive and “qualitative” data-analysis across experimental or quasi experimental studies that have been previously conducted in a certain field. Collected database information from empirical studies enables a meta-analyst to unveil insightful and potential benefits that have not found in previous research attempts.

Glass (1976) defined that meta-analysis is the statistical analysis procedure of a large collection of analysis results for the purpose of integrating the findings. Rosenthal (1995) argued that “meta-analytic reviews are quantitative summaries of research domains that describe the typical strength of the effect or phenomenon, its variability, its statistical significance, and the nature of the moderator variables from which one can predict the relative strength of the effect or phenomenon.” (p. 183)

Accordingly, Lipsey & Wilson (2001) characterize what meta-analysis can do as follows:

The systematic coding procedures of meta-analysis and the construction of a computerized database to record the resulting information have almost unlimited capability for detailed database information from each study and covering large numbers of studies that could be ignored by themselves. (p. 6)

They also illustrate the advantages of meta-analysis as follows:

1. Meta-analysis procedures impose a useful discipline on the process of summarizing research findings.
2. Meta-analysis represents key study findings in a manner that is more differential and sophisticated than conventional review procedures that rely on qualitative summaries or “vote-counting” on statistical significance.
3. Meta-analysis is capable of finding effects or relationships that are obscured in other approaches to summarizing research.
4. Meta-analysis provides an organized way of handling information from a large number of study findings under review.

2.2. Meta-analysis and CALL

With regard to recent meta-analysis studies in technology and second language learning, a relatively few meta-analysis studies have attempted to investigate 1) whether technology use has affected L2 language learning (Pearson, Ferdig, Blomeyer and Moran, 2005; Zhao, 2003) and 2) whether computer-mediated glosses have had an effect on reading comprehension and vocabulary acquisition (Taylor, 2006; Abraham, 2008). The

compounding results of these meta-analysis studies broadly revealed that technology integration has been successful and effective on 2 reading comprehension and vocabulary acquisition. In particular, Abraham (2008) meta-analyzed 11 studies of computer-mediated glosses on second language reading comprehension and incidental vocabulary learning, reporting that a large effect size was found on between an experimental group with access to computer-mediated text glosses and a control group without access to these glosses.

However, none of them has researched the effects of two particular conditions – text-only and text + visual hypertext glosses – on L2 vocabulary acquisition in terms of Mayer’s Multimedia Learning Theory.

2.3. Theoretical Background

Based on Paivio’s dual coding theory (DCT), Mayer (1997, 2001) has further developed Generative Theory of Multimedia Learning that illustrates how both pictorial and written input collaboratively enhance L2 readers’ reading comprehension and vocabulary learning as shown in Figure 1. This model emphasizes the importance of readers’ integrative learning ability to enhance vocabulary acquisition with multimedia input by connecting two verbal and visual systems with written and pictorial cues in the brain. That is, better vocabulary learning with multimedia input is more likely to take place when L2 learners are cognitively capable of dealing with both written and pictorial information at the same time.

Plass, Chun, Mayer and Leutner, (2003) stressed the importance as follows:

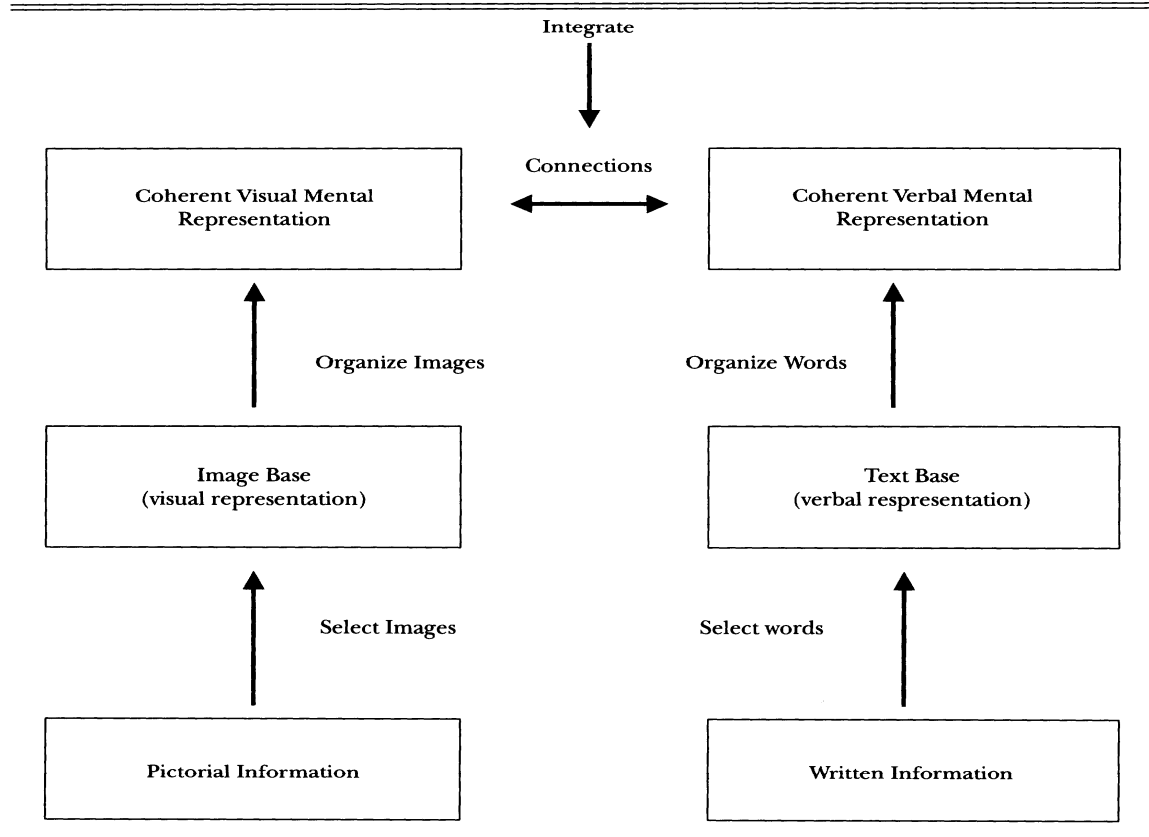
The learner must first select relevant verbal information from a text and visual information from an illustration and then construct a text base in a coherent verbal mental representation and the visual information in the image base into a coherent visual mental representation. Then, the learner must integrate the newly constructed verbal and visual representations by creating connections between the corresponding visual and verbal information. (p. 223)

This multimedia learning theory has been further researched on the aspects of learners' learning styles, preferences and language proficiency such as "perceptual learning styles – different sensor preferences for processing information" and verbal or spatial ability (Yeh and Wang, 2003; Plass, Chun, Mayer and Leutner, 2003). Interestingly enough, low-proficiency learners are less likely to take advantage of multimedia learning environments due to high cognitive load when given two types of annotations for vocabulary learning; multimedia input does not always enhance all learners' performance on vocabulary learning. Thus, research indicates that this fact should be taken into consideration when multimedia is implemented into vocabulary learning.

Figure 1

Mayer's Generative Theory of Multimedia Learning

FIGURE 1
Mayer's (1997, 2001) Generative Theory of Multimedia Learning.



Reprinted from Plass, Chun, Mayer and Leutner (2003)

2.4. Defining a Gloss and a Hypertext Gloss

Traditionally, a 'gloss' refers to short definitions or explanations of the meanings of words at the bottom or sides of a text in order to support learners' reading comprehension (Nation, 1983; Pak, 1996; Lomicka, 1998). This definition might be "*a loose term*" by Roby's taxonomy of glosses (1999). In his article of "What's in a gloss?" it is well illuminated that "glosses are many kinds of attempts to supply what is perceived to be deficient in a reader's procedural or declarative knowledge." (p, 96)

Table 1

Taxonomy of Glosses (Roby, 1999)

- I. Gloss authorship
 - A. Learners
 - B. Professionals
 - 1. Instructors
 - 2. Materials developers
 - II. Gloss presentation
 - A. Priming
 - B. Prompting
 - III. Gloss functions
 - A. Procedural
 - 1. Metacognitive
 - 2. Highlighting
 - 3. Clarifying
 - B. Declarative
 - 1. Encyclopedic
 - 2. Linguistic
 - a. Lexical
 - i. Signification
 - ii. Value
 - b. Syntactical
- IV. Gloss focus
 - A. Textual
 - B. Extratextual
- V. Gloss language
 - A. L1

<ul style="list-style-type: none"> B. L2 C. L3
<ul style="list-style-type: none"> VI. Gloss form <ul style="list-style-type: none"> A. Verbal B. Visual <ul style="list-style-type: none"> 1. Image 2. Icon 3. Video <ul style="list-style-type: none"> a. With sound b. Without sound C. Audio (only)

The term “gloss” “has a more comprehensive meaning. In particular, Roby (1999) specified how glosses can function depending on readers’ knowledge- *procedural* and *declarative*. For example, in terms of gloss functions, glosses not only give linguistic and definitive explanations but also allow readers to consider their deeper metacognitive action of whether or not they are actively reading what they are supposed to read. Table 1 shows Roby’s (1999) taxonomy of glosses in detail.

In terms of the definition of ‘hypertext’, there have been some incomplete explanations over the past decades; however, according to Ted Nelson (1983), “Hypertext is a term for forms of hypermedia, human-authored media that branch or perform on request, that operate textually. Examples include the link-based ‘discrete hypertext’ (of which the Web is one example) and the level-of-detail-based ‘stretchtext’.” Kommers, Grabinger & Dunlap (1996) suggested that “hypertext or hypermedia refers to computer-

based applications that provide information in a nonlinear way through multiple types of resources such as text, graphics, sound, video, and animation.” (p.23) This type of hypertext or hypermedia is very different from traditional or conventional paper-based glossaries in terms of interactivity and nonlinearity to consult words with the help of a variety of comprehensible modes. Since the definition of a hypertext annotation has been debated, Roby (1999) enumerated as follows:

Adjunct aids (Otto & White, 1982), metanotes (Wolfe, 1990), metatext (Lantolf, Labarca, & den Tuinder, 1985), and paratext (Genette, 1987)... Oxford (1995) provides many possibilities under the rubric of assistance: error correction...a pictorial representation of a verbal expression...a cooperative learning activity...an encouraging word at just the right moment. (p. 366)

Stewart and Cross (1991) stressed, “key point and vocabulary glosses represent important statements or provided brief definitions of words.” (p. 6)

Overall, hypertext glosses refer to short definitions or explanations with nonlinearly linked-data associated with graphics, audios, and videos in computerized texts. The potential of hypermedia or hypertext glosses is considered very influential for facilitating L2 learners’ vocabulary learning and reading comprehension (Chun & Plass, 1996). A number of research studies have been conducted to support the use of images in a variety of ways. Visual images have been found effective because they help build L2

learners' background knowledge and schemata appropriate to target texts, and facilitate the contextualization of what is being read (Omaggio, 1979).

2.5. Hypertext Gloss Examples

From one of the definitions mentioned above, for example, Aust, Kelley & Roby (1993) initially conducted a research study of the use of hyper-reference and conventional dictionaries in Spanish, using a technology-embedded electronic book with hyper-reference as shown below in Figure 2:

Figure 2

Screen Shot of Electronic Book with Hyper-Reference

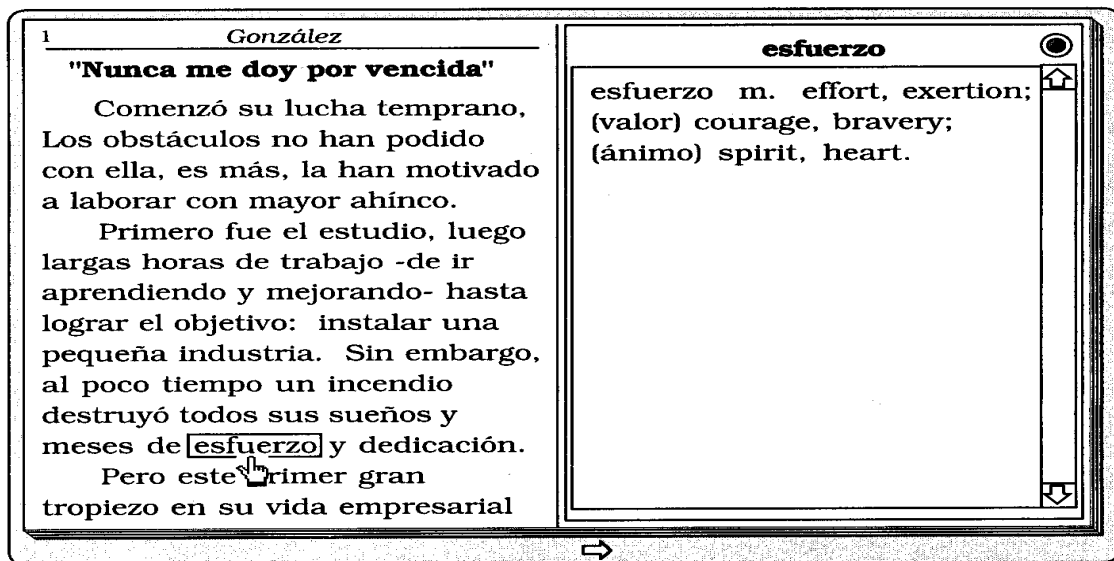


FIGURE 1 □ Electronic Book with Hyper-Reference

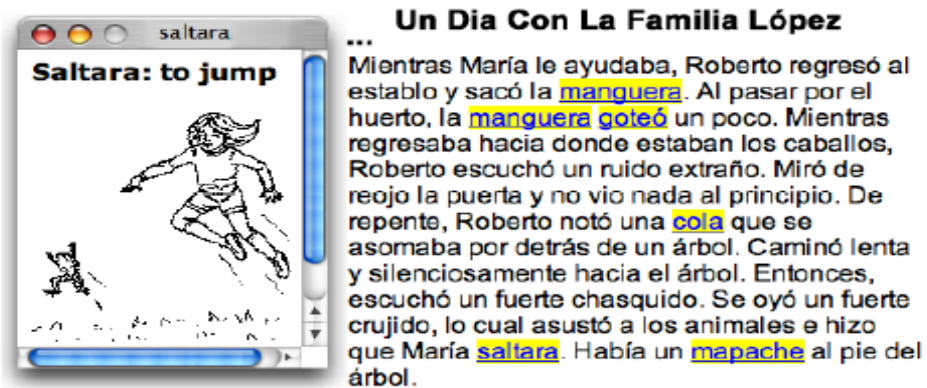
Aust, Kelley & Roby (1993)

When L2 learners click on the unknown vocabulary that they encounter while reading, the meaning comes out with a separate window at the right side so that the learners can differentiate between the text they are reading and the word they want to look at. In this example, researchers tried to create an electronic but paper-looking book with glosses, focusing more on text glosses that are similar to paper ones. In the latest version (2007), the type of hypertext glosses are enhanced with visual and audio input. For example, as shown in Figure 2, L2 learners can listen to the pronunciation of unknown words with illustrations. Yoshii (2006) made a similar gloss that Aust, Kelley & Roby, 1993 used, trying to show how active verbs can be integrated with a pictorial gloss in Figure 3.

Figure 3

Screen Shot of An Electronic Gloss with Picture

Figure 1: An Electronic Gloss with Picture



Note: The gloss window does not cover story text. In audio conditions the word is spoken after the user clicks the word.

Salem & Aust (2007)

Figure 4

Screen Shot of Gloss Types



Figure 2. Gloss Types

Yoshii (2006)

Chun & Plass (1996) in their self-developed program, *CyberBuch* illustrated how vivid pictorial representations can enhance text itself. In Figure 5, it is evident that “a picture can tell a thousand texts.” Just as storytelling can make text reading more fun, pictures depicting words create more interesting outlook for L2 readers, especially low proficient readers.

Figure 5

Screen Shot of *CyberBuch*

Screen shot of *CyberBuch* showing selection of word *droht* (threatens) and available annotation types



APPENDIX B

Screen shot of *CyberBuch* showing picture depicting word *droht*



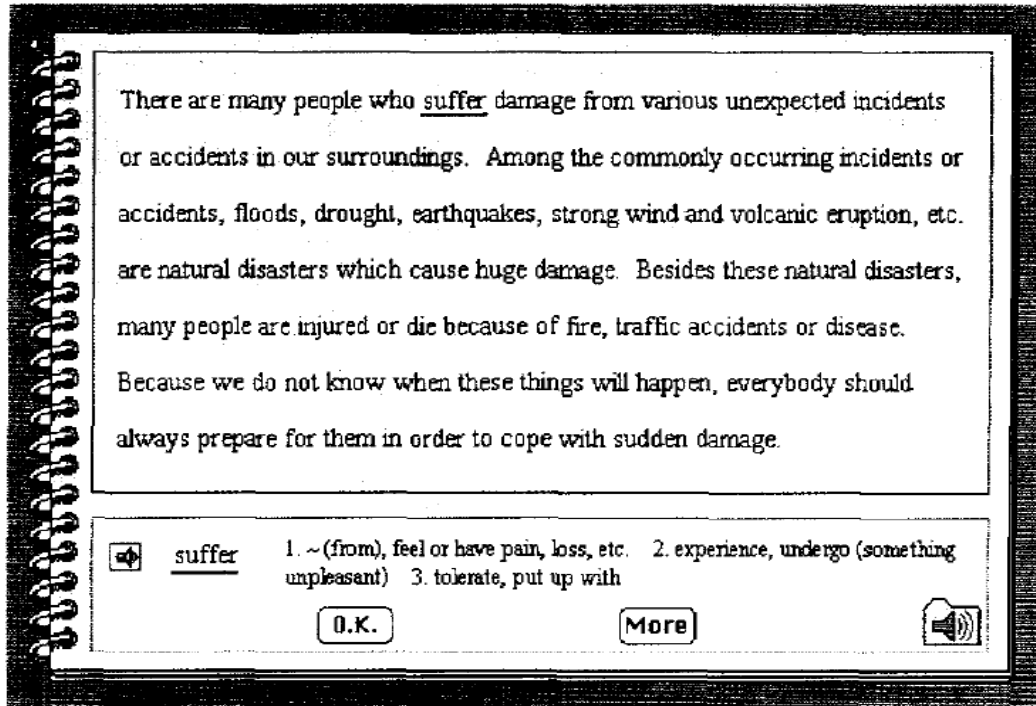
Chun & Plass (1996)

As some other examples of hypertext glosses, hypertext glosses can be located at the bottom of the reading passage similar to traditional paper glosses so that L2 readers can be less confused to use the glosses. Son (1998) represented well in his hypertext gloss-based reading interface in Figure 6.

Figure 6

Screen Shot of Hypertext-based Courseware

Screen Layout from a Card of Hypertext-based Courseware

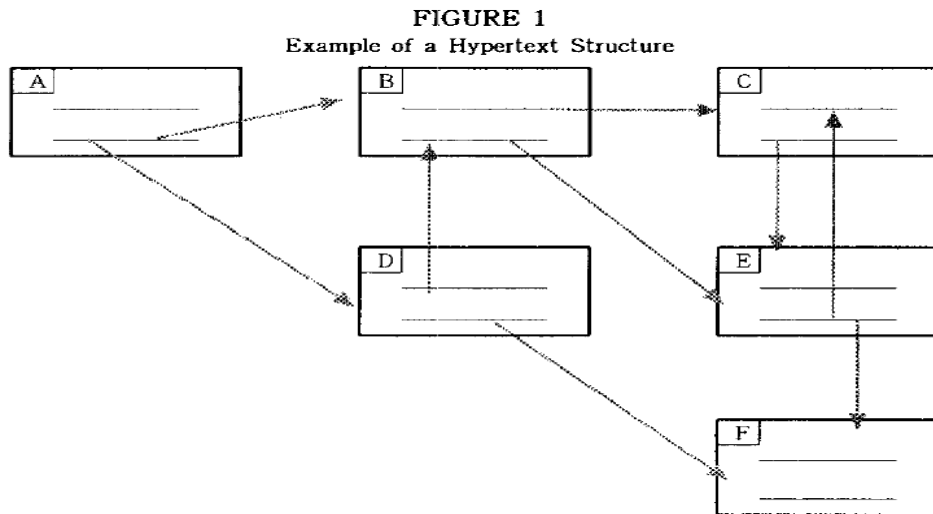


Son (1998)

In addition, Son illustrated how hypertext structures interactively work in the modified model as shown in Figure 7 and 8. Unlike conventional paper glosses, hypertext glosses are nonlinearly linked to one another so that it is not necessary for L2 readers to consult in a word-by-word process. It allows more freedom, time-saving and interactivity while reading.

Figure 7

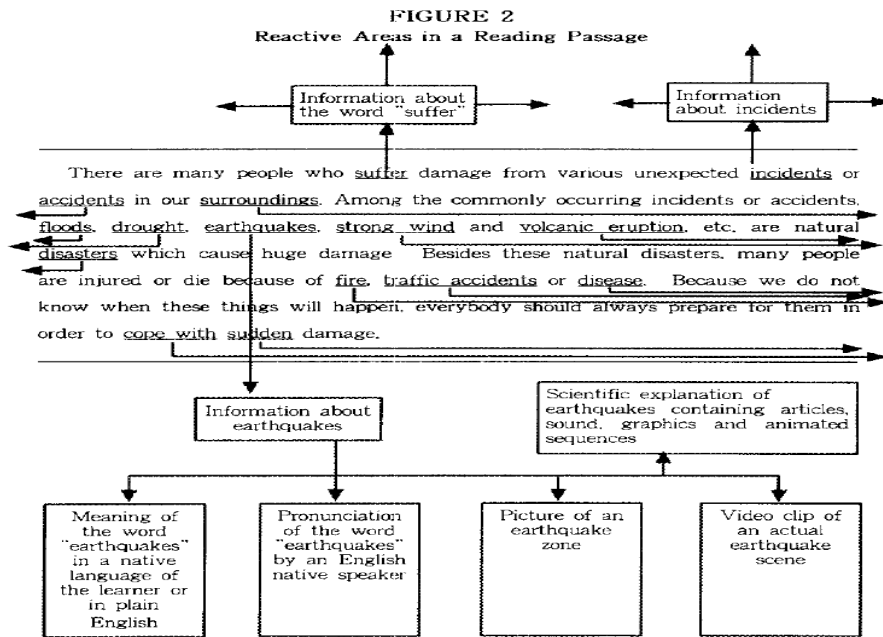
Screen Shot of Example of a Hypertext Structure



Son (1998)

Figure 8

Screen Shot of Reactive Areas in a Reading Passage

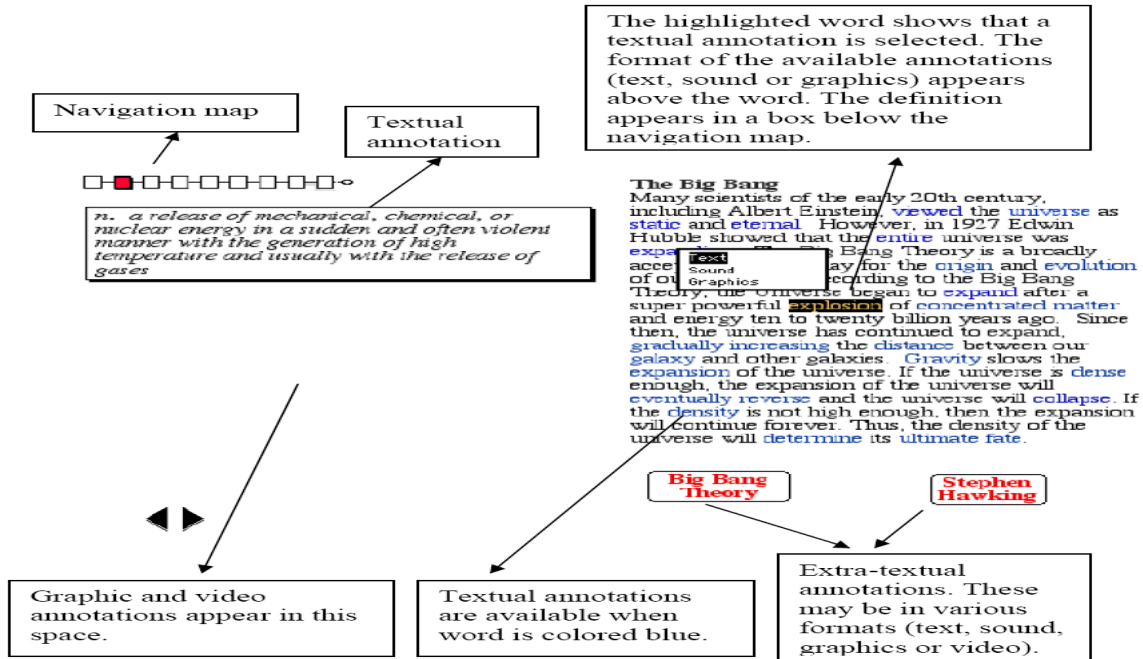


Son (1998)

As technology evolves faster than ever, hypertext glosses are sophisticated with more technology features. Ariew and Ercetin (2004 & 2005) created more learner-centered interface of hypertext glosses, giving separate but select annotations with multimedia to L2 readers in Figure 9 and 10. Regarding the levels of L2 readers, hypertext glosses can be used to help their reading or reduce their cognitive load while reading.

Figure 9

Interactive Hypertext Example 1



Ariew & Ercetin (2004)

Ariew and Ercetin (2004)

Figure 10

Interactive Hypertext Example 2

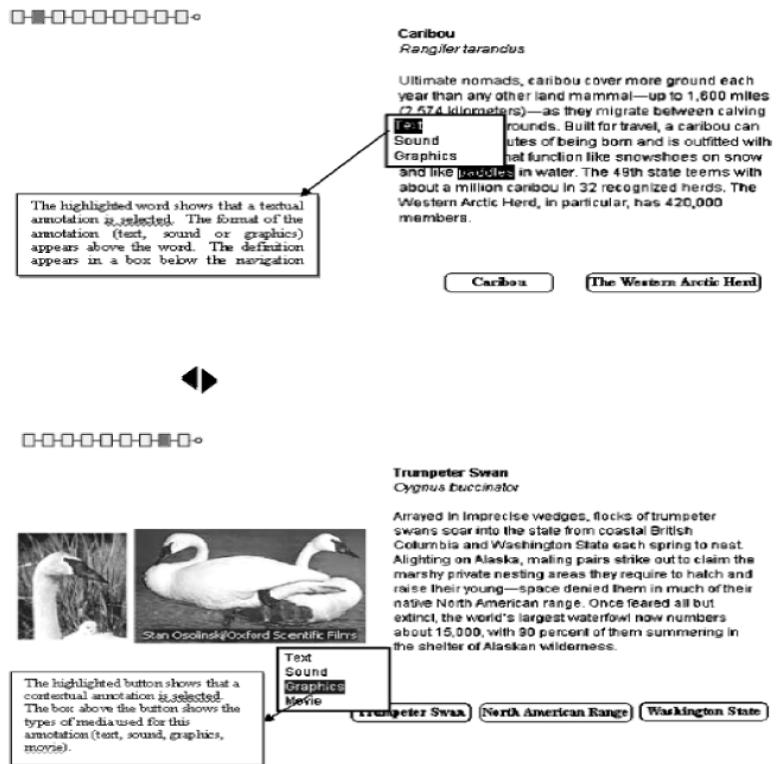


Fig. 2 A screen shot showing a textual annotation.

Ariew and Ercetin (2005)

Hypertext glosses can be easily integrated into authentic materials ideal for L2 readers on the Web or Internet, which empowers extensive readers to acquire more information outside the classroom. It is more common to encounter hyperlinked reading materials on the Web or Internet. When L2 readers encounter difficult or unknown words in their computerized reading, they can be immediately assisted with multimedia-based hypertext or hypermedia glosses with authentic pictorial and audio input as shown in Figure 11.

Figure 11

Screen Shot of Main Course Window

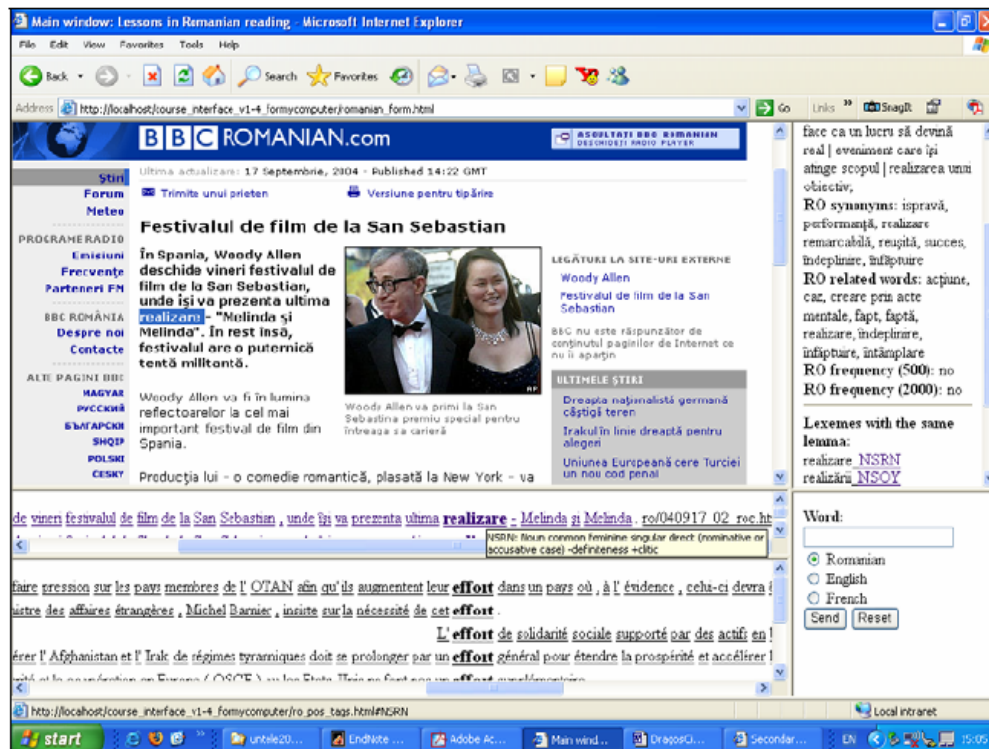


Figure 4: Main course window

Ciobanu, Hartley & Sharoff (2006)

2.6. Hypertext Glosses and SLA

The two key elements of vocabulary acquisition are evident in second and foreign language (L2) reading studies: (1) comprehensible input and (2) exposure to authentic materials of the target language and culture, which leads to comprehensible output. In terms of input theories, Krashen (1985) strongly emphasizes the importance of “*comprehensible input*,” asserting that “one acquires language in only one way- by exposure to comprehensible input. If the input contains forms and structures just a little beyond the

learner's current level of competence in the language ($i + 1$), then both comprehension and acquisition will occur." (reprinted from Lightbown & Spada, 1999, p. 39) Thus, a number of researchers have extensively investigated how comprehensible input should be made to help L2 learners better acquire what they read in CALL (Chapelle, 2005; Kon, 2002; Plass, Chun, Mayer & Leutner, 1998). The result consistently indicates that comprehensible input-rich environments allow L2 learners to acquire more vocabulary in the "*natural environment*" where native speakers of the target language communicate and interact with one another. It is imperative that the *natural environment* is not meant as a decontextualized form-focused environment but a more input-enhanced meaning-focused one.

In addition to comprehensible input for reading, Swain (1985) underscores that the comprehensible input, eventually, leads L2 learners to produce more comprehensible output. That is, comprehensible input increases not only L2 learners' reading skills, but it also enhances speaking skills. In this respect, reading authentic materials through technology-enhanced comprehensible input makes it possible for L2 readers, who are away from the target language and culture, to access the authentic language and culture, overcome language and cultural barriers, and eventually enhance overall language skills (Kim, 2001).

Plass and Jones (2005) also stressed three important factors for language acquisition: comprehensible input, interaction and comprehensible output, defining "second language acquisition with multimedia is the use of words and pictures designed to support the comprehensible input that the learner is exposed to and interacts with, and to elicit and negotiate comprehensible output." (p. 469)

2.7. Hypertext Glosses and Extensive Reading

Extensive reading refers to self-interested or free voluntary reading for readers to find reading materials, depending on their own language proficiency levels and understanding (Day & Bamford, 1998; Krashen, 1993). The notion of extensive reading in second language education has been widely proposed as an ideal way for L2 learners to be independent and lifelong readers in a large body of literature (Day & Bamford, 1998; Grabe, 1991; Grabe & Stoller, 2002; Krashen, 1993, 2004; Palmer, 1969; Simensen, 1987). The emphasis of extensive reading has been on “reading for fun,” so that interesting reading materials are the priority of the selection. This differentiates extensive reading from “conventional” reading for study. For this reason, L2 readers are more encouraged to find authentic and interesting reading materials through all resources, in particular, through the Web or Internet.

On the other hand, the selection of appropriate, interesting, and authentic reading materials, and constructing extensive reading libraries inside the classroom or school has caused such problems as infrastructure, funding and time issues. In addition, dictionary use is discouraged because reading materials are well within the linguistic competence of the readers in terms of vocabulary and grammar. Dictionaries are rarely used while reading because the constant stopping to look up words makes fluent reading difficult (Day & Bamford, 1998, p. 8). In order to reduce this burdensome look-up behavior and focus more on meaning in reading, hypertext glosses as a comprehensible input in SLA can be used to help readers acquire enough vocabulary

to comprehend reading texts. For instance, hypertext glossed reading via technology or on the Web allows L2 learners to take full advantage of extensive reading while they explore and enjoy interesting reading materials at their own levels and pace. With the aid of hypertext glosses, consisting of multimedia-based input such as text, images, sound and video, language learners are more capable of conquering main reading obstacles such as a “look-up behavior” and difficult vocabulary. By clicking a mouse on hypertext glossed words, L2 readers have extra freedom to focus on meaning.

2.8. The Studies of Hypertext Glosses

As a fundamental and essential query, Chun and Plass (1996) questioned as “*How effective are annotations with different media types for vocabulary acquisition?*” (p. 183) They thoroughly examined the effectiveness of multimedia-based annotations associated with pictures and videos, using *CyberBuch*, a multimedia application for German reading texts. 160 second-year German students at three Universities in the United States were measured with different types of hypertext annotations: (1) text definition, (2) text + picture, and (3) text + video. With these 3 studies conducted in different time periods, the results indicated that the group, which consulted the combination of text + picture annotations, significantly outperformed two other groups who consulted text definition and text + video on a vocabulary test while no significant difference was found between the text definition annotation group and the text + video annotation group.

Nagata (1999) investigated the effectiveness of two types of hypertext glosses: (1) a single-gloss that provides a single English (L1) translation and (2) a multiple-choice gloss

with both English (L1) and Japanese (L2). 26 university students taking a Japanese course were measured by a vocabulary pretest and a posttest. The multiple-choice-gloss format group ($M= 13.5$, $SD= 5.5$) outperformed significantly better on a vocabulary posttest than the single-gloss format group ($M= 10.8$, $SD= 4.8$). The result indicates that a multiple-choice gloss format was significantly more effective than a single-gloss format. In the similar vein, Miyasako (2002) also found that an L2 multiple-choice gloss group outperformed a L1 simple gloss group on a vocabulary test.

Yoshii (2006) examined the effects of L1 and L2 glosses on incidental vocabulary acquisition in a multimedia environment. 195 university students learning English as a foreign language were divided into four gloss groups- 1) L1 text only, 2) L2 text only, 3) L1 text + picture and 4) L2 text + picture- and measured by two vocabulary posttests: an immediate test and a two-week delay test. The results show that there was no significant difference between L1 and L2 gloss groups; however, a significant difference between a text + picture group and a text-only group was found only on a definition-supply test.

Lomicka (1998) conducted a study with 12 native speakers of English in undergraduate-level French classes, in which they read a poem in French (L2) while thinking aloud in English (L1). The students were randomly assigned to one of three groups: (1) no access to glosses; (2) access to all glosses of definitions in French and translations in English; (3) access to multiple glosses (definitions, images, pronunciation and translations in English). The results indicate that statistical differences between three groups were not found through think-aloud protocol data even though the students appeared to learn more vocabulary when they chose from a variety of assistive multiple annotations

or glosses. This result supports Mayer's (1997) premise of the Generative Theory of Multimedia Learning.

Aust, Kelley & Roby (1993) examined the magnitude of the relationship between hyper-reference glosses and paper-based glosses with 80 university students taking a fifth-semester university Spanish course in the United States. The 80 participants were divided into four treatments: (1) an electronic article with a bilingual hyper-reference dictionary, (2) an electronic article with a monolingual hyper-reference dictionary, (3) a paper article with a bilingual paper dictionary and (4) a paper article with a monolingual paper dictionary. The mean number of propositions recalled (comprehension) was not statistically significant between the hyper-reference dictionary group ($M= 10.95$) and the paper dictionary group ($M= 12.65$). Reading comprehension was not also significant between the users of bilingual dictionaries ($M= 12.45$) and the users of monolingual dictionaries ($M= 11.15$) even though the hyper-reference group consulted vocabulary and references per minute two times more than the conventional paper group did regarding consultation frequency and efficiency (consultation per minute).

Sakar & Ercetin (2004) conducted a study with 44 (26 males and 18 females) intermediate Turkish students studying English for academic purposes (EAP) at a Turkish university. The study explored two inquiries: 1) whether EAP students prefer hypermedia annotations and 2) whether hypermedia annotations eventually facilitate reading comprehension of EAP students. The results show that the learners preferred visual annotations significantly more than textual and audio annotations; however, reading comprehension was negatively correlated with the frequency of access to annotations ($r= -$

0.42) and the amount of time spent on annotations ($r = -.42$). Especially, it seems that pronunciations, audio-recordings, and videos negatively affected reading comprehension of the participants.

The results of Yeh & Wang's research (2003) also showed that the significance of hypertext annotation use in EFL and vocabulary learning has been influential but inconclusive. Although both text-only and text + still picture groups outperformed a text + still picture + audio group, the compelling result of the second group (text + still picture) was not statistically significant compared to the text-only group.

In sum, the results of the previous studies above point out that hypertext glosses, which include a variety of verbal and pictorial information, appear to increase L2 learners' interests and motivation of L2 reading; texts associated with pictorial representations rather than paper-based linear information interest L2 readers. Nevertheless, whether different types of hypertext glosses enhance vocabulary acquisition of L2 learners was somewhat inconclusive. (see Table 2 for more information)

Table 2
Summary of Hypertext Gloss Studies

Study Author	Research Method	Technology used	Target Language	General Findings
* Akbulut (2007a)	Within subject repeated ANOVA	NA	English	Significant difference on vocabulary tests between text-only and text+visual gloss groups; however, no significant difference on reading comprehension
* Al-Seghayer (2001)	Within subject repeated ANOVA	NA	English	Significant difference between text-only and text+video and text+picture; however, no difference between text-only and text+picture
Aust, Kelly and Roby (1993)	ANOVA	Researcher-invented program	Spanish	Negative between the hyper-reference dictionary and the paper dictionary group
* Chun and Plass (1996)	Within-subject/repeated-measures	<i>CyberBuch</i> , a hypermedia application for reading texts	German	Significant difference between text+visual vs. text-only
Lomicka (1998)	N/A	Researcher-invented program	French	Negative with no-gloss, single-gloss and multiple-gloss
Miyasako (2002)	N/A	Researcher-invented program	Japanese	Significant difference between multiple-choice gloss with both English (L1) and Japanese (L2) vs. a single gloss in English (L1)
Nagata (1999)	N/A	Researcher-invented program	Japanese	Significant difference between multiple-choice gloss with both English (L1) and Japanese (L2) vs. a single gloss in English (L1)

Sakar & Ercetin (2004)	Within-subject/repeated-measures ANOVA	Researcher-invented program	English	Negative correlated with reading comprehension
Salem & Aust (2007)	ANOVA	Researcher-created courseware	Spanish	“Gloss users had significantly higher reading comprehension and vocabulary acquisition scores than non-gloss users,” however, no significant difference between text-only and text+visual+audio
* Yeh & Wang (2003)	ANOVA	Researcher-created courseware	English	Negative between text-only and text+still picture
* Yoshii & Flaitz (2002)	Within subject repeated ANOVA	BANAI READINGS	English	Significant difference between text-only and text+picture
* Yoshii (2006)	Mixed design repeated measure	Researcher-invented program	English	Significant difference between text+picture vs. text-only

* A study used for this meta-analysis

Chapter Summary

Chapter II provided a theoretical rationale, a pertinent literature review and definitions of meta-analysis and hypertext glosses with a variety of authentic examples. It also discussed how different types of hypertext glosses can be integrated into an extensive L2 reading program in order to enhance readers' vocabulary acquisition. The results of the previous hypertext gloss studies discussed showed somewhat inconclusive results in terms of a research method, technology use, target languages and research findings. Thus, meta-analysis is needed for comprehensible insights of hypertext gloss studies on L2 vocabulary acquisition.

Chapter III will provide the procedure of meta-analysis and how it will be conducted in details of the selection of publication, criteria for inclusion and descriptions of coding characteristics.

CHAPTER III

METHOD

3.1. Meta-Analysis Statistics

In order to find out the particular magnitude of effects of hypertext glosses use on L2 vocabulary acquisition, two-variable group contrasts – a treatment (text + visual) and a control (text-only) groups – were applied in the present meta-analysis . These two-variable group contrasts involved an independent variable of hypertext glosses use that was measured on a dependent variable of vocabulary tests, in particular with the combination of both text-only and text + visual hypertext glosses in an experiment or quasi-experiment design. Regarding the effect size statistics, Cohen's d was applied because the operationalization of the meta-analyzed studies varied across instruments (vocabulary tests), research design, samples sizes, technology use (program or software), languages and time on task.

In Formula 1, ES_{sm} (Cohen's d) represents a standardized mean difference effect size. \bar{x}_{G1} is the mean for Group 1 and \bar{x}_{G2} is the mean for Group 2; s_p is the pooled standard deviation.

Formula 1

Cohen's D

$$ES_{sm} = \frac{\bar{x}_{G1} - \bar{x}_{G2}}{s_p} \quad (1.1)$$

$$S_p = \sqrt{\frac{(n_{G1} - 1)S_{G1}^2 + (n_{G2} - 1)S_{G2}^2}{(n_{G1} - 1) + (n_{G2} - 1)}} \quad (2.2)$$

A relatively small sample size of this meta-analysis resulted in selecting Hedges' g correction instead because Lipsey and Wilson (2001) stressed the importance of the unbiased effect size statistics as follows:

“Under such circumstances, it is best to estimate the effect size using only the standard deviation of the control group since it is presumably unaffected by the treatment and, hence, a better estimate of the respective population variance.” (p. 49)

With regard to a small sample bias, Cohen's d was used to estimate Hedges' g simple correction for corrected and unbiased effect size statistics as follows:

Formula 2

Hedge's G

$$ES'_{sm} = \left[1 - \frac{3}{4N - 9} \right] \frac{\bar{x}_{G1} - \bar{x}_{G2}}{s_p} \quad (2.1)$$

$$SE_{sm} = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{ES_{sm}^2}{2(n_{G1} + n_{G2})}}$$

$$SE_{sm} = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{ES_{sm}^2}{2(n_{G1} + n_{G2})}} \quad (2.2)$$

$$W_{sm} = \frac{1}{SE_{sm}^2} = \frac{2n_{G1}n_{G2}(n_{G1} + n_{G2})}{2(n_{G1} + n_{G2})^2 + n_{G1}n_{G2}(ES'_{sm})^2} \quad (2.3)$$

(Lipsey & Wilson, 2001, p. 49)

Where N is the total sample size ($n_{G1} + n_{G2}$), ES_{sm} (Cohen's d) is the biased standardized mean difference as shown in Formula 1. n_{G1} is the number of subjects in Group 1, and n_{G2} is the number of subjects in Group 2. ES'_{sm} , SE_{sm} and W_{sm} refer to a corrected or unbiased standardized mean effect size (Hedge's g), a standard error of Hedge's g and an inverse variance weight of Hedge's g , respectively. An inverse variance weight was applied because a larger standard error corresponds to a less precise effect size value, the actual weights are computed as the inverse of the squared standard error value.

3.2. A Random-effects Model

A random-effects model makes it possible for researchers to detect potential moderator variables to account for systematic between-study variation because this model allows not only within-subject level sampling error but also more variation from between-study level error that represents other sources of variability assumed to be randomly distributed. In other words, this model enables meta-analysts to utilize all sources from a

variety of characteristics of studies, coding and effect sizes in order to explain meta-analysis variability.

As rules of thumb for effect size magnitude, Cohen (1988) reported how effect size magnitude should be interpreted: when the effect size is less than .20, a treatment effect is small while a treatment effect is large when bigger than .80.

Small	Medium	Large
ES \leq .20	ES = .50	ES \geq .80

Overall, it is relatively straightforward to estimate and interpret effect sizes that can be easily computed from empirical descriptive statistics (M , SD and F ratios) in research studies. In addition, effect size statistics allow readers to understand what an entire meta-analyzed study tries to attempt.

3.3. Location and Selection of Publication

After reviewing the literature, key word searches were extensively conducted by using *hypertext or hypermedia gloss, electronic gloss, multimedia annotation, incidental vocabulary learning, reading education, computer-based learning, multimedia, second language learning and teaching and educational technology*, etc (see Appendix B). Main databases used are as follows:

1. **General online search engines:** Linguistics and Language Behavior Abstract (LLBA), *ERIC, ProQuest, DBPIA and Google Scholar.*

2. **Academic and educational online or paper journals searches:** *ACTFL, Academic ASAP, CALICO, SpringerLink, , Ingenta Select, JSTOR, Educational Technology Research and Development, Journal of Educational Technology and Society, Language learning & Technology, Foreign Language Annals, Reading in a Foreign Language, Wilson OmniFile full text select, SAGE Journal Online, TESOL Quarterly, and The Modern Language Journal.*
3. **International journals and databases:** *Asian TEFL, Asian EFL, CALL-EJ online Journal, PacCALL, APACALL, IALLT, and CALL.*

From this first filtering process, selected articles, papers, book chapters, presentation reports and unpublished dissertations were initially analyzed for the inclusion of the present meta-analysis. In the second filtering process, the below inclusion criteria were thoroughly applied to reanalyze the first selected empirical studies. Especially, two of the main criteria were: 1) a study should have both independent variables (text-only and text + visual) and a dependent variable of vocabulary test scores and 2) the outcome measure should group contrasts – a treatment (text + visual) group and a control (text-only) group.

3.4. Criteria for Inclusion

The eligible studies should meet the particular criteria as follows:

1. A hypertext or hypermedia gloss as a key variable should have been included for vocabulary acquisition.

2. Independent variables are different types of hypertext glosses (text-only vs. text + visual), and a dependent variable is vocabulary tests scores.
3. A text + visual hypertext gloss group should have included image, videos, pictures or other visual presentations compared to a control group with a text-only hypertext gloss.
4. Outcome measure should have had group contrasts- a treatment group and a control group.
5. Information should have been sufficient enough for calculating the effect sizes – means, standard deviations, *F* ratios, *t*-values and standardized values.

3.5. Descriptions of Coding Characteristics

There were three major characteristics as shown in Table 3. First, study characteristics included 21 variables such as descriptive statistics and study information. In order to in depth analyze the data available, learner variables such as ethnicity, GPA, gender and first language (L1) were also analyzed if applicable. Secondly, effect size characteristics consisted of effect size types and numbers that were used for calculation. Especially for better understanding meta-analysis, descriptive statistics including a sample size (*N*), a mean (*M*) and a standard deviation (*SD*) were mainly utilized. Regarding technology characteristics, types of software programs and authors were also included.

Table 3
Coding Characteristics

Major Category	Brief Description of the Major Category	No. of Variables	Variables
Study characteristics	Descriptive data about the study	21	Study ID number Author Types of publication The publication year First Language Target population Mean age of sample Students' Ethnicity (RACE) Study years of the target language The sample's mean GPA The number of males The number of females Type of research Sampling assignment Research method Total sample size Total amount of treatment time Control group sample size Duration of the treatment Treatment group sample size Total amount of reading time
Effect size characteristics	Descriptive data about the effect size	14	Study ID number Effect size number Effect size type

			Category of outcome construct Measurement type Category of data effect size Total sample size Treatment (text + visual) group sample size Treatment group mean Treatment group standard deviation Effect Size Control group mean Control group standard deviation Control group (text only) sample size
Technology characteristics	Descriptive information about technology types, authors, and features	6	Study ID Length of technology use Total amount of technology time Category of technology used Computer software Category of hypermedia used

Chapter Summary

Chapter III provided how thoroughly the present meta-analysis study was conducted on the basis on three major procedures: Location and Selection of Publication, Criteria for Inclusion and Descriptions of Coding Characteristics

Chapter IV will provide the overall results of this meta-analysis by analyzing descriptive statistics and study characteristics of effect sizes in details. The research question 1 will be also discussed

CHAPTER IV

RESULTS

4.1. Three Filtering Processes

Approximately 300 articles, reports and papers that had been published or not yet published in between 1990 and 2009 were considered through the first filtering process by extensive key word searching such as *hypertext*, *hypertext gloss* or *annotation*, etc (see Appendix B). From 300 considered, 57 papers, reports, dissertations and articles were selected through the second filtering process by the *Criteria for Inclusion*. The criteria for inclusion of this study were strictly made to focus on a specific combination effect (text-only and text + visual hypertext glosses) on L2 vocabulary acquisition in order to extract exact effect sizes from empirical selected studies. For example, two major inclusions were: 1) a study should have both independent variables (text-only and text + visual) and a dependent variable of vocabulary test scores and 2) the outcome measure should group contrasts – a treatment (text + visual) group and a control (text-only) group. These two criteria were mainly attributed to select a relatively small sample size, but led to high quality meta-analysis. Through the final filtering process, 10 papers, which met the all strict criteria for the inclusion, were selected and applied to extract effect size statistics.

4.2. Research Findings

As shown in Table 4, characteristics of 10 studies (N= 1560) were described in details. For research design, one of the findings was that majority of the 10 studies have

utilized either a within-subject repeated measures or a between-group measures with a short period of a treatment duration (less than two or three weeks) in class session-based quasi-experiment design. One main research question of these studies was to examine the effects of hypertext glosses on reading comprehension and vocabulary acquisition over time (pre, post and delayed time), in order to measure how a treatment effect continues over time. So, it is important to note that the time variable has been playing an important role in deciding on whether the treatment effect of hypertext glosses has influenced on vocabulary learning over time. However, the fact that the time variable has been not clearly defined for pre and post measures might have attributed to the whole inconclusive results of the empirical studies. That is, one question is when a post measure should be appropriately administrated in order to find out treatment effect duration of hypertext glosses on L2 vocabulary learning over time.

With regard to research population, target populations were as diverse as ESL/EFL, German as a foreign language (GFL) and Spanish as a second language (SSL), but limited only to L2 university adult learners. Main instruments were immediate/delayed post vocabulary tests such as a picture or word recognition test, a production test and a think-aloud or recall protocol, which depends on what and how much vocabulary the test takers recall from the texts that they read. In addition, the time of how long the subjects spent and frequently clicked hypertext glossed words was also measured in order to examine the relationships between the time they spent and vocabulary frequency. Table 4 shows more detailed information of the 10 selected studied for this meta-analysis.

Table 4

10 Studies Selected for the 37 Effect Sizes

Study	Number of ES(N_{ES})	Target Language	Learner Proficiency	Text Type/Word Number	Sample Size	Research Method	Duration of the treatment	Research Findings
Al-Seghayer (2001)	1	Various ESL University learners	Intermediate	Narrative	30	Within-subject repeated ANOVA	One week	Significant difference between text-only and text + video and text + picture; however, no difference between text-only and text + picture.
Chun & Plass (1996)	5	English GFL University learners (second-year)	Second year	Narrative(82/762)	36/103/21	Within-subject repeated ANOVA	Two 50-min class periods	Treatment groups who received both text and visual glosses had high scores on vocabulary recall tests. Significant difference between text + picture and text-only or text + video; however, “the difference between the static pictures and the dynamic videos seems to have an impact on vocabulary recall of the annotated words differently.”
Yoshii (2006)	8	Japanese EFL University learners	Diverse background	Narrative (390)	195	Mixed design repeated measures ANOVA	Two weeks	A text + visual group outperformed on a think-aloud protocol than no gloss and text only groups; however, no statistical difference between the groups.
Akbulut (2007a)	4	Turkish EFL University	Advanced	Narrative(42/1330)	69	Within-subject repeated	Two weeks	Significant difference between text + picture and text-only.

learners

ANOVA

Study	Number of ES(NES)	Target Population	Learner Proficiency	Text Type/Word Number	Sample Size	Research Method	Duration of the treatment	Research Findings
Yeh & Wang (2003)	1	Twainese EFL University learners	6-year EFL experienced	NA	82	ANOVA	Two days	“Gloss users had significantly higher reading comprehension and vocabulary acquisition scores than non-gloss users,” however, no significant difference between text-only and text + visual + audio.
Yoshii & Flaitz (2002)	8	Japanese ESL University learners	Beginning/Intermediate	Narrative (14)	151	Within-subject repeated ANOVA	Two weeks	A text + picture combination was the most effective type of vocabulary annotation; however, no statistical difference between text-only and text + picture.
Kost, Foss & Lenzini (1999)	6	English GFL University learners (second-semester)	Beginning	Narrative (20/272)	56	ANOVA	Two weeks	No statistical difference was found on a production task (immediate and delayed, but significant difference on an immediate word recognition task between text gloss, pictorial gloss and text + pictorial gloss groups. Mixed results.
Plass, Chun, Mayer & Leutner (2003)	1	English GFL University learners (second-year)	Intermediate	Narrative(35/762)	152	ANCOVA	Two 50-min class periods	Consistent results with Mayer’s generative theory of multimedia learning and with cognitive load theory; the high-spatial and high-verbal ability learners performed better on word translations; but

								multiple representations did not always help low-ability learners due to high cognitive load.
Jones & Plass (2002)	2	English FFL University learners (second-semester)	Beginning	Narrative(27/331)	171	MANOVA	Two 50-min class periods	Consistent results with Mayer's generative theory of multimedia learning; the learners performed best on reading comprehension and word retention when given both written and pictorial annotations while listening. A larger effect size was detected for pictorial annotations.
Plass, Chun, Mayer & Leutner (1998)	1	English GFL University learners	NA	Narrative(24/762)	103	Mixed ANOVA	Two 50-min class periods	Consistent results with Mayer's generative theory of multimedia learning; the learners performed best on reading comprehension and word retention when given both visual and verbal annotations.

4.3. Descriptive Data of the Present Meta-Analysis

Effect size statistics were summarized in Table 5. *K* is the total numbers of individual studies selected for this meta-analysis; *N* is the total numbers of individual subjects who participated in the selected studies. A *Q* test was conducted to examine the homogeneity test of the variability of standard error in this study; it rejected the null hypothesis, which means the population of the effect sizes is heterogeneous enough to retain at $\alpha .05$. The overall weighted mean effect size of 37 weighted effect sizes (Hedge's *g*) was 0.37 (SE: 0.074). According to Cohen's rules of thumb for effect size magnitude (1988), this effect size magnitude was moderately positive on L2 vocabulary acquisition.

Table 5
Descriptive Effect Size Statistics

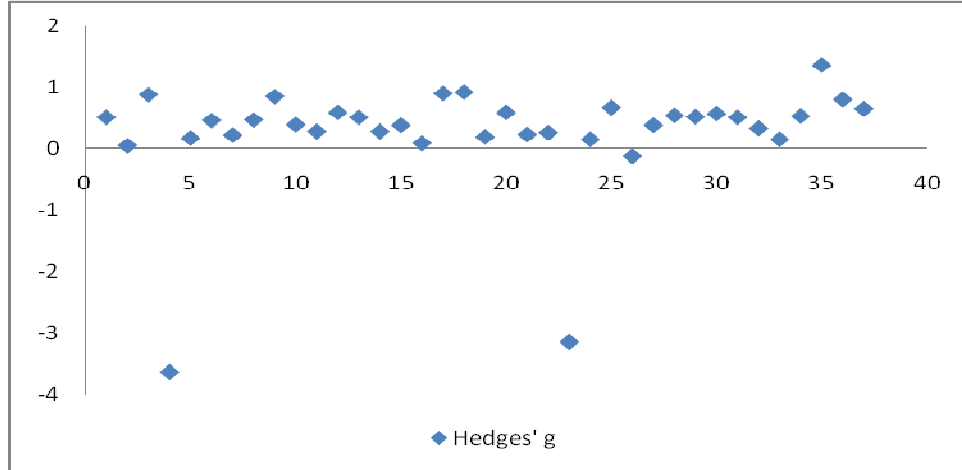
K	N	Number of Effect Size (NES)	Effect Size (g)	SEsm	Z-Value	95% CI	Q	I²
10	1560	37	0.37	0.074	4.91**	0.22 to 0.51	79.96	87.49%

- SE_{sm} = Standard error of standardized mean effect size
- Q: Homogeneity of variance tests
- * Z= 1.96, $p < .05$; **Z= 2.58, $p < .01$

As shown in Graph 1 and 2, majority of 37 effect sizes were equally distributed between 0 and 1; two graphs – scatter diagram and funnel plot – were used to detect a potential publication bias. The results of the scatter diagram and the funnel plot indicated that two possible outliers were detected so that 35 mean effect sizes were again selected for the next statistical procedure instead of 37 effect sizes.

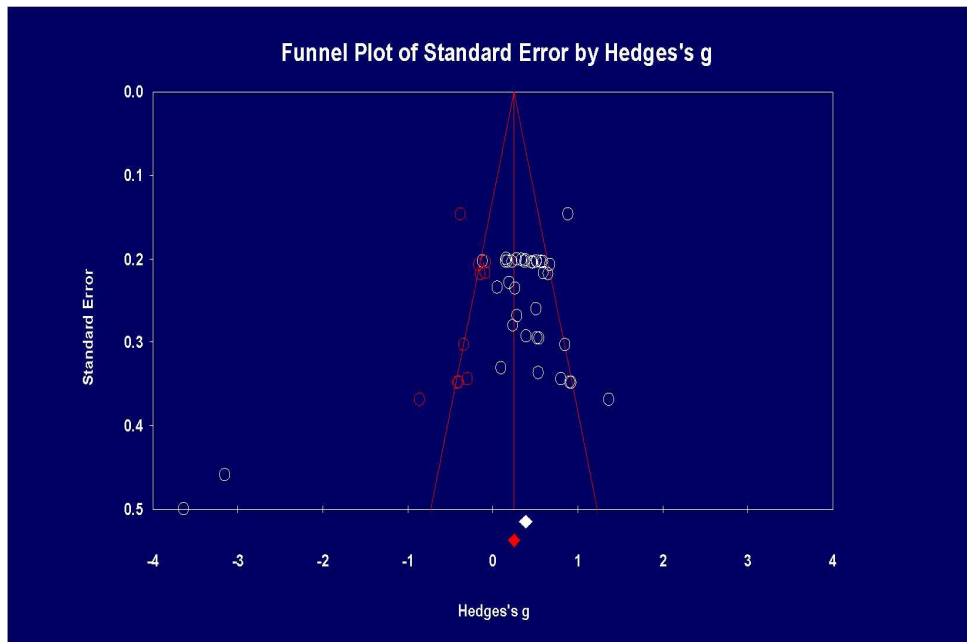
Graph 1

Scatter Plot of 37 Effect Sizes



Graph 2

Funnel Plot



Due to a relatively small sample size ($N_{ES} = 37$), a random-effects model, which allows two error terms (within-subject sampling error and between-study level error), was applied to detect moderator variables which account for between-subject variability for this meta-analysis. On the basis of 35 effect sizes, homogeneity Q test was met as shown in Table 6. The resulting Q -value of 14.34 with 34 degrees of freedom (Number of effect sizes) was less than .05 of the critical value (48.60). Thus, it failed to reject the hypothesis of homogeneity at α .05. It indicates that the variance in this sample of effect sizes is not demonstrably greater than it would be expected from sampling error alone. A weighted mean effect size increased up to 0.46 from 0.37 ($N_{ES} = 37$) previously. According to I^2 (the percent of variance not accounted for by chance variation), 30.25% that remained unexplained might be from either subject-level sampling error or between-study level variability. Moderator variables in the next chapter might be able to help clarify this unexplained variability.

Table 6
Descriptive Effect Size Statistics (without the # 4 and 23 outliers)

K	N	Number of Effect Size (N_{ES})	Effect Size (g)	SE_{sm}	Z-Value	95% CI	Q	I^2
10	1518	35	0.46	0.075	5.242**	0.31 to 0.60	14.34	30.25%

- SE_{sm} = Standard error of standardized mean effect size
- Q : Homogeneity of variance tests
- * Z: 1.96, $p < .05$; **Z: 2.58, $p < .01$

As a whole, overall mean effect size, 0.46, was statistically significant because the 95% confidence interval around the effect size ($0.31 < \mu < 0.60$) did not include zero and reveals the relative precision of the estimate of the mean effect size of the population n of studies from which these 35 were presumably drawn. Correspondingly, the z -test value of 5.24 exceeded the critical value of 2.58 at $p < .01$ so that the weighted mean effect size for this study sample ($N_{ES} = 35$) was statistically significant. That is, the treatment group with access to multiple hypertext glosses performed better than the control group with access to a single gloss on a vocabulary test.

1. Does a group with access to multiple glosses (text + visual) perform significantly better than a group with access to a single gloss (text-only) on a post vocabulary test?

The overall results of this meta-analysis revealed that using a multiple hypertext gloss (text + visual) combination had moderately positive effects on L2 learners' vocabulary learning than using a single text-only hypertext gloss while reading computerized texts. In other words, the overall effect size of 0.46 indicated that various L2 learners with access to a multiple hypertext gloss performed moderately better than those with access to a single text-only gloss on a vocabulary outcome measure. However, the weighted mean effect size ($ES=0.46$) was moderately positive but not conclusively large enough to indicate that the use of text + visual hypertext glosses on L2 vocabulary acquisition is more influential than that of text-only hypertext glosses. Thus, the next question was how characteristics of studies, a research methodology and programs differ from one another.

Overall, effect of multiple (text + visual) glosses on L2 vocabulary acquisition was moderately positive; statistically speaking, the overall results of the 35 effect sizes (N=1518) indicated that there was a statistically significant weighted mean effect size difference between a control (text-only) group and a treatment (text + visual) group on a vocabulary test. In other words, the combination of a text + visual hypertext gloss was more effective on L2 vocabulary acquisition than a text-only hypertext gloss.

Chapter Summary

Chapter IV provided the overall results of this meta-analysis that showed a moderate effectiveness on L2 vocabulary learning with descriptive statistics. A treatment group with access to a multiple gloss combination performed better on a vocabulary test outcome measure than a group with access to a text gloss. During graphical analyses, a scatter diagram showed two outliers with wide variation across 37 effect sizes. As a result, 35 effect sizes without the two potential outliers were analyzed for the present meta-analysis. The final results indicated a strong effect of multiple hypertext glosses on L2 vocabulary acquisition was found.

Chapter V will provide discussion of potential moderators to explain the wide variation of this meta-analysis. Furthermore, implications for future research and research limitations will be followed.

CHAPTER V

DISCUSSION, IMPLICATIONS FOR FUTURE RESEARCH AND RESEARCH LIMITATIONS

5.1. Discussion

Descriptive Results

In the previous chapter, a treatment (text + visual) effect was a statistically significant with moderately positive effect ($ES = 0.46$) on vocabulary learning, but not conclusively large enough as empirical studies indicated in the literature review. Thus, the next step is to analyze characteristics of studies, research methods and technology programs in order to explain the unexplained variation. Regarding effect size characteristics, it is worthy it analyzing between-study variability across the 35 effect sizes.

In order for more in-depth discussion to occur, the characteristics of studies, research methodologies and technology programs were followed respectively to detect some potential moderators that account for variation of the meta-analyzed studies. This following analysis of the variables of meta-analyzed studies implies some insightful findings for the next question of the present meta-analysis.

2. *What are the features of meta-analyzed studies regarding the characteristics of studies, research methodologies and technology programs?*

Study Characteristics

Strict criteria for inclusion ended up yielding 35 effect sizes. 9 out of the 10 meta-analyzed studies were published journal articles which have been peer-reviewed from 1996 to 2007. In order to minimize a publication bias, a rigorous search was conducted but it failed to include more unpublished papers, which tend to have less statistical significances, due to the unavailability of authors or researchers. Major journals included in this meta-analysis were *Language Learning & Technology*, *The Modern Language Journal* and *CALICO Journal* which extensively publish topics of L2 learning and teaching with technology in research.

With regard to time of research conducted, the studies in the 1990s (Chun and Plass, 1996; Kost, et al, 1999; Plass, Chun, Mayer and Leutner, 1998) have mainly been researched on how different hypertext gloss types, including no gloss, visual gloss, verbal and visual gloss in L1 or L2, affect L2 reading comprehension and vocabulary acquisition in a second language multimedia learning environment. However, more recent studies have further examined the effects of hypertext or multimedia glosses not only on reading comprehension and vocabulary acquisition but also on listening comprehension (Jones and Plass, 2002).

In terms of cognitive load theory, the relationships have also been researched on between the effect of hypertext glosses and 1) learning styles (visualizer vs. verbalizer), 2) learners' proficiency levels (low-proficiency ability and high-proficiency ability) and 3) learner differences (verbal vs. spatial ability). It is expected that more research will be following on learners' perspectives and differences based on cognitive theories in the future.

The findings of meta-analyzed studies also indicated that various L2 learners benefit from multiple hypertext glosses regardless of the types of foreign languages: ESL/EFL (Al-Seghayer, 2001; Yoshii, 2006; Akbulut, 2007b; Yeh & Wang, 2003; Yoshii & Flaitz, 2002) SFL (Salem & Aust, 2007) and GFL (Chun, & Plass, 1996). In addition, L2 learners' first languages (L1) varied across English, Japanese, Turkish and French.

Research Methodological Characteristics

Most of the 10 studies have been conducted in an experimental or quasi-experimental design with an average sample size of 86.6 within a short period of research time (less than two or three weeks). Moreover, most studies included a subject population studying various foreign languages over two semesters at the university or college levels; this implies that more L2 learners at K-12 should be researched in order to broaden the scope of hypertext gloss studies in the future. Subject characteristics are as: 1) an averaged mean GPA of the subjects reported was over 3.30, 2) an averaged mean age was 22.1, and 3) gender was relatively equally distributed across selected studies.

For research design, 70% of the studies was conducted with within-subject repeated measures. The studies with within-subject repeated measures had a bigger mean effect size than studies with between-subject measures design.

Program Characteristics

Hypertext gloss programs used in this meta-analysis were researcher-developed by using *Authorware* such as *HyperCard*, *Dreamweaver* (Al-Seghayer, 2001) and *CyberBuch*

(Chun, & Plass, 1996); however, little has been specifically known about the technical algorithm of the author-developed programs in details. This researcher-developed technology may have been attributed to moderate effect sizes of this meta-analysis so that it is necessary to standardize hypertext gloss technology programs for consistent hypertext research results in the future.

L2 reading passages were hypertext glossed with various features: text-only, text + picture and text + picture + audio in L1 or L2. Visuals such as a picture or video clip were most L1 culture-embedded for L2 learners to be familiar with. Thus, using the culture-embedded visuals may have misled L2 readers simply because it is not easy to make visuals universally neutral across different cultures. Especially in hypertext glossed-action verbs, for example, it is not clear of whether pictures of the two verbs, 'jump' and 'dash,' helped L2 readers retain the meanings better in the brain (Salem & Aust, 2007; Yoshii, 2006).

In order to consistently explain more about between-study level variables, the four characteristics coded were analyzed to detect potential moderators that systematically differentiate studies with larger or smaller effect sizes. From the four (coding, study, method and program) characteristics, 8 potential moderators that have accounted for between-subject variation were investigated; findings of the moderator variables revealed some insightful consideration for the next question.

3. *What are some potential moderators to systematically account for the between study variation in the present study?*

In Table 5, the variable, sample size, appeared to be a strong moderator that accounted for the between the two categories (less than 80 and more than 81) with a relatively equal sample size distribution: studies with a less-than-80 sample size generated 17 effect sizes with a mean effect size ($M_{es} = 0.284$) while studies with a more-than-81 sample size produced 18 effect sizes with a mean effect size ($M_{es} = 0.430$), $Q = 3.052$, $p = 0.086$. That is, large sample size studies had more statistic power than small sample size studies: a small sample size was attributed to a small weight while a large sample size tends to produce a large weight. Interestingly enough, this finding contrasted to Liao's results, showing that studies with small samples had more statistical power than those with large samples (1999). He reported that studies with less than 80 samples had a large mean effect size ($ES = 0.6$) compared to those with over 80 samples ($ES = 0.033$). He argued that hypermedia effects on learners' achievement would be questionable when sample size is small or medium. Future research should confirm this contrasting finding.

Learner proficiency was found a statistically significant moderator to affect the treatment effects with $Q = 15.304$, $p < 0.05$; that is, studies with beginning learners had the largest mean effect size, 0.698 while those with intermediate learners had the least mean effect size, 0.233. That is, beginning learners who had access to multiple hypertext glosses most benefited from multiple glosses in reading. This finding contrasts to the results of previous studies showing that low-proficiency learners are less likely to benefit from multiple hypertext glosses than high-proficiency learners do due to high cognitive load.

For example, on the basis of Salem's recent study (2006), the learners who had access to more gloss features, such as text + audio + picture and text + audio + picture +

writing, did not outperform those who had access to a text-only gloss on the word retention test over time. Statistically speaking, there was no statistical difference between the text-only-gloss group, the text-audio group and the text-audio-writing group. Even worse, the simple gloss group that had access to the text-only gloss ($M= 11.75$) slightly outperformed the more features-embedded group with text-audio glosses ($M= 11.38$) on a delayed vocabulary test.

Another significant finding was that mean effect sizes differed statistically across the moderator level of vocabulary test type with $Q= 20.881, p < 0.05$. Recognition (form, meaning, picture and word) multiple-choice format was significantly more used to test L2 learners' vocabulary learning as a dependent outcome measure in the most studies compared to production such as a recall or read-aloud protocol. The format of recognition, consisting of form, meaning, picture or word tests, was preferred across all the studies. A multiple-choice testing type appears to be a fairly reliable and valid instrument to measure test takers' performance at a short period of time.

Target language was not a statistically significant moderator for accounting for the between study variation even though studies with other FLs (French, German, Japanese and Spanish) had a better mean effect size ($N_{es} = 14; M_{es} = 0.405$) than studies with ESL/EFL population ($(N_{es} = 21; M_{es} = 0.379), Q= 0.103, p > 0.05$).

Research design was examined for whether there was a significant mean effect size difference between within-subject and between-study levels; however, no significant difference was found. Studies with between-study measures design ($N_{es} = 6; M_{es} = 0.430$)

had a slightly larger effect than studies with within-subject measures design ($N_{es} = 29$; $M_{es} = 0.380$), $Q = 0.225$, $p > 0.05$.

Such moderator variables as publication year and country were statistically analyzed, but no significant difference was found. The overall results indicated that the treatment effect of multiple glosses tended to disappear shortly after two or three weeks, $Q = 0.663$, $p > 0.05$.

More information of moderator variables analyzed is shown in Table 7.

Table 7

Summary of Moderator Variables

Moderator Variable Level	Number of Effect Size (N_{es})	Effect Size (g)	Lower Confidence	Upper Confidence	Q_B Value
1. Sample Size:					3.052
Less than 80	18	0.284	0.146	0.422	
More than 81	19	0.430	0.341	0.520	
2. Target Language:					0.103
ESL/EFL	22	0.379	0.288	0.470	
Other FLs	15	0.405	0.272	0.538	
3. Learner Proficiency:					15.304*
Beginning	8	0.698	0.491	0.905	
Intermediate	7	0.233	0.058	0.409	
Beginning + Intermediate	8	0.417	0.276	0.557	
Advanced	4	0.579	0.284	0.875	
NA	10	0.294	0.161	0.427	
4. Publication Year:					0.042
1990s	12	0.373	0.216	0.530	
2000s	25	0.391	0.306	0.477	
5. Country:					0.762
USA	24	0.417	0.322	0.513	
Outside USA	13	0.348	0.223	0.472	
6. Research Design:					0.225
Between-subject measures	6	0.430	0.238	0.623	
Within-subject measures	31	0.380	0.298	0.461	
7. Outcome Measure:					0.633
An immediate post test	21	0.413	0.315	0.512	
A delayed post test	16	0.352	0.236	0.467	

8. Vocabulary Test Type:					20.881*
Definition	7	0.313	0.163	0.462	
Production	8	0.435	0.272	0.599	
Recognition	7	0.118	-0.050	0.287	
Recognition + Production	2	0.369	0.029	0.710	
Word recognition	4	0.600	0.358	0.841	
Form Recognition	2	0.689	0.274	1.103	
Meaning Recognition	2	0.455	0.047	0.862	
Picture Recognition	5	0.617	0.410	0.825	

QB values indicate whether effect sizes differ statistically across levels of the moderator variable

**P* < 0.05

5.2. Implications for Future Research

There are some implications for future research from the results of the present study. For research design, hypertext gloss studies have been almost always conducted in the settings of class session-based quasi-experiment design with researcher-developed programs. In other words, an instructional impact has been rarely reported from the empirical studies, which are focusing mainly on multimedia treatments, so that future research should take an instructional effect into consideration in that instructors' effect appears to be a very crucial variable for technology-based reading.

Outcome measure instruments seemed limited to a sort of one-way measurement (measuring outcome values particularly based on learners' performance which reacted to computer programs) such as time on task measured by learners' clicking and multiple-choice recognition tests, which may have not maximized full advantage of the relationship between innovative technology use and individual learners' characteristics. As technology evolves, innovative outcome measuring tools, controlling variability that remained unexplained, could help provide more consistent results of hypertext gloss research in the near future.

In terms of learners' proficiency, the results indicated that low proficient learners are most likely to benefit from multiple glosses than immediate and advanced learners. This finding does not match previous study results. According to cognitive load theory (Chandler and Sweller, 1991), low-ability language learners may have not utilized the whole benefits of multimedia glosses in reading comprehension and vocabulary acquisition

due to their high cognitive (Sweller, 1994; Plass, Chun, Mayer & Leutner, 2003). Future research should verify this finding.

In addition to learners' proficiency, learners' learning preference such as visualizers or verbalizers (Plass, Chun, Mayer and Leutner, 1998) appears to be a critical variable in hypertext gloss studies: learners who prefer visual type annotations tend to benefit most from hypertext glossed reading in particular with a text + visual gloss while learners who prefer verbal or text type annotations tend to benefit most with specialization in a verbal or text only gloss when given a choice either text-only or text + visual glosses. In the next research, applying hypertext combination should be careful depending on learners' learning preferences.

Finally, long-term effects of hypertext glosses on L2 vocabulary learning should be confirmed from longitudinal future research such as HLM because the present study result indicated that treatment effects did not last long enough but decreased shortly after two or three weeks.

5.3. Research Limitations

Even though this study has a higher statistical power than one individual study conducted in the field of hypertext glosses, it should not be ignored that some biased sources such as a publication bias may have not been controlled enough by this meta-analysis procedure due to the limited number of unpublished papers. In addition, a number of significant studies may have not been included in the present study due to critical data unavailability and inaccessibility of the authors. A publication bias means that meta-

analysis tends to heavily rely on published papers which have more statistically significant results. This bias may have increased the overall weighted mean effect size and drawn positive results.

In sum, the relatively small effect sizes ($ES = .35$) may have impacted the whole generalizability of this study in terms of external validity: regarding interpreting the results of this meta-analysis, the efficacy of this particular treatment with a particular type of participants in experimental settings may not necessarily be representative of the effects that occur in routine practice of reading education in non-research settings (Weisz, Weiss, & Donenberg, 1992).

REFERENCES

* *Study chosen for the second stage filtering.*

** *Study used for the final meta-analysis*

*Abraham, L. B. (2007). Second language reading comprehension and vocabulary learning with multimedia. *Hispania*, 90, 98-109.

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 21(3), 199-226.

*Abuseileek, A.F.M. (2008). Preferences and effect on EFL reading comprehension and vocabulary acquisition. *CALICO Journal*, 25(2), 260-275.

Adams, T. W. (1995). What makes materials authentic? *ERIC, Reports-Evaluative/Feasibility* (142)

Aebbersold, J. A. & Field, M. L. (1997). *From Reader to Reading Teacher*. Cambridge University Press.

**Akbulut, Y. (2007a). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, 35. 499-517.

*Akbulut, Y. (2007b). Variable Predicting Foreign Language Reading Comprehension And Vocabulary Acquisition in A Linear Hypermedia Environment. *The Turkish Online Journal of Educational Technology*, 6(1). Retrieved July 20, 2008 from <http://www.tojet.net/articles/615.htm>

- **Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232.
- Al-Seghayer, K. (2003). Technological and pedagogical considerations for a more effective electronic glossary. *The Reading Matrix*, 3(1).
- Al-Seghayer, K. (2005). ESL readers' perceptions of reading in well structured and less structured hypertext environment. *CALICO Journal*, 22(2), 191-212.
- Ariew, R. (2006). A Template to Generate Hypertext and Hypermedia Reading Materials: Its Design and Associated Research Findings. *The Reading Matrix*, 6(3), 195-209.
- *Aust, R., Kelly, M. J., and Roby, W. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research and Development*, 41(4), 63-73.
- Bardovi-Harlig, K. & Hartford, B. (1997). *Beyond Method*. McGraw-Hill, Prentice Hall Regents.
- Beatty, K. (2003). *Teaching and Researching Computer-assisted Language Learning*. Edinburgh England: Pearson Education Limited.
- Bell, F. L. (2006). Comprehension aids, Internet technologies, and the reading of authentic materials by adult second language learners. *UMI*.
- *Bell, F.L. & LeBlanc, L.B. (2000). The Language of Glosses in L2 Reading on Computer: Learners' Preferences. *Hispania*, 83(2), 274-285.
- Bernhardt, E. B. (2005). Progress and procrastination in s second language reading. *Annual Review of Applied Linguistics*, 25, 133-150.
- Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania*, 87, 541-552.

- Brandl, K. (2002). Integrating Internet-based reading materials into the foreign language curriculum: From teacher- to student-centered approaches. *Language Learning & Technology*, 6(3), 87-107.
- Bush, M., & Terry, R. M. (1996). *Technology-Enhanced Language Learning*. ACTFL Foreign language. National Public Company.
- Canal, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*. 1, 1-47.
- Chapelle, C. A. (1998). Multimedia CALL: lessons to be learned from research on instructed SLA *Language Learning and Technology*, 2(1), 22-34.
- Chapelle, C. A. (1999). CALL in the year 2000: Still in search of research paradigms? *Language Learning & Technology*. 1(1), 19-43.
- Chapelle, C. A. (2001). *Computer Application in Second language Acquisition: Foundations for teaching, testing and research*. New York: Cambridge University Press.
- Chapelle, C. A. (2005). Hints About Use From Research. *PacCALL Journal*, 1(1), 1-8.
- Charney, D. (1994). *The impact of hypertext on processes of reading and writing*. Book chapter: Literacy and Computers, 238-263. New York: Modern Language Association.
- Chun, D. (2001). L2 reading on the web: strategies for accessing information in hypermedia. *Computer Assisted Language Learning*, 14, 367-403.
- Chun, D. (2006). Calling on CALL: from theory and research to new directions in foreign language teaching. *CALICO Journal*, 5, 69-98.

- **Chun, D. M. & Plass, J. L. (1996). Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal*, 80(2), 183-198.
- *Cooledge, S. L. (2004). L2 reading and hypertext: A study of lexical glosses and comprehension among intermediate learners of French. Unpublished dissertation at the University of Arizona, USA.
- *Coriano Vela Zquez, A. (2001). Vocabulary acquisition through reading: a study of the effectiveness of different call-based annotations. Master's thesis at the University of Puerto Rico, Mayaguez Campus.
- Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd Ed.)*. Sage Publications, Inc.
- *Davis, J.M. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal*, 73(1), 41-48.
- Day, R. & Bamford, J. (1998). *Extensive reading in the second language classroom*. New York: Cambridge University Press.
- DeCoster, J. (2004). *Meta-analysis Notes*. Retrieved September 20, 2008 from <http://www.stat-help.com/notes.html>
- Doughty, C. & Williams, J. (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge University Press.
- Dunkel, A., Brill, S., & Kohl, B. (2002). The Impact of Self-Instructional Technology on Language Learning: A View of NASILP. In C. A. Spreen (Ed.), *New technologies and language learning: Cases in the less commonly taught languages* (Technical Report #25; pp. 97-120).

- *Ercetin, G. (2003). Exploring ESL learners' use of hypermedia reading glosses. *CALICO Journal*, 20(2), 261-283.
- *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.
- Grabe, W. & Stoller, F. L. (2002). *Teaching and Researching Reading*. Pearson Education Limited.
- Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, 10(2), 77-92.
- *Hew, S. H., & Ohki, M. (2004). Effect of animate graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal*, 21(2), 397-419.
- *Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL*, 41, 87-106.
- *Hong, W. (1997). Multimedia computer-assisted reading in business Chinese. *Foreign Language Annals*, 30, 335-344.
- *Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive online database. *Language Learning and Technology*, 9, 90-110.
- *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. *The Modern Language Journal*, 80(3), 327-339.

- Hymes, D. H. (1971). *On communicative competence*. Philadelphia: University of Pennsylvania Press. Extracts available in Brumfit, C.J. & Johnson, K. (Eds.) (1979), *The communicative approach to language teaching*, pp. 5-26. Oxford: Oxford University Press.
- Johnson, M. (2004). *A Philosophy of Second Language Acquisition*. New Haven and London: Yale University Press.
- Jonassen, D. H., Campbell, J., & Davidson, M.(1994). Learning with media; restructuring the debate. *Educational Technology Research and Development*, 42(2), 31-39.
- **Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-561.
- *Jones, L.C. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO Journal*, 21(1), 41-65.
- *Jones, L.C. (2006). Effects of collaboration and multimedia annotations on vocabulary learning and listening comprehension. *CALICO Journal*, 24(1).
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R.Quirk, & H.G. Widdowson, (Eds.). *English in the world: teaching and learning the language and literatures*. Cambridge: CUP.
- *Kang, H. W. (2005). The effects of the visuality of hyperlink annotation and web-based post vocabulary learning activities on L2 vocabulary acquisition. *Multimedia-Assisted Language Learning*, 8(2), 32-57.
- Khan, B. H. (1997). *Web-based instruction*. New Jersey: Educational Technology

Publications.

- Kim, W. H. (2001). The Test of English for International Communication (TOEIC) as Measure of Korean Adult English Language Oral Proficiency. Unpublished Doctoral Dissertation, University of Kansas, Lawrence, USA.
- Kommers P. A. M., Grabinger S. & Dunlap J.C. (1996). *Hypermedia Learning Environments: Instructional Design and Integration*. Lawrence Erlbaum, Hillsdale, NJ.
- Kon, C. K. (2002). The influence on outcomes of ESL students' performance strategies on a CALL listening comprehension activity. Unpublished MA thesis, Department of English, Iowa State University, Ames, IA, USA.
- *Koren, S. (1999). Vocabulary Instruction through Hypertext: Are There Advantage Over Conventional Methods of Teaching? *TESL-EJ*. 4(1), 1-18.
- **Kost, C., Foss, P., & Lenzini, J. (1999). Textual and pictorial gloss: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1), 89-113.
- Koyama, T. & Takeuchi, O. (2004). How look-up Frequency Affects EFL Learning?: An Empirical Study on The Use of Handheld-Electronic Dictionaries. Paper presented at the 2004 CLaSIC Conference. Retrieved July 20, 2008 from http://www.paccall.org/2004/2004proceedings_papers/koyama.pdf
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Beverly Hills, CA: Laredo Publishing Company. *Implicit and Explicit Learning of Languages*, 45-77. London: Academic Press.

- Krashen, S. (1994). *The input hypothesis and its rivals*. In Ellis, N. (Ed.)
- Laufer, B. & Hill, M. (2000). What Lexical Information Do L2 Learners Select in A CALL Dictionary And How Does It Affect Word Retention? *Language Learning & Technology*, 3(2), 58-76.
- *Leffa, V. (1992). Making foreign language texts comprehensible for beginners: An experiment with an electronic glossary. *System*, 20(1), 63-73.
- *LeLoup, J. & Ponterio, R. (2005). On the Net: Vocabulary support for independent online reading. *Language Learning & Technology*, 9(2), 3-7.
- Leu, D. J. (1994). Designing hypermedia to connect reading and writing through children's literature. Reports presented at the annual national educational computing conference.
- *Levine, A., Bejarano, Y., Carrell, p., & Vered, L. (2004). Comparing dictionary definitions and glosses in hypertext foreign language reading. *The CATESOL journal*, 16(1), 59-68.
- *Liao, Y. C. (1999). Effects of hypermedia on students' achievement: A meta-analysis. *Journal of Educational Multimedia and Hypermedia*, 8(3), 255-277.
- Lightbown, R.M. & Spada, N. (1999). *How Languages are Learned (Revised Ed.)* Oxford University Press.
- Lim, K.M., & Shen, H.Z. (2006). Integration of computers into an EFL reading classroom. *ReCALL*, 18(2), 212-229.
- *Lin, H. & Chen, T. (2007). Reading authentic EFL text using visualization and advance organizers in a multimedia learning environment. *Language Learning & Technology*, 11(3), 83-106.

- Lipsey, M. & Wilson, D. (2001). *Practical Meta-Analysis*. Applied Social Research Methods Series, 49.
- Liou, H. C. (2000). The electronic bilingual dictionary as a reading aid to EFL learners: Research findings and implications. *Computer Assisted Language Learning*, 13, 467-476.
- Liu, S. S. (2007). Electronic Dictionaries And ESL Students. Retrieved July 20, 2008, from [http://www.usc.edu.tw/college/afl/data/%E5%8A%89%E6%B7%91%E7%85%A7%20\(Electronic%20Dictionaries%20And%20ESL%20Students\).pdf](http://www.usc.edu.tw/college/afl/data/%E5%8A%89%E6%B7%91%E7%85%A7%20(Electronic%20Dictionaries%20And%20ESL%20Students).pdf)
- *Liu, M., & Reed, W. M. (1995). The effect of hypermedia-assisted instruction on second language learning. *Journal of Educational Computing Research*, 12(2), 159-175.
- *Lomicka, L. (1998). "To gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41-50.
- Martinez-Lage, A. (1997). Hypermedia technology for teaching reading. In M. Bush & R. Terry (Eds.), *Technology-enhanced language learning*, 121-163.
- Mayer, R. E. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444-452.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19.
- *McDonnell, T. B. (2006). Reading plain text and hypertext on the Internet for native and nonnative speakers of English. Unpublished dissertation at New York University.

- McKnight, C., Dillon, A., & Richardson, J. (1996). *User centered design of hypertext/hypermedia for education*. In: D. Jonassen (ed) Handbook of Research on Educational Communications and Technology. New York: Macmillan, 622-633.
- *Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32 (4), 469-479.
- Nation, I. S. P. (1983). Teaching and learning vocabulary. Wellington: English Language Institute, Victoria University. (Reprinted from "To gloss or not to gloss": An investigation of reading comprehension online, pp. 41, by Lomicka, L. *Language Learning & Technology*, 1(2).
- Nelson, T. (1965). A File Structure for the Complex, the Changing, and the Indeterminate. *Association for Computing Machinery: Proc. 20th National Conference 1965*, 84-100.
- Nikolova, O. R. (2002). Effects of learners' participation in authoring of multimedia materials on student acquisition of vocabulary. *Language Learning & Technology*, 6(1), 100-122.
- *Nikolova, O. R. (2004). Effects of Visible and Invisible Hyperlinks on Vocabulary Acquisition and Reading Comprehension for High-and Average-Foreign Language Achievers. *ALSIC*, 7(1), 29-53, Retrieved July 16, 2008, from http://alsic.u-strasbg.fr/v07/nikolova/alsic_v07_05-rec1v2.htm
- Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly*, 25(2), 279-295

- Oh, S. C. (2005). A practical application of “Writing” hypertext literature in the English education of the elementary school. *English Language & Literature Teaching, 11*(2), 19-34.
- Omaggio, A. C. (1979). Pictures and second language comprehension: Do they help? *Foreign Language Annals, 12*, 107-116.
- Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning & Verbal Behavior, 20*, 532-539.
- Paivio, A. (1986). Mental representation: A dual-coding approach. *New York: Oxford University Press.*
- *Pak, J. (1986). The effect of vocabulary glossing on ESL reading comprehension. Unpublished manuscript, University of Hawaii at Manoa. (Reprinted from “To gloss or not to gloss”: An investigation of reading comprehension online, pp. 41, by Lomicka, L. *Language Learning & Technology, 1*(2).
- Pearson, P. D., Ferdig, R. E., Blomeyer, R. L., & Moran, J.(2005). The effects of technology on reading performance in the middle-school grades: A meta-analysis with recommendations for policy. Reports from Learning Points Associates. Retrieved Feb. 20. 2008, from <http://www.ncrel.org/tech/reading/index.html>
- *Peters, E. (2006). L2 vocabulary acquisition and reading comprehension: The influence of task complexity. *Investigating tasks in formal language learning, 178-198.*
- *Peters, E. (2007). Manipulating L2 learners’ online dictionary use and its effect on L2 word retention. *Language Learning and Technology, 11*(2), 36-58.

- **Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36.
- **Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221-243.
- Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. *The Cambridge handbook of multimedia learning*, 467-488.
- Richards, J. C. (2000). Hypermedia, internet communication, and the challenge of redefining literacy in the electronic age. *Language Learning & Technology, 4*(2), 59-77.
- *Ridder, I. D. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning & Technology, 6*(1), 123-146.
- Robin, R. (2007). Commentary: Learner-based listening and technological authenticity. *Language Learning & Technology, 11*(1), 109-115. *Computer Assisted Language Learning, 17*(5), 517-556.
- *Roby, W. (1991). Glosses and dictionaries in paper and computer formats as adjunct aids to the reading of Spanish texts by university students. Unpublished doctoral dissertation. University of Kansas.
- Roby, W. (1999). "What Is in A Gloss?" *Language Learning & Technology, 2*(2), 94-101.
- Rogers, E.M. (1995). *Diffusion of innovations (4th ed.)*. New York: The Free Press.

- Rogers, D.L. (2000). A paradigm shift: Technology integration for higher education in the new millennium. *Educational Technology Review* 13, 19-27.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1995). Writing Meta-Analytic Reviews. *Psychological Bulletin*, 18(2), 183-192.
- *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, 589-619.
- *Rott, S. (2007). The Effect of Frequency of input-enhancements on word learning and text comprehension. *Language Learning*, 57(2), 165-199.
- Rouet, J. F., Levonen, J. J., Dillon, A., & Spiro, R. J. (1996). *Hypertext and Cognition*. Lawrence Erlbaum Associates (Ed. 1).
- *Sakar, A., & Ercetin, G. (2004). Effectiveness of hypermedia annotations for foreign language reading. *Journal of Computer Assisted Learning*, 21, 28-38.
- Slaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal*, 85(1), 39-56.
- *Salem, E. (2006). The Impact of Electronic Glosses on Word Retention and Reading Comprehension among L2 Spanish Learners. Unpublished doctoral dissertation. University of Kansas.
- *Salem, E. & Aust, R. (2007). The Influence of Feature-Rich Computerized Glosses on Reading Comprehension and Vocabulary Acquisition. *Proceedings of the Sixth*

IASTED International Conference Web-based Education.

- *Smidt, E. & Hegelheimer, V. (2004). Effects of Online Academic Lectures on ESL Listening Comprehension, Incidental Vocabulary Acquisition, and Strategy Use.
- *Son, J. B. (1997). Strategies for reading printed texts and electronic texts: Same or different?
- Son, J. B. (1998). Understanding hypertext: A discussion for TEFL. *English Teaching*, 53(3), 113-124.
- Stakhnevich, J. (2002). Reading on the Web: Implications for ESL professionals. *The Reading Matrix*, 2(2).
- *Stewart, R. A., & Cross, T. L. (1991). The effect of marginal glosses on reading comprehension and retention. *Journal of Reading*, 35, 4-12.
- Swaffar, J. K. (1988). Readers, texts, and second languages: The interactive processes. *The Modern Language Journal*, 72(2), 123-149.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass, S. and Madden, C. (Eds.), *Input in Second Language Acquisition*, pp. 235-256. New York: Newbury House.
- *Taylor, A. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal*, 23(2).
- *Wood, J. (2001). Can software support children's vocabulary development? *Language Learning & Technology*, 5(1), 166-201.
- *Yano, Y., & Long, M. H. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44(2), 189-219.

- **Yeh, Y. & Wang, C. (2003). Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CAICO Journal*, 21(1), 131-144.
- **Yoshii, M. & Flaitz, J. (2002). Second Language Incidental Vocabulary Retention: The Effect of Text and Picture Annotation Types. *CAICO Journal*, 20(1), 33-58.
- **Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85-101.
- *Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, 21(1), 7-27.
- Zhao, Y. (2005). Technology and second language learning: promises and problems. *Working paper*.

APPENDIX A

Studies included in the present meta-analysis

1. Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), pp. 202-232.
2. Chun, D.M. & Plass, J.L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80 (2), 183-198.
3. Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85-101.
4. Akbulut, Y. (2007a). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, 35. 499-517.
5. Yeh, Y. & Wang, C. (2003). Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CAICO Journal*, 21(1), pp. 131-144.
6. Yoshii, M. & Flaitz, J. (2002). Second Language Incidental Vocabulary Retention: The Effect of Text and Picture Annotation Types. *CAICO Journal*, 20(1), pp. 33-58.
7. Kost, C., Foss, P., & Lenzini, J. (1999). Textual and pictorial gloss: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1), 89-113.
8. Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, 19, 221-243.

9. Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-561.
10. Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90(1), 25-36.

APPENDIX B

Keywords Used for Searches

Annotation	Learning vocabulary
CALL	L2 reading
Computer and reading	Multimedia
Cyber reading	Multimedia text
Cyber annotation	Multimedia gloss(ery)
Digital	Multimedia annotation
Digital gloss(ery)	Online
Digital reading and instruction	Online reading
Digital vocabulary learning	Online vocabulary
Dictionary	Online gloss(ery)
Educational technology	Online annotation
EFL	Reading
E-learning	Reading material development
E-learning and education	Second language acquisition
Electronic	Second language learning and technology
Electronic gloss(ery)	Technology
Electronic annotation	Technology use
Electronic reading	TELL
ESL	Technology integration

ESL reading	Technology-embedded language learning
ESL vocabulary instruction	Technology-enhanced language learning
Foreign language learning and technology	and teaching
Gloss(ery)	Vocabulary acquisition
Hypertext	Vocabulary learning
Hypermedia	Word acquisition
Hypertext gloss(ery)	Word retention
Hypertext annotation	Web-based learning and teaching
Hypermedia annotation	Web-based reading
Hypermedia gloss(ery)	Web-based vocabulary learning
Interactive reading	
Interactive gloss(ery)	
Interactive annotation	
Internet reading and vocabulary	
Internet and language	
Instructional technology	
Instruction with technology	

APPENDIX C

Academic and Educational Databases and Journals

<i>APACALL</i>	<i>IALLT</i>
<i>Asian TEFL</i>	Ingenta Select
<i>Asian EFL</i>	<i>Language learning & Technology</i>
<i>ACTFL</i>	Lawrence Erlbaum Journals
<i>Academic ASAP</i>	MetaPress
Blackwell Science Synergy	Ovid
<i>CALICO</i>	<i>PacCALL</i>
<i>CALL</i>	ProQuest Education
<i>CALL-EJ online Journal</i>	PsychInfo
<i>DBPIA</i>	<i>Reading in a Foreign Language</i>
Directory of Open Access Journals	<i>SAGE Journal Online</i>
<i>Educational Technology Research and Development</i>	Sage Publications
ERIC	SpringerLink
<i>Foreign Language Annals</i>	<i>System</i>
<i>Hispania</i>	<i>The Modern Language Journal</i>
<i>Journal of Educational Technology and Society</i>	<i>TESOL Quarterly</i>
<i>Journal of Research on Technology in</i>	Wiley Interscience
	Wilson Education
	<i>Wilson OmniFile full text select</i>

<i>Education</i> <i>Journal of Educational Computing Research</i> JSTOR	
---	--

APPENDIX D

Meta-Analysis Coding Manual for the Effects of Hypertext Annotations on L2 Vocabulary Acquisition

Study Characteristics

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
Study ID	Author	Pub Type	Pub Year	First_Lang	Target Population	Mean Age	RA CE	Study Year	Mean GPA	Male_N	Female_N	R_Type	ASSIGN	Method	Total_N	TX_N	CG_N	Duration	Total_Treat_Tim_e	Total_Readin_g_Tim_e
1	1	Al-Seghayer, K.	2001	6	1			5	1	17	13	2	3	Within-subject/repeated-measures one-way ANOVA	30			1	2	1
2	2.1	Chun, D.M. & Plass, J.L.	2096 (93)	1	3	20.67	3	2	3.29	17	19	1	4	Within-subject	36	36	36	2	4	2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
Study ID	Author	Pub Type	Pub Year	First Lang	Target Population	Mean Age	RA CE	Study Year	Mean GPA	Male N	Female N	R_Type	ASSIGN	Method	Total N	TX_N	CG_N	Duration	Total Reading Time	Total Reading Time	
3	2.2	Chun, D.M. & Plass, J.L.	2	96 (94)	1	3	21.9	3	2	3.42	57	46	1	4	Within-subject	103	103	103	2	4	2
4	2.3	Chun, D.M. & Plass, J.L.	2	96 (95)	1	3	21.05	3	2	3.26	9	12	1	4	Within-subject	21	21	21	2	4	2
6	4	Yoshii, M.	2	06	5	1		2	1					Mixed design repeated measure	195						
7	5	Akbulut, Y.	2	07	7 (Turkish)	1		6 (Turkish)	1		22	47	1	1	ANOVA	69	46	23			
8	6	Salem, E & Aust, R.	2	07	1	2		3	1		32	61	1	1	ANOVA	93	45	15			
9	7	Yeh, Y. &	2	03	7 (Taiwan)	1		2	1				1	1	ANOVA	55	28	27			

		Wang, C.			n)													
1	8	Yoshii, M. & Flaitz, J.	2	02	7 Variou s	1	24.6	5	1		1	1	ANOVA	100	50	50		

Study Level Coding Manual

- A. Study ID number – assign a unique identification number to each study. If a report presents two independent studies, add a decimal to the study ID number to distinguish each study within a report and code each independent study separately.
- B. Author – report last name, first (e.g., Yun, Jeehwan)
- C. Types of publication: The priority is as follows:
 - 1. book
 - 2. journal article or book chapter
 - 3. thesis or doctoral dissertation
 - 4. conference paper
- D. The publication year – if two separate reports are being used to code a single study, code the publication year of the more formally published report.

❖ **Sample Descriptions**

- E. First Language – English = 1; Spanish = 2; French = 3; German = 4; Japanese = 5; mixed = 6; other = 7.
- F. Target population – ESL/EFL = 1; SSL/SFL = 2; GSL/GFL = 3; FSL/FFL = 4; JSL/JFL = 5; other second language learning = 6.
- G. Mean age of sample – Unspecified = 0; write down exactly the mean age.
- H. Students' Ethnicity (RACE) – Unspecified = 0; Hispanic = 1; Asian = 2; White = 3; European = 4; Mixed = 5; Others = 6.
- I. Study years of the target language.
- J. The sample's mean GPA (Mean_GPA).
- K. The number of males (Male_N).
- L. The number of females (Female_N).

❖ **Research Design Descriptors**

- M. Type of research (R_Type) – Experimental = 1; Quasi-experimental = 2.
- N. Sampling assignment – Random = 1; Nonrandom = 2; Matching = 3; unspecified = 4.

- O. Research method (e.g., *t*- test, ANOVA, Repeated Measure, Regression, Correlation Coefficient) (Method).
- P. Total sample size (Total_N).
- Q. Treatment group sample size (TX_N).
- R. Control group sample size (CG_N).
- S. Duration of the treatment – less than one day = 1; between one day and seven days = 2; more than a week = 3; less than a month = 4; more than a month = 5.
- T. Total amount of treatment time – less than 30 minutes = 1; 30 to less than 60 minutes = 2; one hour to less than two hours = 3; more than two hours = 4.
- U. Total amount of reading time– less than 30 minutes = 1; 30 to less than 60 minutes = 2; one hour to less than two hours = 3; more than two hours = 4.

APPENDIX E

Statistics for Effect Sizes & Characteristics in the Analysis

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Study ID	ES_N	ES_T YPE	OUT COM E	M-Type	ES_ CA T	Total_ N	TX _N	TX_Mea n	TX_SD	CG_ N	CG_Mean	CG_ SD	T_Val ue	F_ Val ue	P- Valu e	ES (g)	ES (r)	Z- Value
1	1	2	2	1	2 (21)	1	90	60	5.4	.856	30	4.03	1.58 6		1.19 52	.513	1.659 0651	
2	2.1	2	2	1	2 (15)	1	72	36	1.29	1.075	36	1.31	.89	.09	0.02 03	-.01	- 0.085 3141	
3	2.2	2	2	1	2 (36)	1	206	103	3.255	1.85	103	2.15	1.72		.618 6	.2955	4.316 6947	
4	2.3	2	2	1	2 (36)	1	42	21	7.095	1.23	21	13.52	2.36	.04	3.41 42	-.8629		
5	3	1	2	1	1	1	76	38	25.4	4.9	38	24.6	3.3		.191 5	.0953	0.824 6194	
6	4	1	2	1	2	1	98	50	2.64	1.97	48	1.78	1.74		.462 1	.2251	2.239 8486	

7	5	1	2	1	2	1	69	46	36.482	8.098	23	29.81	9.66	.772	.3601	2.897 0361
8	6	1	2	1	2	1	58	39	13.9	3.033	19	14.1	2.1	-.072	-.0361	0.254 5242
9	7	1	2	1	2	1	55	28	23.41	3.4	27	22.44	3.8	.269	.1335	0.979 9409
10	8	1	2	1	2	1	100	50	5.105	2.29	50	4.025	2.40 5	.459	.2241	2.253 037

Effect Size Level Coding Manual

- Study ID number (STUDYID) – identification number of the study from which the offset size is coded.
- Effect size number (ES_N) – assign each effect size within a study a unique number such as 1, 2, 3, 4.....

❖ Dependent Measure Descriptors

- Effect size type (ES_TYPE) – pretest comparison = 1; posttest comparison = 2; follow-up comparison = 3.
- Category of outcome construct (OUTCOME) – vocabulary learning = 1; reading comprehension = 2; reading skills = 3; study time = 4.

- Measurement type (M_type) – recall protocol = 1; vocabulary test = 2; reading comprehension test = 3; survey = 4; Interview = 5.

❖ **Effect Size Data**

- Category of data effect size based on (ES_CAT)
 1. Means and standard deviations
 2. *t*-value or *F*-value
 3. *chi*-square (*df* = 1)
 4. Other
- Total sample size (Total_N).
- Treatment (text + visual) group sample size (TX_N).
- Treatment group mean (TX_Mean).
- Treatment group standard deviation (TX_SD).
- Control group (text only) sample size (CG_N).
- Control group mean (CG_Mean).
- Control group standard deviation (CG_SD).

- t -value (T_Value).
- F -value (df for the numerator must = 1) (F_Value).
- P - value (P-Value).
- Effect Size (d).
- Effect Size (r).
- Z-Value.

APPENDIX F

Technology Characteristics

	A	B	C	E	F
	Stud y ID	Leng_T ech_Use	TIME	Hyper_ CAT	Software
1	1		2	NA	Dreamw eaver 2.0
2	2.1		2	NA	HyperCa rd
3	2.2		2	NA	NA
4	2.3		1	<i>CyberB uch</i>	Author
5	3		1	<i>CyberB uch</i>	Author
6	4		1	<i>CyberB uch</i>	Author
7	5		2	GALT	Author
8	6		2	BANAI READI	Author

NGS				
9	7	2	NA	Author
10	8	2	NA	Author

Technology Level Coding Manual

- Study ID
- Length of technology use (Leng_Tech_Use)
- Total amount of technology treatment time (TIME) – less than 30 minutes = 1; 30 to less than 60 minutes = 2; one hour to less than two hours = 3; more than two hours = 4.
- Category of technology used (Tech_CAT) –
- Category of hypermedia used (Hyper_CAT) –
- Computer software (Software)

APPENDIX G

Effect Size Statistics

DATA ENTRY								RAW DIFFERENCE						STANDARDISED EFFECT SIZE			
Outcome measure	Treatment group			Control group			pooled standard deviation	p-value for difference in SDs	Mean Difference	p-value for mean diff (2-tailed T-test)	Confidence Interval for Difference		Effect Size	Bias corrected (Hedges)	Standard Error of E.S. estimate	Confidence Interval for Effect Size	
	mean	n	SD	mean	n	SD					lower	upper				lower	upper
001) Immediate test	4.7	30	0.952	4.03	30	1.586	1.31	0.00	0.67	0.05	-0.01	1.35	0.51	0.51	0.26	-0.01	0.51
study Immediate test	1.36	36	1.1	1.31	36	0.89	1.00	0.11	0.05	0.83	-0.42	0.52	0.05	0.05	0.24	-0.41	0.05
study Immediate test	3.75	103	1.89	2.15	103	1.72	1.81	0.17	1.60	0.00	1.10	2.10	0.89	0.88	0.15	0.60	0.89
study Immediate test	6.86	21	0.94	13.52	21	2.36	1.80	0.00	6.66	#####	-7.78	-5.54	3.71	3.64	0.50	-4.62	3.71
Immediate test	25.4	38	4.90	24.6	38	3.30	4.18	0.01	0.80	0.41	-1.11	2.71	0.19	0.19	0.23	-0.26	0.19
.1- nition Immediate test	3.15	50	2.33	2.76	47	2.20	2.27	0.35	0.39	0.40	-0.52	1.30	0.17	0.17	0.20	-0.23	0.17
.2- ion Immediate test	2.64	50	1.97	1.78	48	1.74	1.86	0.20	0.86	0.02	0.11	1.61	0.46	0.46	0.20	0.06	0.46
.1- Immediate test	8.54	50	3.14	7.87	47	2.78	2.97	0.20	0.67	0.27	-0.53	1.87	0.23	0.22	0.20	-0.18	0.23

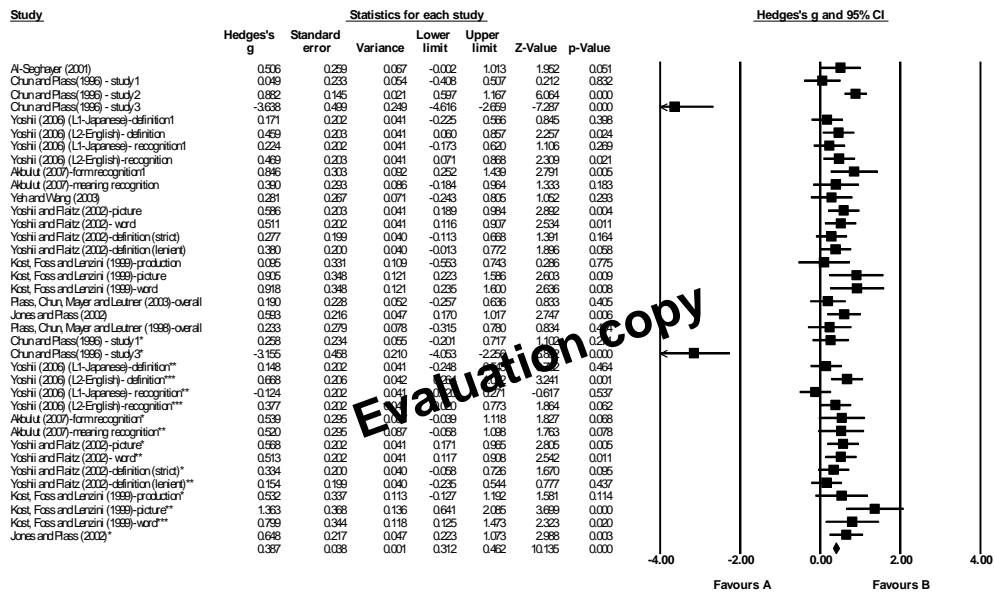
2- ition form	Immediate test	9.36	50	2.73	8.08	48	2.68	2.71	0.45	1.28	0.02	0.19	2.37	0.47	0.47	0.20	0.07
	Immediate test	35.3	23	5.04	30.17	23	6.76	5.96	0.09	5.13	0.01	1.59	8.67	0.86	0.85	0.31	0.24
ition	Immediate test	28.91	23	4.00	26.78	23	6.45	5.37	0.01	2.13	0.19	-1.06	5.32	0.40	0.39	0.30	-0.19
	Immediate test	23.41	28	3.40	22.44	27	3.40	3.40	0.50	0.97	0.29	-0.87	2.81	0.29	0.28	0.27	-0.25
z	Immediate test	7.46	50	2.53	5.98	50	2.48	2.51	0.44	1.48	0.00	0.49	2.47	0.59	0.59	0.20	0.19
z	Immediate test	7.58	50	2.60	6.12	50	3.05	2.83	0.13	1.46	0.01	0.34	2.58	0.52	0.51	0.20	0.11
z n	Immediate test	1.86	50	1.80	1.38	50	1.63	1.72	0.24	0.48	0.17	-0.20	1.16	0.28	0.28	0.20	-0.12
z n	Immediate test	3.52	50	2.24	2.62	50	2.46	2.35	0.26	0.90	0.06	-0.03	1.83	0.38	0.38	0.20	-0.02
	Immediate test	2.88	17	4.28	2.44	18	4.77	4.54	0.33	0.44	0.78	-2.68	3.56	0.10	0.09	0.34	-0.57
	Immediate test	8.47	17	3.04	5.33	18	3.69	3.39	0.22	3.14	0.01	0.81	5.47	0.93	0.90	0.36	0.21
word ayer (03)-	Immediate test	11.53	17	2.18	8.61	18	3.78	3.11	0.02	2.92	0.01	0.78	5.06	0.94	0.92	0.36	0.22
s	Immediate test	25.4	38	4.90	24.6	38	3.30	4.18	0.01	0.80	0.41	-1.11	2.71	0.19	0.19	0.23	-0.26
ayer (98)-	Immediate test	19.75	44	3.20	17.02	44	5.60	4.56	0.00	2.73	0.01	0.80	4.66	0.60	0.59	0.22	0.17
	Immediate test	40.4	25	30.00	33.5	25	28.30	29.16	0.39	6.90	0.41	-9.68	23.48	0.24	0.23	0.28	-0.32
study	Delayed test	1.61	36	1.23	1.33	36	0.89	1.07	0.03	0.28	0.27	-0.22	0.78	0.26	0.26	0.24	-0.21

study	Delayed test	7.29	21	0.82	13.51	21	2.61	1.93	0.00	-	####	-7.43	-5.01	-	-	0.46	-4.06
.1-	Delayed test	2.16	50	1.67	1.91	47	1.69	1.68	0.47	0.25	0.47	-0.43	0.93	0.15	0.15	0.20	-0.25
.2-	Delayed test	2.42	50	1.55	1.44	48	1.35	1.46	0.17	0.98	0.00	0.40	1.56	0.67	0.67	0.21	0.26
.1-	Delayed test	7.6	50	3.22	7.98	47	2.81	3.03	0.17	-	####	-1.60	0.84	-	-	0.20	-0.52
.2-	Delayed test	8.02	50	2.78	6.96	48	2.8	2.79	0.48	1.06	0.06	-0.06	2.18	0.38	0.38	0.20	-0.02
form	Delayed test	30.43	23	7.39	26.48	23	7	7.20	0.40	3.95	0.07	-0.33	8.23	0.55	0.54	0.30	-0.05
hition	Delayed test	27.17	23	5.23	24.13	23	6.22	5.75	0.21	3.04	0.08	-0.38	6.46	0.53	0.52	0.30	-0.07
z	Delayed test	6.48	50	2.67	4.92	50	2.78	2.73	0.39	1.56	0.01	0.48	2.64	0.57	0.57	0.20	0.17
z	Delayed test	6.06	50	3.11	4.62	50	2.42	2.79	0.04	1.44	0.01	0.33	2.55	0.52	0.51	0.20	0.11
z	Delayed test	1.14	50	1.63	0.68	50	1.04	1.37	0.00	0.46	0.10	-0.08	1.00	0.34	0.33	0.20	-0.06
z	Delayed test	1.98	50	2.2	1.68	50	1.61	1.93	0.02	0.30	0.44	-0.47	1.07	0.16	0.15	0.20	-0.24
z	Delayed test	2.59	17	3.24	1.11	18	2.11	2.72	0.05	1.48	0.12	-0.39	3.35	0.54	0.53	0.34	-0.14
h	Delayed test	8.12	17	2.29	4.78	18	2.49	2.40	0.37	3.34	0.00	1.69	4.99	1.39	1.36	0.38	0.63
word	Delayed test	8.59	17	2.53	5.78	18	4.11	3.44	0.03	2.81	0.02	0.45	5.17	0.82	0.80	0.35	0.11
s	Delayed test	14.08	44	4.02	11.15	44	4.9	4.48	0.10	2.93	0.00	1.03	4.83	0.65	0.65	0.22	0.22

APPENDIX H

Forest Plot for 37 Effect Sizes

Meta Analysis



Meta Analysis