

IMPLICIT CORRECTIVE FEEDBACK IN COMPUTER-GUIDED INTERACTION:  
DOES MODE MATTER?

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

By

Kenneth A. Petersen, M.A.

Washington, DC  
March 22, 2010

## ABSTRACT

Interaction research on recasts has been largely confined to examining the effects of recasts delivered in an oral, face-to-face modality. The challenge of creating task conditions in a written interactive mode, in which the dynamism and immediacy of recasts can be brought to bear on L2 development, has been a formidable obstacle to advancing the exploration of written recasts. This study examines the developmental effects of recast-intensive interaction on ESL question formation and morphosyntactic accuracy. It compares the effects of recasts delivered in oral (face-to-face) tasks with recasts delivered in analogous written (computer-guided) tasks. In a pretest/posttest design, 56 high school ESL learners engaged a series of communicative tasks with either a native English speaker or an intelligent virtual interlocutor. The results indicated that participation in recast-intensive interaction was a significant predictor of both ESL question development and improvement in morphosyntactic accuracy. The mode of interaction—oral or written—had no effect on development.

## ACKNOWLEDGMENTS

*In loving memory of Esther Mae Petersen (1932-1994).*

I am humbled to think of the enormous amount of support that I have received from the many people that have made this research possible. I am thankful foremost for having had the opportunity to work with Dr. Alison Mackey. Dr. Mackey's insightful guidance over the many years that she served as my mentor was critical in shaping the theoretical and methodological design of the study. Thank you, Alison, for having faith in my ability to execute such a complex thesis and for standing by and supporting me throughout the duration of my work.

I would also like to thank the other members of my dissertation committee for their invaluable insights and their continuous support. I owe an enormous debt of gratitude to Dr. Detmar Meurers for the many nights spent Skyping and days spent huddled over my laptop at CALICO conferences—thanks especially for pushing me towards the end. Thanks also go to Dr. David Yarowsky for his insights and feedback on the computational design of the study. I am also grateful to Dr. Kendall King and Dr. Jenifer Philp for their guidance during the proposal phase of this research.

In addition to the committee members, there are many other individuals in the Georgetown community that have contributed to this research. Thank you to Dr. Rusan Chen and Dr. Jeff Connor-Linton for their advice in the statistical analysis of the L2 assessment data. Thanks as well to Rebecca Sachs for her consistently brilliant advice. Big thanks, also to Erin Esch for putting up with me the past couple of years.

Throughout the pursuit of my doctoral degree, I have been employed at American Councils for International Education. There are several key individuals at American Councils that deserve thanks and praise for their support. A simple thank you does not begin to express the depth of gratitude I feel towards Dr. Dan Davidson and Dr. Masha Lekic for supporting my doctoral research and having the patience and flexibility to allow me to complete this enormous endeavour. Thanks also goes out to Shabnam Tafreshi for standing strong and holding down the fort throughout the data collection as well as to William Morse for his ongoing advice about the design and implementation of Sasha.

A special thanks is due to Mr. Bruce MacPherson, *artiste extraordinaire*, for providing the amazing illustrations used in the assessment and treatment tasks. Bruce, your enormous talent and unbounded generosity compete to define you. Please claim your reward. Thanks also to Dana Aeschliman for the early advice on running the statistics.

The data collection for this study was facilitated by a number of people at T.C. Williams High School and the Alexandria City Public School System. Thank you, Taki Sidley, for acting as my advocate, liaison and *amba yaluwa*. Appreciation is also extended to Susannah Courand (head of ESL), Dr. Mel Riddel (Principal) and Dr. Monte Dawson (Executive Director of Accountability, ACPS) for approving the administration of this research. Special thanks goes out to Carmen Hernan for her excellent work as a research assistant and to Diane Hixson and Susannah Courand for helping coordinate the scheduling of the research sessions. I am also profoundly grateful to the many teachers who graciously allowed their classrooms to act as research labs during their free periods.

Finally, I would like to thank my family. I would like to thank my father, Harold Petersen, for a lifetime of love and support. To my brother, John Petersen, and my sister, Kathy Rigdon—I am eternally in your debt for all that you do for dad. I would never have had the opportunity to pursue this direction in my life without the two of you.

Saving the best for last, I turn to thank Lara Atella—my wife, my best friend and my star research assistant. Thank you, Lara, for being such a patient and supportive partner, engaged mother and spot-on professional data collector.

## TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>i</b>
<b>ACKNOWLEDGMENTS</b>	<b>ii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xiii</b>
<b>LIST OF FREQUENTLY USED TERMS AND ABBREVIATIONS</b>	<b>xiv</b>
<b>CHAPTER I. INTRODUCTION</b>	<b>1</b>
1.1. Reconciling CALL and SLA	1
1.2. Oral Recasts in Interaction Research	4
1.3. Written Recasts in CALL/SLA Research	6
1.3.1. Computer-based feedback research	6
1.3.2. Computer-based recast research	8
<b>CHAPTER II. LITERATURE REVIEW</b>	<b>10</b>
2.1. Discourse in Computer-Centered Language Tasks	10
2.1.1. Overview of computer-centered discourse contexts	10
2.1.2. Locating discourse contexts	13
2.2. L2 Development and Feedback in CALL Research	33
2.2.1. The interface between CALL and SLA	33
2.2.2. L2 Development in CMC Research	35
2.2.3. Recasts in CMC Research	37
2.2.4. Feedback and Development in non-communicative (I)CALL	41
2.3. Recasts in Oral SLA Research: Saliency, Ambiguity and Context	47
2.3.1. Saliency of recasts	49
2.3.2. Ambiguity in learner perception	51
2.3.3. Interactional context	53
2.4. Research Question	54
2.4. Hypothesis	54
<b>CHAPTER III. METHOD</b>	<b>56</b>
3.1. Participants	56
3.2. Materials	58
3.2.1. Computers and lab space	58
3.2.2. Lab setup for oral tasks	59
3.3. Procedures	62
3.3.1. Target form	63
3.3.2. Experimental groups	65

- 3.3.3. Experimental schedule 66
- 3.3.4. Assessment tasks 67
- 3.3.5. Treatment tasks 72

## **CHAPTER IV. SASHA: SYSTEM DESIGN**

**79**

- 4.1. Overview 79
- 4.2. Preparing L2 Input for Analysis 81
  - 4.2.1. Filtering ‘noisy’ data 81
  - 4.2.2. Spell checking 81
  - 4.2.3. Disambiguating possessives and contractions 85
  - 4.2.4. Removing superfluous words and phrases 85
- 4.3. Syntactic Parsing 86
  - 4.3.1. L2 retraining of the syntactic parser 86
  - 4.3.2. Daemonized implementation of Collins Parser 87
- 4.4. Enriching the Input Question Data 90
  - 4.4.1. Post-hoc checks on parser output 92
- 4.5. The Input Question Module 92
- 4.6. Feedback Generation 95
  - 4.6.1. The Recast module 95
  - 4.6.2. Repairing syntactic errors 97
  - 4.6.3. Repairing morphological errors 103
  - 4.6.4. Recast confirmation 103
- 4.7. Generating a Semantic Response 104
  - 4.7.1. The activity model 104
  - 4.7.2. The Semantic Response module 111
- 4.8. Managing Discourse 116

## **CHAPTER V. SASHA: SYSTEM ANALYSIS**

**118**

- 5.1. Overview 118
- 5.2. Spell Checking 120
- 5.3. Recast Provision 120
  - 5.3.1. Good recasts 120
  - 5.3.2. Partial recasts 121
  - 5.3.3. Missed recast opportunities 123
  - 5.3.4. Bad recasts 128
- 5.4. Semantic Response Provision 132
  - 5.4.1. Valid responses 133
  - 5.4.2. Invalid responses 135
  - 5.4.3. Incomplete responses 138
- 5.5. Improving System Performance for Future Use 139

5.6. Conclusion *143*

## **CHAPTER VI. CODING**

**144**

6.1. Introduction *144*

6.1.1. Emergent stage-level development *144*

6.1.2. Accuracy *145*

6.2. Coding Procedures *147*

6.2.1. Coding written assessment data *147*

6.2.2. Coding oral assessment data *148*

6.2.3. Valid questions *150*

6.2.4. Coding variables *152*

6.2.5. Coding for stage level emergence *152*

6.2.6. Coding for stage-level accuracy *155*

6.2.7. Coding for morphosyntactic accuracy *158*

6.3. Interrater Reliability *160*

## **CHAPTER VII. RESULTS**

**165**

7.1. Introduction *165*

7.1.1. Mean distribution of questions *165*

7.1.2. Mean distribution of questions by modality *165*

7.2. Emergent Stage-level Development in ESL Question Formation *166*

7.2.1. Emergence: Descriptive statistics *166*

7.2.2. Emergence: Test for statistical significance *169*

7.3. Stage-level Accuracy in ESL Question Formation *171*

7.3.1. Stage-level accuracy: Descriptive statistics *171*

7.3.2. Stage-level accuracy: Test for statistical significance *175*

7.4. Morphosyntactic Accuracy *177*

7.4.1. Morphosyntactic accuracy: Descriptive statistics *177*

7.4.2. Morphosyntactic accuracy: Test for statistical significance *178*

7.5. Comparing increases in emergence and stage-level accuracy *180*

7.6. Summary *183*

## **CHAPTER VIII. DISCUSSION**

**184**

8.1. Summary of Results *184*

8.1.1. ESL question formation *184*

8.1.2. Morphosyntactic accuracy *184*

8.2. Explanation of Results *185*

8.2.1. Developmental gains in oral interaction *186*

8.2.2. Developmental gains in written interaction *188*

8.3. Implications	190
8.3.1. Implications for recast research	190
8.3.2. Implications for CALL research	193
8.3.3. Pedagogical implications	198
<b>CHAPTER IX. LIMITATIONS AND FUTURE RESEARCH</b>	<b>201</b>
9.1. Generalizability of Results	201
9.1.1. Duration	201
9.1.2. Target structures	202
9.1.3. Sample size	202
9.2. Interactional Context	203
9.2.1. Psychological factors of human-computer interaction	203
9.2.2. Written vs. oral input	205
9.2.3. Written vs. oral output	205
9.3. Recasts and Control	206
9.4. Limitations of Sasha	207
9.4.1. Recast provision	207
9.4.2. Semantic response provision	208
9.5. Future Instantiations of Sasha	209
<b>CHAPTER X. CONCLUSION</b>	<b>210</b>
10.1 Benefits and Challenges of Computer-Based Language Instruction	210
10.2 The Future of NLP-Driven Technologies in Language Instruction	213
10.3 The Future of CALL in Second Language Research and Instruction	217
<b>REFERENCES</b>	<b>219</b>
<b>APPENDIX A.1: EXAMPLE INTERACTION WITH EMDODIED AGENTS</b>	<b>234</b>
<b>APPENDIX A.2: EXAMPLE INTERACTION WITH HEIFT'S <i>E-TUTOR</i></b>	<b>238</b>
<b>APPENDIX A.3: EXAMPLE INTERACTION WITH SASHA</b>	<b>239</b>
<b>APPENDIX B: PARTICIPANT CONSENT FORM</b>	<b>241</b>
<b>APPENDIX C: RESEARCHER MANUAL</b>	<b>245</b>
<b>APPENDIX D: EXAMPLES FROM RESEARCHER TASK BOOKLET</b>	<b>277</b>
<b>APPENDIX E: EXAMPLES FROM STUDENT TASK BOOKLET</b>	<b>282</b>
<b>APPENDIX F: SASHA TREATMENT TASK EXAMPLES</b>	<b>287</b>
<b>APPENDIX G: DEBRIEFING QUESTIONNAIRE</b>	<b>290</b>
<b>APPENDIX H: SASHA'S RULE SET FOR ESL QUESTIONS</b>	<b>291</b>



## LIST OF FIGURES

2.1.	Computer-centered discourse contexts in CALL research.	13
2.2.	Discourse factors in communicative CALL environments.	14
2.3.	Communicative ICALL.	24
2.4.	Interfacing CALL with SLA findings on interaction and L2 development.	34
3.1.	Materials for oral data collection and sample task layout.	60
3.2.	Developmental stages and question forms.	64
3.3.	Experimental schedule.	67
3.4.	Example imitation task item.	69
3.5.	Example fill-in task item.	71
3.6.	Example email task item.	72
3.7.	Partial and complete recasts.	74
3.8.	Sasha.	76
4.1.	System overview.	80
4.2.	Spell checking and correction.	84
4.3.	Interaction between Sasha and the Collins Parser daemon.	89
4.4.	Preparing L2 question for analysis.	91
4.5.	Sasha::InputQuestion.	93
4.6.	Sasha::Recast.	96
4.7.	Recast generation.	98
4.8.	Rule-set matching example.	100
4.9.	Sasha: Semantic markup interface.	105
4.10.	Activity model: Database table: <i>senseMap</i> .	107
4.11.	Activity model: Database table: <i>locations</i> .	108
4.12.	Activity model: Database table: <i>possessions</i> .	109
4.13.	Activity model: Database table: <i>thematicRoles</i> .	109
4.14.	Sasha::Activity model.	110
4.15.	Semantic response.	112
4.16.	Semantically-enriched nouns in the Input Question.	113
4.17.	Example semantic responses.	116
5.1.	Example question from treatment data.	118
5.2.	System performance overview.	119

5.3.	Recast performance overview.	121
5.4.	Partial recasts.	122
5.5.	Recast missed opportunities.	124
5.6.	Missed recasts: Mal-rules.	125
5.7.	Missed recasts. System inaccuracy.	126
5.8.	Bad recasts.	128
5.9.	Bad recasts: System inaccuracy.	129
5.10.	Semantic response performance overview.	133
5.11.	Example semantic responses.	134
5.12.	Invalid responses.	135
5.13.	Invalid responses: Coding inaccuracy.	136
5.14.	Incomplete responses.	138
5.15.	Summary of coverage: Word sense disambiguation.	140
5.16.	Summary of coverage: Noun classification.	141
5.17.	Summary of coverage: Existential questions.	141
5.18.	Summary of coverage: Total mal-rules.	142
5.19.	Summary of coverage: Breakdown of mal-rules.	142
6.1.	Coding interface: Written assessment tasks.	148
6.2.	Coding interface: Oral assessment tasks.	150
6.3.	Coding formulaic questions for sample learner.	152
6.4.	Assessment Coding: Variables.	152
6.5.	Assessment coding: Target stage examples.	156
6.6.	Disambiguating Yes-No target stage.	157
6.7.	Sample stage-level accuracy score.	158
6.8.	Coding interface: Blind interrater arbitration.	162
6.9.	Three rater percentage agreement.	163
7.1.	Means of total valid questions by group.	165
7.2.	Means of total valid questions by modality.	166
7.3.	Mean distribution of stage-level question production.	167
7.4.	Stage assignments by group and testing interval.	169
7.5.	Stage-level attempts.	172
7.6.1.	Mean accuracy scores for stage 3.	173

<b>7.6.2.</b>	Mean accuracy scores for stage 4.	174
<b>7.6.3.</b>	Mean accuracy scores for stage 5.	174
<b>7.7.1.</b>	Group accuracy gains for stage 4.	175
<b>7.7.2.</b>	Group accuracy gains for stage 5.	176
<b>7.8.1.</b>	Mean accuracy scores for morphology.	177
<b>7.8.2.</b>	Mean accuracy scores for syntax.	178
<b>7.9.1.</b>	Group accuracy gains for morphology.	179
<b>7.9.2.</b>	Group accuracy gains for syntax.	180
<b>7.10.1.</b>	Mean accuracy scores for stage 4 by emergence gain.	181
<b>7.10.2.</b>	Mean accuracy scores for stage 5 by emergence gain.	182
<b>8.1.</b>	Debriefing question about noticing.	193

## LIST OF TABLES

<b>7.1.</b>	Means for stage-level productivity by experimental group.	167
<b>7.2.</b>	Mean increase for stage-level productivity by experimental group.	168
<b>7.3.</b>	Percentage stage increase by group.	168
<b>7.4.</b>	Stage-level movement by group.	169
<b>7.5.</b>	Pretest stage 5 learners by group.	170
<b>7.6.</b>	Stage-level increase for eligibles by group.	170
<b>7.7.</b>	Mean stage-level attempts by experimental group.	171
<b>7.8.</b>	Means for stage-level accuracy by group.	172
<b>7.9.</b>	The effect of treatment on stage-level accuracy.	175
<b>7.10.</b>	Means for morphosyntactic accuracy.	177
<b>7.11.</b>	The effect of treatment on morphosyntactic accuracy.	179
<b>7.12.</b>	Stage-level accuracy: developed vs. not developed.	182

## FREQUENTLY USED TERMS AND ABBREVIATIONS

### CALL

Computer-Assisted Language Learning (CALL)	1
Computer-Mediated Communication (CMC)	11
Asynchronous Computer-Mediated Communication (ACMC)	11
Synchronous Computer-Mediated Communication (SCMC)	11
Intelligent Computer-Assisted Language Learning (ICALL)	12
Communicative ICALL	23
Natural Language Processing (NLP)	7
Non-communicative (traditional) ICALL	23

### EXPERIMENTAL GROUPS

Computer-Guided Recast (CGR)	75
Control (CTL)	65
Face-to-Face (FTF)	73

### HCI

Artificial Intelligence (AI)	26
Conversational agent	25
Disembodied conversational agent	26
Embodied conversational agent	25
Full-duplex	16
Half-duplex	16

### SASHA

Activity model	104
Dialogue queue	116
Mal-rules	99
Part of speech (POS) tagger (Brill)	82
Semantic network (Wordnet)	107
Syntactic parser (Collins)	86

### SLA

Accuracy	145
Morphosyntactic accuracy	146
Stage-level accuracy	145
Developmental stages	64
Emergence	145
Interaction	4
Modality	11
Recast	4

## **CHAPTER I: INTRODUCTION**

This study comparatively explored the developmental effects of recast-intensive interaction on L2 development in two distinct modalities—oral, face-to-face interaction and written, computer-guided interaction. The oral treatment tasks occurred in a dyadic (face-to-face) laboratory setting while computer-guided interaction was carried out by an intelligent conversational agent (Sasha) that was capable of sustaining text-based conversational discourse with learners in a manner comparable to the discourse sustained in face-to-face interaction. The results of this study suggest that recast-intensive interaction is equally effective in the two modalities. These findings have implications for computer-assisted language learning (CALL) research and recast research as well as pedagogical implications for distance and blended learning classrooms.

The design of this study addressed many of the methodological challenges in equating written and oral conversation that have consistently plagued CALL research as it has tried to ground itself in second language acquisition (SLA) theory.

### **1.1. Reconciling CALL and SLA**

Computers have been viewed as a potentially beneficial tool for second language learning for several decades. As early as the 1970's, language research began exploring the nature of computer-based instruction. With the explosion of Web-based communication tools and advances in natural language processing (NLP) technologies, a broad array of computer-based language learning contexts has emerged. Yet, CALL research has largely been relegated to the margins of SLA research (Chapelle, 2004; Hulstijn, 2000) as it has struggled to contextualize itself within a larger theoretical framework of SLA (Chapelle, 1998, 2001; Doughty & Long 2003; Ortega, 2009;

Salaberry, 2000). The difficulty in reconciling CALL and SLA research stems, in large part, from the pronounced contextual differences between computer-centered and human-centered tasks. A major challenge for CALL research is to operationalize task conditions in such a way that there is, at least, qualified comparability to tasks performed under more traditional language learning conditions (Chapelle, 2001; Hulstijn, 1997; Ortega, 2009).

Any arguments one might make about language learning or development in computer-centered language tasks can, and indeed should, be informed by empirical findings derived from research in more traditional language learning conditions such as face-to-face interaction. However, the task conditions inherent in any form of computer-based interaction are likely to alter many of the underlying assumptions about interactional features and instructional techniques under exploration.

Early CALL research, in which tasks were generally conceptualized as sophisticated correlates of workbook exercises, was limited to discussing the effects of computer-centered tasks vis-à-vis self-monitored, written drill and practice activities. In recent years, however, there has been a strong shift in CALL research and design towards reframing computer-centered task contexts upon models of conversational interaction. This trend has been motivated by two significant technological advances—(1) the rapid proliferation of Internet-based communication tools and (2) the incorporation of increasingly sophisticated natural language processing (NLP) tools into CALL tasks. Each of these developments has enabled computer-centered language tasks to bear a closer resemblance to oral conversation than the more static, discrete-point tasks that typified earlier CALL applications. Yet, a host of empirical questions remains about the

potential mediating role of (1) presentational mode (written vs. oral) of input and output, (2) the allocation and sequencing of turns, (3) increased time for planning, production, repair and review, (4) decreased social pressure and (5) the absence of paralinguistic cues and facial expressions.

Task conditions in the experimental treatment tasks employed in the current study differed solely in that learners in the face-to-face (FTF) group orally asked questions of a native speaking (NS) researcher, while learners in the computer-generated recast (CGR) typed their questions and received visual, text-based feedback and answers via Sasha's dialogue bubble. Because they interacted with a computer, learners in the CGR condition produced questions in a context absent the social pressure and paralinguistic cues that may have been present in the FTF condition. Otherwise, the discourse conditions and provision of feedback in the two interactional modes were largely isomorphic.

Learners in the two conditions engaged in identical meaning-focused language tasks that required them to ask numerous questions about series of illustrated prompts. In both conditions, the recasts provided attempted to correct any and all errors found in the learner's question. No explicit opportunity was provided for learner uptake (i.e., learner response to feedback) because both Sasha and the NS researchers provided a semantic response to the learner's question immediately following the provision of the recast.

Sasha was able to sustain written conversational interaction that was temporally and sequentially more similar to oral discourse conventions than the interaction evidenced in synchronous computer-mediated communication (SCMC) research. The rapid response time of the system and the dyadic nature of the interaction in the CGR condition ensured that, unlike in SCMC, the adjacency of learner errors and corrective

feedback reflected that found in oral interaction. Interaction with Sasha also differed from SCMC in that learners did not have access to a written transcript of the conversation. Feedback and responses were presented for 4 seconds in Sasha's dialogue bubble and then cleared from the screen. As a result, learners in both conditions had to rely solely on their working memory abilities to make any cognitive comparisons between their output and the feedback they received.

## **1.2. Oral Recasts in Interaction Research**

The role of oral corrective feedback in promoting interlanguage development has engendered an extensive body of SLA research in recent decades. Within the context of conversational interaction in particular, a wide range of feedback types has been investigated in both classroom and laboratory contexts. The theoretical foundation of this research has largely been founded upon the Interaction Hypothesis (Long, 1996, Gass, 1997, Gass & Mackey, 2005), which suggests that corrective feedback obtained during conversational interaction plays a facilitative role in L2 development.

In particular, recasts, or the corrective reformulations of erroneous utterances (Farrar, 1990, 1992), have received considerable attention in SLA research. In communicative L2 classroom contexts, recasts are among the most common type of form-focused feedback and are generally provided incidentally in the course of meaning-focused interaction (Long, 2007; Lyster & Ranta, 1997). Unlike other, more explicit, forms of corrective feedback, recasts preserve the meaning of the learner's original utterance while making necessary lexical, morphosyntactic or phonological corrections (Long, 1996). To date, scores of experimental and quasi-experimental studies have examined recasts in both laboratory and classroom contexts (for a comprehensive review,

see Long, 2007). Converging evidence from this research suggests that recasts may be an effective instructional technique in facilitating L2 development in the context of communicative language tasks.

The inherent efficacy of recasts for L2 development is claimed to lie in the immediate juxtaposition of the learners' error and the correct reformulation provided by the more advanced speaker (Saxton, 1997, 2005). This immediate juxtaposition, it is argued, enhances the salience of the corrected linguistic form in the feedback (Ishida, 2004; Saxton, 1997) and provides an opportunity for learners to make a cognitive comparison between the targetlike model in the recast and their own non-targetlike production, thus promoting restructuring of their interlanguage representation of the form (Long & Robinson, 1998; Mackey & Philp, 1998; Oliver, 1995; Philp, 2003). Recasts are multi-faceted discourse structures that potentially represent negative feedback as well as positive evidence for the learner. It has been argued that the implicit negative feedback contained in recasts may not be as crucial to L2 development as the enhanced salience of the positive evidence they provide (Leeman, 2003).

Recasts are considered to be a relatively implicit form of negative feedback in that they provide no meta-linguistic information about the type of error that has occurred and, by and large, do not interrupt the natural flow of conversational interaction. Yet, the implicit nature of recasts has led some researchers to argue that the negative feedback entailed in recasts may not be as readily noticed by learners as other, more explicit, forms of corrective feedback (e.g., Lyster, 1998b).

Increasingly, recasts have come to be viewed as an elastic interactional move whose efficacy for second language learning is contingent upon the degree of elaboration

with which it is delivered (Doughty & Varela, 1998), the type of form-meaning connection it is intended to facilitate (Mohan & Beckett, 2001) and the pedagogical context in which it appears (Ellis, Basturkmen & Loewen, 2001; Nabei & Swain, 2002; Nicholas, Lightbown & Spada, 2001; Oliver, 1995; Sheen, 2004).

### **1.3. Written Recasts in CALL/SLA Research**

#### **1.3.1. Computer-based feedback research.**

A small body of research has begun to explore the efficacy of written, computer-based feedback in several distinct contexts—SCMC, intelligent CALL (ICALL) and traditional CALL tasks. This research draws heavily on findings from the exploration of feedback techniques native to oral interaction and is subject to a wide range of questions concerning ecological and construct validity of the feedback techniques under investigation. In each distinct computer-centered context, the quality of discourse is often constrained by technology and task design in ways that have distanced written feedback techniques from their oral corollaries beyond the question of the modality.

#### ***Feedback in SCMC tasks.***

Early CMC research analyzed language production from a largely socio-cultural perspective and focused on qualifying and quantifying the provision of feedback in text-based, networked learner-learner and learner-teacher interactions. In recent years, a number of SCMC studies have begun to examine the developmental benefits of written feedback from an interactionist perspective. This work has explored the extent to which SCMC interactional feedback leads to lexical development (Smith, 2004, 2005) and promotes L2 development and linguistic awareness (Chun et al., 2008; Lai & Zhou, 2006; Loewen & Erlam, 2006; Sachs & Suh, 2007). Findings from this research largely suggest

a positive relationship between synchronous written feedback and L2 development (cf. Loewen & Erlam, 2006). However, comparability to oral interaction has been constrained by the disjointed turn-taking sequences common in chat interaction and the presence of written transcripts.

### ***Feedback in ICALL.***

ICALL applications employ sophisticated NLP techniques in order to provide dynamic, customized feedback in response to learners' errors (Nagata & Swisher, 1995). In ICALL tasks, interaction is guided by computational mechanisms for parsing the syntactic and/or semantic content of a learner's utterance and, in turn, producing feedback and semantic responses. Within the ICALL literature, there has been a dearth of empirical research on the developmental benefits that text-based feedback provision may entail in the context of human-computer interaction (cf. Nagata, 1995, 1997a, 1997b; Heift, 2002, 2003, 2004, forthcoming; Heift & Rimrott, 2008). Yet, ICALL tasks have tended to be extremely restrictive in the amount of conversational opportunities they provide the learner, making any theoretical links to the oral literature on feedback somewhat tenuous. In ICALL tasks, learner production has largely been constrained by task design and the focus of the interaction has been on form and not meaning (but see Bailey and Meurers, 2008).

### ***Feedback in traditional CALL.***

While a few studies have explored oral feedback techniques in traditional CALL activities, the interactional context engendered by the task environments in this line of research bears little resemblance to spontaneous oral interaction. In traditional CALL tasks, both learner output is highly constrained and system responses must be pre-

scripted. In this context, interaction is overwhelmingly focused on form. Because traditional CALL interaction is neither communicative nor meaning-focused, it is difficult to operationalize instructional techniques such as recasts in a manner comparable to oral interaction.

### **1.3.2. Computer-based recast research.**

Almost everything that we know about recasts has been derived from evidence obtained in oral research. Though, as Long (2007) notes, there is a need for further research on the effects of written recasts because (1) the written modality is a robust environment for manipulating the degrees of saliency of target items and (2) findings from this line of research may have important pedagogical implications for teachers and materials developers in distance language instruction programs. Several recent studies have begun to explore the effects of recasts in written, computer-based interaction with mixed results. To date, written recast research has occurred in two interactional contexts—SCMC (Loewen & Erlam, 2006; Sachs & Suh, 2007) and non-communicative CALL (Ayoun, 2001,2004). Sachs & Suh (2007) found that the provision of visually enhanced (underlined and bolded) written recasts in dyadic chat NS-NNS interaction had a positive effect on levels of reported awareness of recast, though no relationship was found between the enhancement of a recast and subsequent developmental gains. Loewen & Erlam (2006) found recasts to be less effective than more explicit meta-linguistic feedback in large group SCMC. In contrast, Ayoun (2001, 2004) employed traditional CALL methodologies to demonstrate that written recasts, compared to explicit grammatical instruction, lead to a significant increase in pre-posttest gain scores.

Collectively, these findings provide only tentative preliminary evidence in regards to the nature and potential benefits of written recasts. Many questions remain concerning the validity of computer-delivered recasts and the extent to which empirical claims are generalizable between oral and written interaction research.

The results of the current study showed that participation in recast-intensive conversational interaction in both face-to-face and computer-generated communicative tasks promoted similar developmental gains in ESL question formation and morphosyntactic accuracy. These results have theoretical and methodological implications for both CALL research and oral SLA research. Specifically, this study provided empirical evidence that recast-intensive interaction may be effective in promoting L2 development in autonomous, computer-guided interaction. This positive finding has pedagogical implications for materials developers and educators evaluating whether recasts may be considered an appropriate and effective means of correcting learner errors in blended and distance-learning language courses.

## CHAPTER II: LITERATURE REVIEW

### 2.1. Discourse in Computer-Centered Language Tasks

#### 2.1.1. Overview of computer-centered discourse contexts.

Over the past 20 years, advances and innovations in communication technology have created a wide range of unique discourse contexts that may be potentially suitable environments for L2 development. The past two decades has witnessed the emergence of a host of synchronous and asynchronous communication tools that facilitate remote conversational interaction in both text-based and oral modalities. Tools such as email, chat/instant messaging, discussion boards, blogs, social networking sites have increasingly become well-established cultural artifacts among high school and college populations (Thorne, 2003) and have found a growing role in the modern L2 classroom (Rhodes & Pufahl, 2009; Thorne & Payne, 2005). Recognizing this shift in communicative practice in second language learning, many researchers have attempted to quantify and qualify the characteristics of L2 production and development in a variety of computer-mediated discourse contexts.

However, the discourse contexts produced by communicative, computer-centered L2 tasks are difficult to characterize vis-à-vis traditional models of either written or oral language production. Computer-mediated communication is an increasingly diverse construct and engenders a broad range of genres of L2 production (Thorne & Payne, 2005). In computer-mediated language tasks, the mode of communication is often visual and text-based—yet, it is subject to many online conversational constraints that are more commonly associated with oral interaction (Beauvois, 1992; Kitade, 2000; Tudini, 2004; Warschauer, 1998). This has led many to characterize computer-centered communication

as a distinct, hybrid form of communication (Abrams, 2003; Beauvois, 1992; Payne & Whitney, 2002; Smith, 2003; Sotillo, 2000; Sykes, 2005; Warschauer, 1998).

There is no single model of discourse that can capture all of the modal variations evidenced in the diverse forms of computer-centered communication. The specific discourse characteristics of any computer-centered communicative language task are determined by a number of contextual variables, including: synchronicity in turn sequencing, mode of interaction (written/oral), and a wide variety of context-specific constraints. The discourse contexts engendered by computer-mediated language tasks explored in the CALL/SLA research to date can broadly be placed in the following three categories: (1) asynchronous computer-mediated communication (ACMC); (2) synchronous computer-mediated communication (SCMC); and (3) intelligent computer-assisted language learning (ICALL). The current study employs ICALL techniques to sustain written conversation.

The predominance of computer-centered L2 interaction research has been conducted in the context of computer-mediated communication (CMC). CMC research focuses on the written discourse that occurs as learners engage in human-human interaction through the use of some form of networked communication technology. In ACMC, students engage in written L2 tasks that utilize asynchronous communication tools (such as email or discussion boards) to facilitate L2 communication. In SCMC, synchronous communication tools such as text or voice chat or instant messaging services are employed to elicit L2 interaction.

The majority of the research on CMC has been conducted in synchronous contexts. The few studies that have examined L2 communication in asynchronous

environments have been conducted, almost solely, in a written modality (cf. Poza, 2005). In SCMC research, the focus has largely been upon communication in a written modality—though some researchers have begun to explore the characteristics of L2 production in synchronous voice chat (e.g., Blake, 2005; Jepson, 2005; Satar & Özdener, 2008; Sykes, 2005; Volle, 2005; Wang, 2004).

In a wholly separate line of research, a growing number of studies have employed ICALL systems to explore human-computer interaction in L2 tasks and activities. Although there have been scores of articles and book chapters that have reported on ICALL projects, many are simply system descriptions and do not attempt to examine the nature of L2 production or development in human-computer interaction (Bailin, 1995; Oxford, 1995; Schulze, 2008).

Most ICALL systems are incapable of sustaining any form of communicative discourse and typically provide a model of interaction that might best be characterized as an enhanced extension of traditional workbook practice. These systems use natural language processing (NLP) to provide rich grammatical feedback in response to individual learner errors (Nagata & Swisher, 1995) in the context of grammar-focused language activities.

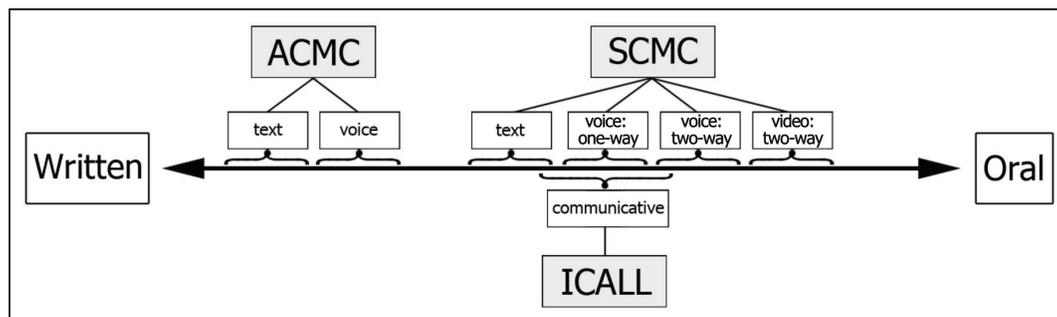
Only a few publications have described ‘communicative’ ICALL systems that employ a conversational agent capable of processing the semantic content of learner input and sustaining meaning-focused interaction (Bernstein, Najmi & Ehsani, 1999; DeSmedt, 1995; Ehsani, Bernstein & Najmi, 2000; Felshin, 1995; Hamburger, 1995; Kaplan & Holland, 1995; Kaplan, Sabol, Wisner & Seidel, 1998; Murray, 1995; Zacharski, 2003).

Yet, throughout the extant research, there are few empirical claims that might help explain the nature or potential benefits of communicative ICALL interaction.

### 2.1.2. Locating discourse contexts.

Figure 2.1 illustrates a rough approximation of where each of the distinct computer-centered discourse contexts explored in the CALL literature lies on a modal continuum between traditional written and oral task contexts. The predominance of the research has been conducted in SCMC. The discourse conditions in these studies varied widely and can be categorized in the following four sub-categories: (1) text-based chat or instant messaging (2) one-way online voice communication where only one speaker at a time can hold the floor, similar to an intercom system (3) two-way voice communication like voice-over IP teleconferencing and (4) two-way video communication where interaction is face-to-face.

Figure 2.1. Computer-centered discourse contexts in CALL research.



Note that each of the contexts illustrated in Figure 2.1 represents a range of possible modal instantiations that are characterized by the specific technologies and task conditions employed in the research methodology. The distribution of contexts on the

proposed continuum is (by no means) definitive, but is based upon the environmental variables listed in Figure 2.2.

Figure 2.2. Discourse factors in communicative CALL environments.

	ACMC		SCMC			ICALL
	<i>Text</i>	<i>Oral</i>	<i>Text</i>	<i>Half Duplex</i>	<i>Full Duplex</i>	<i>Communicative</i>
<b>Immediacy</b>	Low	Low	Moderate	Moderate	High	Moderate-high
<b>Planning/ Processing Time</b>	No limit	No limit	Somewhat limited	Somewhat limited	Limited	Somewhat limited
<b>Sequencing of turns</b>	Long Delay	Long Delay	Slight Delay	Slight Delay	Oral-like	Oral-like (strict)
<b>Conversational overlap</b>	Highly unlikely	Highly unlikely	Highly likely	Impossible	Oral-like	Highly unlikely
<b>Use of outside resources</b>	Very Possible	Very Possible	Difficult	Difficult	Very Difficult	Possible, if unproctored
<b>Role of computer</b>	Closed conduit	Closed conduit	Closed conduit	Closed conduit	Open conduit	Interlocutor

In traditional, face-to-face, oral interaction there is a high level of social immediacy that requires learners to produce spontaneous language in the immediate presence of an NS language expert and/or NNS peers. The social and performance pressures inherent in spoken conversation afford the learner a very limited amount of time to plan and produce language. This pressure to perform has been argued to be fundamental to the benefits of oral interaction (Bygate & Samuda, 2009; Gass, 2003). Also, as no written record is available to the learner, spoken interaction forces learners to rely on working memory capacity to process complex, multi-turn negotiation routines.

The sequencing of oral conversational turns and the likelihood of conversational overlap in oral interaction is governed by traditional social conventions for turn-taking—

where only one interlocutor may generally hold the floor at a time and any number of paralinguistic cues may mediate the turn-taking process. The use of outside language resources (e.g., pocket dictionaries) is extremely difficult to manage in oral conversation, as it tends to break the natural flow of interaction.

***Asynchronous computer-mediated communication.***

In both text and voice ACMC, conversational turns occur in a broad temporal window that obviates much of the social immediacy that is characteristic of synchronous conversation by allowing learners ample time to plan, process and review (Sauro, 2009). The delayed nature of interaction in asynchronous communication also minimizes the likelihood of conversational overlap and permits unlimited access to outside linguistic resources (Abrams, 2003). In any communicative task interaction, the use of external linguistic resources, such as dictionaries, grammar textbooks or online translators could have an unexpected impact on the quality L2 learning facilitated by the task. In delayed, online task contexts, it is extremely easy for learners use outside resources to enhance and refine any spontaneous output that they might produce.

***Synchronous computer-mediated communication: Written.***

In text SCMC, communication is also written, though conversational turns occur within a relatively short temporal window that more closely resembles that found in oral communication (Abrams, 2003). Yet, compared with oral interaction, textual SCMC affords learners considerably more time to process and review the utterances of their interlocutors and to plan and produce their L2 output (Sauro, 2009; Warschauer 1995). Conversational overlaps and interruptions are highly likely in text SCMC contexts (Lai &

Zhou, 2006; Smith & Sauro, 2008) —especially in large group conversations. In text SCMC, the use of outside resources is feasible, yet potentially cumbersome.

***Synchronous computer-mediated communication: Voice.***

Some more recent SCMC tools have been designed to facilitate oral online communication—although research employing these tools remains sparse. Depending on the technology employed, spoken online communication may be restricted to one-way (half-duplex<sup>1</sup>) exchanges or the open two-way (full-duplex) exchanges. In one-way voice communication it is only possible for one interlocutor to hold the floor at a time—similar to a walkie-talkie or intercom. Two-way communication, such as that enabled by tools like Skype or Apple’s iChat allow open, two-way exchanges in a discursive context similar to that found in telephone or video conferencing conversations.

In one-way voice CMC, the sequencing of turns is largely governed by the restrictions imposed by the technology. Speakers in this context have a physical device for taking and holding the conversational floor and do not have to rely upon any knowledge of the social and discursive cues found in traditional oral interaction. In this context there is no possibility for conversational overlap and the use of outside resources would likely be cumbersome.

In two-way CMC, many of the characteristics of interaction are nearly isomorphic with those found in traditional dyadic and small group oral interaction. In this context, the computer serves simply as an open conduit for communication—placing no technological constraints on the sequencing of turns other than a possible network delay. In full-duplex video, visual cues—such as gestures and facial expressions—are also

---

<sup>1</sup> In telecommunications, the term *duplex* simply refers to the ability of an oral communication device to transfer information through language in real time.

transmitted in real-time. To date, research in this area remains largely unexplored. Full-duplex CMC tools have only become widely available with recent global increases in processing speed and bandwidth capabilities. While full-duplex technologies present the optimal context for emulating face-to-face interaction, they are more cumbersome for use in SLA research than text-based forms of CMC. Unlike text-based communication tools, which produce, *de rigueur*, an accurate transcript of all interaction, oral CMC interaction needs to be recorded and transcribed by hand—as in face-to-face interaction.

In sum, CMC is not a unitary form of communication. Rather, it can take many disparate forms that affect the quality and quantity of interaction in which learners engage. It is not a trivial matter to explain how the communicative conditions present in any given CMC task affect instruction and learning.

#### ***Comparative modality studies in CMC research.***

A growing body of CMC research has begun to comparatively explore the effects of discourse conditions on L2 production and development. To date, research has compared text-based CMC to the following modalities: (1) face-to-face oral interaction (Abrams, 2003; Beauvois, 1992; Bump, 1990; Fitze, 2006; Kern, 1995; Payne & Whitney, 2002; Payne & Ross, 2005; Sykes, 2005; Warschauer, 1996), (2) oral SCMC (Jepson, 2005; Satar & Özdener, 2008; Sykes, 2005) and (3) text-based ACMC (Abrams, 2003; Sotillo, 2000).

#### ***Text vs. oral, face-to-face.***

Several early descriptive CMC studies examined the effect of modality on L2 production in group discussions by comparing text-based CMC interaction with face-to-face interaction. The central claims put forward by the early comparative CMC research

argue that synchronous written conversation provides a beneficial environment for language learning in several ways. Firstly, CMC promotes greater, more equal, participation and leads to a greater amount of L2 production (Beauvois, 1992; Bump, 1990; Kern, 1995; Warschauer, 1996). Second, the quality of L2 production may be superior in CMC group discussions than that produced in similar face-to-face contexts. CMC production has been found to contain a broader range of morphosyntactic features (Kern, 1995) and a wider variety of discourse functions (Kern, 1995; Warschauer, 1996). The increased quantity and quality of learner output in early CMC research was attributed to the ability of group participants to simultaneously take the conversational floor (Kern, 1995) as well as to the anonymous (Bump, 1990) and student-centered (Kern, 1995; Warschauer, 1996) nature of CMC discourse and students positive attitude towards written interaction (Beauvois, 1992; Warschauer, 1996).

In recent years, several more descriptive and quasi-experimental studies have gone on to continue exploring differences between text SCMC and oral conversation. This strand of CMC research has examined: (1) the quantity of L2 production (Abrams, 2003; Fitze, 2006), (2) the extent to which benefits acquired through interaction in text-based SCMC might transfer to a learner's oral performance (Abrams, 2003; Payne & Whitney, 2002; Payne & Ross, 2005), (3) the noticing of errors and feedback (Lai & Zhou, 2006) and (4) the effect of modality on pragmatic development (Sykes, 2005). The claims from these studies suggest that modality has little effect on overall quantity of L2 production in treatment and/or subsequent assessment tasks—neither Abrams nor Fitze found any effect for modality on quantity of production—calling into question earlier claims on greater productivity in CMC. The results related to transferability

(discussed in more detail below) are mixed in regards to the effect of treatment task modality on subsequent performance on oral assessment tasks. While Payne and Whitney (2002) found that language practice in text-based SCMC provides benefits for subsequent oral performance, Abrams (2003) failed to find a similar transfer effect. In terms of pragmatic development, Sykes (2005) found a significant effect for modality on the acquisition of speech acts, wherein SCMC intervention outperformed oral.

Payne and Whitney (2002) investigated the effect of written CMC interaction on production in subsequent oral production. Students from four intact Spanish classes engaged one of two conditions: +SCMC and –SCMC. The study was carried out over the course of an entire semester, during which the +SCMC classes conducted two of their four weekly group discussion in written chat. The –SCMC group received identical instruction throughout the semester, only all of the whole-class discussions occurred in an oral modality. Payne and Whitney operationalized oral proficiency development on a broad 50-point scale that considers comprehensibility, fluency, lexical usage, morphosyntax and pronunciation. Comparing pre and post-test oral proficiency scores through the prisms of Levelt's (1989) model of language production and working memory theory, the authors found greater gains in oral proficiency for participants in the +SCMC condition.

Payne and Ross (2005) employed discourse and corpus analysis techniques to reanalyze the longitudinal learner output in the SCMC tasks carried out in Payne & Whitney (2002) in order to examine the extent to which individual differences in working memory capacities constrain overall language output and lexical repetitions. The findings of this study suggested a positive relationship between working memory

capacity and overall language production in SCMC, though working memory differences could not account for a decline in the use of lexical repetitions that occurred as the semester progressed.

Abrams (2003) also examined the effects of CMC versus oral group interaction on subsequent oral performance. Comparing pre and post oral assessment task performance on three experimental groups—SCMC (text chat), ACMC (threaded discussion board) and oral, face-to-face group interaction—Abrams found a significant effect for modality on the overall quantity of L2 German production. Learners in the SCMC group outperformed those in both the ACMC and oral groups. Analysis of lexical density and complexity in learners' oral assessment task performance did not reveal any significant effect between the groups, though a descriptive analysis of the data suggested that: (1) learners in the oral condition exhibited greater gains than their counterparts in the CMC conditions in terms of lexical richness and (2) learners in the SCMC group outperformed the other groups in terms of lexical density. Abrams presents some preliminary evidence that suggests that ACMC may not be as beneficial for subsequent oral performance as either mode of synchronous communication. However, no effect was found that might help explicate the difference between text SCMC and oral interaction. This may be due, in part to the relatively limited time of exposure in the treatment tasks—learners only engage in a single 50-minute group discussion.

Lai and Zhou (2006) employed a within-group design on a group of 12 ESL learners in which each student engaged in one face-to-face and one SCMC chat session. The following day, a stimulated recall session was administered to each learner to explore the extent to which interactional feedback (negotiation of meaning and recast) and self-

correction were noticed in each modality (see below for more on noticing in Lai and Zhou). This exploratory study makes some modest claims regarding the effect of modality on the noticing of interactional features and linguistic errors in synchronous communication. The authors found that learners reported noticing a significantly greater number of instances of self-correction in their written transcripts than in the videotape recordings of their oral interaction.

Sykes (2005) compared the effects of voice and text SCMC vs. oral group interaction on the acquisition of pragmatic competence in triadic group interaction in Spanish L2 tasks. Sykes focused on the acquisition of speech acts related to the refusal of invitations and found that learners in the oral, face-to-face condition significantly outperformed their counterparts in the voice SCMC condition—though a similar relationship between oral interaction and text SCMC failed to emerge.

Fitze (2006) examined the quantity and lexical complexity of discourse produced in large group ESL discussions conducted in both oral and SCMC conditions. Two intact classes engaged in four 20-minute discussions over the course of four weeks. In a counterbalanced design, each class engaged in two SCMC (text chat) and two oral whole-class discussions. Contrary to the earlier comparative research, Fitze found no difference in the overall quantity of production between the two groups. However, a comparative analysis of lexical density found that learner speech was lexically more complex in CMC interaction than in similar oral group discussions.

*Text vs. voice in computer-mediated communication.*

A few recent CMC studies have begun to comparatively explore the characteristics and effects of text and half-duplex voice SCMC. Sykes (2005)

experimentally exposed learners to oral, text SCMC and voice SCMC interaction in order to compare the extent to which discursive context affects pragmatic development (see above). Sykes found no significant effect for modality in SCMC.

Jepson (2005) compared text and voice ESL chat interaction by analyzing a wide range of repair moves (clarification requests, confirmation checks, self-repetitions, recasts, explicit corrections, and uptake) in NNS-NNS dyads. Jepson found a greater amount of repair moves (negotiation and negative feedback) in voice chats (largely related to pronunciation) than in text chat.

Satar & Özdener (2008) examined the comparative effects of text and voice-based SCMC on (1) the development of oral speaking skills and (2) learner anxiety levels in dyadic, computer-centered NNS-NNS interaction in a series of four 45-minute tasks spanning four weeks. Similar to Payne and Whitney (2002), spoken ability was broadly measured according to a custom rubric, developed by Hughes (2003), that considered pronunciation, lexis, morphosyntax, fluency and comprehension. Based on ANOVAs comparing between-group differences on pre and post-test oral performance among learners who engaged in voice and text chat sessions in the course of their regular instruction, Satar & Özdener concluded that voice chat significantly outperformed control, but no difference was found between voice and text chat. The authors also reported a decrease in learner anxiety in the text condition as self-reported by the learners in a debriefing questionnaire.

*Synchronicity in written computer-mediated communication.*

Two studies have directly explored the role of synchronicity in the quality and potential benefits of CMC interaction. Abrams (2003), discussed above, found that the

asynchronous condition lead to significantly fewer gains in the quantity of L2 production subsequent oral tasks. Abrams also found descriptive evidence that the ACMC group used more subordinate, relative, and infinitive clauses during their computer-mediated communication, though this usage did not transfer to their oral exchanges. Sotillo (2000) explored qualitative and quantitative differences between language produced in ACMC and SCMC as ESL learners engaged in group discussions of assigned reading texts. Findings from this study suggest that ACMC interaction is more constrained in terms of discourse functions, yet syntactically more complex than SCMC.

***'Communicative' ICALL.***

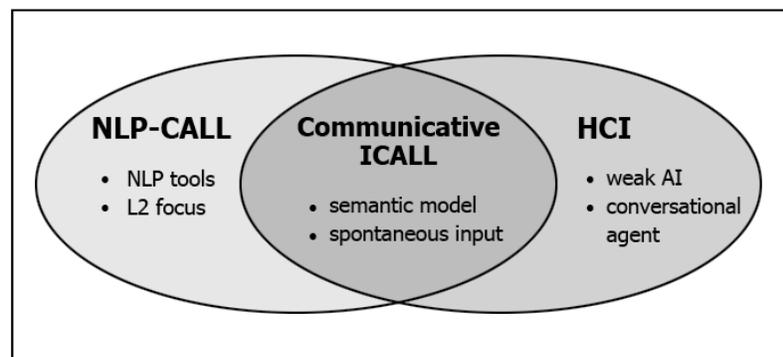
Unlike CMC language tasks, ICALL tasks do not simply facilitate human-human interaction in a computer-based context. Rather, they provide opportunities for self-instructed learning in which a learner interacts with a linguistically 'intelligent' computational system. Garrett (1995) identified two distinct types of ICALL systems—those that focus on language as a linguistic system and those that focus on language as communicative behaviour and argued that, in either case, ICALL development needs to operate within well-attested theoretical frameworks (see also: Oxford, 1993, 1995).

The nature of the interaction facilitated by the majority ICALL systems explored in the existing literature cannot be characterized as *communicative* in any way comparable to oral interaction. This is primarily due to the fact that many ICALL task constructs have been modeled upon self-instructed learning paradigms typically found in workbook exercises and traditional CALL applications and do not provide opportunities for spontaneous conversational interaction. In these non-communicative contexts, it is

impossible to create contexts for incidental focus on form since there is no primary focus on meaning.

In order to create and sustain communicative discourse, an ICALL system must employ a broad range of NLP technologies and techniques and go beyond processing just lexis and morphosyntax. To qualify as communicative, an ICALL system must additionally provide mechanisms for handling semantics and discourse management and should be capable of (1) providing appropriate responses to a learner's questions or propositions and (2) managing an ongoing conversation with the learner. Ideally, communicative ICALL tasks allow the learner to produce spontaneous input.

Figure 2.3. Communicative ICALL.



As Figure 2.3 illustrates, communicative ICALL—what Thorne & Payne (2005) refer to as ‘intelligent computer-mediated communication’ (ICMC)—combines L2 parsing techniques found in non-communicative ICALL with the use of a conversational agent, similar to those found in human-computer interaction (HCI) research.

*Communicative ICALL as an extension of Human-Computer Interaction.*

HCI is a broad discipline that examines how people interact with computers and to what extent computers might facilitate successful interaction with human beings. HCI

research has explored a broad range of physical, ergonomic and linguistic factors related to the way humans interact with computers. From a linguistic point of view, HCI has focused primarily on computer interaction with native speakers. Communicative ICALL employs methods and techniques similar to those used in HCI research, but focuses interaction in an L2 context.

One strand of HCI research of particular interest to communicative ICALL is the use of conversational agents. Conversational agents have been employed extensively in HCI research and have been shown to be capable of carrying out extended conversations with human interactors (Cassell, 2001; Cassell & Tartaro, 2007). There is also evidence that humans are willing to apply face-to-face-like communication strategies when engaging with competent conversational agents (Gieselmann & Stenneken, 2006; Kopp, 2006). Yet, conversational agents are restricted by the sophistication and accuracy of the systems that govern their understanding and use of the language. It has been argued that learners may adapt their language to meet the constraints of the dialogue system with which they are engaging (Bell, 2003; Le Bigot, Terrier, Amiel, Poulain, Jamet & Rouet, 2007).

Conversational agents may engage with a human interlocutor either an embodied or a disembodied form. In the past decade, many HCI experiments have been conducted that have employed visual, embodied characters that serve as virtual interlocutors for human conversational partners in dyadic interaction. Individual embodied agents have been designed to respond to either spoken or written input and may be capable of producing spoken, written or even multimodal output. The visual representations of embodied agents in HCI research have varied widely and have included both human and

non-human (robotic) incarnations capable of a wide range facial expressions and physical gestures. Visual paralinguistic cues have been simulated using scripted animations and micro-video clips that attempt to express mood and emotion and simulate the facial appearance of language of oral language production (see Appendix A.1 for examples embodied agents).

HCI experiments have also extensively explored human interaction with non-visual, disembodied conversational agents. A disembodied agent generally has a name (e.g., ELIZA, ALICE, etc.), but no visual representation. The utterances produced by the agent are presented to the human conversant in a manner similar to that found in CMC interaction.

HCI models have been driven by both strong and weak forms of artificial intelligence (AI). In its strong form, AI attempts to produce language that is spontaneous and context-independent. Strong AI systems employ sophisticated machine learning techniques to create agents that are designed to be capable of handling any input produced by a human interlocutor. Weak AI systems, on the other hand, rely upon restricted models of domain knowledge that limit the range of information that the conversational agents must account for. To be certain, human-computer interaction is still far from perfect. To date, no model of computational interaction, in either the private or public sector, can be argued to consistently pass the Turing test—wherein a computationally-driven conversational interaction is indistinguishable from human-human interaction.

In communicative ICALL contexts, conversational agents are (by necessity) guided by weak forms of AI, in which the conversational agent is governed by a

restricted domain knowledge and is incapable of spontaneous or independent thought (Heift & Schulze, 2007, p. 116.). Strong AI-driven conversational agents are wildly unpredictable and do not lend themselves to productive conversational interaction in the context of a language learning task (Zacharski, 2003).

*Conversational agents in communicative ICALL.*

A small number of ICALL projects have employed either embodied or disembodied conversational agents to facilitate conversational interaction in computer-centered L2 tasks. The Hypercard-based system Subarashii (Bernstein, Najmi, & Ehsani, 1999; Ehsani, Bernstein & Najmi, 2000) is the only documented use of embodied conversational agents in an ICALL context. Subarashii was designed to match restricted spoken Japanese input to a dialogue model that uses a finite-state architecture to generate pre-scripted video responses to learners' written input. The published discussions of Subarashii are largely system descriptions and stop short of offering any evidence-based findings that might help describe the nature or benefits of interaction with an embodied conversational agent.

Disembodied conversational agents have been employed in ICALL projects to facilitate role-play tasks (DeSmedt, 1995) or the manipulation and description of objects in virtual microworlds (Hamburger, 1995; Felshin, 1995; Kaplan & Holland, 1995; Kaplan, Sabol, Wisher & Seidel, 1998; Murray, 1995). The language produced by these disembodied agents is presented to the learner in a text-based format that resembles the way messages are displayed to a learner in a chat or instant messaging context.

Several ICALL projects have explored the feasibility of dynamic visual microworlds as a vehicle for promoting communicative L2 interaction. Microworlds are

visual representations of complex discrete environments, such as a virtual kitchen or office. L2 tasks in microworld environments ask students to interact with the computer to describe and manipulate the state of the visible objects. The semantic framework for controlling interaction within these worlds relies upon a model of knowledge representation that describes the visible objects and potential actions that may be performed upon them.

The FLUENT system (Hamburger, 2005) was developed at MIT in the early 1980's and later incorporated into the larger Athena Language Learning Project (Felshin, 1995; Murray, 1995). L2 tasks in the FLUENT system afford learners the opportunity to collaborate with a disembodied agent to complete complex goals (e.g., boiling a pot of water) within a visual microworld. Semantic and pragmatic knowledge representation in FLUENT relies on a system of interaction schemas and object relationships to facilitate sustained goal-oriented conversational interaction. Learners are not, by themselves, capable of manipulating the microworld environment—they must rely on language to communicate their intentions to a disembodied agent. The agent may respond by either completing the requested action or commenting on the state of the target objects (e.g., 'The cup is full.').

The MILT system (Kaplan & Holland, 1995; Kaplan et al., 1998), developed for self-instructed language training by the U.S. Army Research Institute, also makes use of a disembodied agent in a microworld environment. In its early years, MILT was capable of purely text-based interaction (Kaplan & Holland, 1995) and later went on to employ automatic speech recognition (ASR) technology to accept and interpret spoken Arabic sentences and phrases (Kaplan et al., 1998). The use of ASR in spoken human-computer

interaction is, indeed, a precarious undertaking. Even in L1 contexts, it is extremely difficult to train a system to make accurate sense of a wide variety of spoken language. When spoken input is non-native, the challenge becomes exponentially more difficult. Spoken input in MILT tasks is therefore highly restricted in order to limit the demands placed on its underlying ASR processing architecture. Learners are not asked to engage in spontaneous spoken interaction with the agent—rather, they are asked to produce spoken output that corresponds to any one of three potential target-like L2 models (commands) presented to them on the screen in a written format. The system needs only be capable of handling input that corresponds to a selected model. Task interaction advances when a learner’s spoken input is matched to the selected model command. As in the FLUENT system, the agent moves the task forward by either completing the requested action or commenting on the state of the target objects.

While FLUENT and MILT have been extensively described in several research reports, none of the published works provides any empirical evidence about learner output or the potential benefits of L2 interaction with a disembodied agent in a microworld learning environment.

Another early ICALL project, Herr Commisar (DeSmedt, 1995), features a pair of conversational agents with which L2 German learners interact. Herr Commisar situates learners in a role-play detective game wherein learners type questions about a crime and a disembodied agent (‘Hoffman’) provides a semantic response. Hoffman’s responses appear to the learner exactly as if he were a human interlocutor in a chat session. A second agent, ‘Sgt. Schulz’ intervenes between Hoffman and the learner and may interrupt the conversation to point out morphosyntactic errors. Schulz’s feedback is

presented in a popup dialogue box, which includes a visual icon of a male face—providing the agent Schulze with some minimal visual embodiment. As with all of the other communicative ICALL research, DeSmedt focuses largely on system description and provides no empirical data relevant to SLA.

*Locating communicative ICALL.*

ICALL research, on whole, has always struggled to establish itself within a well-grounded theoretical SLA framework (Oxford, 1993, 1995; Schulze, 2008). To date, communicative ICALL research, in particular, has yet to empirically explore either L2 production or development in a human-computer context. Thus, the location of communicative ICALL on the modality continuum proposed in Figure 2.1 is largely speculative and may vary according to myriad methodological constraints, such as the mode of input, the realism of the conversational agent and the speed, scope and accuracy of the system. In its idealized form, ICALL interaction might be barely distinguishable from oral conversation—an optimal ICALL system might accept spontaneous spoken learner input and use a lifelike agent to provide semantically and pragmatically appropriate oral responses in a manner that closely resembles human speech—both visually and linguistically.

In its current state, however, ICALL methodology remains limited in its ability to realize such an idealized model. The criteria in Figure 2.2, help to explain the placement of communicative ICALL on the written-oral continuum in Figure 2.1.

1. Immediacy.

The immediacy of conversational interaction in ICALL task environments must be considered along two separate dimensions—temporal and social. Depending on the

speed of the particular system, the temporal immediacy of ICALL interaction might best be described as resembling highly controlled task-based dyadic oral interaction with increased time for planning and production. In many modern ICALL systems, responses are delivered with almost no delay. Unlike text-based SCMC, ICALL interaction is not contingent upon the effortful intervention of a human interlocutor. ICALL systems are thus capable of maintaining oral-like response intervals while attending to both form and meaning.

There are many lingering empirical questions regarding the social immediacy (Abrams, 2003) of ICALL interaction. For example, what social pressures might a learner experience when engaging with a non-human agent and how might those pressures affect the learner's L2 performance? To date, no evidence has been provided that might help characterize L2 learner perceptions when engaging in human-computer interaction. When they communicate with an ICALL system, learners are obviously aware that they are interacting with a piece of software, yet little is known about how the absence of the 'human element' of a conversational partner may affect an L2 learner's performance anxiety or willingness to test hypotheses in the target language.

## 2. Planning/processing time.

ICALL tasks afford learners an ample amount of planning time as they compose their output. Unlike oral communication or SCMC, interactional sequences in ICALL tasks are always initiated by the learner—she therefore need not fear that the agent might move the conversation forward if she spends too much time composing any particular utterance. The sense of urgency in the production of L2 output in ICALL task interaction may, however, be motivated by task external conditions such as a running timer or game-

like task goals. As in text-based SCMC, if a transcript of the interaction is made available, the learners are also afforded additional time for reviewing and processing language.

### 3. Sequencing of turns/conversational overlap.

Communicative ICALL systems are capable of facilitating extended turn-taking sequences that closely resemble those found in strictly controlled dyadic oral interaction. By design, most ICALL systems feature interaction that is learner-initiated. In other words, the system sits idle until the learner triggers a sequence of responses by submitting some form of language input. Thus, conversational overlap in ICALL interaction is highly unlikely—especially when system responses are promptly delivered.

### 4. Use of outside resources.

In communicative ICALL, the use of outside resources such as online translators is possible given the low temporal immediacy of the discourse. However, in proctored conditions, it is relatively simple to control for the use of outside resources.

As this review illustrates, computer-centered discourse may occur in a wide variety of disparate contexts. Each of these distinctive contexts represents a unique, emergent genre of L2 interaction that differs significantly from traditional forms of expression such as essay writing or face-to-face speech. When examining the effects of instructional interventions such as feedback in a computer-centered context, it is important to consider how the ecological constraints of the selected context might affect the underlying characteristics of the instructional technique and its relationship to cognition and learning.

## **2.2. L2 Development and Feedback in CALL Research**

### **2.2.1. The interface between CALL and SLA.**

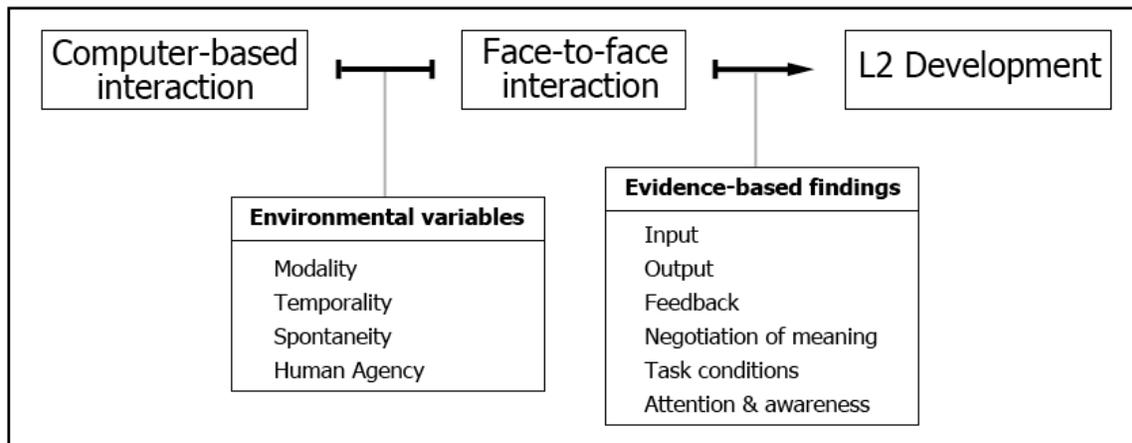
From its inception, CALL research has been faced with challenge of explaining language use and development vis-à-vis linguistic research conducted in non-technical learning environments (Bailin, 1995; Chapelle, 1997, 2001; Oxford, 1995; Schulze, 2008;). CALL tasks share the same pedagogical goals as traditional SLA tasks—to provide language practice and instruction that results in some form of measurable developmental gains. It is logical to assume, therefore, that CALL research might draw upon the findings from experimental research conducted in more traditional SLA settings. However, when situating CALL research within larger theoretical frameworks of SLA, it is incumbent upon the researcher to address the environmental variables that distinguish the selected form of CALL interaction from that which occurs in a traditional SLA environment.

In any context, linking features of conversational interaction to L2 development presents a considerable challenge. Mackey (1999) notes that, to make this connection, one must devise tests that can directly measure development of the structures that occur in interaction and find a means to operationalize second language development. It is also vital to target specific linguistic structures and create tasks that provide numerous contexts for the target structures to occur. Only a small subset of CALL research has attempted to move beyond describing the characteristics of computer-centered interaction and explore the effects of this interaction on L2 development.

In recent years, a number of CALL studies have turned to oral interaction research as a theoretical framework for explaining developmental gains in computer-centered

language tasks. Figure 2.4 illustrates areas where claims have been made about L2 development in face-to-face interaction and some of the environmental variables that CALL research must address when discussing instructional interventions and language development based upon the assumptions and findings of face-to-face interaction research.

Figure 2.4. Interfacing CALL with SLA findings on interaction and L2 development.



Any of the environmental variables listed in Figure 2.4 might arguably mediate the effects of interaction on L2 development. For example, in synchronous online communication (SCMC and ICALL), the contingency and juxtaposition of utterances varies widely in the extent to which it might be considered comparable to oral interaction. In the oral literature, these temporal features have been argued to be essential for creating optimal conditions for instructional interventions such as recasts (Cowan, 1988; Long, Inagaki, & Ortega 1998; Ohta, 2000; Saxton 1997, 2005). If the contingency and juxtaposition of utterances in a computer-centered task differ widely from what one might expect to find in oral interaction, it becomes a challenge to reconcile any developmental claims about computer-delivered recasts with the large body of evidence

on orally-delivered recasts that has consistently assumed these features to be somewhat stable (Chun, Fei & Roots, 2008).

### **2.2.2. L2 Development in CMC Research.**

Early CMC research largely set out to describe and analyze the quantity and quality of L2 production in various text-based CMC contexts. However, this work stopped short of providing any empirical, theoretically grounded claims about the potential developmental effects of computer-based interaction. By the turn of the century, many salient features of L2 production in interactive written discourse had begun to emerge, including: (1) CMC interaction occurs in a relatively equitable, non-threatening and student-centered context (Kern, 1995; Warschauer, 1996, 1997), (2) relative to oral conversation, CMC environments facilitate greater amount (Beauvois, 1992; Kelm, 1992; Kern, 1995) and quality (Chun, 1994; Kern, 1995; Warschauer, 1996) of language, (3) CMC provides ample opportunities for review and reflection (Kroonenberg, 1995; Sotillo, 2000), (4) CMC may be assumed to provide an optimal environment for drawing learners' attention to linguistic form (Blake, 2000; Kelm, 1992; Pellettieri, 2000; St. John & Cash, 1995; Warschauer, 1997) and (5) students exhibit improved attitudes towards foreign language learning as a result of engaging in CMC interaction (Beauvois, 1992; Chun, 1994; Kern, 1995; Warschauer, 1996).

In more recent years, CMC research has begun to examine the potential developmental benefits of interactional feedback in SCMC using theoretical frameworks borrowed from oral interaction research. This work has explored the extent to which SCMC interactional feedback leads to lexical development (Smith, 2004, 2005) and

promotes L2 development and linguistic awareness (Chun et al., 2008; Lai & Zhou, 2006; Loewen & Erlam, 2006; Sachs & Suh, 2007).

Recent work by Smith (Smith, 2004, 2005), has empirically explored the link between CMC interaction and lexical development among intermediate-level ESL students. Smith (2004) found that the acquisition of previously unknown lexical items is facilitated by negotiation of meaning in chat-based, learner-learner interaction. New lexical items that had been the subject of negotiation in the experiment were retained significantly better in post-test recognition tasks than those items where preemptive input alone was provided. Smith (2005) went on to examine the relationship between negotiated interaction and learner uptake in SCMC interaction. He found no relationship between the amount of uptake and the acquisition of target lexical items. His findings also suggest that learner uptake in response to lexical errors may be less likely to occur in a CMC environment than in face-to-face interaction.

Smith (2009) identified a critical methodological issue that could have an impact on how learner interaction is evaluated in CMC. In oral interaction, learner repair moves are an essential part of any recorded and/or transcribed conversation. Because learners who are engaged in chat interaction are free to compose and edit their utterances in private before submitting them to the discussion, self-initiated repair moves are hidden from the researcher upon review of the interaction. In a descriptive study that used dynamic screen capturing to record each keystroke of L2 German learners as they engaged picture sequencing tasks, Smith found descriptive evidence that a considerable amount of language occurs in SCMC interaction that is not reflected in the final chat transcript.

### **2.2.3. Recasts in CMC Research.**

Several interactionist studies have begun to investigate the nature and efficacy of recasts in written chat interaction. Several of these studies have employed concurrent and/or offline attentional measures to examine the extent to which learners notice text-based recasts and the source of their L2 errors. Two studies to date—Loewen & Erlam, (2006) and Sachs & Suh (2007)—have explored the relationship between recasts and L2 development in CMC.

Lai & Zhou (2006) (discussed above) grouped 12 ESL learners into six mixed-proficiency dyads that engaged in two consecutive spot-the-difference tasks—one via online chat and the other through face-to-face conversation. Each of these conversations was recorded and learners were later asked to sit with a researcher in a stimulated recall session where the researcher isolated instances of interactional feedback (negotiation of meaning and recast) and invited the learner to discuss what she was thinking at that moment. Lai and Zhou found tentative evidence that text-based online chat promotes more noticing than face-to-face conversations in terms of learners' noticing of their own linguistic mistakes. However, there was insufficient data to provide any claims about the relationship between recasts and noticing.

Loewen & Erlam (2006) was a partial replication of a previous oral study, Ellis, Loewen, and Erlam (2006), that had found that explicit metalinguistic feedback was more effective than recasts in promoting pre-posttest accuracy gains for English past tense verbs in oral imitation and grammaticality judgment tasks. Employing a quasi-experimental design, Loewen & Erlam randomly assigned 31 ESL learners to one of three conditions: metalinguistic feedback, recast and control. Each of the learners in the

experimental groups participated, for approximately one hour, in two information-gap tasks in NS-NNS group chat. A comparison of pre-posttest performance on two computer-based grammaticality judgment tasks (one timed and one untimed) revealed that neither experimental group made any significant gains in past-tense accuracy.

The results of this CMC study run counter to those from the earlier face-to-face study upon which it was modeled. The authors argue that this may be due to a variety of influences, including: learners' comparative lack of developmental readiness to acquire the target form, the reduced temporal immediacy of the feedback due to conversational overlaps and the inability of the native-speaking researcher to prevent participants from slipping off-task in their submissions.

It is difficult to make the case that large group SCMC interaction might create discourse conditions similar to those found in face-to-face group interaction. In oral group discussions, only one person may hold the floor at a time. In NS-NNS group conversation, the native speaking expert controls the allocation of turns and can usually ensure that corrective feedback occurs in immediate juxtaposition to any non-target utterances. In group chat, on the other hand, any number of learners may take the floor in the time that is required for the native speaker to receive, process and respond to any ill-formed utterance. Thus, it may be extremely difficult for learners to recognize that they are being corrected in group chat contexts.

Sachs and Suh (2007) used a pretest-treatment-posttest design to compare the efficacy of textually enhanced (highlighted) and unenhanced recasts in promoting noticing and L2 development in SCMC interaction. The treatment session engaged 30 Korean EFL learners in recast-intensive NS-NNS dyadic chat-based interaction in four

experimental conditions, based on the following dependent variables: the provision of textual enhancement and the requirement to think-aloud. In the textual enhancement condition, the L1 English interlocutor applied on-the-fly formatting (underlining and bolding) to each written recast in order to highlight the requisite morphosyntactic features of backshifting in reported speech. In the think-aloud condition, learners were equipped with microphones and asked to verbalize any thoughts that came to them as they were engaged in the chat interaction.

Sachs and Suh found that exposure to textually enhanced recasts lead to greater amounts of reported awareness and that reported awareness was related to post-test performance. However, no relationship was found between the enhancement of a recast and subsequent developmental gains.

Chun et al. (2008) explored the ways in which the contingency between an ill-formed utterance and a recast might mediate the amount of noticing that is promoted by text-based recasts in SCMC interaction. Responding to the fact that SCMC turn-taking sequences are often disjointed and syncopated, the authors examined the extent to which recasts delivered immediately adjacent to ill-formed utterances promoted a greater amount of noticing than those that appeared several turns after the utterance. Seventeen intermediate ESL learners were asked to engage in two 15-minute spot the difference tasks via MSN chat with a native-speaking interlocutor. In the chat sessions, the learners were equipped with microphones and asked to think-aloud throughout the written interaction. Following the chat sessions, learners participated in stimulated recall sessions in which they reviewed an animated screen capture of their chat interaction with a researcher and were asked to explain their mental processes at any point that any recast

had occurred. Upon completion of the study, learners' working memory capacities were assessed using on a 30-item reverse digit span test.

Analyzing the data obtained in the think-aloud protocols and stimulated recall sessions, a Wilcoxon Signed Rank Test showed that learners noticed contingent recasts significantly more often than non-contingent recasts. In the vast majority of the cases, recasts were noticed when they simply corrected lexical errors.

Contrary to similar work on oral recasts (Mackey et al., 2002), Chun et al. did not find a significant correlation between learners' working memory capacities and the amount of noticing that occurred in regards to recasts. The authors puzzled over the inability to interface these findings with those from oral research, but did not address the potential role of the chat transcript in obviating the need for strong working memory skills. Throughout the chat interaction, a transcript of the conversation was made available to each learner. This transcript was visible in the stimulated recall sessions as well. In oral interaction, working memory capacity is critical for making the cognitive comparisons necessary for noticing a recast as negative feedback. Learners must accurately remember the lexis, morphology and syntax of their original utterance and compare all of its features to the reformulated utterance put forth by the interlocutor. However, when a transcript is available, learners may make these comparisons based upon the visual record of the discourse rather than needing to rely on what is available in their short-term memory store.

In sum, preliminary explorations of recasts in text-based chat interaction has suggested that (1) recasts occur more frequently in NNS-NNS chat than in face-to-face interaction (Lai & Zhou, 2006), (2) recast-intensive written chat interaction may lead to

L2 development (Sachs & Suh, 2007; cf. Loewen & Erlam, 2006) and (3) the salience of a recast may be mediated by textual enhancement (Sachs & Suh, 2007) and by the contingency between L2 errors and recasts (Chun et al., 2008).

#### **2.2.4. Feedback and development in non-communicative (I)CALL.**

In CMC, technology serves simply as a conduit for spontaneous human-human communication. The challenge of interfacing CMC with oral research lies only in accounting for the modal and temporal characteristics of computer-centered discourse and explaining how these characteristics may impact learning in a manner comparable to face-to-face interaction. However, in self-instructed computer contexts, learners do not engage in spontaneous human-human communication, but interact with a computer program in the pursuit of language development.

#### ***Feedback in non-communicative ICALL.***

Corrective feedback is a central feature of many ICALL systems. Yet, there is a dearth of evidence that might help test the relationship between computer-generated feedback and L2 development. Only two notable figures, Noriko Nagata and Trude Heift, have made any attempt to measure and describe the potential developmental outcomes promoted by ICALL feedback provision. The systems employed in this line of investigation can be described as non-communicative in that the primary focus of task interaction was on linguistic form, since no model of knowledge representation was present to facilitate meaning-focused exchanges.

Throughout the 1990's Nagata published a series of articles describing her evolving ICALL system for self-instructed Japanese language practice—from the Hypercard application Nihongo CALI to the Web-based BANZAI system. Nagata's

methodology engaged learners in NLP-driven drill and practice activities in which explicit grammar instruction and explicit grammatical feedback were provided to the learners. Nagata (1997) comparatively examined the developmental effects of deductive (explicit metalinguistic) feedback and inductive feedback (L1 translations). Based on a two-sample dependent t-test, Nagata found that metalinguistic feedback is superior to L1 translations in promoting pre-posttest development of Japanese particles.

Nagata (1998) compared the effects of treatment task conditions—input (learners read a grammar note) vs. output (learners read a grammar note, produce a sentence and receive metalinguistic feedback). Though no pretest was administered to establish a baseline, Nagata reported that (based on a two-sample dependent t-test) learners in the output condition significantly outperformed their counterparts in subsequent reading comprehension and production tests.

More recently, Heift has brought an interactionist perspective to the study of ICALL feedback by exploring the effects of various types of feedback in eliciting uptake among L2 learners. Like Nagata, Heift embedded her research into a functioning intelligent language tutoring system (ILTS) that had been designed and employed for use in university-level language classes. Though the interaction facilitated by Heift's system cannot be considered communicative, its complex feedback delivery system is capable of sustaining extended form-focused exchanges that can lead learners, step-by-step, towards targetlike production (see Appendix A.2 for a sample form-focused exchange with the E-Tutor).

Heift's task design consistently required learners to build German sentences from proscribed templates that provided the learner with root forms of all of the required

lexical items (Heift, 2002, 2003, 2004, forthcoming). Learners were required to use all of the words in the displayed template to construct a morphosyntactically correct sentence. For each sentence, learners could make an unlimited number of attempts at the target form. Learners could submit their proposed sentences for review and receive various forms of feedback as many times as necessary until the output was deemed targetlike. They also had the option of giving up on any sentence by requesting the target form or simply skipping to the next sentence before all errors were corrected.

Heift (2004) explored the relationship between feedback type and the amount of elicited uptake. Based roughly on Lyster & Ranta's (1997) rubric for feedback types, Heift identified two distinct feedback conditions capable of being produced by her ICALL system—(1) meta-linguistic feedback and (2) repetition of learner output with highlighting applied to erroneous words. The results of this study suggest that, in an ICALL context, explicit metalinguistic feedback (with or without additional highlighting) may be more effective in promoting self-correction than highlighted repetitions.

Heift and Rimrott (2008) examined the relationship between the explicitness of feedback and successful learner uptake in response to corrective feedback targeting spelling errors.

The authors found that learners were more likely to successfully correct typographical errors when feedback included lists of spelling suggestions. Heift (forthcoming)

continued the investigation of the effects of feedback type on learner uptake. This longitudinal study tracked ten German students as they engaged with the E-Tutor system over the course of three academic semesters. Throughout the course of instruction, each learner was exposed, in alternating chapters, to meta-linguistic feedback and repetition with highlighting. Heift again found that the more-explicit metalinguistic feedback

promotes greater amounts of uptake than repetition with textual enhancement—for both lexical and morphosyntactic errors. Heift (forthcoming) also suggested that L2 level mediates the relationship between feedback type and amount of learner uptake and that the amount of uptake in response to metalinguistic feedback may increase over time.

***Recasts in non-communicative CALL.***

The developmental effects of recasts have been examined in a small number of non-communicative CALL studies that have drawn heavily on the findings from oral interaction research to provide a theoretical explanation of the benefits attributed to recasts in a computer-based context.

Ayoun (2001) compared the developmental effects of explicit instruction, recasts and written models on the learning of French verb aspect (*passé composé* and *imparfait*). Using a mixed-design ANOVA, Ayoun's results showed that recast-provision produced a significant increase in pre-posttest gain scores compared to explicit grammatical instruction. Ayoun also found that, relative to written models, exposure to written recasts produced greater (though not significant) gains. In a follow-on study, Ayoun (2004) re-examined the assessment data collected in Ayoun (2001) and found that learners improved in producing well-formed compound past tense verbs, but there was no improvement in the production of the aspectually more complex imperfective verb forms. It must be noted, however, that there are myriad methodological variables that pose a problem for the construct validity of Ayoun's operationalization of recasts.

The first methodological issue with Ayoun's design is that learner production was not spontaneous and was only minimally meaning-focused. Participants in the recast condition were asked to describe an illustration and were presented with a template for

producing the target utterance. For example, learners were shown a picture of a girl playing guitar, and were provided with the following sentence template: *Elle / jouer / \_\_\_\_\_ tour l'après midi. (She/to play/ \_\_\_\_\_ all afternoon long (translation not displayed))*. The learners' task was to conjugate the main verb (presented in its infinitive form) and to provide a Spanish word to fill in the missing noun. Ayoun argued that this task type elicits both form-focused (verb conjugation) and meaning-focused (lexical selection) attention. The interactionist construct of 'focus on form' presupposes that attention to linguistic form occurs incidentally as learners are engaged in attention to meaning (Long, 1991). However, simple lexical selection does not constitute a primary focus on meaning in any way comparable to the co-construction of meaning that occurs in conversational tasks such as information gap and jigsaw tasks.

Another methodological concern with Ayoun's design relates to the provision of recasts in the treatment tasks. In the recast condition, the target sentence was displayed in response to every learner attempt—regardless of whether or not it contained an error. Thus, if the learner made an error, the computer response served as implicit negative feedback; otherwise, it served as non-corrective repetition. It is not clear whether learners were able to distinguish between the two types of responses or to what extent the non-corrective responses contributed to the developmental benefits attributed to recast provision. As the author acknowledges, this indiscriminate repetition makes the claims about the beneficial effects of written recasts in the study somewhat tenuous (Ayoun, 2001, p. 238).

Sagarra (2007) investigated the extent to which computer-administered spoken recasts might facilitate the development of L2 Spanish noun-adjective agreement and

promote uptake in the form of modified output. A single treatment tasks presented two groups—recast ( $n=35$ ) and control ( $n=30$ ) —with 32 fill-in-the blank questions. Learners were instructed to type the properly inflected form of the missing adjective for each sentence (e.g., *la puerta es \_\_\_\_ (red)*). In the recast condition, incorrect adjectival forms prompted the system to present to the learner a pre-scripted spoken model of the target sentence. Following the spoken recast, learners had the choice of either modifying their output by trying the sentence again or moving on to the next sentence.

Following the treatment session, Sagarra administered a series of written post tests as well as a communicative oral task to examine the developmental effects of recasts on accuracy in Spanish adverbial morphology. In all of the post-test tasks, the recast group significantly outperformed control—although, learners’ ability to benefit from recasts appeared to be constrained by individual working memory capacity. The amount of targetlike modified output was also found to be constrained by working memory as well as the type of agreement error that had occurred—learners made more successful second attempts for number agreement than for gender.

As with Ayoun (2001, 2004), it is theoretically challenging to situate Sagarra’s findings in relation to empirical claims from oral interaction research due to several ecological and methodological constraints. Sagarra’s model of recast provision is similar to Ayoun’s in that recasts are delivered in response to clearly proscribed sentences—outside of the context of spontaneous, meaning-focused interaction. Sagarra’s treatment task design also altered the level of explicitness of recasts in a manner quite foreign to the methodology of oral interaction research. One of the central characteristics of a recast is that it is a relatively implicit form of feedback and, therefore, does not necessarily

interrupt meaning-focused conversation or elicit a required response from the learner. Sagarra, however, introduced a task condition that required learners to write down numeric codes they heard at the end of each pre-recorded recast. Learners were thus forced to explicitly acknowledge that they had listened to the recording of the model sentence. It is unclear what effect the explicit instruction to attend to and produce the numeric code spoken in immediate juxtaposition with the recast might have had on learners' attentional resources or the salience of the recast itself.

Both Ayoun and Sagarra altered several underlying ecological characteristics of recast that have informed current theories on the relationship between recasts and L2 development. It is, therefore, not a trivial theoretical challenge to explain the benefits of recasts in this strand of research vis-à-vis the findings from oral interactionist research.

### **2.3. Recasts in Oral SLA Research: Salience, Ambiguity and Context**

Within oral SLA research, mounting evidence suggests that the effectiveness of recasts, as a means of promoting grammatical development, is contingent, in part, upon the salience of the recast provided. It has been widely argued that, in order for corrective feedback to effectively promote L2 development, it must be perceived by the learner as an indication of non-targetlike usage. One factor that has been implicated in affecting the salience and ambiguity of a recast is the interactional context in which it is provided.

In SLA research, an increasing focus has been given to questions regarding how selective attention and awareness mediate input and acquisition as the learner is exposed to linguistic input (Long, 1996). The most widely cited claims about the benefits of recasts are founded on Schmidt's (Schmidt, 1990, 1995, 2001; Schmidt & Frota, 1986) "noticing hypothesis". The noticing hypothesis claims that, in order to become potential

candidates for intake and subsequent learning, formal linguistic features in the input must first be noticed by the learner (Schmidt, 1995, p. 20). By consciously attending to formal features in the input, learners increase the likelihood that those features will enter into their developing interlanguage. According to Schmidt, noticing (or conscious attention to form) is a necessary, though not sufficient, condition for second language acquisition.

It has been suggested that input salience affects the internal cognitive process of attentional selection and possible subsequent error detection (Schmidt, 1990, 1995, 2001; Sharwood Smith, 1993). Although it is difficult to accurately define precisely what factors contribute to the salience of a linguistic form, Sharwood Smith claims that the salience of formal L2 features is derived from both learner-internal factors that make certain features intrinsically more salient to the learner (e.g., the learner's readiness to acquire a particular form) as well as from external factors, such as typographical enhancement, metalinguistic instruction or corrective feedback. Others have argued that the salience of formal features is, in part, determined by the communicative value of the form (Van Patten, 1990, 1996; Shook, 1994;) or its structural prominence. Free morphemes, for example, have been found to be more salient than bound morphemes in visual language processing (Koda, 2000; Leow, Egi, Nuevo & Tsai, 2003). In oral interaction, salience may also be determined by phonetic substance, syllabicity and sonority (Long, 2007).

The extent to which recasts constitute an effective technique for drawing learners' attention to the formal properties of language has been the subject of much debate within SLA literature. Some have suggested that recasts lack the salience entailed by other, more explicit, forms of feedback and that the corrective intent of recasts may be

ambiguous to learners. Other research has shown that the salience and ambiguity of recasts may be affected by a number of internal and external variables and that interactional context, in particular, may play a crucial role in determining the extent to which recasts may facilitate the noticing of non-targetlike forms.

### **2.3.1. Salience of recasts.**

Recasts have been argued to increase the salience of linguistic forms by providing an immediate and direct comparison of two slightly differing utterances—the learner's original utterance and the corrected form provided by the more advanced speaker (Farrar, 1990; Long, 1996; Saxton 1997, 2005). When the learner's utterance and the recast convey the same meaning and differ only in the use of a particular linguistic form, that form is said to be made perceptually more salient to the learner (Farrar, 1990). However, there are a number of intervening variables that might mediate the extent to which a recast might enhance the salience of corrected linguistic forms.

Philp (2003), for example, found that the salience of recasts is subject to variation due to learners' developmental readiness vis-à-vis the corrected form (see also: Han, 2002; Iwashita, 2003; Mackey & Philp, 1998), the length of the recast provided and the number of corrections made in the reformulation. Her findings suggest that effectiveness of recasts in oral interaction is closely tied with working memory constraints and that recasts may be more salient when (1) learners have *a priori* knowledge of the forms being recast and (2) the recasts are shorter and contain a fewer number of changes. Similarly, Sheen (2004) found that reduced/partial recasts are more salient than more elaborate recasts. When only the incorrect portion of the learner's utterance is recast, learners are better able to focus their attention on the targeted linguistic form.

Han (2002) has also implicated several other factors that may affect the salience of recasts. Specifically, Han cited individualized attention, consistent focus, developmental readiness, and intensity as contributing factors to the success of recasts in improving ESL learners' tense consistency in oral and written production tasks. These four interdependent factors, she argued, jointly created salience, thus making the target linguistic feature more noticeable.

Recasts have also been found to be more salient when they are delivered along with other form-focused techniques or interactional moves. Doughty and Varela (1998), for example, operationalized “corrective recasts” in a classroom setting as the teacher’s repetition of the learner’s incorrect utterance (adding emphasis to the error) followed by a recast where emphasis was again added to corrected target feature. The positive results Doughty and Varela found for recasting in classroom-based interaction suggests that when the technique involves an “initial attention-getting phase” (Doughty, 1999), they may be more salient than simple recasts. Similarly, Ellis, Basturkmen and Loewen (2001) found that the complexity or length of negotiation in which a recast is provided may help determine the saliency of that item. Recasts, they argue are potentially more salient they are part of “focus on form episodes” (FfEs) with more than three turns.

Leeman (2003) explored the possible roles that enhanced salience and negative evidence may play in contributing to the efficacy of recasts. Most applied research on recasts has assumed a substantial role for negative evidence, wherein the learner makes a cognitive comparison of the targetlike and non-targetlike forms and rejects the non-targetlike alternative. Leeman questions whether such comparison and rejection are necessary for recasts to be effective in promoting language development. Negative

evidence, she argues, may not play as crucial a role in explaining the benefits of exposure to recasts as the enhanced salience of positive evidence they provide.

### **2.3.2. Ambiguity in learner perception.**

There has been considerable debate regarding the extent to which recasts are perceived by learners as corrective feedback. Because recasts are delivered in a manner that is nearly isomorphic with non-corrective repetitions, it has been argued that recasts may not be recognized as corrective feedback. Lyster and Ranta (1997) first pointed out the problem of ambiguity in a French immersion classroom-based study that observed the feedback behavior of four teachers and examined subsequent learner uptake in response to various types of feedback. They found that while recasts were most frequently provided by the teachers, they elicited the lowest rate of learner uptake and only led to very low levels of repair. Lyster (1998b) later conducted further analysis on the interactional data and found that recasts (corrective repetition) and non-corrective repetition had occurred in parallel, in terms of both discourse function and distribution. In such a context, Lyster argues, it is plausible that learners would be unable to disambiguate recasts from non-corrective repetition and topic-continuation moves. Although, counter-arguments have also been made that the issue of ambiguity does not negate the utility of recasts for second language acquisition, as demonstrated by the broad range of positive empirical claims in the literature (see Long, 2007).

Lyster (2004) provided additional support to the notion that recasts may not be the most effective means of providing young L2 learners with negative feedback in Immersion classrooms where the primary focus is on subject matter. Lyster examined the differential effects of recasts and prompts (i.e., clarification requests, non-corrective

repetitions, metalinguistic clues and elicitation). He found that prompts were more likely than recasts to engage learners in productive interaction and to improve their ability to assign grammatical gender in subsequent written tasks. The reason for this disparity was attributed, in part, to the ambiguous nature of recasts. However, as in Lyster's previous research, no empirical measures were employed to provide evidence of ambiguity in learners' perceptions about recasts. Instead, Lyster draws conclusions about learners' interpretations of recasts based upon the lack of uptake and repair that occurs in response to the provision of recasts (see also: Lyster & Ranta, 1997; Lyster, 2002; Lyster, 2004; Panova & Lyster, 2002).

Oliver and Mackey (2003) examined learner uptake and repair (modified output) in a descriptive study of classroom discourse that explored the relationship between interactional contexts and the provision and incorporation of various feedback types. Their findings revealed that, in explicit language-focus exchanges (vs. content, management or communication exchanges), recasts were the most common form of feedback provided by the teachers. This evidence suggests that recast may, in fact, prompt users to detect the errors in their output, but that the extent to which this may be true is dependent upon the context of the interaction in which the recasts occur.

Discounting the role of uptake in measuring the efficacy of recasts, Loewen and Philp (in press) also offer counter-evidence to Lyster's claims about the ineffectiveness of recasts in classroom situations. Based upon accuracy of recall in post tests, Loewen and Philp found that the ambiguity of recasts may be mitigated by factors such as classroom context, the discourse context in which the recast appears (see also: Ellis et al., 2001) and variable elements of the recast itself, such as segmentation, declarative intonation, and

prosodic emphasis provided by the teacher. Consistent with Philp (2003), the authors claim that shorter recasts with few changes may increase the likelihood that recasts will be perceived as corrective feedback.

### **2.3.3. Interactional context.**

The salience and ambiguity of recasts may also be dependent upon the interactional context in which they appear (Nabei & Swain, 2002; Nicholas, Lightbown & Spada, 2001; Oliver, 1995; Oliver & Mackey, 2003; Sheen, 2004). In communicative, immersion language classrooms, where the instructional focus is primarily on meaningful language production, recasts and repetitions are often difficult for learners to disambiguate (Lyster, 1998b, Lyster, 2004; Lyster & Ranta, 1997). In more form-focused classroom environments (e.g., Ellis, Basturkmen & Loewen, 2001; Ohta, 2000; Sheen, 2004, but see: Havranek, 1999; Lochman, 2000) and especially in dyadic laboratory contexts (e.g., Han, 2002; Iwashita, 2003; Long, Inagaki and Ortega, 1998; Mackey & Philp, 1998, 2003) recasts may be more likely to be noticed as corrective feedback.

Oliver and Mackey (2003) posited that the abundance of teacher feedback provided in language-focused interaction might make the teachers' expectations more transparent to the learners. Thus, learners may be less likely to misinterpret recasts as non-corrective repetitions. Similarly, it might be argued that in laboratory-based dyadic interaction, the didactic nature of recasts is made evident to the learner by the highly instructional nature of the interaction.

In the current study, treatment tasks were carried out in two distinct interactional contexts: dyadic oral and written interaction. In the oral condition, learners engaged in

face-to-face, meaning-focused interaction with a native-speaking researcher who provided recasts in response to learner errors. In the written condition, meaning-focused interaction and recast provision were carried out through the use of an intelligent conversational agent—where it was highly unlikely that learners would interpret a recast produced by the system as a form of non-corrective repetition.

#### **2.4. Research Question**

This study compared the developmental effects of recast-intensive interaction in oral and written modalities and explored the following research question:

*Is the effectiveness of recast-intensive interaction constrained by the modality in which it occurs?*

#### **2.5. Hypothesis**

On the basis of findings from oral interaction and CALL research, the following hypothesis was established:

*Recast-intensive interaction will promote developmental gains in ESL question formation among the learners in both oral and written modalities.*

The tasks employed in this study were modeled on well-attested task types in the oral literature that have been shown to be effective in facilitating meaning-focused communication—in which, incidental focus on form might promote L2 development. There is growing evidence that recasts are effective at promoting learning in dyadic face-to-face interaction. In particular, recasts-intensive dyadic interaction has consistently been found to promote development in ESL question formation (McDonough, 2005; McDonough & Mackey 2006; Mackey, 1999; Mackey & Oliver, 2002; Mackey & Philp, 1999; Mackey & Silver, 2005; Philp, 2003; Silver, 2000). Therefore, it is logical to

assume that recasts may promote developmental gains among the learners in the oral, face-to-face condition.

It is also logical to assume that if learners in the written condition were given similar opportunities to produce ESL questions and receive intensive recasts, they might exhibit developmental gains similar to those of the learners in the oral condition. Yet, given the dearth of empirical research examining the effects of modality on L2 development, it is not possible to predict which environment might be more conducive for learning.

## CHAPTER III: METHOD

### 3.1. Participants

The participants in this study comprised 56 high school students (30 male and 26 female) enrolled in a special ESL track at a large public school in the Washington D.C. metropolitan area. Students were recruited from a large pool of learners who were participating in the school's *Theater Arts for ESL* program. The ESL learners who participated in the study were between the ages of 15 and 18 and were nearly all L1 Spanish speakers (n=50). The remaining learners came from various L1 backgrounds, including Arabic, Amharic, French, Thai and Urdu.

The Theater Arts for ESL program was comprised of three intermediate-level classes and one class for beginners. Participants were drawn from all four of these classes. Forty-three of the participants came from the three intermediate classes and 13 came from the beginner class. Proficiency levels were based upon spoken and written placement tests administered by the ESL department upon enrollment at the school.

#### 3.1.1. Attrition.

In total, 61 learners signed up to participate in the study. Throughout the course of the study, three learners dropped out for various reasons and had to be excluded. Another two learners failed to produce even a small number of questions at pretest and posttest and data from these learners were thus not coded.

#### 3.1.2. Incentives for participation.

Participation in the study was voluntary, though some non-academic incentives were offered to those who chose to participate. Participants who completed all assessment and treatment tasks received a \$25.00 gift card to a local bookseller and were

invited to a pizza party on the last day of school to celebrate the completion on the study and the school year.

### **3.1.3. Approval and informed consent.**

In addition to establishing IRB compliance with the author's home research institution, official approval for this project was obtained from the participating high school's ESL department, principal and the local public school system. A proposal of the study was first presented to the entire ESL department to determine whether administration would be (1) feasible in terms of scheduling, space and computer resources and (2) potentially beneficial to the learners. Because the design of the study involved pulling students from their regularly scheduled ESL classes, it was critical to establish that participation in the study would be beneficial to the students' language development. Following a formal presentation and follow-up discussion, the teachers in the ESL department unanimously agreed that the individualized language instruction provided in the proposed tasks would be beneficial for the students.

The study was then presented to the school's principle, the head of ESL and team of literacy specialist. With the recommendation of this committee, a formal application for approval was submitted to and approved by the Executive Director of Monitoring and Evaluation in the local public school system.

Student recruitment included a presentation of the study by the author to each of the four classes. In this presentation, students were informed that their participation in the study was voluntary and would have no impact on their grade in their Performance Arts class. Following the presentation, interested students were given the chance to put their names on a sign-up sheet that was left at the front of the room after the author left.

On the first day of the study, each learner signed a consent form (see Appendix B) after hearing a detailed explanation of the study and its potential risks and benefits. The author presented this information in English and called on the teacher to provide Spanish translations where necessary. Based on the recommendations of the principle, which were confirmed by the school system, informed consent came from the students and not their parents.

#### **3.1.4. Native speaker interlocutors.**

All oral assessment and treatment tasks were carried out by three native-speaking researchers. The team of researchers included the author, his wife (a professional researcher with an advanced degree in Cognitive Psychological research) and a teacher from the participating school with an advanced degree in Education. The latter was paid \$30 an hour—the former earned the author’s undying gratitude. The two additional researchers were given extensive training in task administration that included extensive review of a detailed 35-page researcher manual (see Appendix C) and a 1-2 hour one-on-one training session with the author. Ideally, the assessment and treatment sessions would have been counterbalanced so that all learners would have interacted with all of the native speakers. However, because of scheduling constraints, it was impossible to ensure that this balance would be achieved for all learners. Participants were assigned to interact with as many of the native speakers as possible, though, in some cases, learners were paired with a single researcher for many of the oral tasks.

### **3.2. Materials**

#### **3.2.1. Computers and lab space.**

Data collection for the current study was embedded in a highly regimented

academic context. Daily activity in a public high school is regulated by strict schedules, bells and hall monitors. Throughout the five-week data collection period, groups of learners were escorted to various research classrooms. Research classrooms were graciously provided by non-ESL teachers in the school whose rooms were not occupied by a class. Most of the classrooms had large banks of Internet-connected computers (both Macs and PCs). When assessment tasks were performed in rooms with no computers, students were asked to bring their school-issued laptops with them to the sessions. Whenever multiple learners were engaging in oral tasks in the same room, the lab tables were separated as much as the layout of the room would allow. In general, at least 10 feet separated each of the lab tables in the oral tasks. Learners working on the computer tasks were separated as much as the lab configuration allowed. Student computers were typically separated by one to two feet.

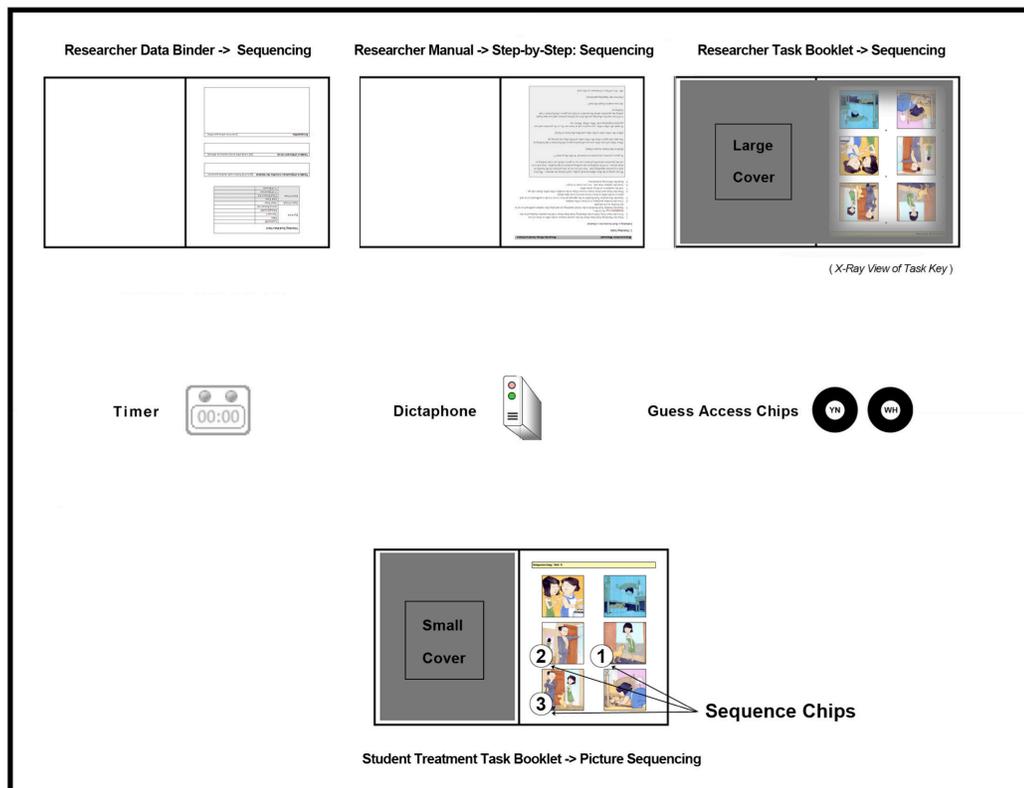
### **3.2.2. Lab setup for oral tasks.**

In order to complete data collection, the design of this study needed to account for a dizzying array of logistical considerations (bell schedules, state testing schedules, the availability of free space/computers, absenteeism, etc.) imposed by carrying out extra-curricular research embedded in the working day of a large public high school. The research materials thus needed to be highly mobile and easy to setup and break down at a moments notice. Each researcher was equipped a portable research box that contained the items listed in Figure 3.1.

Figure 3.1. Materials for oral data collection and sample task layout.

1. Researcher Manual
2. Researcher Data Binder
3. Researcher Task Booklet (Answer Keys)
4. Student Assessment Task Booklet (Picture Sets)
5. Student Treatment Task Booklet (Picture Sets)
6. Guess access chips
7. Count chips
8. Cardboard page covers
9. A digital timer
10. A dictaphone
11. Spare AAA batteries
12. No. 2 pencils
13. Pencil sharpener

Researcher



Student

***Researcher manual.***

The researcher manual contained detailed information about the administration of assessment and treatment tasks as well as specific guidelines for recast provision. The manual included a step-by-step protocol for each task type that informed the researcher how to setup and initiate the task, record the data, deliver recasts, etc. (see Appendix C).

***Researcher data binder.***

The researcher data binder contained simple data sheets to log attendance and note any task irregularities. The data binder was also updated daily with the schedule for the day.

***Researcher task booklet.***

The researcher task booklet contained the prompts for the oral assessment tasks as well as the answer keys for the oral treatment tasks (see Appendix D). Assessment prompts included the model questions (paired with thumbnails of the visual prompts displayed to the learner) for the elicited imitation tasks (discussed below) and detailed scenarios for the role-play tasks. Answer keys for the treatment task consisted of (1) target prompts for the matching tasks, (2) alternate pictures for the spot the difference tasks and (3) correct story sequences for the picture sequencing tasks.

***Student assessment task booklet.***

The student assessment task booklet contained the picture prompts presented to the learners during the oral imitation tasks. All of the pictures used in the assessment and treatment tasks were illustrations created by Bruce MacPherson—a professional illustrator whose work has appeared in major national newspaper and magazines as well as in several children’s books (see Appendix E for references).

### ***Student treatment task booklet.***

The student treatment task booklet contained the visual prompts presented to the learners throughout the three tasks in the treatment sessions. These prompts were presented to learners in the face-to-face condition and were identical with those presented digitally to learners in the computer-guided condition.

### ***Guess access chips & count chips.***

Throughout the oral treatment tasks, a series of small round chips were used to indicate task progress to the learner. The guess access chips visually indicated to the learner whether or not they were allowed to guess the target picture based on the number and type of questions they had asked. The count chips were used to (1) display the number of differences found in the spot the differences task and (2) label the order of the prompts in the sequencing task.

### ***Cardboard page covers.***

Cardboard page covers were used throughout the tasks to keep the question prompts and answer keys in the Researcher Task Booklet hidden from the learner. Since there was no physical divider to prevent learners from viewing the researcher's answer keys, the researchers used the cardboard covers to keep the keys hidden.

## **3.3. Procedures**

This study employed a pretest, posttest design to comparatively explore the developmental effects of recast-intensive communicative interaction in comparable face-to-face oral and computer-guided written contexts.

### **3.3.1. Target form.**

The primary target structure in this study was ESL question formation. The effects of instruction on ESL question formation are well attested in SLA research (e.g., Adams, 2004; McDonough, 2005; McDonough & Mackey 2006, 2008; Mackey, 1999, 2006; Mackey & Oliver, 2002; Mackey & Philp, 1999; Mackey & Silver, 2005; Pienemann, 1998; Pienemann & Johnston, 1987; Pienemann, Johnston & Brindley, 1988; Philp, 2003; Silver, 2000; Spada & Lightbown, 1993,1999; White, Spada, Lightbown & Ranta, 1991). In this study, ESL question development was operationalized as a learner's movement to a higher stage in Pienemann and Johnston's (1987; Pienemann, Johnston, & Brindley, 1988) developmental sequence for ESL question formation (see Figure 3.2).

As a target form, question formation was also ecologically well suited to computer-guided interaction. HCI models that comprise embodied conversational agents usually require the user to pose questions to the embodied agent. Thus, question formation might be assumed to occur naturally in this type of computer-guided context.

Figure 3.2. Developmental stages and question forms.

Stages	Constructions	Examples
3: Fronting	<i>Do + SVO?</i>	<i>Does the man have glasses?</i> <i>Do you have a telephone next to the woman?</i> <i>*Does the lady is writing?</i> <i>*Does is he under a tree?</i>
	<i>Wh + (be/do) SVO?</i>	<i>What flies over the man?</i> <i>Who sits on the red chair?</i> <i>*Who she is?</i> <i>*Who are drink milk?</i> <i>*How many people there are?</i> <i>*What is the man is doing?</i> <i>*What kind of clothes she is wearing?</i>
	<i>Be + SVO?</i>	<i>*Is he take a picture?</i> <i>*Is take the boy the money?</i> <i>*Are they sit down?</i>
4: Pseudo-inversion; yes/no inversion	<i>(Wh) + copula + S</i>	<i>What color is the house?</i> <i>Who is that guy?</i> <i>How many people are there?</i> <i>Are you looking at a woman?</i> <i>Is he alone?</i>
	<i>Aux/modal + SV</i>	<i>Can you help me?</i> <i>Will it be difficult?</i> <i>*Can you teaching me after school?</i> <i>*Has the man close the eye?</i>
5: Aux second	<i>Wh + aux/do</i>	<i>What is she doing?</i> <i>What does the woman have in her hand?</i> <i>Where is the ugly man sitting?</i> <i>Who is the man talking to?</i> <i>Why is the man looking at the floor?</i> <i>*How does she looks like?</i> <i>*Why does the man want take a picture?</i>

Source: adapted from McDonough (2005)

### **3.3.2. Experimental groups.**

#### ***Group assignment.***

Each participant was randomly assigned to one of two experimental groups or to control. Group assignments were determined by the author prior to the study based on class rosters of unfamiliar students—the author had no previous contact with any of the students. In order to minimize between-group effects, distribution was balanced in terms of gender and proficiency level. Each group comprised similar ratios of males to females and novice to intermediate learners.

#### ***Face-to-face.***

Participants in the face-to-face (FTF) group participated in dyadic, communicative treatment tasks with a native speaker and received intensive recasts in response to morphosyntactic errors in their questions.

#### ***Computer-generated recast.***

Participants in the computer-generated recast (CGR) group participated in similar communicative tasks on a computer and received intensive recasts generated by an embodied conversational agent.

#### ***Control.***

The control group served simply as a testing control group. Participants in the control group did not receive any instruction between the pre and posttests. Comparisons to control in this study, therefore, examined the relationship between engaging in recast-intensive interaction and doing nothing. The control group ensured that there was no testing effect, but because learners in the control group did not participate in any form of

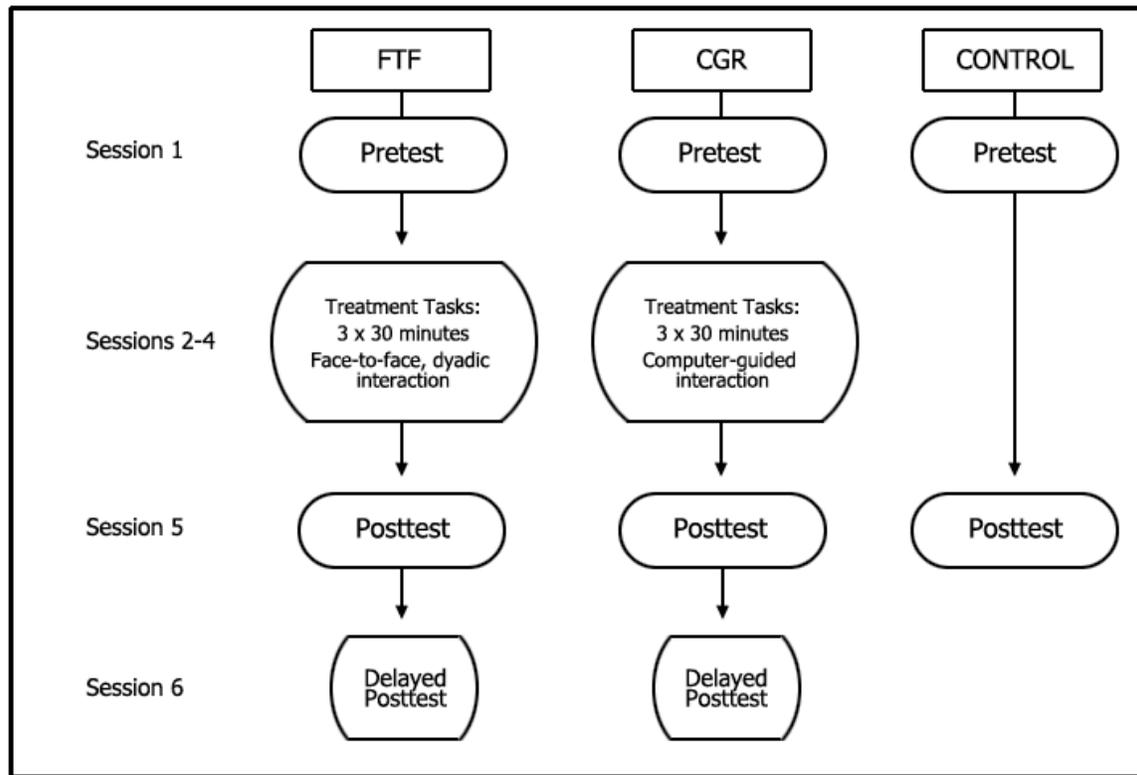
interaction between the pretests and posttests, it is difficult to isolate the effects of recasting on L2 gains promoted by treatment.

The decision to simplify control was one of sheer necessity. By the time all of the administrative hurdles had been cleared and data collection was ready to begin, only five weeks remained in the school's academic calendar. A preliminary design, in which learners in the control group would have engaged in oral interaction with no recasts, would have required a minimum of six weeks to complete. Control was thus reduced to a testing control in order to complete the data collection in the limited time available. With this reduced schedule, data collection successfully concluded on the final day of classes at the school.

### **3.3.3. Experimental schedule.**

Participants in all groups received a battery of oral and written pretest tasks to assess their developmental level in question formation and their ability to form English questions. Following the pretest, learners in the two experimental groups engaged in a series of three 30-minute treatment sessions. Learners were then given a battery of posttest s similar to those administered in the pretest. Roughly two weeks after the posttest, a delayed posttest was administered (see Figure 3.3). On the day of the final session, students were also asked to complete an exit survey that asked them to characterize their perceived abilities in speaking and writing English and to indicate how useful and/or enjoyable they found the tasks in the study (see Appendix F).

Figure 3.3. Experimental schedule.



### 3.3.4. Assessment tasks.

Each testing session lasted approximately 20-30 minutes and consisted of two oral and two written tasks. Within each modality, tasks consisted of one open and one closed task. In the closed activities, there was a correct answer (or limited range of answers) that the learner was expected to produce in response to each prompt provided. The open tasks did not require any pre-specified correct answer, but rather relied on the learners' linguistic abilities and creative skills to produce appropriate questions. Because learners in this study possessed a wide range of oral and written skills, equal time on task did not necessarily lead to equal amounts of production. Therefore, a count of all questions produced in each group was made for each task and the amount of production was found

to be highly comparable in all groups for each of the assessment tasks (see Figures 7.1 and 7.2).

In each assessment session, learners were brought to the lab space on a 2:1 ratio of learners to available researchers. Each researcher was able to administer the assessment tasks to two learners simultaneously due to the fact that the written tasks were timed and guided by an online assessment system. In a few instances, students taking the written tests had urgent questions or technical issues. In these cases, the researcher paused the timer for the oral task, answered the learner's question and then returned to resume the oral task.

### ***Oral tasks.***

The two oral tasks were conducted at a table where the researcher and the learner sat facing one another (see Figure 3.1). The researcher began each task by reading a set of instructions, confirming that the learner was ready to begin and starting the dictaphone recording and the timer.

### ***Imitation task.***

Elicited imitation has widely been argued to be an effective means for assessing interlanguage development (e.g., Bley-Vroman & Chaudron, 1994; Ellis, 2005, 2008; Erlam, 2006; Vinthner, 2002). Ellis (2008), in particular, argued that oral imitation tasks may be effective in providing a means of determining a learner's developmental stage. However, it remains unclear whether learner production in oral imitation tasks reflects learners' interlanguage state or their ability to reproduce language from their phonological store (Vinthner, 2002).

The closed oral imitation tasks employed in the current study, were designed to ensure that learner production reflected a reformulation of the question (based on the learner's developmental level) rather than rote repetition by (1) requiring attention to meaning between the prompt and repetition and (2) using structurally complex question prompts. In the imitation tasks, learners were presented with a series of 10 pictures by the researcher. With the presentation of each picture, the researcher read aloud a question that the learner was required to first answer, and then repeat back to the researcher. A combined average of 7.41 questions was elicited from each learner in the pre- and posttest imitation tasks.

Learners were required to produce a semantic response to the question prior to repetition in order to impair their ability to produce a repetition from their short-term phonological store. To further reduce the likelihood of rote repetition, each of the question prompts contained many prepositional phrases and adjectives to add extraneous complexity unrelated to the morphosyntactic rules necessary for producing a stage-appropriate question. Figure 3.4 illustrates an example of an imitation task item.

Figure 3.4. Example imitation task item.

---



---

Model Question: *Who is running between the mountains behind the men?*

*Role-play task.*

In the open role-play task, learners were presented with a brief scenario in which they were required to play a role and ask the researcher questions. Unlike the more discrete imitation task, the role-play is a dynamic, open task in which the amount and type of questions produced by the learner is subject to considerable variability. The mean number of total questions produced in the role-play tasks was comparable across groups in both the pretest and posttest. In total, the role-play tasks elicited an average of 13.22 learner questions per five-minute task.

Throughout each role-play task, the researcher was responsible for continuously generating contexts for potential learner questions. One of the role-play tasks, for example, created a scenario that cast the learner as a detective and the researcher as the sole witness to a crime. As the detective, the learner needed to ask the witness numerous questions to uncover the details of the crime. Whenever the learner was at a loss for what kind of question to ask next, the researcher prompted the learner to ask a question about a specific detail. For example, the researcher might say, “I saw what was in the bag” or “I know what the thief looked like”. If the learner is still had trouble coming up with a question, the researcher would deliver a more direct question prompt, such as, “Try asking me something about his clothes.”

***Written tasks.***

*Fill-in the blank task.*

In the closed fill-in the blank tasks, learners were presented with a series of 10 pictures by the computer. The learners’ task was to complete a question-answer sequence for each picture presented. Below each picture, a text field was presented for the student

to type in a question that would be appropriate for the answer that was written out below. Each question was also accompanied with a specific instruction indicating what type of question the learner was expected to ask. Figure 3.5 illustrates an example of a fill-in the blank question exactly as it was presented to the learners.

Figure 3.5. Example fill-in task item.

---



Your Question:

Answer:

*He is behind the old lady.*

Instructions:

*Start Your Question with 'Where'*

---

The fill-in the blank tasks elicited a combined average of 5.7 questions by all learners in pre- and posttest intervals.

#### *Email Task.*

The email task was an open written task in which the students were asked to read a short email message and write a response in which they needed ask many questions. Each of the email scenarios was designed to elicit numerous questions by presenting the students with an urgent situation in which there were large information gaps. For

example, the student has received the following email from her hypothetical English teacher (see Figure 3.6).

Figure 3.6. Example email task item.

---

You have just received an email from your English teacher about an important exam that will be given to your class.

Read the email message below and write your response in the text box at the bottom of the page.

- Please try to ask AT LEAST 10 questions about the exam.
- When you are finished, click on 'Save your work'.

From:	Your Teacher <teacher@school.edu>
To:	Your Name <student@school.edu>
Subject:	Getting Ready for the Big Test.

**Message:**

I hope you are ready for the big test coming up! If you have any questions about what will be on the test, where it will be given, how it will affect your grade, etc, feel free to ask me.

Good luck!

Mr. X.

---

The total number of questions elicited by the email task was relatively low compared with the other assessment tasks. Though the learners were asked to try to produce at least 10 questions, the email task only elicited an average of 4.72 questions from each learner.

### 3.3.5. Treatment tasks.

In both experimental groups, learners engaged in a series of three 10-minute communicative tasks. The oral (FTF) and written (CGR) task constructs were designed to be isomorphic, with the exception of the discourse context in which they were conducted. Learners in both groups were informed that they would be engaging in a series of tasks that would require them to ask many questions. Learners were not explicitly told that they would receive corrective feedback, but were informed that the

purpose of the activities was to help them with their communication skills—in particular their ability to form questions in English.

*Oral recast provision.*

The oral tasks were conducted in face-to-face NS-NNS dyads where the learners asked numerous questions and the researcher provided recasts in response to morphosyntactic errors. The following guidelines were strictly adhered to when formulating recasts in the oral treatment tasks:

Only questions that contained grammatical (syntactic and morphological) errors were responded to with a recast. Recasts were delivered as complete questions and all morphosyntactic errors—not just those that are implicated in question formation—were corrected. Although it has been argued in the oral literature that segmented or focused recasts may be more salient than complete recasts (Long, 2007), partial recasts of English questions were deemed prohibitively difficult to model in computer-guided interaction. Question formation is a complex construct that requires an array of morphosyntactic competencies to form a wide variety of different question types.

The current study needed to create an operational model of recasts that could be sustained with broad equivalence by both a native speaker and a computer. In any given non-targetlike question, a number of morphosyntactic errors may be present. In the example shown in Figure 3.7, the learner attempted to produce a stage 5 question, but failed to insert ‘do’ in the second position. She also uttered an extraneous pronoun following the subject, misapplied inflection on the main verb and failed to provide a preposition.

Figure 3.7. Partial and complete recasts.

---

Learner question:	<i>*What the man he puts his bag?</i>
a. Focused recast:	<i>*What does the man he puts his bag?</i>
b. Segmented recast:	<i>*What does the man?</i>
c. Complete recast:	<i>What does the man put in his bag?</i>

---

In order to meet the basic criteria for a stage 5 question (*Wh* + *aux/do*), as proposed by Pienemann and Johnston (see Figure 3.2), only the first error would need to be repaired. Thus, a focused recast would still have contained three errors. A partial or segmented recast might only reformulate the first portion of the utterance and ignore any segments of the question that did not relate to developmental stage. Not only is it a challenge to explicitly model all of the interdependencies between morphosyntax and question stage, it is also unclear how learners might respond to non-targetlike or incomplete input generated by a computer program.

Recasts were not provided in response to lexical errors—if the learner produced a grammatically well-formed question, but one or more of the words used in the question did not correspond semantically with the picture in question, no recast was provided. Sentence fragments and unintelligible utterances were responded to with clarification requests, rather than recasts.

Recasts were always given immediately contingent to any non-targetlike questions asked by the students. Following Long, Inagaki and Ortega (1998) and Leeman (2003), no opportunity for modified output was provided in either modality. In both the oral and computer guided interaction conditions, recasts were immediately

followed by a semantic response that obviated the need for the learner to produce a response for the recast. The decision to discourage any uptake or explicit repair was motivated primarily by the desire to control and isolate the effect of recasts from those of modified output. The effectiveness of recasts in this study was measured by subsequent L2 development outcomes and not by the immediate response they may provoke.

Finally, recasts were delivered without any special emphasis (pitch, stress, facial expressions, etc.) that might indicate if or where an error had occurred. These paralinguistic cues were suppressed in order to help ensure equivalency between the oral and written groups.

*Written recast provision.*

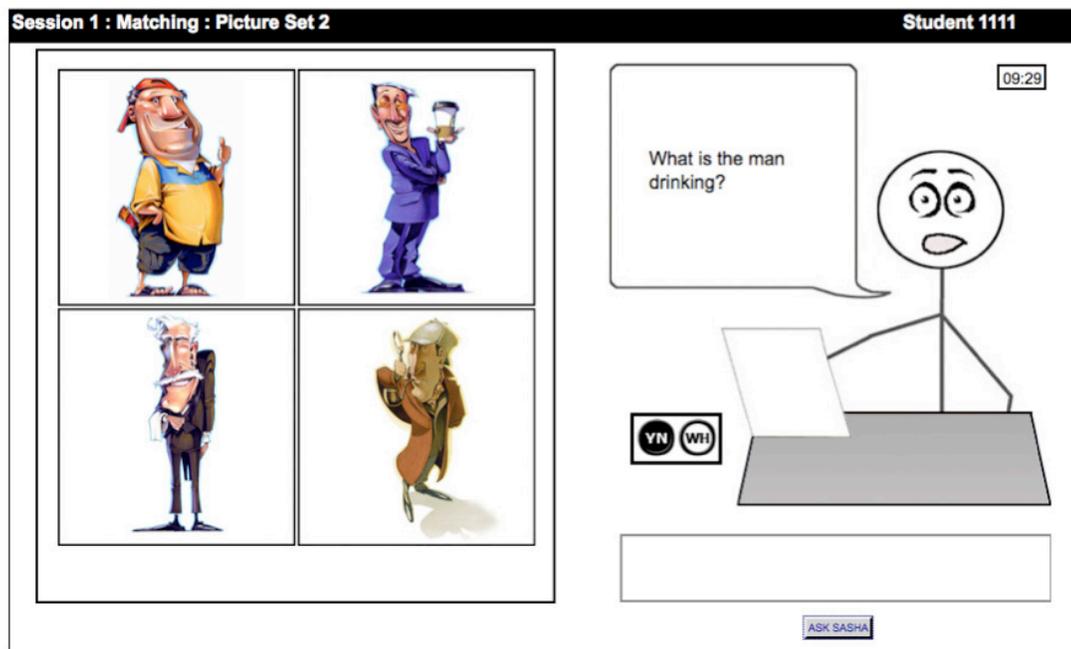
The recasts delivered in the written communicative tasks were generated by a communicative ICALL system created for this study, described in detail in the following chapter. In each task, learner-computer interaction was facilitated by an embodied conversational agent who was introduced to each learner at the outset of the study simply as 'Sasha'. As an interlocutor, Sasha was intended to be discursively, though not visually anthropomorphic. As Cassell & Tataro (2007) argue, the goal of human-computer interaction design should not be creating a physically believable agent, but one that is capable of facilitating believable interaction. Attempting to provide a visually anthropomorphic interlocutor can have unintended effects on learners' motivation and task performance (Schneiderman, 1992; Schultze, 2003).

Sasha appeared to learners as a simple stick figure that was capable of providing feedback and semantic responses and managing task interaction. As a virtual research assistant, Sasha was responsible for (1) analyzing the syntactic/semantic structure of the

learner's input (2) providing corrective feedback and semantic responses (3) managing the tasks and (4) tracking learner data.

Figure 3.8 shows an example of the written treatment task interface where Sasha is delivering a recast in a matching task.

Figure 3.8. Sasha.



Note that the learner's question (*\*What the man he drinking?*) was cleared from the learner's input box before the presentation of the recast. Thus, the comparison between the recast and the non-targetlike utterance relied solely upon the learner's working memory capacity (see Appendix A.3 for an example of learner interaction with Sasha).

#### *Picture matching tasks.*

In the picture matching tasks (see Appendix F for an example of a written picture matching task), the learner was presented with a series of picture sets that contained four

pictures. In each set, one of the pictures had been selected as the target picture that the researcher or Sasha was looking at. The task of the learner was to ask questions in order to identify the target picture that Sasha was looking at. Before the learner was allowed to guess which picture was the intended target, she needed to ask at least one Yes/No question and one WH- question. Once this requirement was met, the learner was allowed to guess the target picture. If the correct picture was identified, the task advanced to the next set of pictures. If the guess was incorrect, the learner needed to ask more questions before being allowed to guess again.

#### *Spot the differences tasks*

Spot the difference tasks have been widely employed in interaction research to create contexts for form-focused interaction (e.g., Chun et al., 2008; Lai & Zhou, 2006; Mackey et al., 2000). In the spot the differences tasks employed in the current study, learners were presented with a single picture depicting a complex scene (see Appendix F). The learner was told that the researcher/Sasha was looking at a similar picture that contained a number of distinct differences from the picture that was visible. The learners' task was to ask questions prompted by the displayed picture in order to determine what the specific differences between the two pictures were. The learner received a point for each difference she identified. The identification of a difference was determined by the researcher/Sasha by comparing the two pictures.

#### *Picture sequencing tasks.*

Picture sequencing tasks have been used in several computer-mediated interaction studies (Sachs & Suh, 2007; Smith, 2005, 2009). In the picture sequencing tasks in the current study, the learners were presented with a series of six pictures that represented a

short story, displayed in scrambled order (see Appendix F). The task of the learners was to ask questions to determine the correct order of the six pictures. The learners accomplished this by identifying one picture at a time – beginning with the first picture and moving on to the second, third, etc. Once the learner had identified the correct order of the sequenced story, the task advanced to the next sequence. As in the matching tasks, learners were required to ask at least one Yes/No question and one WH-question before they were allowed to guess a target picture.

### **Comparability in face-to-face and computer-guided interaction.**

Before discussing the comparative effects of recasts in oral and computer-guided interaction, it is necessary to explain the mechanics of Sasha and how it was capable of emulating oral-like interaction. The following two chapters provide a detailed description of the design and performance of the computational system that facilitated computer-guided interaction in this study.

## CHAPTER IV: SYSTEM DESIGN

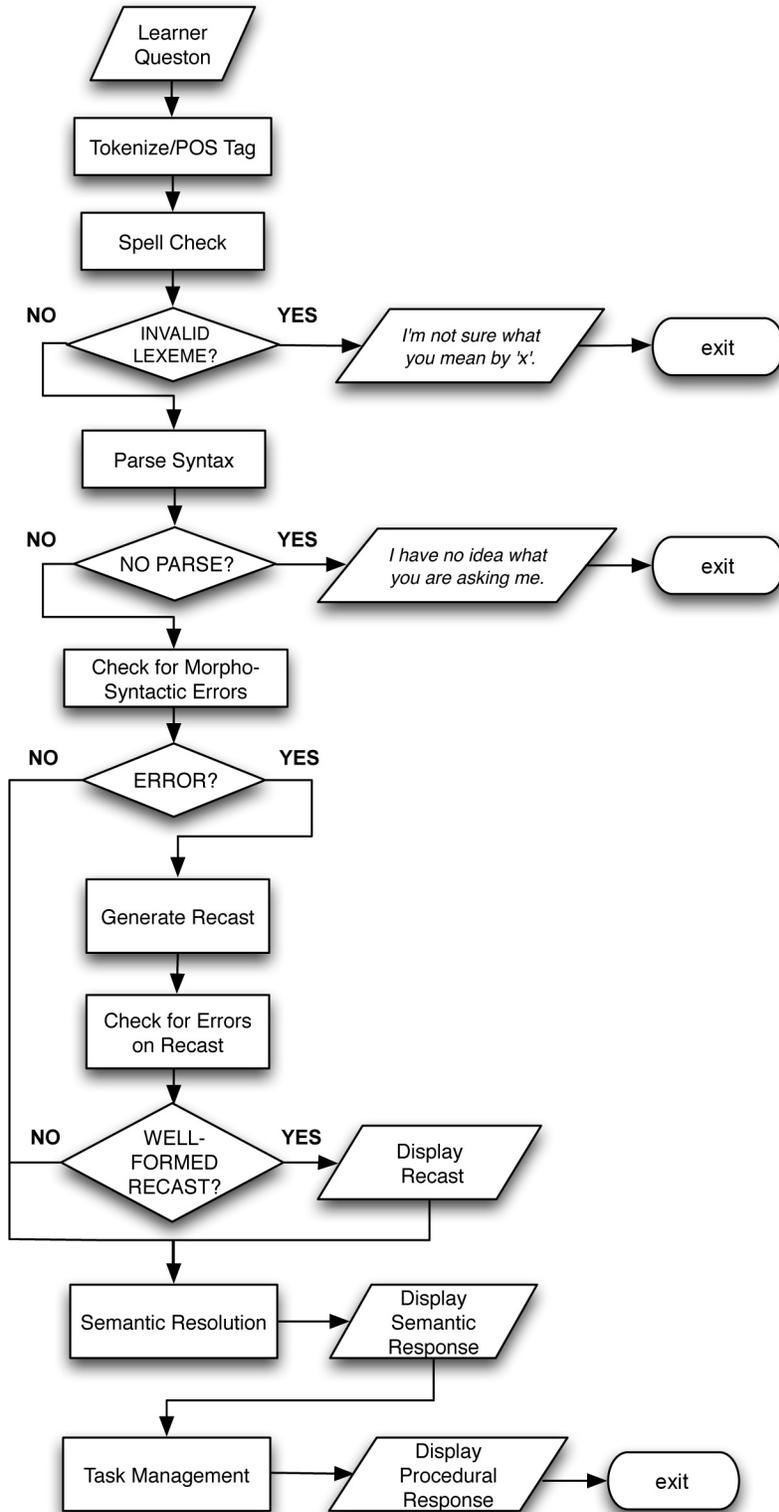
### Introduction

All of the task-based, computer-guided conversational interaction in this study was carried out using an 'intelligent' computational system that was designed by the author to (1) generate feedback in response to L2 learner errors, (2) provide semantically appropriate responses to learners' questions and (3) manage task procedures. All three of these functions were performed by a minimally embodied conversational agent—hereafter referred to simply as 'Sasha' (see Figure 3.9). Sasha's linguistic intelligence was founded upon the interweaving of a broad range of open-source natural language processing tools into a complex system for analyzing grammatical and semantic structures in written ESL questions.

### 4.1. System Overview

Figure 4.1 illustrates the general algorithm carried out by Sasha on all input questions. Each time a learner posted a question in one of the conversational tasks, the incoming question was checked for spelling errors, tokenized, lemmatized and assigned parts-of-speech (POS) tags. The tags and tokens were then parsed into a syntactic tree. The resulting structural data was then analyzed to identify any morphosyntactic errors in the input. If errors were found, a recast of the input question was generated. The recast generated by the system was then analyzed for morphosyntactic errors. If the recast was found to be error-free, it was presented to the learner, otherwise, no recast was displayed.

Figure 4.1. System overview.



Following morphosyntactic analysis and feedback generation, the question was analyzed for semantic intent and an answer was generated by comparing the semantic representation of the learner's question with that of the pictures being displayed. Finally, the visual cues on the page were updated to indicate progress in the task and, where appropriate, a procedural response (e.g., "Try asking me another question.") was generated to move the task forward.

## **4.2. Preparing L2 Input for Analysis**

### **4.2.1. Filtering 'noisy' data.**

Questions were entered into the system through an unrestricted text area by non-native speakers and were thus prone to a considerable amount of noise (i.e., imprecise use of punctuation, spaces and carriage returns). Therefore, each question was first stripped of all such typographical imprecisions. Additionally, the input question was analyzed to ensure that the input constituted of a minimally well-formed clause—containing a subject and a predicate. If a question did not contain a subject and a main verb, the system returned the statement, "I have no idea what you were asking me." In these cases, feedback and response processing were cancelled and Sasha waited for the learner to reformulate a new question.

### **4.2.2. Spell checking.**

In order to handle the large amounts of typographical errors inherent in L2 input, it was necessary to develop a robust system of spell checking. Before each question was sent through the system for syntactic and semantic processing, spell checking was employed to both find *and correct* typos made by the learner. It is important to note that the spell checker was not only responsible for identifying typographical errors, but for

correcting them as well. This means that, without confirmation from the learner (e.g., 'Did you mean *frog?*'), typos were forcibly replaced in the input string with their spelling suggestion before any analysis was conducted on the learner's question.

Figure 4.2 illustrates the procedure employed in Sasha's spell checker. The spell checking module relied upon the *suggest()* method available through Google's open API (Application Programming Interface) in conjunction with (a) the lexicon in Wordnet (Fellbaum, 1998) and (b) the semantic representations generated in Sasha's activity model for the pictures being displayed in the task (see 4.7.1. *The Activity Model*, below, for a detailed description of the representation of semantic knowledge in the target picture prompts). Once the original input string had been tokenized, the tokens were passed as a contiguous string to Google's Web-based API to see if the *suggest()* method would return a context-sensitive spelling suggestion. If a suggestion was returned, the original input and the suggestion were compared to find the token-to-token differences between input question and suggested question.

For each token-to-token mismatch, the suggested token was POS tagged and both of the tokens were stemmed. The mismatched lemmas were then validated in the Wordnet lexicon. In order to be valid (in both Wordnet and the Activity Model) the lemma needed to exist in the lexicon with the same POS assignment as that determined by the Brill Tagger.

If Google offered a suggestion and the input lemma was not valid in Wordnet, it was likely to have contained a typographical error or misspelling. In this case, if the suggested lemma was valid in the Wordnet lexicon, the suggestion was accepted. Otherwise, Google *suggest()* offered a word that could not be found in the Wordnet

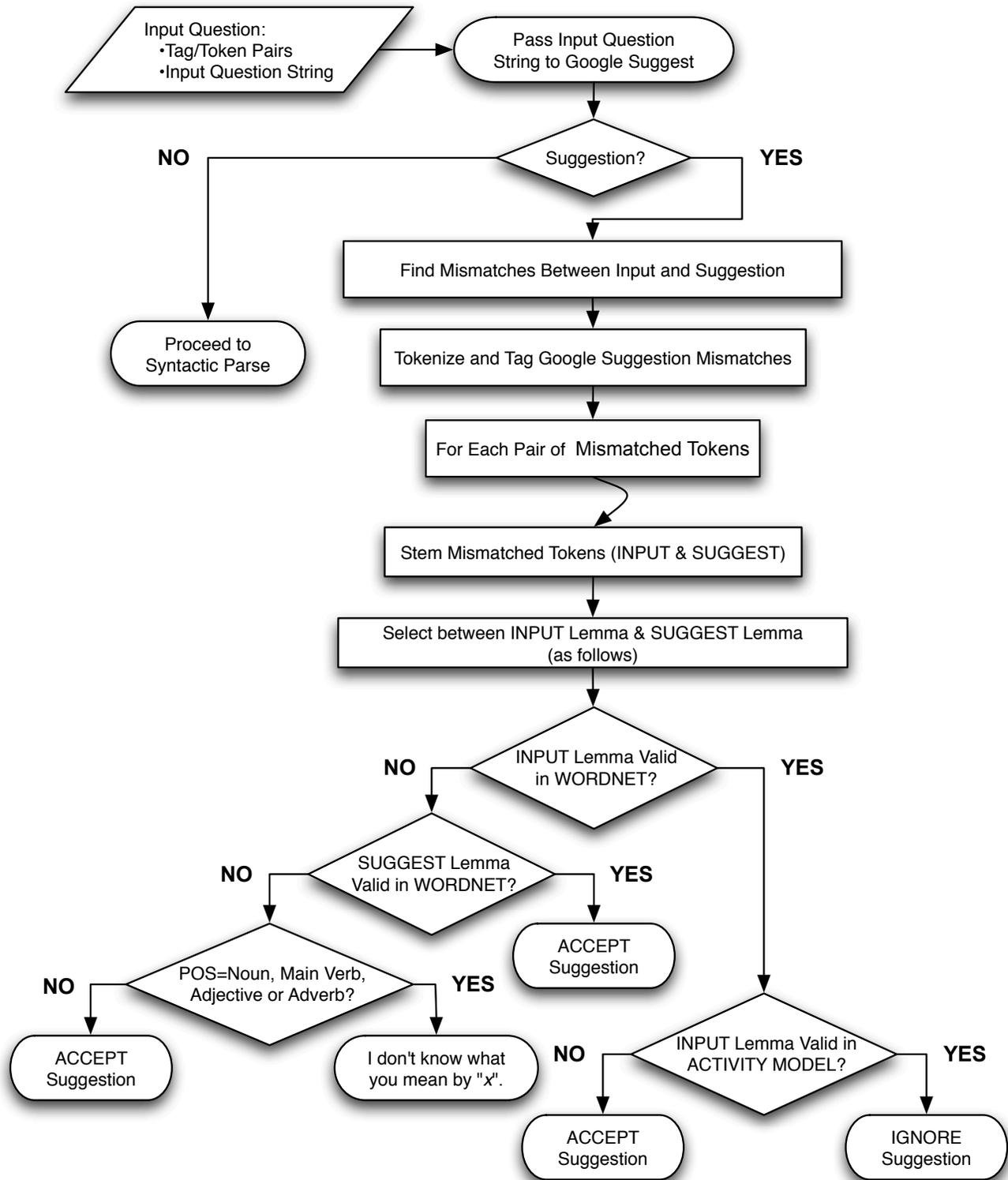
lexicon. If the POS assignment of the suggestion was not one of the five Wordnet POS categories (e.g., determiner, preposition, pronoun, etc.), the suggestion was accepted. Otherwise, the system returned the statement, *I don't know what you mean by 'x'*.<sup>2</sup> and all further processing was cancelled.

If a suggestion was offered and yet the input lemma was valid in Wordnet, the system turned to the lexicon in Sasha's Activity Model to determine whether or not to accept the suggested token. In these cases, the learner submitted a low-frequency word for which Google offered a higher frequency alternative that was similar in spelling. For example, if the learner had asked a question about a 'buoy' and Google suggested that the token be replaced with 'boy', the system looked for the presence of the input lemma in the Activity Model's microworld representations for all of the pictures that were on display at that point in the task (for more on microworld representations, see 4.7.1. *The Activity Model*). If the input question was valid in the Activity model (i.e., there was a buoy in the picture) the suggestion was ignored. Otherwise, where there was no evidence that the learner was actually trying to use the low frequency form (i.e., there was no buoy in the picture) the suggestion was accepted.

---

<sup>2</sup> Where *x* is the token submitted by the learner.

Figure 4.2. Spell checking and correction.



### 4.2.3. Disambiguating possessives and contractions.

Apostrophe-s constructions presented a source of considerable ambiguity in parsing the structure and semantics of a statement. For example, the token *man's* in the following pair of L2 questions have radically different implications on the syntactic and semantic representation of the question.

*Examples:*

1. What color the man's shirt?
2. What the man's doing?

During the tokenization routine, apostrophe-s constructions were marked for further analysis. Following POS tagging, apostrophe-s constructions that were (1) not immediately followed by a determiner and (2) followed at some point in the string by a noun with no intervening verb (e.g., *Example 1*) were assigned a POS tag of 'POSS'—indicating that the apostrophe is a possessive marker. Otherwise, the apostrophe-s construction was inserted prior to syntactic parsing as a separate token ('is'), and assigned the POS tag 'VBZ'.

### 4.2.4. Removing superfluous words and phrases.

The final preparatory procedure was to remove any initial words or phrases that might be considered superfluous to the structural and semantic analysis of the input. Superfluous phrases included initial interjections and parentheticals (e.g., *oh, so, well, then, maybe*) as well as given material in terms of the discourse context and information structure (e.g., *in this picture, in your opinion*). These words and phrases provided no structural or semantic value to Sasha and could have served to produce unnecessary variation in the parsing of the syntactic tree. They were, therefore, cropped from the input

before parsing.

### **4.3. Syntactic Parsing**

In order to provide an accurate account of the syntactic structures produced by beginning and intermediate ESL learners, Sasha began by analyzing the output of a pair of open-source NLP tools—the Brill POS Tagger (Brill, 1992) and Collins Parser (Collins, 1996). The POS tagger assigned each token a POS tag from the Penn Treebank tag set (Marcus, Santorini & Marcinkiewicz, 1993). The parser generated a syntactic tree based upon the token/POS tag pairs present in the input. Crucially, each of these tools was designed to analyze well-formed English input—not the wildly complex variation one might expect from a non-native speaker. Sasha therefore used the morphosyntactic analysis provided by these tools merely as a jumping off point for analyzing the L2 input, generating feedback and responding to learners' questions.

#### **4.3.1. L2 retraining of the syntactic parser.**

Sasha's syntactic analysis relied foremost upon the production of a syntactic tree by the Collins Parser. The Collins Parser is a lexicalized probabilistic context-free grammar (PCFG) parser that was trained on a set of hand-parsed sentences and attempts to produce the most likely structural analysis of a given sentence. In order to instruct the parser how to produce reliable syntactic trees for L2 input, it was retrained to consider a number of canonically ill-formed structures upon which it might otherwise fail to produce a consistent parse. Prior to retraining, the library of example sentences was populated with hundreds of derivative examples of a wide array of canonically ill-formed question structures. The canonical structures were abstracted from the entirety of example sentences reported in the question formation literature as well as from the complete

transcripts of the data collected in an oral SLA study.

The retraining of the Collins Parser was made possible through the use of the *TrainerEventToCollins* class of the Bikel Parser (Bikel, 2005). This class converted customized Penn Treebank merge files—which contained trees exemplifying ESL questions—into L2-sensitive event files to be used by the Collins parser during interaction.

#### **4.3.2. Daemonized implementation of Collins Parser.**

Sasha was initially built to utilize the Collins Parser through the CPAN Perl Module *Lingua::CollinsParser*. However, this module was designed for parsing large corpora—not individual statements. Because PERL scripts are not persistent and the module needed to load the statistical model from scratch for each sentence, the response time proved to be prohibitively slow to facilitate realistic conversational interaction.

The use of the Perl Module *Lingua::CollinsParser* was dropped in favor of a daemonized version of the Collins Parser created by Kan (2003). In Kan's implementation of Collins, the hash table of events was loaded only once at the start of a session, and then the executable forked to place a copy of itself in the background. Kan's source code was modified to reduce the daemon's wake interval from 15 seconds to 100 microseconds—thus allowing instant response to learners' questions.

Using the daemonized version of the Collins parser posed a number of challenges. First, the interaction between the Sasha's Perl modules and the parser daemon were asynchronous. In other words, a mechanism needed to be devised to bridge the communication between Sasha and the Collins daemon. Secondly, Sasha needed to be

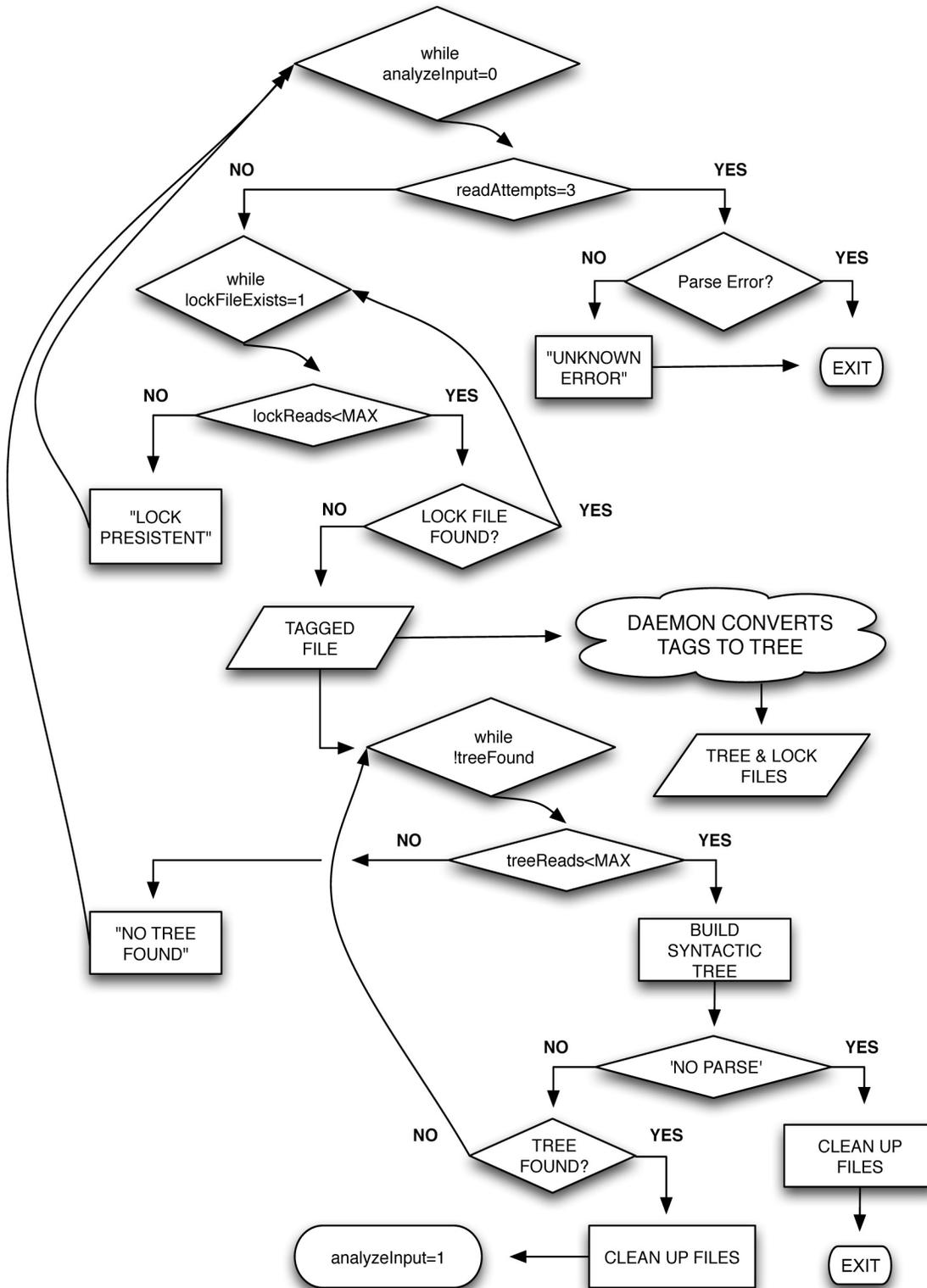
prepared to interact with multiple simultaneous users—so it was vital to ensure that each syntactic tree matched the corresponding input.

Figure 4.3 shows the interaction between Sasha and the Collins daemon. Once Sasha had prepared the input for syntactic parsing, it produced a file containing the tagged, tokenized input and produced a lock file to prevent any other users from processing the syntactic tree produced from the tagged file. Sasha then waited for the daemon to produce a file containing the complete syntactic tree.

The production of a tagged file triggered the Collins daemon to initiate a syntactic parse. Once the daemon had completed the parse, it replaced the tagged file with a file containing the syntactic tree generated by the parser.

After Sasha retrieved the syntactic tree file produced by the daemon, it cleaned up the files on the system and removed the file lock. As soon as the lock was removed, the next request in the dialogue queue was ready to be processed.

Figure 4.3. Interaction between Sasha and the Collins Parser daemon.

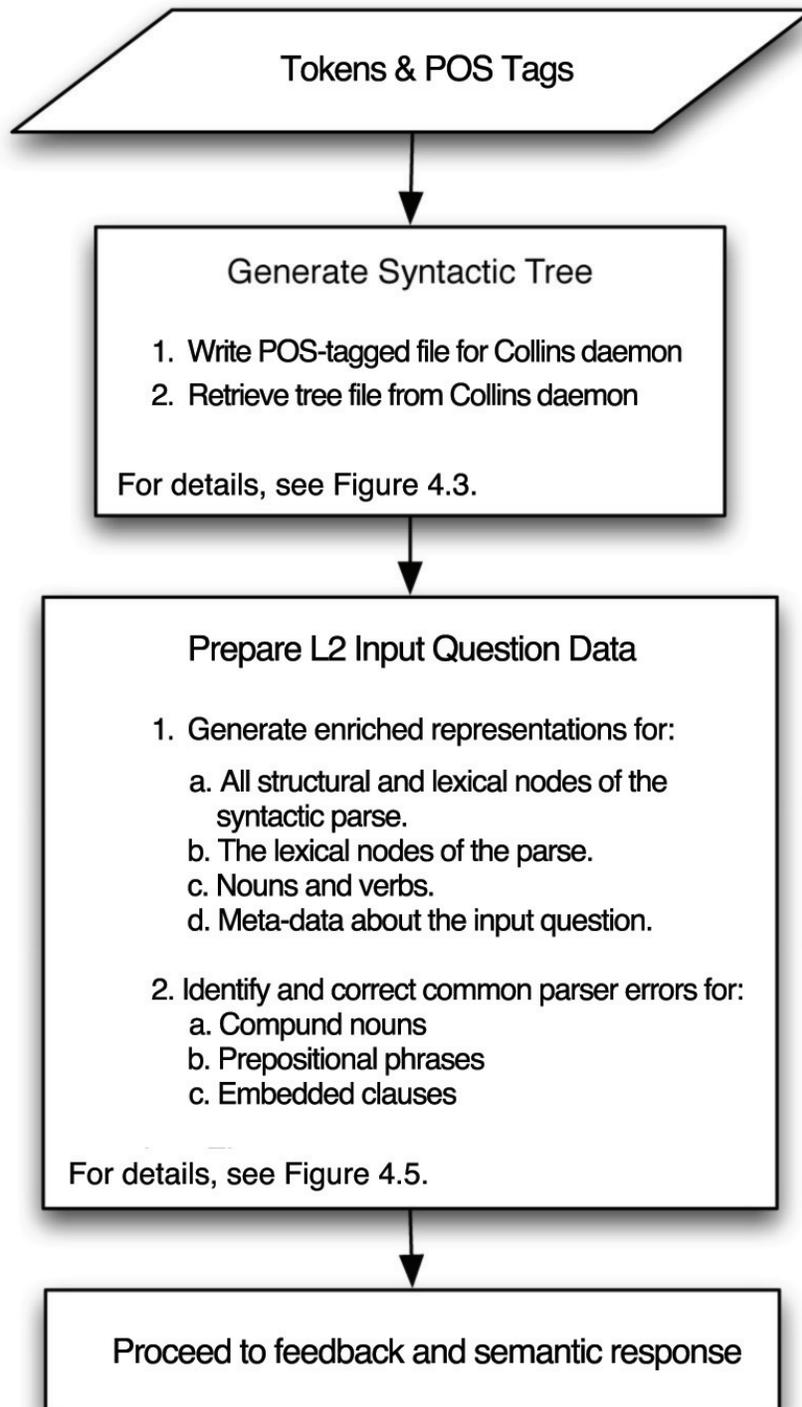


#### 4.4. Enriching the Input Question Data

Once a syntactic parse was received from the Collins Parser, Sasha began analyzing the input question in preparation for the generation of corrective feedback and a semantic response (see Figure 4.4). Prior to feedback generation and semantic processing, the system constructed a series of enriched data structures to assist in the process of syntactic and semantic analysis. Preparation for analysis began by converting the output of the syntactic parse into an enriched representation of each node (both lexical and structural) found in the tree. This syntactic node data (see *allNodes*, below) was, in turn, used to generate a further enriched representation of each lexical node in the tree (see *lexicalNodes*, below), along with separate representations for all of the nouns and verbs present in the input (see *nouns* and *verbs*, below). Finally, Sasha used the information contained in these enriched structures to analyze the question and produce a set of properties describing the data of the input question as a whole (see *questionData*, below). The enriched data structures for the input question were used as the basis for all subsequent feedback generation and response provision.

Figure 4.4. Preparing L2 question for analysis.

---



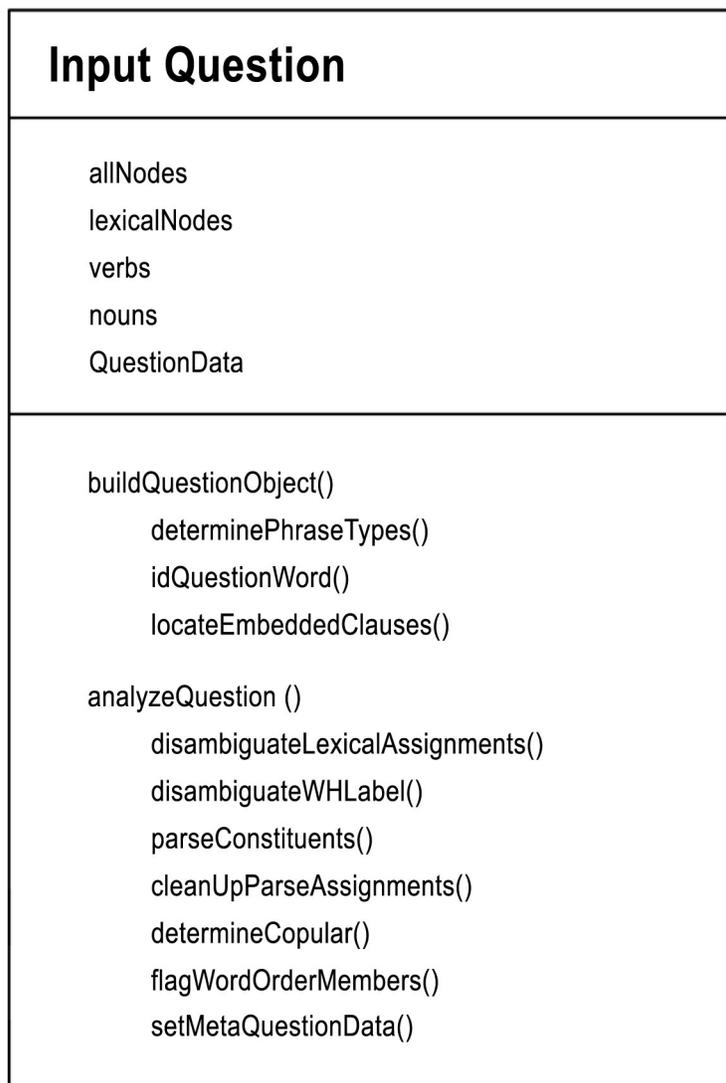
#### **4.4.1. Post-hoc checks on parser output.**

The Brill Tagger and Collins Parser were not designed to interpret non-targetlike L2 input. It was therefore necessary to employ robust measures to reconcile L2 input with syntactic structures predicted by the parser. In addition to retraining the parser to anticipate canonically non-targetlike usage, it was necessary to perform post-hoc checks to determine that the integrity of clausal and phrasal assignments were not corrupted by non-targetlike omissions or errors. Sasha performed post-parse checks on the integrity of the structural representation of compound nouns, prepositional phrases and embedded clauses.

#### **4.5. The Input Question Module**

Prior to feedback generation, syntactic and semantic data representations were generated for each question submitted to the system. Figure 4.5 presents a UML (Unified Modeling Language) table of Sasha's *InputQuestion* module. The items in the top cell represent the enriched data structures for the input question, the bottom cell lists the individual functions responsible for accurately populating the data.

Figure 4.5. Sasha::InputQuestion.



The following is a brief explanation of the enriched data representations built by the *inputQuestion* module:

***All nodes.***

The *allNodes* data structure was a flattened list of all lexical and structural nodes in the syntactic tree. Each item in the list contained detailed information about the node's

structural assignment and served primarily to build the *lexicalNodes* data structure. Each member of *allNodes* contained the following properties:

1. *nodeKey*: The unique identifier for the node.
2. *token*: If present, the token associated with the node.
3. *POS tag*: The word-, phrase- or clause-level tag associated with the node.
4. *phraseKey*: The unique identifier for the phrase associated with the node.
5. *phraseType*: The type of phrase in which the node appeared (e.g., NP, etc).
6. *clauseKey*: The unique identifier for the clause associated with the node.
7. *parentNodeKey*: A pointer to the parent node.
8. *parentPhraseType*: The type of phrase in which the parent node appeared.

### ***Lexical nodes.***

After all of the properties in *allNodes* were assigned, those nodes that contained lexical data (as opposed to those nodes that were purely structural) were copied to the *lexicalNodes* data structure. Each lexical item was further enriched by adding the following properties:

1. *serialPosition*: The serial position of the word in the string.
2. *lemma*: The stemmed form of the token.
3. *determinesStageType*: Indicates whether or not the token is essential for determining the stage-level assignment of the question (according to Pienemann and Johnston, 1987).

### ***Verbs.***

Because the information about the verbs and nouns in the input question needed to be accessed repeatedly in syntactic and semantic processing they were stored in separate data structures that were optimized for faster processing and analysis. The *verbs* data structure was a subset of *lexicalNodes* representing only the verbs in the input. Each verb contained additional information about whether it was a main verb or auxiliary verb.

### *Nouns.*

The *nouns* data structure was a subset of *lexicalNodes*, representing only the nouns in the input. Each noun contained additional information about its number and person features.

### *Question data.*

The *questionData* data structure contained general information about the type of question that was being asked. This data structure contained information on:

1. Sentence type: (e.g., SBAR,SBARQ,SINV,SQ,S).
2. Question word: (e.g., *who, what where, which, how much, how many*).
3. Copular status: Indicates whether or not the question is copular.
4. Main verb tense and aspect: The tense and aspect of the main verb in the main clause.
5. Subject number and person: Number and person features of the subject of the main clause.

## **4.6. Feedback Generation**

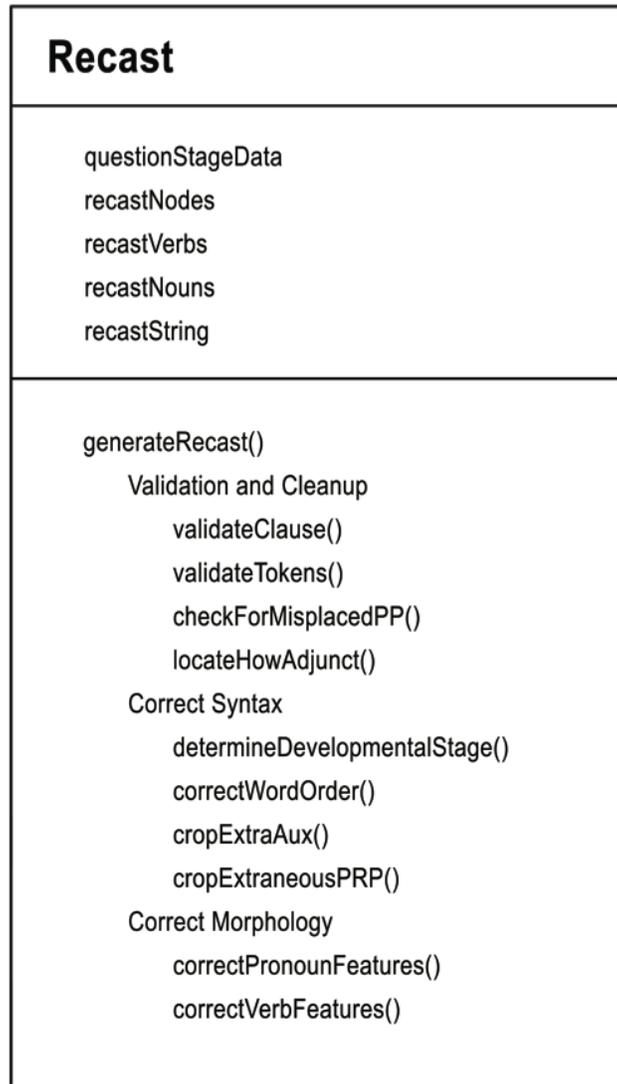
Once the input had been analyzed, Sasha began the task of generating feedback. The feedback type being investigated in this study (recast) was particularly challenging because (like the spell checker) the feedback generator needed not only to identify errors, but to *correct* them as well. Recast generation required the system to (1) identify a canonical rule for the input structure, (2) identify the target (well-formed) structure for rules that were deemed ill-formed and (3) execute the requisite morphosyntactic modifications for transforming ill-formed input into a well-formed recast.

### **4.6.1. The Recast module.**

Figure 4.6 presents a UML table of Sasha's *Recast* module. The data in the recast module initially represented the learner's question—exactly as it was formed.

Throughout the process of generating a recast, this data was transformed as errors were detected and fixed until a final, well-formed recast string was produced.

Figure 4.6. Sasha::Recast.



The following is a brief explanation of the dynamic data structures that were built and manipulated by the *Recast* module:

### ***Recast nodes, verbs and nouns.***

The initial values of all items in these data structures were copied from the final state of their corresponding data structures in the *inputQuestion* module. Throughout the process of recast generation, the values in these data structures were manipulated extensively to transform ill-formed input structure into its target-like form.

### ***Question stage data.***

The *questionStageData* data structure contained general information about the type of recast that was generated. This library contained information on:

1. The syntactic phrase structure rule on which the input question matched.
2. The error type of the input (morphological, syntactic, none or both).
3. The developmental stage level of the input.
4. The developmental stage level of the recast.

### ***Recast string.***

The *recastString* was the reformulated question that was constructed as the output of the recast module and potentially presented to the learner.

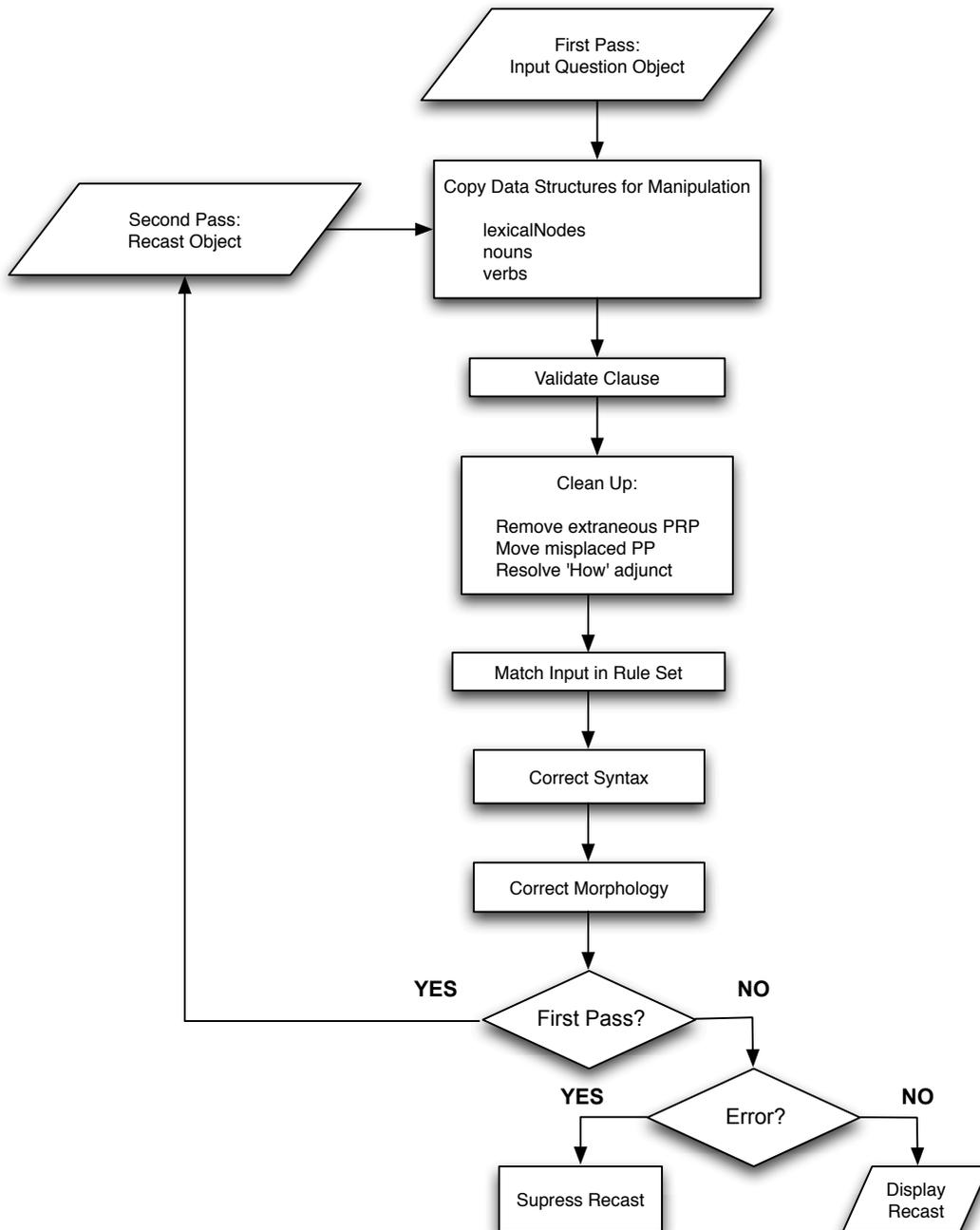
#### **4.6.2. Repairing syntactic errors.**

Figure 4.7 illustrates the general algorithm employed for detecting and repairing morphosyntactic errors. Recast generation took as input (1) the information on the lexical nodes resulting from the analysis of the learner response, and (2) the enriched representations obtained for the verbs and the nouns. The system first checked whether there was at least one noun and a main verb in the response. If the clause was not valid, all further processing was cancelled.

Prior to rule-set matching, several canonical L2 errors were corrected, if detected in the input. When any of these errors was present, the input was flagged as containing a syntax error. First, extraneous pronouns were deleted (e.g., *What the man he is doing?* ->

*What the man is doing?*). Next, misplaced prepositional phrases were moved to the end of the question (e.g., *Who in the chair is reading a paper?* -> *Who is reading a paper in the chair?*). Finally, the adjacency of the compound question words *how much* and *how*

Figure 4.7. Recast generation.



*many* was checked and corrected if the words were separated by any other words (e.g., *How is much flowers on table? -> How much flowers is on table?*). If this error was found, the word that separated the compound words was moved to the position following the subject of the question.

### ***Rule-set matching.***

Sasha's recast generator was equipped with a rule set that contained both well-formed and mal-formed rules (mal-rules) about English question formation. Before beginning the process of recast generation, Sasha needed to match the learner's input with one of the rules present in the rule set. Based on the data in the input question, a match string was produced to be used in regular expression matching (see Figure 4.8 for example match string). The match string consisted of the phrasal Penn Treebank tag for each noun, verb and question word in the input. Determiners, adjectives, adverbs and prepositional phrases played almost no role in rule matching since the rules are meant to exemplify possible combinations of question words, noun phrases and verb phrases. WH- and VP phrases contained additional information, in square braces, necessary for regular expression matching in the rule set. Each WH question word and verb listed its token and POS tag. Verbs were additionally marked as being either main or auxiliary.

The match string was compared to a set of over 100 rules that were meant to represent, as well as possible, the entirety of predictable ESL question structures. These rules were designed by the author based on sample data obtained from the existing research on ESL question formation. Each rule was expressed as a regular expression for matching against the input match string.

For ease of development, a shorthand system was created for that allowed complex regular expressions to be created by arranging a few simple shorthand variables (see Figure 4.8 for example). During rule matching, each shorthand variable produced a regular expression that matched the format of the input match string. For an example of an ill-formed question, its match string and the regular expression in the rule set on which it matched, see Figure 4.8. For a complete list of shorthand variables and rule sets, see Appendix G.

Figure 4.8. Rule-set matching example.

---

**Question:**            *What the man doing?*

**Match String:**    WH[what,WP] NP VP[1,doing,VBG]

**Shorthand Rule:** (\$WHADVP|\$WHNP) ?(\$AUX|\$ISNT|\$DONT|NOT)? \$NP \$VPG

**Expanded Rule:** (WH[\w+,WRB]| WH[\w+,WP]) ?(VP[0,\w+,VB(Z|P|D)?]|  
 VP[(0|1)+,be, VB(Z|P|D)?] NOT| VP[0,do VB(Z|P|D)?] NOT|  
 NOT)? NP( PP)?( CC NP( PP)?)? ((VP[(0|1)+  
 (gets?|getting|becomes?|becoming)+  
 (VB(Z|P|D)?|VBG)+])?VP[1,\w+,VBG])+

**Sample Variables:** \$word: "\w+"; # a contiguous string of letters  
 \$VPG: "VP[1,\$word,VBG]"; # any progressive verb  
 \$AUX: "VP[0,\$word,\$VB]"; # any auxilliary verb

---

In the example in Figure 4.8, the learner submitted a question with a WH-question word, a noun phrase and progressive verb. The match string included additional information about the question word (the token and the POS tag) as well as the verb (its designation as main verb, its token and POS tag). The question matched on a mal-rule that represented the templated structure found in the input match string. When matching input to rules, the system expanded each of the shorthand variables in the rule to form a complex regular expression string against which the input was matched. In other words,

the shorthand rule was the more human-readable representation of the long and unreadable expanded rule that the system used in regular expression matching.

If no rule was matched, the system cancelled any attempt at feedback generation and moved on to semantic processing.

***Correcting word order.***

Each rule was coded as being targetlike or non-targetlike. Each non-targetlike mal-rule triggered one or more specific word order manipulation functions that modified the data structures of the recast. The available word-order functions were:

*insertAux()*.

This function inserted an auxiliary into the specified place in the recast string. Depending on the context specified by the rule, the inserted aux may or may not have overwritten the word in the specified position. For example, the question, *What the man doing?* would not overwrite the word in the second position. Whereas the question *What does the man doing?* would overwrite the misused auxilliary *does*.

*frontAux()*.

This function raised a misplaced auxiliary verb in the recast. AUX could be raised to either the initial position of the string (Yes/No questions) or to the position immediately following the question word (WH- questions). For example, the non-targetlike Yes/No question, *The man is reading the paper?*, triggered this function to move *is* to the initial position of the question. While the non-targetlike WH- question, *What the man is reading?*, caused the auxiliary to be moved into the second position of the question.

*frontQuestionWord()*.

This function accounted for mal-formed constructions like, *The man is where?*. In this case, the question word was relocated to its proper position at the beginning of the question.

*frontNot()*

This function accounted for mal-formed constructions such as, *Why he not buying the watch?* In this case, *frontNot()* was triggered along with *insertAux()*. The system would first insert the 3<sup>rd</sup> person form of auxiliary *to be* in the second position and then call on this function to raise the negative particle to the position immediately following the verb. The recast string appeared to the learner as, *Why isn't<sup>3</sup> he buying the watch?*.

*unFrontMainVerb()*

This function accounts for mal-formed constructions like, *Writes the man a letter?* In this case, *unFrontMainVerb()* was also triggered along with *insertAux()*. After the inflected form of *to do* was inserted at the beginning of the question, this function moved the main verb into its proper position following the subject noun phrase and removed the 3<sup>rd</sup> person inflection. The recast string appeared as, *Does the man write a letter?*

*cutNode()*

There were times when an extraneous word in the input question was the cause for syntactic ill-formedness. For example, in a question like, *What was the man was doing?*, the learner has raised the auxiliary verb, but still allowed AUX to surface in its trace position. Another common example of a problematic extraneous node was when

---

<sup>3</sup> All negative auxiliaries were, as a matter of course, converted to contractions before being presented in recast. This was done to provide a natural sounding question in the recast.

the learner used both a concrete noun and a pronoun to indicate the same referent—for example, *What was the man he doing?* In cases such as these, the extraneous node was completely removed from the data structures in the *Recast* module. Note that the latter example was triggered during the cleanup process prior to rule set matching (discussed above).

#### **4.6.3. Repairing morphological errors.**

Following the correction of syntax, a number of repairs were made on the morphological inflection of the pronouns and verbs present in the data structures of the *Recast* module. For pronouns, Sasha examined and corrected the case assignments for number and person on all subject, object and possessive pronouns. (e.g., *What is her wearing?* → *What is she wearing?*) For auxiliary verbs, Sasha corrected the number and person features to correspond with the features assigned to the subject of the main clause (e.g., *Is the women happy?* → *Are the women happy?*).

#### **4.6.4. Recast confirmation.**

A central guiding principle for Sasha's recast generating mechanism was to first, do no harm to the learners' interlanguage development. The purpose of a recast was to advance development by providing a *well-formed* example of the structures that the learner was attempting to produce. Sasha was much more likely than a human, native-speaking interlocutor to provide ill-formed feedback. It was therefore necessary to provide a means for militating against the delivery of ill-formed recasts. As Figure 4.6 illustrates, each recast that was generated was subjected to a second pass through the recast generator before it was licensed for presentation to the learner. The purpose of this second pass was to ensure that the question generated by the system did not, itself,

contain any morphosyntactic errors. Essentially, the data structures generated by the first-pass recast formation were treated exactly as if they had been submitted as input by the learner. If the recast was found to contain any morpho-errors, it was suppressed; otherwise, it was licensed for delivery.

#### **4.7. Generating a Semantic Response**

In addition to supplying, form-focused, morphosyntactic feedback, Sasha was capable of providing meaning-focused, semantic responses to the learners' questions about the pictures on the treatment tasks. Following feedback generation, Sasha moved on to the task of trying to provide an appropriate response to the learner's question based on the semantic representation provided for the target picture(s). To answer the semantic content of a given question, Sasha's semantic response module compared the semantic relationships in the input question with those found in the semantic representations provided in the activity model (see below) for the picture set on display in the task. The *semanticResponse* module read in the semantic representation in the activity model and generated a semantic representation for the input question. All learner questions were answered based upon a comparison between the semantic representations generated for the input question and those specified in the activity model.

##### **4.7.1. The activity model.**

Sasha's ability to accurately answer incoming questions was facilitated by the presence of an activity model that was developed by the author prior to the treatment tasks. The activity model provided a semantic representation of the microworld visible in each of the pictures displayed during the tasks. As learners engaged with Sasha, they were instructed to ask questions about picture sets that they saw displayed in the

whiteboard section of the interface (see Appendix F). Sasha's semantic universe was limited to the microworlds represented in the pictures visible to the learner.

The creation of the semantic microworld representations in the activity model was made possible through a custom interface designed by the author for describing entities present in the task pictures and their relationship to one another. Figure 4.9 presents an example of the semantic markup interface.

Figure 4.9 Sasha: Semantic markup interface.

**Semantic Web: Prompt Markup Tool** Global Adjectives Global Characters

Go To Prompt Editor

Task Type: spotDiff

**Junkyard [B] (611868)**

[select prompt by name]

Nouns Verbs Ask Sasha:

Saved Nouns:

- bandana
- birdcage
- card[1]
- card[2]
- chip
- cinder block
- drink
- eyepatch
- game
- hand[1]
- hand[2]
- head
- Junkyard Bill
- pail
- parrot
- shoulder
- table
- tennis racket
- umbrella
- wall
- Junkyard Betty

Noun: table Save Adjectives

Senses Adjectives Locations Possessions Alternates

Saved Adjectives:

- gray
- grey
- round
- wooden

Adjective: old

<input type="checkbox"/>	<b>(used especially of persons) having lived for a relatively long time or attained a specific age</b>
<input checked="" type="checkbox"/>	<b>of long duration; not new</b>
<input type="checkbox"/>	<b>(used for emphasis) very familiar</b>
<input type="checkbox"/>	<b>skilled through long experience</b> <i>Synonyms: [older]</i>
<input type="checkbox"/>	<b>belonging to some prior time</b> <i>Synonyms: [erstwhile, formerone-time, onetime, quondam, sometime]</i>
<input type="checkbox"/>	<b>(used informally especially for emphasis)</b> <i>Synonyms: [honest-to-god, honest-to-goodnesssure-enough]</i>
<input type="checkbox"/>	<b>of a very early stage in development</b>
<input type="checkbox"/>	<b>just preceding something else in time or order</b> <i>Synonyms: [previous]</i>

This Web-based interface connected the pictures that appeared in the treatment tasks to the databases underlying in Sasha's activity model and Wordnet to allow rapid markup of the characteristics of the entities and events present in the task pictures.

In Sasha's activity model markup tool (see Figure 4.9), nouns, verbs and adjectives could easily be described so as to match on a broad range of appropriate lexical choices in the input. As new elements (nouns, verbs and adjectives) were added to any given picture, the interface presented all possible senses provided by Wordnet for the lemma being added. For each sense selected, all of the synonyms and hypernyms assigned to the word by Wordnet were included in the search scope for the lemma. The range of words included in the search scope for any word could be extended by manually adding alternate words (and their synonyms/hypernyms). For example, the picture in Figure 4.9 shows a woman with her feet in a pail. As the pail was added to the system, it was mapped to the following Wordnet sense:

**pail:** a roughly cylindrical vessel that is open at the top

*Synonyms:* [bucket]

*Hyponyms:* [cannikin, dinner bucket, dinner pail, dredging bucket, kibble, slop jar, slop pail, wine bucket, wine cooler]

This mapping would allow the entity of the pail to match on words like *bucket* and *jar*, but the entity might also be referred to as a *garbage can*. This term can be added as an alternate, broadening the search scope to include the following:

**garbage can:** a bin that holds rubbish until it is collected

*Synonyms:* [ash bin, ash-bin, ashbin, ashcan, dustbin, trash barrel, trash bin, trash can, wastebin]

The semantic markup interface, illustrated by Figure 4.9, allowed for rapid, yet thorough, description of the entities portrayed in the task pictures, along with their relationship to one another and the actions in which they were engaged.

In the semantic markup interface, nouns could further be described by an unlimited number of adjectives, as well as by their location and possession relationships vis-à-vis the other entities in the picture. Verbs could be further elaborated by describing the thematic roles of the nouns that they might have taken as arguments.

***Activity Model database.***

The semantic representation in Sasha's *activityModel* object was constructed from the values stores in a series database tables that were populated using the semantic markup interface. These tables comprised a thin lattice of indices that mapped the semantic properties, relationships and actions of the conceptual elements depicted in the displayed pictures with their semantically-rich referents in the Wordnet MySQL database (Bou, 2005. See Appendix H for the complete database schema).

*Database table: senseMap.*

The *senseMap* was essentially a network of pointers to words and synonym sets in the Wordnet database. Each record included a stemmed *word*, its Wordnet values—*wordid* and *synsetid*—and POS type (see Figure 4.10).

Figure 4.10. Activity Model: Database table: *senseMap*.

---

word	[text]
wordid	[integer]
synsetid	[integer]
posType	[set] {'noun', 'verb', 'adjective'}
quantity	[integer]
parentid	[integer]

---

The synsetid value was used for generating synonym and hyponym sets for all nouns, verbs and adjectives. This allowed the system to cast a broad net in matching elements in the input question with those in the activity model.

Each noun could additionally have a *quantity* value specified. For example, if there was a pile of 5 indistinguishable books in the picture, the word 'book' would have a *quantity* of value of 5. Each adjective contained a *parentid* to indicate the id of the noun that it described. The value of *parentid* was also used to map nouns and verbs to a specific task picture.

*Database table: locations.*

Locational relationships were stored in a concise table that identified a subject, an object and a location value (see Figure 4.11).

Figure 4.11. Activity Model: Database table: *locations*.

---

promptid	[integer]
subjectid	[integer]
objectid	[integer]
location	[set] { 'at', 'above', 'front', 'left' }

---

The potential number of values for the location field was limited to a reduced set of locational relationships. The value 'at', for example, subsumes ('in', 'at', 'on', 'inside', 'near', etc.). This reductive approach was intended to mitigate the difficulty in marking up and analyzing the full range of locational relationships.

The locational relationships stored for the pictures were also bi-directional. In other words, if *x was* to the left of *y*, it would hold that *y* is to the right of *x*. For example, if a picture depicted a man standing on the left of a woman, the relationship expressed in the database was defined as the subject *man* was left of the object *woman*. If the question

was asked, *Where is the woman?*, the system performed a reverse lookup on the locational relationship to determine that she was to the right of the man.

*Database table: possessions.*

Possessional relationships were also stored in a concise table that identified a subject and an object of possession. Any noun that possessed another was mapped as the subject of possession. The possessed noun was mapped as the object of possession (see Figure 4.12).

Figure 4.12. Activity Model: Database table: *possessions*.

promptid	[integer]
subjectid	[integer]
objectid	[integer]

*Database table: thematicRoles.*

The final database table contained information about the thematic roles of the scenes depicted in the pictures (see Figure 4.13). This table mapped any number of nouns to a specific verb and identified the thematic role of each noun.

Figure 4.13. Activity Model: Database table: *thematicRoles*.

verbid	[integer]
nounid	[integer]
role	[set] {'agent', 'theme', 'recipient', 'goal', 'source', 'beneficiary', 'instrumental', 'location' }

### ***Building the Activity Model.***

The semantic representation in the *activityModel* module was dynamically generated for each question based upon the elements found in the *senseMap* table that corresponded with the pictures being displayed to the learner.

Figure 4.14 presents a UML table of Sasha's *activityModel* module. The items in the cell on the left represent the data structures that comprised the semantic representation of the picture set being displayed, on the right are the individual functions responsible for generating the semantic libraries in the activity model.

Figure 4.14. Sasha::Activity model.

<b>Activity Model</b>	
nouns	generateActivityModel(promptid)
direct_lexical_selection	buildNouns()
elementid	buildMatchLemmas()
synsetid	buildAdjectives()
matchLemmas	buildLocations()
weight	buildPossessions()
lemmas	buildVerbs()
adjectives	buildThematicRoles()
direct_lexical_selection	getAlternates()
matchLemmas	assignParentIDs()
weight	getSenses()
lemmas	getSynonyms()
locations	getHyponyms()
location	
objectid	
possessions	
elementid	
role	
verbs	
direct_lexical_selection	
elementid	
synsetid	
matchLemmas	
weight	
lemmas	
thematicRoles	
role	
nounid	

The builder functions retrieved data stored in the Sasha and Wordnet databases and generated a series of semantic representations that were used to match against those

in the input. As in the semantic representation of the input question, the activity model consisted of multi-dimensional data structures that provided semantic information about nouns and verbs in the displayed pictures.

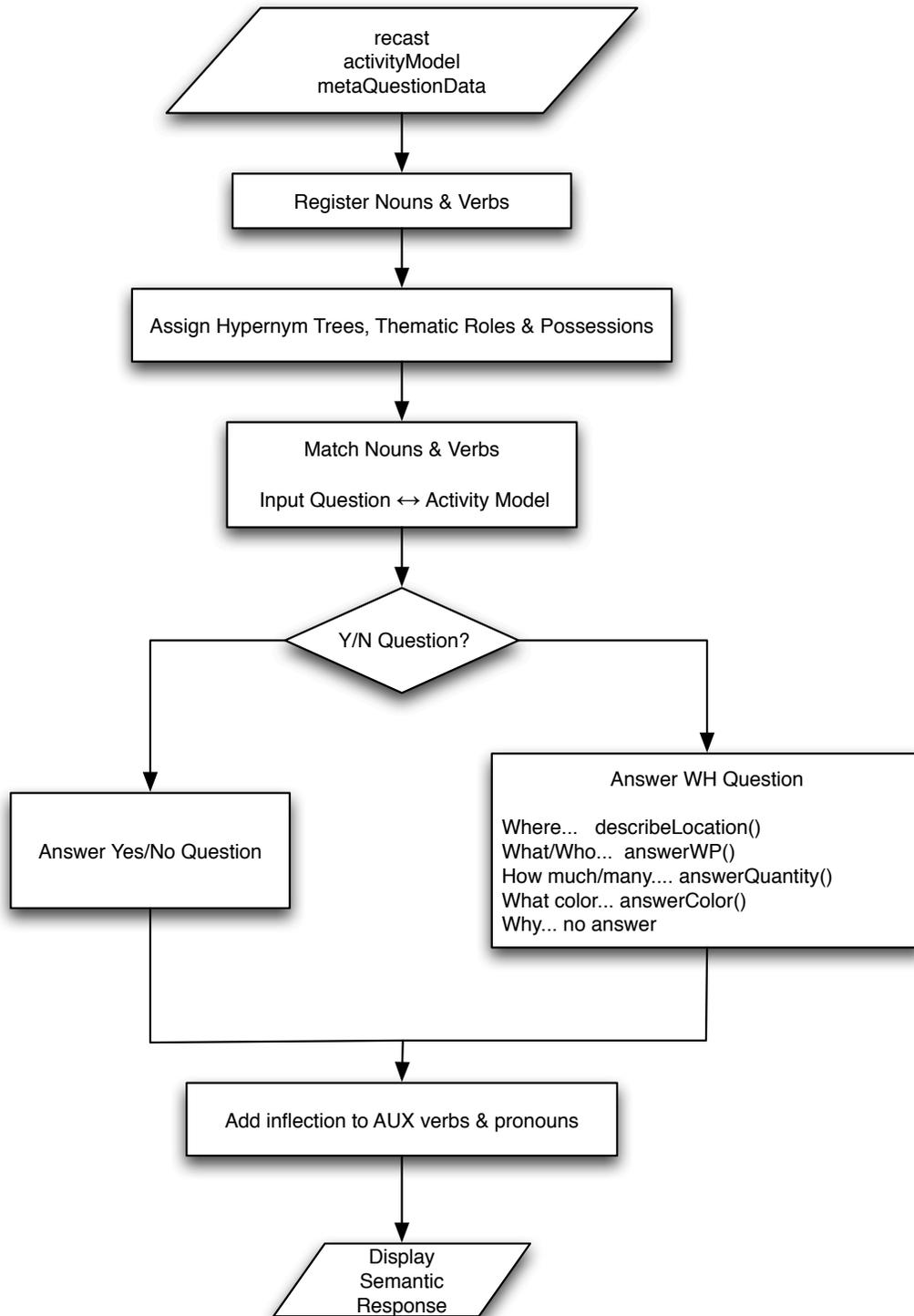
#### **4.7.2. The Semantic Response module.**

The *semanticResponse* module was responsible for answering questions about the pictures in the computer-guided tasks. The module started by reading in the libraries from (a) the *Recast* module (corrected syntactic input libraries) and (b) the *activityModel* (semantic representation of the displayed pictures). It then generated a semantic representation for the input question and matched the elements therein with those of the semantic libraries in the activity model. Once all elements had been matched, the question was answered and a response was generated. Figure 4.15 illustrates the process of semantic response generation.

##### ***Producing a semantic representation for the Input Question.***

The semantic response module began by copying the nouns and verbs data structures from the *Recast* object and converting them into semantically-enriched data structures that served as the basis for semantic representation of the input question. For both *nouns* and *verbs*, each lemma was matched in the Wordnet database and the resulting synsetids (senses) were added to the corresponding element in the semantic representation of the input question.

Figure 4.15. Semantic response.



Through a series of diagnostic functions, the *nouns* data structure was further enhanced by adding hypernyms trees, inner-phrasal relationships, possessions, locations and thematic roles. Figure 4.16 shows the structure of the semantic representation of nouns in the learner's question.

Figure 4.16. Semantically-enriched nouns in the Input Question.

---

Nouns	
lemma	
elementid	
senses	
gender	{'masc','fem','neut'}
number	{'sing','plur','noncount'}
hypernyms	
adjectives	
PP_semHead	
PP_objects	
isPossessive	
possessions	
locations	{
	location {'left','right','on',etc}
	objectid
	}
thematicRoles	{
	role {'agent','theme','source', etc.}
	verbLemma
	verbid
	}

---

***Matching Input Question elements in the Activity Model.***

The most critical task of semantic response generation was accurately matching elements in the input question with their counterparts in the activity model. This task becomes especially difficult when there were multiple elements in the activity model upon which the input element might match. For instance, a picture might contain two or

three men. One may be old, appear happy and be reading a book, another might be young, angry and talking on the telephone, and so on. When the question was asked, *What is the man doing?*, Sasha attempted to disambiguate which man the learner was referring to. In order to disambiguate elements the system looked at adjectives, thematic roles, possessions and locations to determine which lexically-matched element was the most appropriate referent. In the example above, Sasha would be unable to determine an indisputable referent in the activity model and would produce a response such as, *One of them was reading a book.* Had the question been posed as, *What is the old man doing?*, the system would have matched exclusively on the old man and produced the response, *He is reading a book.*

#### ***Answering questions.***

Once Sasha had matched input elements with their counterparts in the activity model, it moved on to attempting to answer the question being asked. In order to answer the learner's question, Sasha simply compared the properties of the matched elements in the input question with those of their counterparts in the activity model. First, however, it was necessary to determine what type of question was being asked.

#### *Yes/No Questions*

Yes/No questions were answered by analyzing the noun, verb and adjective structures of the matched elements in the activity model. Yes/No analysis considered adjectives, locations, possessions and thematic roles and returned a Boolean yes/no value. Example

Yes/No questions include:

1. *Is the man happy?*
2. *Is the man next to the table?*
3. *Does the man have a bird?*
4. *Is the man playing tennis?*

### *WH Questions*

There were a wide range of WH question types asked of the system, each requiring a specific type of answer. The WH question type was determined by the question word that initiated the question. Question words (and phrases) such as *Who, What, Where, How much/many*, etc. each triggered their own unique analysis. Note that Sasha was not designed to answer *When* and *Why* questions. The semantic representation in the activity model did not contain the temporal or inferential information that would be required to answer such questions. The task design employed in this study specifically attempted to limit the number of opportunities for *When* and *Why* questions. Not a single *When* question was asked throughout the three treatment sessions. When *Why* questions were asked, Sasha skipped the semantic analysis and simply said, *Sorry, I don't do well with 'Why' questions.*

*Who* and *What* questions required that a lexical value be returned to fill the semantic void represented by the question word. For example, if the learner asked, *What is the man eating?*, the system looked for the matched *man* as an agent of the verb *eat* and returned the lexical value of the theme of the verb. If there was no match found in the activity model, Sasha responded by saying, *I am not sure.*

*Where* questions required a location reference. For a question such as, *Where was the man?*, the system returned the location references found for the matched lemma *man*.

*How much/how many* questions required a quantity value to be returned for the noun in question. The system used the quantity value of the nouns data structure in the activity model to answer this type of question.

### *Formulating a response string.*

The last step in generating a semantic response involved transforming the core semantic response into a simple, well-formed statement to present to the learner. Rather than answering a question such as *What was the girl doing?* with only a verb (e.g., *running*), Sasha embedded the answer in a complete sentence (*She is running.*).

Figure 4.17. Example semantic responses.

<b>Question</b>	<b>Response</b>
1. Is the woman running?	Yes, she is.
2. Are the cards on the table?	No, they are not.
3. Is the boy fat?	I don't think there is a boy.
4. What is he reading?	He is reading a book.
5. Who is eating an apple?	The old woman.
6. What color is the book?	It is red.
7. Where is the parrot?	It somewhere near a birdcage.
8. How many people are there?	I see 2 of them.
9. What is the citten eating?	I don't understand what you mean by 'citten'.
10. He asd?	I am not sure what you were asking me.

### **4.8. Managing Discourse**

Sasha was designed to manage highly structured conversational interaction in which the learner asked a question and Sasha provided some combination of feedback, semantic response and procedural prompt. The management of Sasha's speech was controlled by a dialogue queue that received instructions about what to say and how long to keep the utterance visible to the learner. The queue only allowed one statement at a time to appear in Sasha's dialogue bubble. Once a statement had been displayed for its allotted time, Sasha paused for one second and then displayed the next message in the queue. All recasts and responses were allotted four seconds of display time with a one-

second pause in between. If a procedural response was required (e.g., the learner clicked on the picture she thought was the target), the system displayed the short response for three second.

The following chapter discusses in detail how well Sasha performed the tasks of delivering well-formed recasts and providing appropriate answers to the learners' questions in the treatment tasks.

## CHAPTER V: SYSTEM ANALYSIS

### 5.1. Overview

In order to explain the gains demonstrated in the Computer Guided Recast (CGR) group, it is necessary to provide a detailed analysis of the performance of the computational system employed in the experimental tasks. The task design employed in this study represents a giant leap in anchoring ICALL tasks to oral SLA theory. In order to emulate the open-ended conversational discourse inherent in oral, dyadic SLA tasks, the system accounted for largely unrestricted L2 input, provided targeted feedback for morphosyntactic errors and responded to the semantic content of the learners' questions.

In this experiment, each of the 19 learners in the CGR group engaged with the system for three 30-minute task sessions—during which, a total of 3,192 queries were presented to the system. For each question asked, the system provided a recast if an error was detected and then provided a response to the question based upon a comparison between the semantic representation of the input question with that of the microworld representations present for the pictures displayed in the task (see Figure 5.1).

Figure 5.1. Example question from treatment data.

---



Learner 43:           waht's the man is doing?  
Sasha (recast):       What is the man doing?

Sasha (response): He is jumping.

Throughout the treatment tasks, the system attempted to recognize and correct all lexical, morphological and syntactic errors that appeared in the free response L2 input and answer all questions accurately. Following the experiment, each question was coded to evaluate the accuracy of the system in delivering recasts and semantic responses.

In terms of feedback, the system provided a correct or partially correct recast in response to 65% of all questions asked containing morphosyntactic errors during the CGR treatment tasks. In terms of semantic responses, the system provided a valid response in for 86% of all questions asked by the learners.

Figure 5.2. System performance overview.

---

<b>Total Questions</b>	
Number of Queries	3192
Number of Valid Questions	2912
<b>Spell Checking</b>	
Accurate Instances	122
Inaccurate Instances	17
<b>Recast Provision</b>	
Good	481
Partial	78
Inaccurate	53
Missed	245
<hr/>	
Total Opportunities	857
<b>Semantic Response</b>	
Valid	2487
Invalid	281
Incomplete	144
<hr/>	
Total Opportunities	2912

---

In order to provide a detailed descriptive analysis of Sasha's overall performance, each question was coded to evaluate system accuracy (in terms of precision and recall) in the following categories: (1) spell checking, (2) recast provision and (3) semantic response (see Figure 5.2). Note that, given the unrestricted nature of the input, many of the queries posted in the treatment tasks were not valid questions. Queries were coded as invalid if they failed to meet the minimal NP + VP requirement of a valid question.

## **5.2. Spell Checking**

The spell checker proved very effective in the treatment tasks—doing far more good than harm. There were 139 instances where the spell checker accepted the replacement provided by *Goggle::Suggest*. The spell checker was accurate in 88% of these instances.

## **5.3. Recast Provision**

Of the 2912 questions asked, 857 questions contained some sort of morphosyntactic error that should have prompted the system to provide a recast.

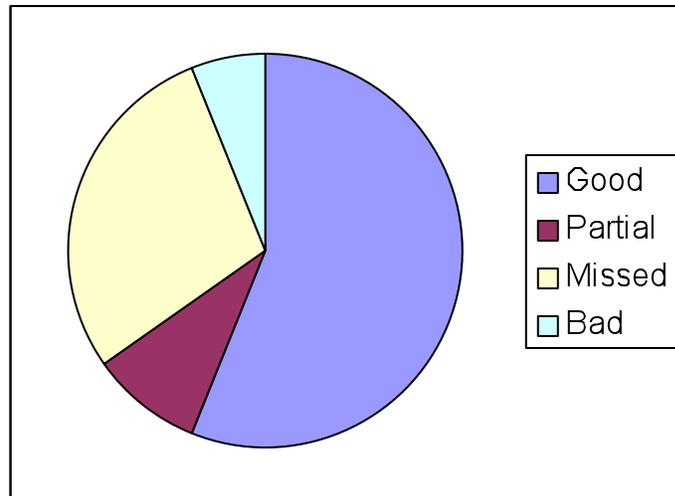
### **5.3.1. Good recasts.**

A recast is considered good if the system has accurately corrected all morphosyntactic errors and maintained the intended structure and meaning of the question. It should be noted that the system strives to emulate the real-life performance of a human interlocutor. In a typical oral task, the researcher does not provide recasts in response to 100% of the opportunities presented.

Figure 5.3. Recast performance overview.

---

Good	481	56%
Partial	78	9%
Missed	245	29%
Bad	53	6%
Total	857	



### 5.3.2. Partial recasts.

A recast was considered partial if the system accurately corrected some, *but not all*, morphosyntactic errors and maintained the intended structure and meaning of the question. In other words, in partial recasts, the system corrected one or more major errors, but incorrectly handled some other minor errors in the question. In the example recast in Figure 5.4, the system correctly reformulated the syntax of the question and the morphology of the subject and verb. It failed, however, to correct the morphology of the object 'a pants'.

Figure 5.4. Partial recasts.

---

*Example (Morphological error persists):*

Learner Question    he has a pants?  
Recast                Does he have a pants?

*Uncorrected Error Types:*

Morphological	69	88%
Syntactic	6	8%
Lexical	3	4%
Total	78	

---

***Uncorrected morphological errors.***

The majority of the uncorrected errors found in the partial recasts consist of morphological agreement errors such as that shown in the example in Figure 4. This is due largely to a lack of a well-articulated strategy for noun classification. For more on noun classification, see 5.5.1. *Noun Classification*.

***Uncorrected syntactic errors.***

Very few partial recasts contained syntactic errors. Generally, the presence of a syntactic error resulted in the in recast being coded as 'bad'. There were a few cases where minor syntactic errors persisted that did not invalidate the other corrections made by the system.

*Example:*

Learner Question: what the man is touching it?  
Recast:                What is the man touching it?

*Uncorrected lexical errors.*

In a few cases, spell checking failed and the recast contained a nonsensical lexical item:

<p><i>Example:</i></p> <p>Learner Question: he has a ha? Recast: Does he have a ha?</p>
---

**5.3.3. Missed recast opportunities.**

A recast was considered missed if there were errors present in the learner's question, but no recast was provided. Missed recast opportunities stemmed from three general shortcomings of the system: (1) mal-rule inaccuracy, (2) system inaccuracy, and (3) word sense ambiguity (see Figure 5.5).

Figure 5.5. Recast missed opportunities.

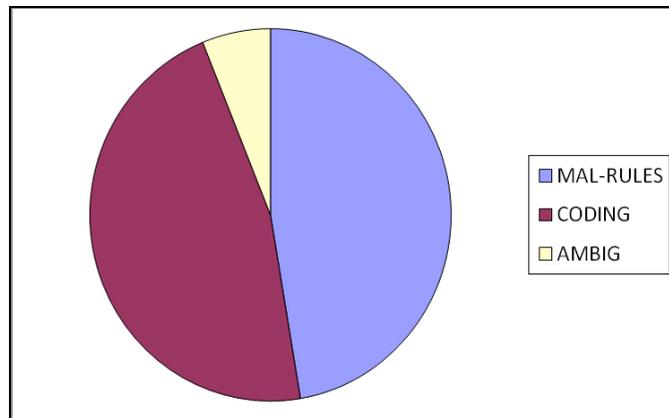
---

*Example:*

Learner Question Is a women?  
Recast none

*Reason for Suppression (245 Total):*

Mal-rule Inaccuracy	116	47%
System Inaccuracy	114	47%
Word Sense Ambiguity	15	6%
Total	245	



---

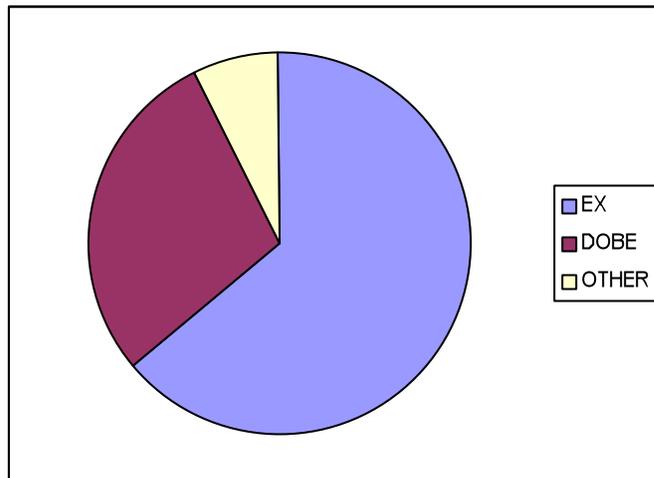
***Missed recasts: Mal-rule inaccuracy.***

There were 116 instances where the system failed to provide a recast because there was a failure to match on a mal-rule that would trigger the transformations required in order to recast the input question (see Figure 5.6).

Figure 5.6. Missed recasts: Mal-rules.

---

Existential Questions	70	56%
DO + BE + S	32	9%
Other	14	29%
Total	116	



---

*Mal-rule: Existential questions.*

Figure 5.6 illustrates that there were no rules in place to trigger the necessary insertion of the missing EX 'there' in ill-formed existential questions. For more on existential questions, see 5.5.1. *Existential Questions*.

*Mal-rule: DO + BE + S*

Another commonly occurring L2 pattern that was not represented sufficiently in the mal-rules was DO + BE + S.

*Example:*

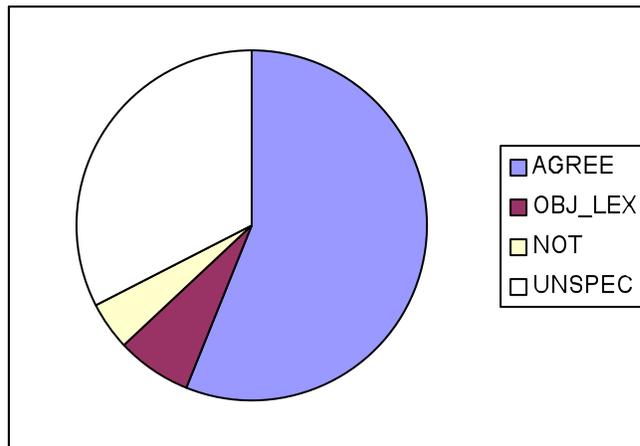
Learner Question: does is he writing?  
Recast: none

***Missed recasts: System inaccuracy.***

Figure 5.7. Missed recasts. System inaccuracy.

---

Agreement	60	53%
Human Referents	8	7%
Negative- 'or not'	5	4%
Unspecified	41	36%
Total	114	



There were 114 instances where the system failed to provide a recast because there was a failure by the code to properly account for the errors present in the input. There are three specific areas that account for 68% of these instances:

*Undetected agreement errors.*

Because the system was not able to robustly account for noun classes, many agreement errors went undetected. For more on noun classification, see 5.5.1. *Noun Classification.*

*Lexical inconsistency in object human referents.*

Human referents (*man, woman, person, someone, etc.*) were not properly resolved when they appeared in the object position of a phrase. The system treats human referents

as a special case in order to provide broader coverage for complex task of describing people. When human referents appeared in object position, the system encountered a bug that caused the referent to be dropped.

*Negation.*

The phrase 'or not' caused a failure to match on an appropriate mal-rule.

*Unspecified system limitations that lead to missed recasts*

There were 41 instances of missed recasts where an immediately clear systemic solution fails to emerge. This means that the system failed to deliver a recast where no clearly remediable reason emerges in response to less than 5% of the total number of recast opportunities.

These questions often contain wildly idiosyncratic constructions that are difficult, if not impossible, to categorize. In most cases (73%), these instances contain syntactic errors that caused the system to produce an error-riddled recast that the system rightly repressed.

*Examples:*

1. There are a women have wear a red shirt?
2. What the dog is during?
3. Where are is the many?

***Missed recasts: Word sense ambiguity.***

There were 15 instances where the system failed to provide a recast because of word sense ambiguity. For example, in many cases, the word 'play' was erroneously considered a noun, which led to an incorrect syntactic parse. For more on ambiguity,, see 5.5.1.

*Word Sense Disambiguation.*

### 5.3.4. Bad recasts.

A recast was considered bad if it contained morphosyntactic errors and/or it failed to maintain the intended meaning or structure of the question. The provision of a bad recast might potentially be considered the most critical error that system could make. Bad recasts provide examples of incorrect usage in the form of corrective feedback and would likely never occur in oral conversation with a native speaker as in the FTF condition.

Figure 5.8. Bad recasts.

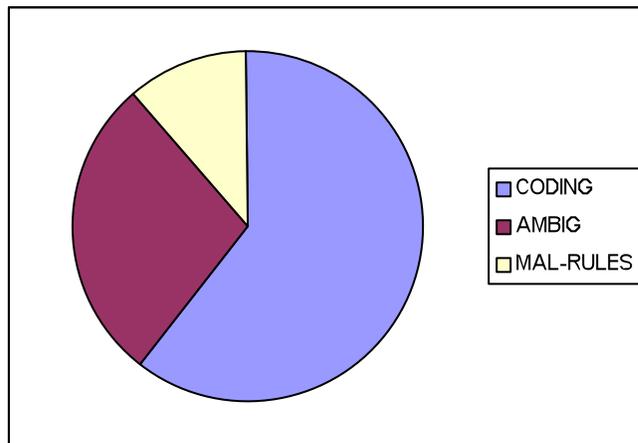
---

*Example:*

Learner Question    Are there two me?  
Recast                    Am there two I?

*Reason for Bad Recast:*

Coding Inaccuracy	32	60%
Sense Ambiguity	15	29%
Mal-rule Inaccuracy	6	11%
Total	53	



---

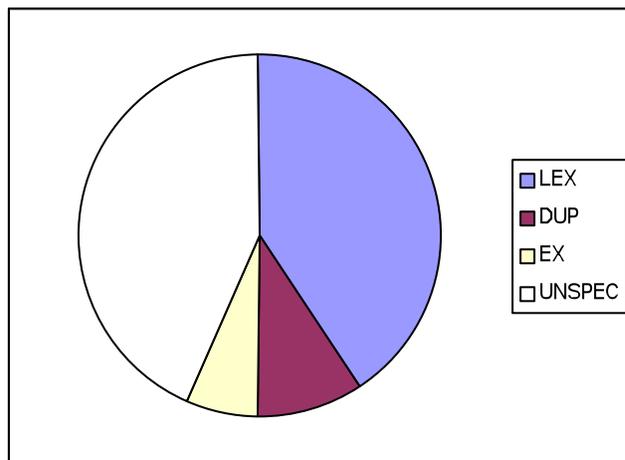
Sasha attempted to militate against the delivery of bad recasts by providing morphosyntactic error analysis on the recast before it is licensed for delivery (see 4.6.4.

*Recast confirmation*). Yet, because the system accepted largely unrestricted L2 input, there were instances where a syntactically or semantically non-targetlike recast was delivered to the learner. In total, 53 bad recasts were delivered—due largely to three general shortcomings of the system: (1) coding inaccuracy, (2) word sense ambiguity and (3) mal-rule inaccuracy.

***Bad Recasts: System inaccuracy.***

Figure 5.9. Bad recasts: System inaccuracy.

Lexical Validation	13	41%
Split Word Duplication	3	9%
Existential Questions	2	6%
Unspecified	13	41%
Total	32	



***Lexical validation.***

Nearly half of the bad recasts generated as a result of coding inaccuracy were the result of a shortcoming in the system's lexical validation procedure. Specifically, lexical validation failed in cases where the spell checker returned an invalid word (one that cannot be found in the Wordnet lexicon).

*Example:*

Learner Question: women is suiomin?  
Spell Checker: women is suimin?  
Recast: Are women suiminning?

In this example, lexical validation should have failed on the word 'suiomin' and returned the response, "I don't understand what you mean by 'suiminning' ". However, since the spell checker first returned the word 'suimin' the lexical validator assumed that it must be a valid lemma. In short, the system failed to properly account for cases where both the input and the spell check suggestion are, in fact, invalid.

*Split word duplication.*

A peculiar error occurred when the system was presented with a question that contained either (1) conjoined words (not separated by a space) or (2) split words, where the lexical string contained an extraneous space.

*Examples:*

1. Question: what **doyou** ben?  
Recast: What does ben ben?
2. Question: he run and **look ing** bak?  
Recast: Does he run and look back bak?

When the system corrected syntax, it relied on a series of pointers to know exactly where to insert and delete words in the corrected string. When split words were corrected early in the processing of the question, the total word count and position data was not properly updated . The pointer was thus off by 1 position and produced recasts that contained duplication.

*Existential questions.*

In a few cases, the absence of rules that triggered the necessary insertion of the missing existential lexeme 'there' causes the system to deliver a bad recast. For more on existential questions, see 5.5.1. *Existential Questions*.

*Unspecified system limitations that lead to bad recasts.*

There were 13 instances of bad recasts where an immediately clear systemic solution fails to emerge. In total, the system delivered an ill-formed recast where no clearly remediable reason emerges in response to less than 5% of the total number of recast opportunities.

These questions generally contained idiosyncratic constructions in which the system failed to suppress the ill-formed recast it generated.

*Examples:*

1. Question: hes this bad?  
Recast: Does this bad have?
2. Question: Why the gave some money to the other man?  
Recast: Why is the gave some money to the other man?

***Bad recasts: Word sense ambiguity.***

Word sense ambiguity lay at the root of 29% of the bad recasts delivered (for more on ambiguity, see 5.5.1. *Word Sense Disambiguation*).

***Bad recasts: Mal-rule inaccuracy.***

All 6 of the bad recasts generated as a result of a missing mal-rule suffered from the absence of the same type of rule. Each of these questions began with 'What' followed immediately by an NP (e.g., *What time..., What kind..., what color...*). This type of error

can easily be corrected by allowing the mal-rules to treat WHAT + NP the same way they do WHICH + NP.

#### **5.4. Semantic Response Provision**

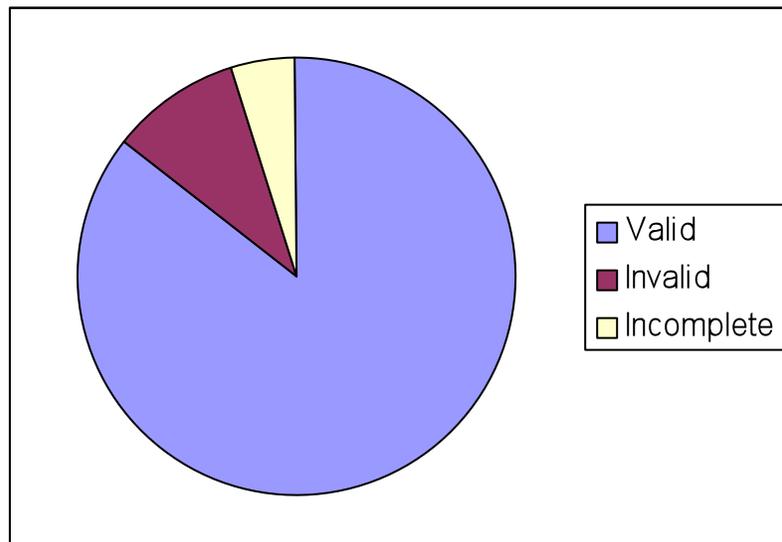
In order to sustain prolonged interaction with the learners in the CGR condition, Sasha needed to provide accurate and reliable answers to learners' questions about the pictures used in the treatment tasks. The decisions the learners needed to make to advance any task forward (e.g., guessing a picture or identifying a difference) relied solely upon the answers provided by Sasha. If the system had regularly provided incorrect information or failed to provide an answer, it is likely that the learners would have lost interest in engaging in collaborative interaction with Sasha.

Of the 2912 questions asked, the system delivered a valid, task-appropriate response for approximately 86% of the questions. The system gave outright incorrect answers in response to slightly less than 10% of the questions (see 5.4.2. *Invalid Responses*). In response to 4% of the questions, the system erroneously answered "I'm not sure" (see 5.4.3. *Incomplete Responses*).

Figure 5.10. Semantic response performance overview.

---

Valid Responses	2490	86%
Invalid Responses	281	10%
Incomplete Responses	141	4%
Total	2912	



---

#### 5.4.1. Valid responses.

Valid responses cover a broad range response types. It should be noted that a valid response is not necessarily the optimal response that a native speaker could have given, but rather one that is appropriate to:

1. The information available (without implicature) in the activity model for the pictures presented in the task.
2. The information provided in the learner's question.

Figure 5.11. Example semantic responses.

---



*Valid*

1. Learner Question: Where are they playing?  
Response: They are somewhere near the table.
2. Learner Question: What is the person doing?  
Response: I'm not sure which person you are talking about.

*Invalid*

1. Learner Question: Is the man playing cards?  
Response: I don't think there are any cards in this picture.

*Incomplete*

1. Learner Question: What is the woman holding?  
Response: I'm not sure.
- 

In 127 of the responses marked as valid, there was insufficient information provided in the learner's question to disambiguate which object is being asked about. For example, in *valid* example (2) in Figure 5.11, the system is not able to tell whether the learner is talking about the man or the woman, since no disambiguating information has been provided in the question. Had the learner asked, "What is the person with glasses doing?", the system would have replied, "She is reading a book."

In these ambiguous cases, the response "I'm not sure which  $x$  you are talking about." is not necessarily optimal, yet it was considered valid.

### 5.4.2. Invalid responses.

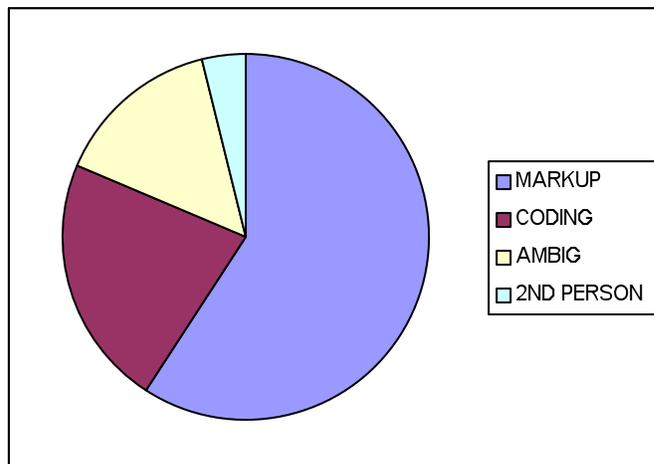
The system provided an invalid response to less than 10% of all questions asked. A response was considered invalid if the information in the response was simply incorrect.

Figure 5.12. Invalid responses.

---

*Reason for Invalid Response:*

Incomplete Markup	166	59%
System Inaccuracy	62	22%
Sense Ambiguity	42	15%
2 <sup>nd</sup> Person	11	4%
Total	281	



#### ***Invalid response: Incomplete markup.***

The majority of the incorrect responses were the result of insufficient markup in the activity model. This was due largely to the lack of adequate time available to comprehensively markup the microworld representations for each of the 124 individual pictures that appeared in the three computer-guided tasks.

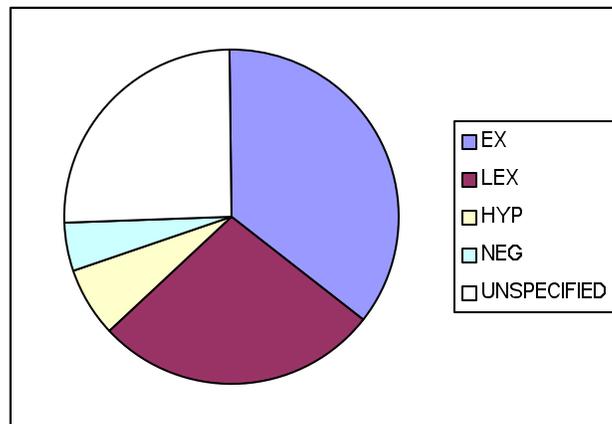
***Invalid response: System inaccuracy.***

Approximately one quarter of the inaccurate responses stemmed from coding inaccuracy in the system's semantic module. Of the 62 inaccurate responses, 46 fall into clearly definable patterns.

Figure 5.13. Invalid responses: Coding inaccuracy.

---

Existential Questions	22	36%
Lexical Exceptions	17	27%
Hypernyms	4	6%
Negative- 'or not'	3	5%
Unspecified	16	26%
Total	62	



---

***Existential questions.***

Of the 62 invalid responses due to coding inaccuracy, 36% can be attributed to the improper handling of existential questions. For more on existential questions, see 5.5

***Existential Questions.***

### *Lexical exceptions.*

There are a number of lexical exceptions that were not accounted for and which led to invalid responses. These include:

1. 'look + like' : e.g., *How she **looks like**?*
2. 'picture' : e.g., *What kind of **picture** are you looking at?*
3. Count/non-count nouns: e.g., *Is he playing **cards**?*

### *Hypernyms.*

The semantic module provided only limited support for hypernyms. In a few cases in this experiment, the system failed to resolve questions like, *Is there an animal?*, where the microworld representation for the target prompt included elements like *cat*, *dog* or *parrot*.

### *Negation.*

In 3 cases, the phrase 'or not' caused a failure to provide a valid response.

### *Unspecified system limitations that lead to invalid responses.*

There were 16 instances of invalid responses where an immediately clear systemic solution fails to emerge. In total, the system delivered invalid responses for no clearly remediable reason in less than 1% of the total number of questions asked.

### ***Invalid response: Word sense ambiguity.***

Word sense ambiguity caused of 22% of the invalid responses provided.

For more on ambiguity, see: 5.5. *Word Sense Disambiguation*.

### ***Invalid response: Second person.***

Approximately 4% of the invalid responses were due to learners using the second person in their question.

*Example:*

Learner Question: Are you looking a woman?

Response: No, it is not.

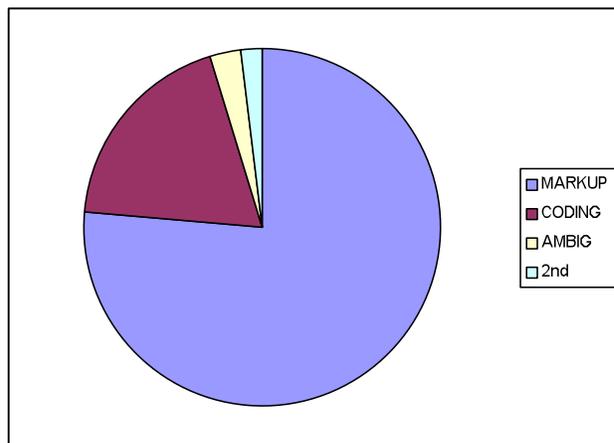
These types of questions should be systematically responded to with a statement such as, “Please don't ask about what I am doing, ask about the things in the picture.”

### 5.4.3. Incomplete responses.

Figure 5.14. Incomplete responses.

---

Incomplete Markup	110	78%
System Inaccuracy	24	17%
Sense Ambiguity	4	3%
2 <sup>nd</sup> Person	3	2%
Total	141	



***Incomplete response: Incomplete markup.***

As with the invalid responses, the vast majority of the incomplete responses were the result of insufficient markup in the activity model.

***Incomplete response: System inaccuracy.***

Of the 24 incomplete responses due to coding inaccuracy, 14 incomplete responses can be attributed to the improper handling of existential questions (see 5.5.1. *Existential Questions*). The remaining 10 incomplete responses do not fall into a broadly definable pattern and remain unspecified.

***Incomplete response: Word sense ambiguity.***

Word sense ambiguity caused of 3% of the incomplete responses provided. For more on ambiguity, see: 5.5. *Word Sense Disambiguation*.

***Incomplete response: Second person.***

Approximately 3% of the incomplete responses were due to learners using the second person in their question.

## **5.5. Improving System Performance for Future Use**

It is important to note that this study was an alpha run for an ICALL system that is highly extendable and reusable. As the system matures, it needs to provide increased accuracy in feedback and response and account for broader linguistic contexts. Based on the evidence provided by this analysis, there are several clearly defined areas where adjustments can be made to dramatically improve the performance of feedback generation and semantic response provision.

### 5.5.1. Refactoring of code.

Based upon the findings of this analysis, it is clear that there are several broad categories of systemic shortcomings that can easily be addressed to increase precision and recall in subsequent studies. Some specific areas where code performance can be dramatically improved include:

#### *Word sense disambiguation.*

Figure 5.15. Summary of coverage: Word sense disambiguation

---

Category	Number	% of Total in Category
Missed Recasts	15	6%
Bad Recasts	15	29%
Invalid Responses	42	15%
Incomplete Responses	4	3%
Total	76	-

---

A simple technique, similar to that used in spell checking (see 4.2.2. *Spell checking*), can be applied to use the activity model for the displayed prompts to help the system correct for part of speech errors by the POS tagger. Validation can be applied to the lemma/POS tag pairs which queries the activity model to find POS matches for each lemma. In the case of a mismatch, the system should rely on the POS data found in the activity model. For example, if the lemma 'play' has been tagged as NP by the POS tagger, yet it is found as a VP in the activity model of any of the pictures being shown to the learner, it should be changed to VP before it is passed to the syntactic parser.

***Noun classification.***

Figure 5.16. Summary of coverage: Noun classification.

---

Category	Number	% of Total in Category
Partial Recasts	69	88%
Missed Recasts	60	24%
Total	129	-

---

The system is currently ill-equipped to resolve agreement errors in determiners. In order to fix this problem, it will be necessary to incorporate some sort of system for noun classification.

***Existential questions.***

Figure 5.17. Summary of coverage: Existential questions.

---

Category	Number	% of Total in Category
Missed Recast	70	29%
Bad Recasts	2	6%
Invalid Responses	22	8%
Incomplete Responses	14	10%
Total	108	-

---

***Recasting ill-formed existential questions.***

In 72 instances, learners used a common ill-formed construction to attempt to ask an existential question. In these cases, the learners omitted the existential lexeme 'there' (e.g., 'Is an old man?', 'Are pets in the picture?', etc.). This construction was not present in the mal-rules and led to the generation of ill- formed recasts (97% of which were suppressed). The entry of a single mal-rule will prevent this type of system error.

*Answering existential questions.*

In 34 instances, learners asked an existential question about an element in the target picture for which there was more than one potential match (e.g., *What is the man doing?*, where the picture shows more than one man). The system responded by saying, “I'm not sure which *x* you are talking about.” This can be fixed by overriding element disambiguation when the learner is merely asking about the existence of the element in question.

***Mal-rules.***

Figure 5.18: Summary of coverage: Total mal-rules.

---

Category	Number	% of Total in Category
Missed Recast	116	47%
Bad Recasts	6	11%
Total	122	-

---

The absence of mal-rules accounts for nearly half of all missed recasts and more than 10% of all bad recasts.

Figure 5.19. Summary of coverage: Breakdown of mal-rules.

---

Mal-rule	Number	% of Total
Existential	70	57%
DO + BE + S	32	26%
What + NP	6	5%
Other	14	12%
Total	122	100%

---

A full 88% of the total number of instances where the lack of mal-rules led to either missed or bad recasts can be fixed by implementing the 3 specific rules listed in

Figure 19. For details on missing mal-rules, see: 5.3.3. *Missed recast opportunities* and 5.3.4. *Bad recasts*.

### **5.5.2. Further markup of visual prompts.**

The microworld representations that inform the activity model for the 124 pictures used as task prompts need to be more painstakingly marked up to provide more comprehensive support for semantic responses.

### **5.5.3. Open source licensing.**

The entire system will be made open source under GNU licensing. Each of the modules in the system will be made available for use either as a part of the system or independently. All of the professional illustrations used in the experimental tasks (with their extensible microworld representations) will also be made available for reuse in future Applied and Computational Linguistic research.

## **5.6. Conclusion**

Despite its shortcomings, the performance exhibited by the ICALL system employed in this study proved effective in sustaining extended, unrestricted conversational interaction with intensive feedback provision that was comparable to similar interaction in oral, face-to-face conditions with a native speaker. As the results in Chapter 7 will demonstrate, interaction with Sasha was as effective as that with a native speaker in promoting developmental gains in ESL question formation.

## CHAPTER VI: CODING

### 6.1. Introduction

This study examined development in ESL question formation according to three separate measures: (1) emergent stage level development, (2) stage level accuracy and (3) morphosyntactic accuracy.

Emergent stage level development was considered according to Pienemann's emergence criterion (Pienemann, 1998) and looked for evidence of initial productive usage of higher-level stage questions. Stage level accuracy scores reflected learners' ability to successfully produce stage 3, 4 and 5 questions in required contexts. Morphosyntactic accuracy scores reflected the learners' general ability to produce morphologically and syntactically well-formed questions.

#### 6.1.1. Emergent stage-level development.

Following on numerous studies that have measured L2 development according to Pienemann and Johnston's developmental sequences, this analysis began by measuring stage-level development in ESL questions on the basis of Pienemann's emergence criterion. This approach, in which L2 development is identified at the point of first emergence, was first developed in the early 1980's in response to contemporary accuracy-based approaches that were criticized as being overly ambiguous (Meisel, Clahsen and Pienemann, 1981). In recent years, considerable evidence has begun to mount suggesting that recast-intensive oral interaction has a positive effect on promoting the emergence of higher-level stage development in ESL question formation (Adams, 2005; McDonough, 2005; Mackey, 1999, 2006; Mackey & Oliver, 2002; Mackey & Philp, 1999; Mackey & Silver, 2005; Philp, 2003).

### **6.1.2. Accuracy.**

Skehan (1998) defined *accuracy* as the extent to which “learners try to use an interlanguage systems of a particular level to produce correct, but possibly limited, language” (Skehan, 1998, p. 5). In the current study, accuracy was measured in two distinct areas: stage-level question production and morphosyntax.

#### ***Stage-level accuracy.***

Stage-level accuracy scores measured the extent to which learners successfully used their interlanguage systems for ESL questions to produce questions at the higher stage levels. Although Pienemann’s emergence criterion has proven to be widely effective in measuring ESL question development, it is a somewhat reductive measure in that it reduces a learner’s ability to form questions to a single stage-level score at each testing interval. Since the focal point of emergence-based scoring is limited to the first productive use of a particular rule, it fails to make any predictions about the rates or routes with which productive accuracy of stage-qualifying rules may develop (Norris, 2005, p. 27). Emergence of developmental stages occurs relatively early in ESL learners’ language development. Target-like usage (or mastery) of these stages, on the other hand, develops much more gradually (Ellis, 1989; Hudson, 1993; Jansen, 2000; Norris, 1996; Norris, 2005).

Pienemann’s model specifies a low threshold of development separating non-emergence from emergence, but no upper threshold to distinguish emergence from mastery (Pallotti, 2007). In order to explore this upper threshold, the current study used the stage-level sequences proposed by Pienemann & Johnston (1987) as the basis for a rubric to classify learner questions according to stage and to measure how accurately they

were able to produce higher-level stage questions throughout the assessment tasks. For each testing interval, learners were assigned an accuracy score for stage 3, 4 & 5 questions. Importantly, developmental stages are not based upon target-like usage. Even at higher-level stages, questions may be rife with errors. Stage level qualification is only concerned with a discrete set of grammatical competencies related to question formation, such as verb movement and auxiliary support. Stage-level accuracy was operationalized as the percentage of questions attempted at each stage where the learner question met the requirements of the target stage (see below for more on target stage). This measure represented the extent to which the learners exhibited mastery of the requisite grammatical competencies for higher-level stage production.

***Morphosyntactic accuracy.***

Morphosyntactic accuracy scores measured the extent to which learners successfully used their interlanguage systems to produce ESL questions that were free of morphological and syntactic errors. This measure is wholly independent of Pienemann's theory of developmental sequences and is modeled on the measurement of grammatical accuracy in Foster & Skehan (1996, 1997).

Because the recasts provided in the current study broadly corrected morphosyntactic errors (i.e., were not limited to stage-qualifying errors), this analysis considered the effects of recast-intensive interaction on morphosyntactic development in addition to stage-level development. Several previous interactionist studies have found a positive effect for recasts on pre-post test gains in morphosyntactic accuracy (e.g., Ayoun, 2004; Ishida, 2002; Iwashita, 1999).

## **6.2. Coding Procedures**

All of the coding for this analysis was performed through a Web-based coding system that was developed by the author to facilitate efficient markup of both written and oral data by multiple raters.

### **6.2.1. Coding written assessment data.**

All of the written assessment tasks were administered using the Web-based A-CLASS assessment system. This system, developed and built by the author, has been used to administer thousands of online language exams in multiple languages. In the written tasks, learners were seated at computers and guided through the assessment tasks by the computer-based testing system. When displaying written data, the Web-based coding tool presented each rater with a list of all questions produced by a particular learner at a chosen interval (see Figure 6.1).

Figure 6.1. Coding interface: Written assessment tasks.

Sasha: Question Coding: Assessment
Download Documentation
Ken Petersen

NAVIGATE
GROUP: CTL
LEARNER: 6
EXAM: PRE
MODE: WRITTEN

Questions for: LEARNER 6 >> PRE >> WRITTEN TASKS
Save Values

**TASK: FILL**

Student-Produced Questions	CODING				UTILS
	Stage	Target	Morph	Syntax	Flag Ques.
1. he is happy?	2	4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0
2. the man reading the fiction?	2	4	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0
3. what the man cuching he's hand?	3	5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0
4. the wite hair weare he sithing?	3	4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0
5. what have the man in he's hand?	3	5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0

**TASK: EMAIL**

Student-Produced Questions	CODING				UTILS
	Stage	Target	Morph	Syntax	Flag Ques.
1. hey there i have question for your company	---	---	<input type="checkbox"/>	<input type="checkbox"/>	0
2. my question is how i can get the job from your campny	3	5	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0
3. what kind of job you have it can be for my age	3	5	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0

Save Values

Questions in the discrete written task (FILL) were written and submitted one at a time by the learners. In the open-ended written task (EMAIL), learners wrote a series of questions, in the form of a fully composed email message, and submitted the entire passage. Prior to coding, the email messages were parsed into individual utterances (splitting on line breaks and terminal punctuation (periods, question marks and exclamation marks)).

### 6.2.2. Coding oral assessment data.

The oral assessment tasks were conducted in face-to-face dyads and recorded onto dictaphones. Audio files were transferred daily onto a PC and backed up on the testing server. When moving audio files from dictaphone to computer, each file was tagged by

learner id and testing interval. Upon completion of the data collection, all files were compressed into .mp3 files and uploaded to a Web server where they could be streamed into the coding interface.

The oral data was not fully transcribed. Rather, the exact starting point of each question in the selected audio file was tagged so that the question could be instantly played with the click of a button. On the initial pass of the oral data, the author carefully marked the exact starting point for each valid question produced in the two oral tasks.

In the oral tasks, secondary and tertiary raters were presented with lists of active and inactive 'Listen' buttons (see Figure 6.2, but note that the 'Mark audio' functionality is only available to the primary rater). Clicking on an active 'Listen' button cues the audio to the start of a valid question. In the discrete IMITATION task, 10 oral prompts were presented which the learner was asked to repeat. For each of these questions, the coding interface displayed the target prompt along with the coding categories for the question. In the open-ended ROLE PLAY task, there was no specific set of questions elicited. Learners produced a broad range of questions in this free-form conversational task. For the ROLE PLAY task, the coding interface displayed a maximum of 25 questions for each learner. Each valid question was indicated by an active 'Listen' button. In addition to the standard coding categories, each ROLE PLAY question was accompanied by a text field to allow for per-question transcription.

Figure 6.2. Coding interface: Oral assessment tasks.

Sasha: Question Coding: Assessment
Download Documentation
Ken Petersen

NAVIGATE
GROUP: CTL
LEARNER: 6
EXAM: PRE
MODE: ORAL

Questions for: LEARNER 6 >> PRE >> ORAL TASKS
Save Values

**TASK: IMITATION**

Target Questions	CODING				UTILS		
	Stage	Target	Morph	Syntax	Flag Ques.	Mark Audio	Listen
1. Does the woman have an orange pumpkin in her hands?	2	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. What is the guy in the crazy hat pouring into the bowl?	---	---	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Is the man relaxing under a tree?	4	4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Who is running between the mountains behind the 2 men?	5	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Is the man with the microphone a happy guy?	3	4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. What is the woman in the blue shirt looking at?	---	---	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. How tall is the boy with the dog?	---	---	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. What is the woman wearing on her head?	3	5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Is the old woman on the bench smiling?	---	---	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. What does the blonde-haired woman want to buy?	3	5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**AUDIO FILE:** Jump By: 5 Sec 00:00 00:00

**TASK: ROLEPLAY**

Per-Question Notes	CODING				UTILS		
	Stage	Target	Morph	Syntax	Flag Ques.	Mark Audio	Listen
1. who did this guy is broke in the locker	3	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. so, he is big man	2	4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. ...							

### 6.2.3. Valid questions.

In order to qualify as a valid (codeable) question, each utterance had to conform to the following criteria:

**1. Utterance contains a minimally well-formed clause.**

The utterance needed to contain some sort of subject-predicate structure.

*Example of Disqualified Utterance:*  
Participant: *Is good for learn?*

## ***2. Utterance is a question.***

The utterance needed to be a question. In the open-ended tasks, there were many declarative statements produced that were, in no way, intended to be questions. These utterances are distinct from stage 2 questions, which have declarative syntactic structure, yet elicit some sort of response.

*Example of Disqualified Utterance:*

Participant: *I think the question are so difficult for me.*

## ***3. Utterance is not a stage 1 question.***

Stage 1 questions were, as a rule, not coded. Stage 1 questions consist of two types of utterances:

### *i. Single Words and minimal phrases.*

*Examples:*

*Why?, What?, Where?, How?, When?, How much?, etc.*

*Can you?*

*Did he?*

### *ii. Formulaic questions.*

Formulaic Questions are somewhat difficult to code in that they require keen observation of long-distance patterns that occur across multiple turns or utterances. The following guidelines were followed when formulaic usage was detected: If a learner exhibited an obvious formulaic pattern of repetition in her questions, only the first instance used of this formula was coded for each task.

Figure 6.3 demonstrates how formulaic questions were selected for coding in the data:

Figure 6.3. Coding formulaic questions for sample learner.

Task	Ques. #	Question	CODE?
FILL	3	What he do?	Yes
FILL	6	What he say?	No
FILL	8	What he took?	No
FILL	10	Where did it happen?	Yes
EMAIL	2	What he say?	Yes
EMAIL	3	Where did it happen?	Yes
EMAIL	9	Where did it happen?	No

#### 6.2.4. Coding variables.

All valid questions (n=3478) produced by the participants in the four assessment tasks, at both pre (n=1722) and posttest (n=1756), were coded as shown in Figure 6.4:

Figure 6.4. Assessment Coding: Variables.

<i>Coding Variable</i>	<i>Possible Values</i>	<i>Description</i>
1. Stage Level	{2,3,4,5,6}	The developmental stage of the learner's question.
2. Target Stage	{3,4,5,6}	The target stage of the learner's question.
3. Error Type	{none,m,s,ms}	The type of morphosyntactic error in the question.

#### 6.2.5. Coding for stage level emergence.

Each valid question was assigned a stage level according to the stages in Pienemann and Johnston's (1987) developmental sequence for ESL question formation. The following coding guidelines were adhered to when determining the stage level for each question spoken or written by the participants:

##### ***Stage 2: Declarative statements.***

Stage 2 questions are simply declarative statements with rising intonation (oral) or a question mark (written) at the end.

**Stage 3: Fronting.**

*Do + SVO?*

Yes/No questions, starting with 'Do', followed by a declarative [Subject + Verb] or [Subject + Verb + Object] clause.

*Do you know him?*

*Did it work?*

*\* Does a man and a woman married?*

*\* Does he happy?*

*\* Does they in the classroom?*

*WH + (be/do) SVO?*

WH question that may or may not be followed by 'be' or 'do' and are followed by a declarative [Subject + Verb] or [Subject + Verb + Object] clause.

*Who plays the banjo?*

*What sits under the chair?*

*\* What he is do/does/doing?*

*\* What he do/does/doing?*

*\* What is he his going to do?*

*\* What does he is doing?*

*\* What reads the man?*

*\* Who chasing the dragon?*

*\* Where he is wants to go?*

*Be + SVO?*

Yes/No questions that start with 'Be', followed by a declarative [Subject + Verb] or [Subject + Verb + Object] clause.

*\* Is he wants to go?*

*\* Are the people eat lunch?*

**Stage 4: Pseudo inversion - Y/N inversion.**

*(Wh) + copula + S.*

Copular questions, where 'Be' is the main verb. In other words, there is no verb in the question other than 'Be'. The copula may or may not be preceded by a WH- question word.

*Is he happy?*  
*Is she a teacher?*  
*Are the cookies in the jar?*  
*Where is the teacher?*  
*Who are they?*  
*How tall is the boy?*

*Aux/Modal + SV.*

Yes/No questions that contain a main verb and an auxiliary or modal, which appears at the start of the question.

*Are the children playing?*  
*Can you tell me about it?*  
*Are you going to ask her?*  
*Would you mind telling me?*  
*Will he be happy?*

***Stage 5: Aux second.***

*Wh + Aux/Do.*

These are questions that (1) Start with a Wh- question word, (2) usually have an auxiliary or 'do' in second position and (3) have a main verb (i.e., are non-copular).

*What does he play?*  
*What is he playing?*  
*Where is he playing (the banjo)?*  
*Who is he playing (the banjo) for?*  
*Who is playing the banjo?*  
*\* What (do/is) he play?*  
*\* What do(es) he playing?*

***Stage 6.***

*Cancelled inversion.*

Verb raising is cancelled in subordinate clauses.

*Can you tell me where the cat is?  
Do you know what he took?*

*Negative questions.*

*1. (Wh) Aux + N'T + SVO?*

*Didn't you see him?  
Won't you tell me?  
Why didn't you stop him.*

*2. (Wh) Aux +S + NOT + VO?*

*Will you not tell me?*

*Tag questions.*

These are declarative statements, followed by a well-formed question tag.

*You understand what I mean, don't you.  
These examples are good, aren't they?  
That will do, won't it?*

***Emergence scores.***

Learners were assigned a stage level for each test based on the highest stage for which they demonstrated productive usage. Productive usage was operationalized as 2 or more questions in 2 or more tasks.

#### **6.2.6. Coding for stage-level accuracy.**

***Target stage.***

In order to consider stage-level accuracy in question formation, it was necessary to code each learner question to indicate the stage level it represented as well as its intended target stage. The target stage represented the stage level of the question that the learner was attempting to produce. Put another way—the target stage value is the stage level of the recast that the learner's question would elicit from a native speaking researcher or instructor. This type of target-form inference is, it should be noted, a

requisite step in the provision of any corrective recast. In order to provide an appropriate recast, a researcher must first determine the intended, well-formed, construction of the learner's question. Figure 6.5 presents a list of sample learner questions and their target stage assignments.

Figure 6.5. Assessment coding: Target stage examples.

<i>Input Question</i>	<i>Stage Level</i>	<i>Error Type</i>	<i>Target Stage</i>
1. *The man standing there is sweeping?	2	s	4
2. *The guy who look in the picture look happy?	2	ms	3
3. Do I have to work on the weekend?	3	none	3
4. Is the man reading?	4	none	4
5. *When I will start the work?	3	s	5
6. When does the company open?	5	none	5
7. *What does the man holding with his hand?	5	m	5

Crucially, without target stage values, it would impossible to disambiguate between questions such as the following two stage 3 examples:

1. *Does the man enjoy eating sushi with his girlfriend?*
2. *\*What he say?*

In example 1, the learner formed a complex, target-like stage 3 question that achieved the intended target stage. In example 2, the learner attempted to ask a stage 5 question (*What does/did he say?/What is he saying?*), but, because no auxiliary support was provided, the question was ill-formed and only qualified as a stage 3 question.

Though it may appear to require a certain amount of inference, determining the target stage for an ESL question is a fairly straightforward proposition (were this not the case, it would be impossible to effectively provide recasts that might lead to stage-level development). In most cases, the target stage of the question is obvious and not subject to any ambiguity. There were certain cases, however, where the target may prove to be slightly ambiguous. The most common case for this type of ambiguity is in determining

a target stage value for ill-formed non-copular Yes-No Questions. These types of questions may have one of two targets—stage 3 (*Do + SVO?*) or stage 4 (*Aux/Modal + SV*). Figure 6.6 presents a list of prototypical ill-formed Yes-No questions and their target stage values.

Figure 6.6. Disambiguating Yes-No target stage.

<i>Prototypical L2 Yes/No Questions</i>	<i>Target Stage</i>
1. <i>*The man work? (habitual)</i>	3
2. <i>*Is the man does work? (habitual)</i>	3
3. <i>*Is does the man work? (habitual)</i>	3
4. <i>*The man working? (progressive)</i>	4
5. <i>*Does the man is working? (progressive)</i>	4
6. <i>*Does the man working? (progressive)</i>	4
7. <i>*Is does the man working? (progressive)</i>	4
8. <i>*Does the man can work? (modal)</i>	4
9. <i>*Does the man will work? (modal)</i>	4
10. <i>*Does the man happy? (copular)</i>	4

When the target stage of an ill-formed Yes-No was ambiguous, target stage assignment was governed by either the aspect of the main verb or the presence of a modal. When no modal was present in the learner's question, the target stage was determined by the aspect of the main verb. If the main verb was expressed in the habitual aspect, it was considered to have a target of stage 3 (examples 1-3). If the main verb was expressed in the progressive aspect, it was considered to have a target of stage 4 (examples 4-7). When a modal was present in the main clause, it was considered to have a target of stage 4 (examples 8-9). Copular Yes-No questions were always considered to have a target of stage 4 (example 10).

***Stage-level accuracy scores.***

Each learner was assigned three stage-level accuracy score (for stages 3, 4 & 5) at each testing interval. The accuracy scores were calculated as the percentage of questions

asked for each stage in which the learner’s question accurately achieved the target stage. To be considered achieved, the stage-level assignment of the question needed to be equal to the intended target stage. The final accuracy score was the percentage of total attempts in which the learner achieved the target stage (see Figure 6.7 for an example score).

Figure 6.7. Sample stage-level accuracy score.

<b>Variable</b>	<b>Value</b>
Total achieved	6
Total attempted	10
Accuracy score	60%

### **6.2.7. Coding for morphosyntactic accuracy.**

There are two broad types of errors that were coded for this study: morphological and syntactic errors. Simple lexical errors were not coded. Each question was coded with a pair of binary values for both morphological and syntactic well-formedness. A question was flagged for morphology or syntax if one or more errors of either type were found in the learner's question.

#### ***Morphological errors.***

A question was flagged as containing morphological errors if inflection had been misassigned or unassigned. Morphological errors included such things as:

1. *Verbs: Tense, aspect and number.*

*\*What will he did?*

*\*Where does he going?*

*\*Where do he go?*

2. *Nouns: Number.*

*\*Are the mens sitting at the table?*

3. *Pronouns: Number, gender and case.*

- \*What is they?*
- \*Is him happy?*
- \*Does the mother love his baby?*

4. *Articles: Number and specificity.*

- \*Does he have a dogs?*
- \*Is a doctor happy?*

5. *Prepositions: Incorrect use of preposition*

- \*Is he sitting in the table?*
- \*Is he looking from the mirror?*

6. *Incorrect AUX Verb: Using 'do' instead of 'be' or vice-versa.*

- \*Does he sitting in the table?*
- \*Is he like ice cream?*

***Syntactic errors.***

A question was flagged as containing syntactic errors if the word order was not correct or required lexical items (e.g., prepositions) were missing. Syntactic errors included such things as:

1. *Failed verb raising.*

- \*Where he is sleeping?*
- \*He is a doctor?*
- \*Where does walk the man?*

2. *Lingering AUX in trace position.*

- \* Where does he is sleep?*
- \* What are they are doing?*

3. *Missing lexical items.*

- \*What does the man put the table?*
- \*Is the man taking?*

***Both (morphological and syntactic).***

Some errors, such as the lack of do insertion, generally entail both a morphological and syntactic error.

***No auxiliary support.***

Where required, auxiliary support (*do* or *be*) is necessary for the syntax of the question and is also responsible for carrying the verb inflection. Questions that failed to provide required auxiliary support were flagged for both morphology and syntax.

*\*What the man doing?*

*\*Where the man sleeps?*

*\*Where the man sleeping?*

*\*Where sleeps the man?*

*\*Where the man he sleeps?*

*\*What you gonna do?*

*\*Why he walking like that?*

***Morphosyntactic accuracy scores.***

In the present analysis, morphosyntactic accuracy scores were calculated for each learner in each task/testing interval as follows:

1. *Morphological Accuracy:* The percentage of total questions produced that were free of morphological errors.
2. *Syntactic Accuracy:* The percentage of total questions produced that were free of syntactic errors.

**6.3. Interrater Reliability**

All of the oral and written data were initially coded by the author for stage level and error type. In an effort to maximize the potential of the Web-based coding tools prepared for this study, a team of 5 graduate students was tasked with providing a second rating on the data. The second raters were provided with a one-hour training session on the coding criteria and working with the interface. They were then each asked to code

approximately 10% of the data and were provided with a detailed set of coding guidelines.

Each second rater had a unique login that allowed them to code the data independently of any other raters (including the first rater). The second raters were each randomly assigned 5 learners from among the three groups to code. They then coded all written and oral data for both PRE and POST in terms of stage-level assignment and morphosyntactic errors. In all, the team of 5 second raters coded 1386 questions (61% of total). Due in part to lack of experience and insufficient training time, the performance of the second raters proved to be less than satisfactory. Combined agreement percentages were woefully inadequate—coming in at only 63%.

A careful analysis of the second rater performance in the written, pretest tasks revealed that 80% of the disagreements were due to obvious coding errors on the part of the second raters (almost entirely in regards to stage level assignment). Rather than discard the entire set of second rater data, it was decided that a third, more experienced, rater would be employed to blindly arbitrate the 37% of the double-coded data where disagreements were found.

In order to facilitate this blind third rating, the Web-based coding interface was extended to support interrater arbitration (see Figure 6.8).

Figure 6.8. Coding interface: Blind interrater arbitration.

Sasha: Question Coding: Assessment
Download Documentation

NAVIGATE
2nd RATER: 6
EXAM: PRE
MODE: WRITTEN

Questions for: RATER 6 >> PRE >> WRITTEN TASKS
Save Values

**TASK: FILL**

Student-Produced Questions	RATING 1	RATING 2	RATING 3			Flag
			Stage	Morph	Syntax	
1. he is happy because is one champion?	Stage: 3 Error: syn	Stage: 2 Error: both	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. what is he reading a story?	Stage: 5 Error: none	Stage: 4 Error: syn	<input type="text" value="4"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. what are did he reading?	Stage: 5 Error: syn	Stage: 3 Error: both	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. what is in his hand?	Stage: 3 Error: none	Stage: 5 Error: none	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Where is the man?	Stage: 5 Error: none	Stage: 4 Error: none	<input type="text" value="4"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. where is the man with the paper?	Stage: 4 Error: none	Stage: 5 Error: none	<input type="text" value="5"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. where is the tall man?	Stage: 5 Error: none	Stage: 4 Error: none	<input type="text" value="4"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. what has the man in is head?	Stage: 5 Error: morph	Stage: 3 Error: both	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. who is this man?	Stage: 5 Error: none	Stage: 3 Error: none	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**TASK: EMAIL**

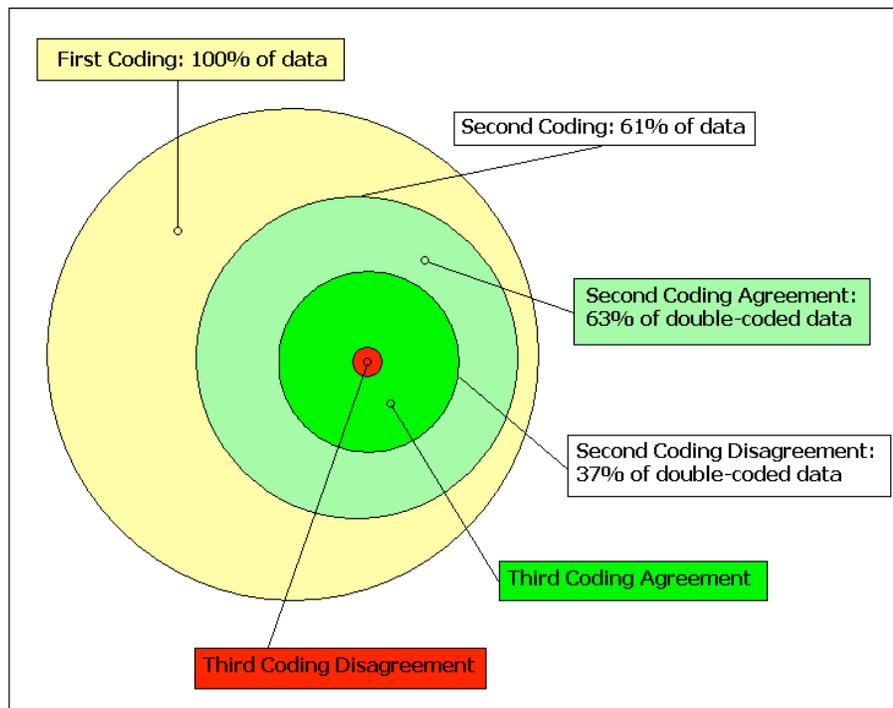
Student-Produced Questions	RATING 1	RATING 2	RATING 3			Flag
			Stage	Morph	Syntax	
1. (6:CTL 8 PRE) how many dollar gonna be my pay	Stage: 3 Error: both	Stage: 5 Error: both	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. (6:CTL 8 PRE) how many hour i have to stay there	Stage: 3 Error: both	Stage: 5 Error: none	<input type="text" value="3"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. . . .						

The third rater was presented with all written and oral questions for which there was a disagreement between the first and second rater. For each question, the third rater could choose between the first two ratings or add a new rating of her own if neither of the two previous ratings was deemed accurate. In order to keep the identity of the initial raters completely blind, the coding interface randomly shuffled the order of the first two ratings for each single question. This way, the third rater could not develop any sort of

bias based upon who provided the rating. The final percentage agreement, where either the second rater or the third agreed with the first, was 97%.

Figure 6.9 represents how final percentage agreement was operationalized in this study.

Figure 6.9: Three rater percentage agreement.



The outermost region represents the entirety of the data that was coded by the first rater. The first containing circle represents the 61% of the data that was double-coded by the five 2<sup>nd</sup> raters. The light green region therein represents the 63% of agreed-upon ratings between raters 1 and 2. The second contained circle represents the 37% of the questions for which there was disagreement between raters 1 and 2. The dark green region represents the total number of questions where the 3<sup>rd</sup> rater blindly agreed with the

1<sup>st</sup> rater. The final, dark red circle is the total number of double-coded questions for which the 3<sup>rd</sup> rater disagreed with the 1<sup>st</sup> rater.

**Tests for normal distribution for parametric statistics.**

In order to ensure that the variables employed in the accuracy-based analysis were suitable for use in parametric statistics (repeated measures ANOVAs), the normal distribution of all accuracy scores was verified using Kolmogorov-Smirnov test for goodness of fit. Pre and posttest accuracy scores for stages 3, 4 and 5 and accuracy scores for morphology and syntax were all found to have a normal distribution.

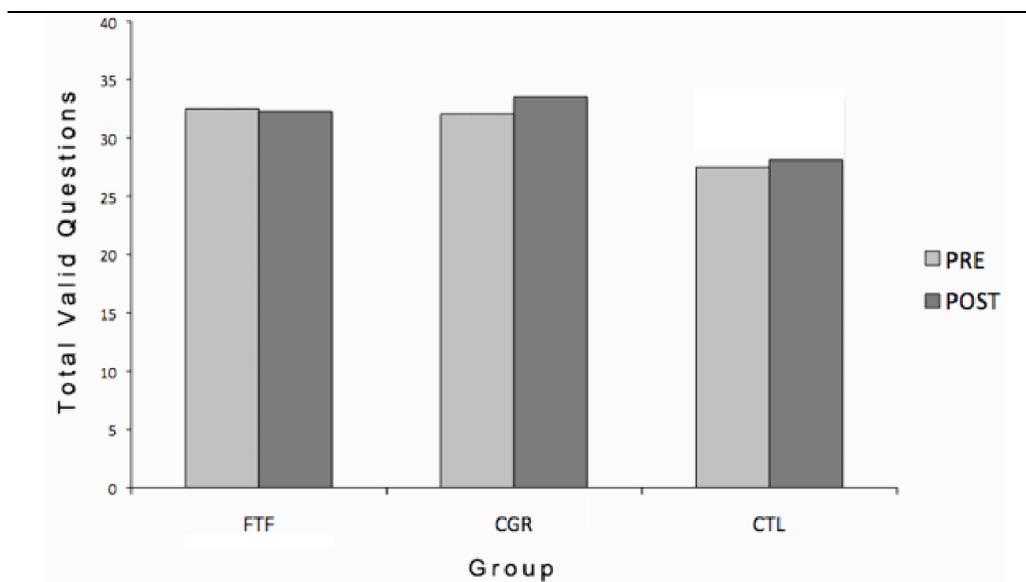
## CHAPTER VII: RESULTS

### 7.1. Introduction

#### 7.1.1. Mean distribution of questions.

In total, 3478 valid questions were produced in the pre (n=1722) and post (n=1756) assessment tasks. The questions were distributed evenly among the three groups at both pre and posttest (see Figure 7.1).

Figure 7.1. Means of total valid questions by group.



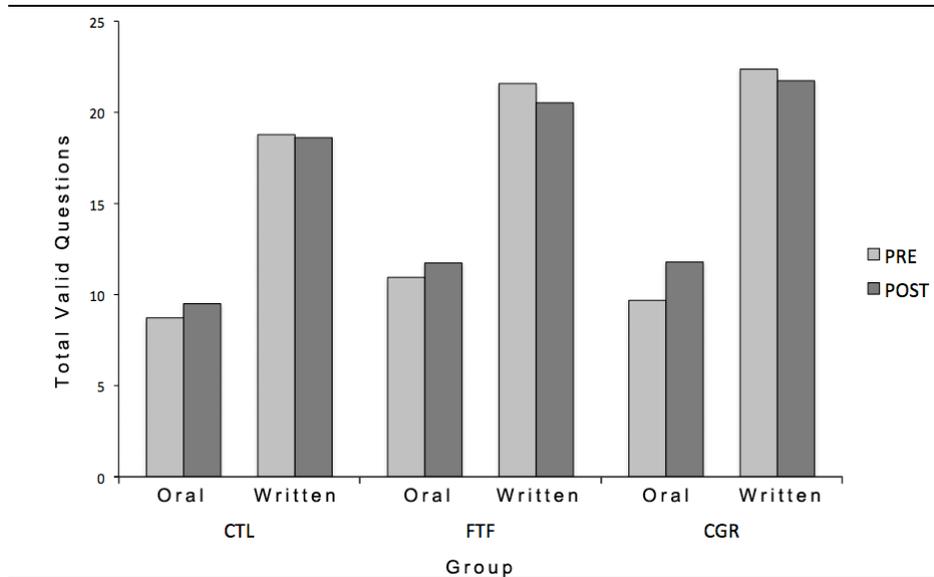
A one-way ANOVA on the total number of questions revealed no significant differences in the total number of questions produced at PRE or POST within or between groups. A repeated measures ANOVA revealed no significant group differences in the total number of valid questions asked in the two testing intervals.

#### 7.1.2. Mean distribution of questions by modality.

Though the mean total number of questions produced at each testing interval was relatively constant, the total number of questions produced in each modality varied

greatly. As Figure 7.2 shows, the oral tasks elicited many more questions than did the written tasks for all groups at both testing intervals.

Figure 7.2. Means of total valid questions by modality.



There were no significant differences, however, between groups in terms of the per-task distribution of questions at either PRE or POST. One-way and repeated measures ANOVAS on total number of valid questions produced by task at PRE and POST revealed no significant group difference.

## 7.2. Emergent Stage-level Development in ESL Question Formation

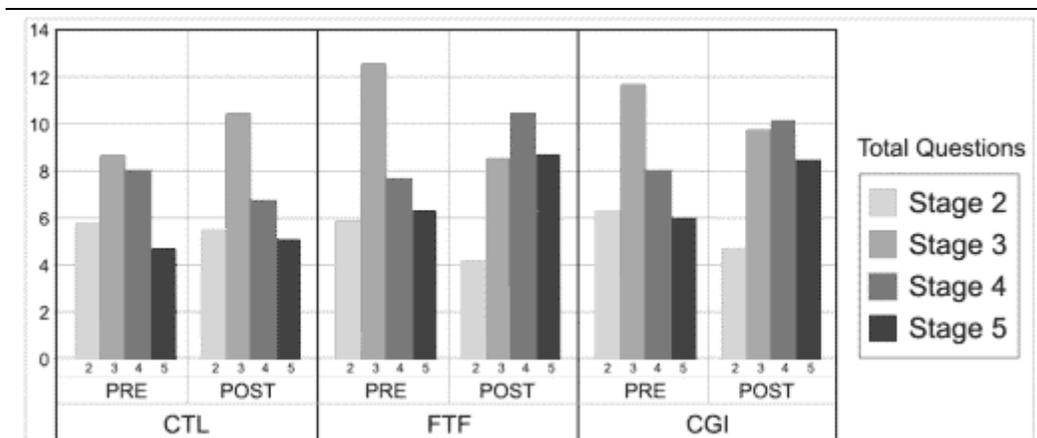
### 7.2.1. Emergence: Descriptive statistics.

Table 7.1 and Figure 7.3 show the mean distributions for stage 2, 3, 4 & 5 questions for each group at pre and posttest.

Table 7.1. Means for stage-level productivity by experimental group.

Interval	Group	Stage 2		Stage 3		Stage 4		Stage 5	
		M	SD	M	SD	M	SD	M	SD
PRE	CTL	5.78	3.46	8.67	4.68	7.94	4.12	4.72	4.70
	FTF	5.89	4.65	12.58	5.96	7.68	3.62	6.32	4.84
	CGR	6.32	4.26	11.68	5.79	8.00	4.23	6.00	4.26
	Total	6.00	4.09	11.02	5.67	7.88	3.93	5.70	4.57
POST	CTL	5.50	2.68	10.50	4.90	6.78	3.64	5.11	4.38
	FTF	4.21	3.46	8.58	4.82	10.47	5.17	8.74	4.31
	CGR	4.74	3.26	9.79	5.34	10.21	3.92	8.53	3.70
	Total	4.80	3.15	9.61	5.00	9.20	4.55	7.50	4.39

Figure 7.3. Mean distribution of stage-level question production.



Each of the experimental groups demonstrated consistent mean increases in the number of higher-level stage questions produced. This corresponded with consistent decreases among the lower-level stages. Table 7.2 illustrates the mean group increase in productivity for each stage.

Table 7.2. Mean increase for stage-level productivity by experimental group.

Group	Stage 2	Stage 3	Stage 4	Stage 5
	PRE-POST	PRE-POST	PRE-POST	PRE-POST
CTL	-0.28	+1.83	-1.16	+0.39
FTF	-1.68	-4.00	+2.79	+2.52
CGR	-1.52	-1.89	+2.21	+2.53

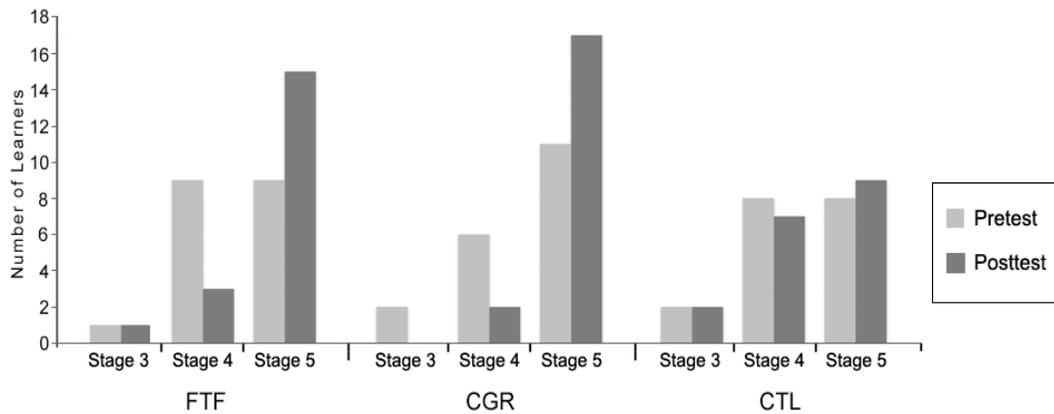
Table 7.3 and Figure 7.4 illustrate that a large number of learners in both experimental groups demonstrated stage-level gains from pretest to posttest. In the face-to-face group, six learners moved from stage 4 to stage 5. In the computer-guided interaction group, two learners moved from stage 3 to stage 4 and another six learners moved from stage 4 to stage 5. In the control group, which received only pre and posttests with no intervening treatment, only one learner demonstrated any measurable gain, moving from stage 4 to stage 5<sup>4</sup>.

Table 7.3. Percentage stage increase by group.

	FTF	CGR	CTL
Percentage	32%	42%	6%
Ratio	6/19	8/19	1/18

<sup>4</sup> This particular learner produced far more total questions in the posttest tasks than in the pretest tasks (pretest:21 questions; posttest: 32 questions). Moreover, this learner attempted only 3 stage 5 questions at

Figure 7.4. Stage assignments by group and testing interval.



### 7.2.2. Emergence: Test for statistical significance.

A chi-square analysis on the number of learners in each group who advanced in stage-level assignment suggests a significant relationship between group assignment and stage movement:  $X^2(2) = 6.633, p < .05$ . (see Table 7.4).

Table 7.4. Stage-level movement by group.

	CTL	FTF	CGR	Total
Developed	1	6	8	15
Not developed	17	13	11	41
Total	18	19	19	56

$X^2(2) = 6.633, p < .05$

A Mann-Whitney two-sample rank-sum test between the oral (FTF) and written (CGR) groups indicated that there were no significant differences between the two groups:  $Z(1,38) = -.664, p = .583$ .

### *Considering ceiling effect and eligibility for stage movement.*

pretest, whereas he attempted 13 stage 5 questions at post. We might assume that this learner was, in fact, a stage 5 question former at pretest, but failed to perform up to his level at pretest.

In all three groups, many of the learners qualified as stage 5 learners at pretest and were thus nearly incapable of moving to a higher stage at post since the assessment tasks were designed largely to elicit stage 3, 4 and 5 questions. While several learners produced stage 6 questions in the open-ended tasks (role-play and email), none produced enough stage 6 questions to qualify them as stage 6 learners at either pre or post. As a result, only 50% of the learners were eligible for stage movement from pretest to posttest. Table 7.5 shows the percentage of learners in each group who hit the ceiling of stage 5 at the pretest interval.

Table 7.5. Pretest stage 5 learners by group.

	FTF	CGR	CTL	Total
Percentage	47%	58%	44%	50%
Ratio	9/19	11/19	8/18	28/56

A chi-square analysis on only those learners who were eligible for stage movement suggests a strong relationship between group assignment and stage movement:

$X^2(2) = 14.732, p < .001$  (see Table 7.6).

Table 7.6. Stage-level increase for eligibles by group.

Group	Pretest Level		Posttest	
	3	4	Developed	Not Developed
FTF (n=10)	1	9	6	4
CGR (n=8)	2	6	8	0
CTL (n=10)	2	8	1	9

$X^2(2, 28) = 14.732, p < .001$

Again, a Mann-Whitney two-sample rank-sum test between the eligibles in the oral (FTF) and written (CGR) groups indicates there is no significant differences between the two groups:  $Z(1,20) = -1.971, p = .173$ .

### **Emergence in question formation: Summary.**

To sum up the emergence-based analysis, a positive relationship was found between recast-intensive conversational interaction and the emergence of higher-level question formation. However, no significant difference was found between the oral and written groups.

### **7.3. Stage-level accuracy in ESL Question Formation**

#### **7.3.1. Stage-level accuracy: Descriptive statistics.**

Table 7.7 shows the mean number of attempts for stage 3, 4 & 5 questions for each group at pre and posttest.

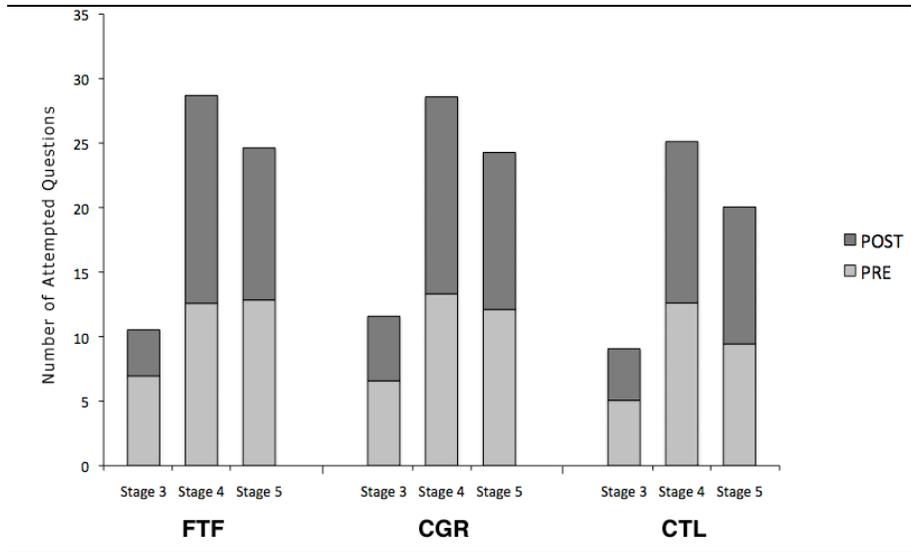
Table 7.7. Mean stage-level attempts by experimental group.

Interval	Group	Stage 3		Stage 4		Stage 5	
		M	SD	M	SD	M	SD
PRE	FTF	6.95	4.28	12.58	4.31	12.84	5.44
	CGR	6.58	3.58	13.32	4.16	12.11	5.86
	CTL	5.06	2.80	12.61	3.93	9.44	5.16
	Total	6.21	3.64	4.16	4.08	11.50	5.59
POST	FTF	3.58	2.57	16.11	6.02	11.79	4.20
	CGR	5.00	2.60	15.26	4.67	12.16	3.67
	CTL	4.00	2.169	12.50	3.47	10.61	4.89
	Total	4.20	2.49	14.66	5.01	11.54	4.25

There were only moderate pre to posttest differences in the total number of stage-level attempts for each group. This means that the accuracy scores for all three stages were calculated from a similar number of attempts at each interval. Figure 7.5 shows the distribution of mean number of attempts of stage 3, 4 and 5 questions at both testing intervals. This chart illustrates a common pattern among the three groups in which the

posttest tasks elicited a slightly greater number of stage 4 and 5 questions and fewer stage 3 questions.

Figure 7.5. Stage-level attempts.



Accuracy scores were calculated for each learner at stage 3, 4 and 5 as described in the previous chapter. Table 7.8 shows the distribution of group mean accuracy scores for stage 3, 4 and 5 questions at both testing intervals.

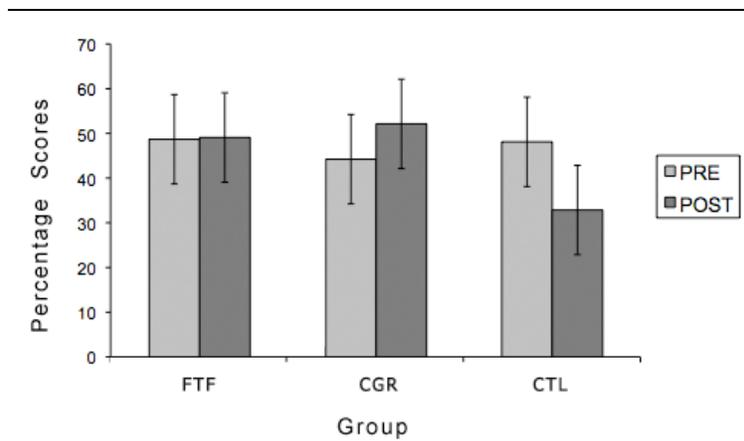
Table 7.8. Means for stage-level accuracy by group.

Interval	Group	Stage 3		Stage 4		Stage 5	
		M	SD	M	SD	M	SD
PRE	FTF	48.69	35.71	60.34	20.21	45.25	29.96
	CGR	44.23	32.66	54.52	21.85	46.20	27.26
	CTL	48.11	32.94	60.50	22.14	40.34	28.64
	Total	46.99	33.25	58.42	21.20	43.99	28.24
POST	FTF	49.08	44.40	68.17	18.84	64.23	27.67
	CGR	52.13	37.07	67.51	16.19	64.21	21.21
	CTL	32.86	30.12	55.49	25.60	37.33	25.51
	Total	44.90	38.04	63.87	20.94	55.58	27.56

The high standard deviations displayed in Table 7.9 can be explained by the fact that the accuracy scores are percentages and that many learners only attempted a small number of questions at a particular stage. For example, if a learner only attempted one stage 5 question at pretest, the only possible values for the stage 5 accuracy score would be 0 or 100. With each additional question asked, the array of possible percentage scores increases by one and the distance between each possible value diminishes (e.g., 2 questions: {0,50,100}; 3 questions: {0,33,67,100}; 4 questions: {0,25,50,75,100}).

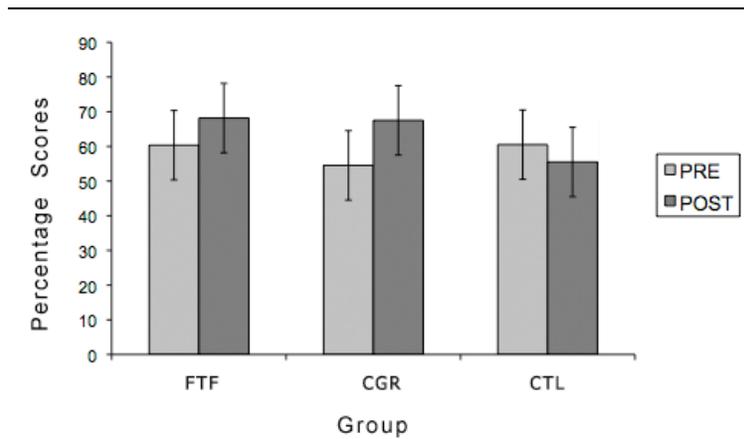
The three graphs in Figure 7.6 illustrate accuracy gains by group for stage 3, 4 and 5 questions.

Figure 7.6.1. Mean accuracy scores for stage 3.



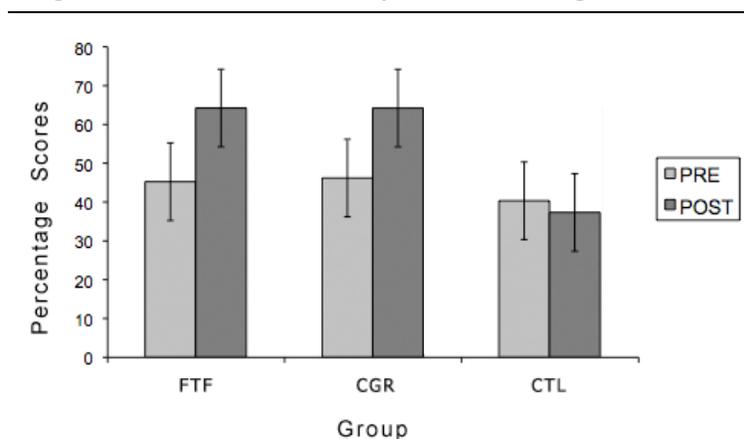
No clear patterns in accuracy gain emerge at stage 3 (Figure 7.6.1). Note that there were, in total, relatively few stage 3 questions attempted. The mean number of stage 3 questions asked across all groups and intervals was only 5.21 (less than half of either stage 4 or stage 5). This is due to the fact that the only possible constructions that can be considered as having a target level of stage 3 are *Do + SVO?* Questions and *WH + VO?* questions. These questions were generally not elicited in the discrete tasks and were used somewhat infrequently in the open-ended tasks.

Figure 7.6.2. Mean accuracy scores for stage 4.



At stage 4, there appears to be a trend that suggests improvement among the experimental groups and a slight decline for control (Figure 7.6.2). Stage 4 questions were the most frequently attempted level of questions in all of the tasks (mean=13.75).

Figure 7.6.3. Mean accuracy scores for stage 5.



Both experimental groups demonstrated noticeable gains in producing stage 5 questions (Figure 7.6.3). The mean number of stage 5 questions attempted in all of the tasks was 11.52.

### 7.3.2. Stage-level accuracy: Test for statistical significance.

One-way repeated measures ANOVAs on stage-level accuracy scores revealed a significant relationship between participation in recast-intensive conversational interaction and L2 development at stages 4 and 5 (see Table 7.9).

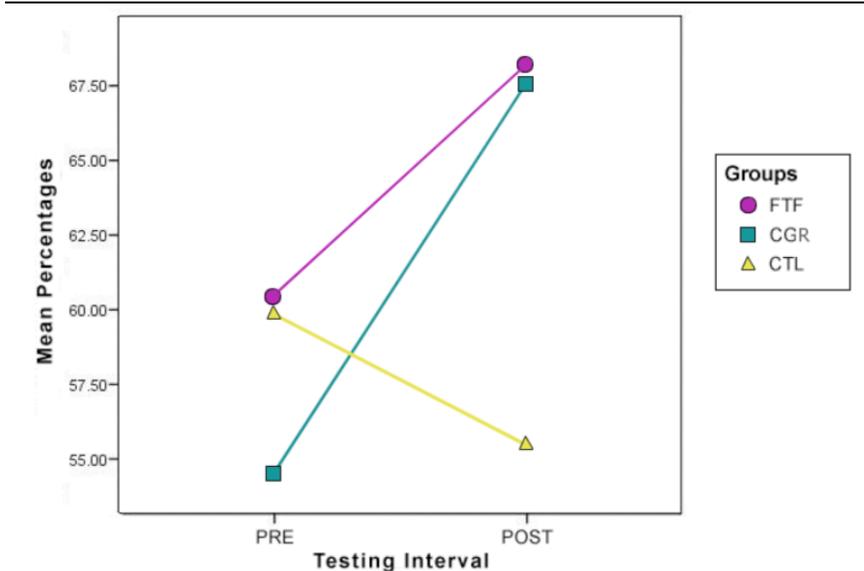
Table 7.9. The effect of treatment on stage-level accuracy

Stage Level	Within Subjects			Between Subjects		
	<i>df</i>	<i>F</i>	$\eta_2$	<i>df</i>	<i>F</i>	$\eta_2$
Stage 3	2	2.522	.087	2	.386	.014
Stage 4	2	4.683*	.150	2	.511	.019
Stage 5	2	8.683**	.247	2	2.489	.086

*\*p* < .05 *\*\*p* < .001.

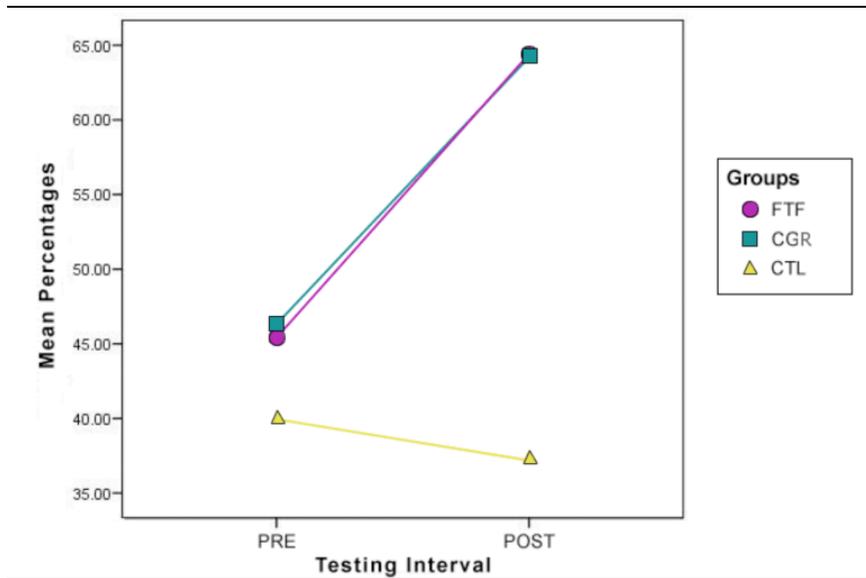
Post-hoc analyses revealed significant differences in accuracy gains between each experimental group and control for stage 4 and stage 5 question formation. However, no significant differences were found between the experimental groups. The plots in Figure 7.7 provide a graphic representation of mean accuracy gains by groups for stage 4 and stage 5 performance.

Figure 7.7.1. Group accuracy gains for stage 4.



As a post-hoc measure, a series of three pair-wise repeated measures ANOVAs (CTL-FTF, CTL-CGR and FTF-CGR) revealed significant gains in Stage 4 accuracy relative to control for both FTF ( $F(1,35) = 4.25, p < .05$ ) and CGR ( $F(1,35) = 9.59, p < .05$ ). No significant differences were revealed between FTF and CGR.

Figure 7.7.2. Group accuracy gains for stage 5.



Pair-wise repeated measures ANOVAs between individual group pairings also revealed significant gains in Stage 5 accuracy relative to control for both FTF ( $F(1,35) = 12.88, p < .001$ ) and CGR ( $F(1,35) = 12.76, p < .001$ ). Again, no significant differences were revealed between FTF and CGR.

#### **Stage-level accuracy: Summary.**

To sum up the accuracy-based analysis of stage-level development, a positive relationship was found between recast-intensive conversational interaction and accuracy gains for higher-level questions. However, no significant difference was found in the amount of improvement between the oral and written groups.

## 7.4. Morphosyntactic Accuracy

### 7.4.1. Morphosyntactic accuracy: Descriptive statistics.

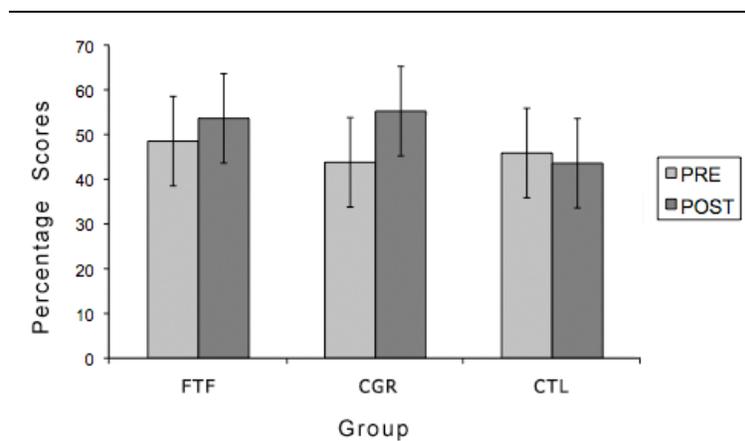
Table 7.10 shows the distribution of group mean accuracy scores for morphology and syntax.

Table 7.10. Means for morphosyntactic accuracy.

Interval	Group	Morphology		Syntax	
		M	SD	M	SD
PRE	FTF	48.52	14.72	45.47	23.56
	CGR	43.78	19.30	46.59	22.27
	CTL	45.87	14.73	45.69	18.39
	Total	46.06	16.24	45.92	21.18
POST	FTF	53.61	18.71	59.36	21.83
	CGR	55.22	16.89	58.49	20.31
	CTL	43.58	14.71	37.51	20.10
	Total	50.93	17.36	52.04	22.76

The charts in Figure 7.8 illustrate mean trends in morphosyntactic accuracy gains for the experimental groups and control.

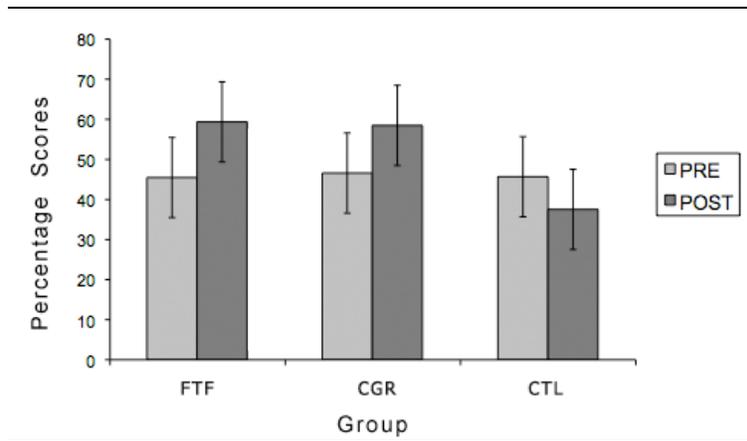
Figure 7.8.1. Mean accuracy scores for morphology.



Both experimental groups showed some gain in morphological accuracy from pretest to posttest (Figure 7.8.1). Learners in both experimental groups demonstrated

mean increases in morphological accuracy. Learners in the control group showed a slight decrease.

Figure 7.8.2. Mean accuracy scores for syntax.



Both experimental groups also showed some gain in syntactic accuracy from pretest to posttest (Figure 7.8.2). Similar to morphological accuracy, learners in both experimental groups demonstrated mean increases in syntactic accuracy. Learners in the control group again showed a moderate decrease.

#### 7.4.2. Morphosyntactic accuracy: Test for statistical significance.

One-way repeated measures ANOVAs on morphosyntactic accuracy scores revealed a significant relationship between participation in recast-intensive conversational interaction and L2 development in terms of morphosyntactic accuracy (see Table 7.11).

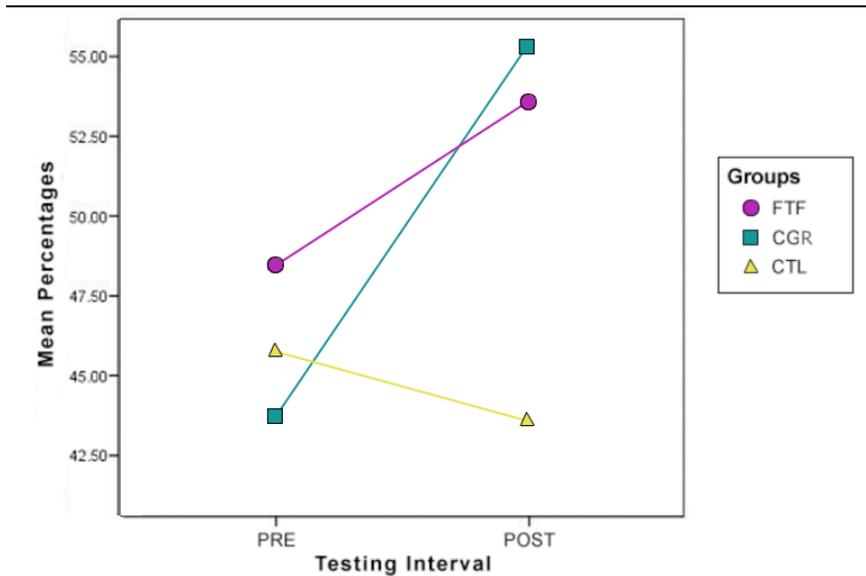
Table 7.11. The effect of treatment on morphosyntactic accuracy.

Error Type	Within Subjects			Between Subjects		
	<i>df</i>	<i>F</i>	$\eta_2$	<i>df</i>	<i>F</i>	$\eta_2$
Morphology	2	4.669*	.15	2	.868	.032
Syntax	2	13.942**	.345	2	1.808	.064

\* $p < .05$  \*\* $p < .001$ .

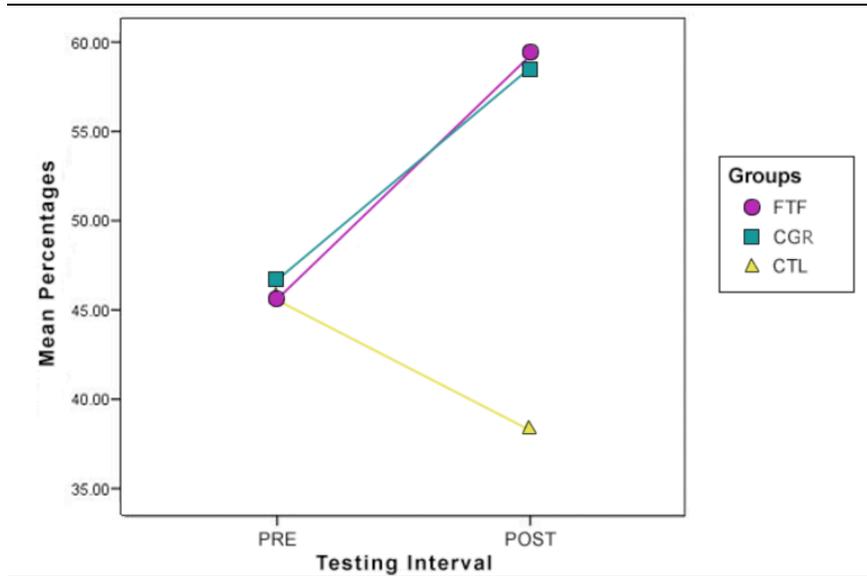
Post-hoc analyses revealed significant differences in morphosyntactic accuracy gains between each experimental group and control for stage 4 and stage 5 question formation. However, no significant differences were found between the experimental groups. The plots in Figure 7.9 provide a graphic representation of mean accuracy gains by groups for morphosyntactic well-formedness.

Figure 7.9.1. Group accuracy gains for morphology.



Pair-wise repeated measures ANOVAs between individual group pairings (1. CTL-FTF; 2. CTL-CGR; 3. FTF-CGR) revealed significant gains in morphological accuracy relative to control only for CGR ( $F(1,35) = 10.90, p < .05$ ). No significant differences were revealed between FTF and either CTL or CGR.

Figure 7.9.2. Group accuracy gains for syntax.



Pair-wise repeated measures ANOVAs revealed significant gains in syntactic accuracy relative to control for both FTF ( $F(1,35) = 19.60, p < .001$ ) and CGR ( $F(1,35) = 23.98, p < .001$ ). Again, no significant differences were revealed between FTF and CGR.

### 7.5. Comparing increases in emergence and stage-level accuracy.

In measuring L2 development in question formation, the relationship between emergence and accuracy is less than certain. Emergence-based accounts focus only on the initial productive evidence of rule acquisition, while accuracy-based accounts measure incremental proficiency gains beyond the point of initial productive use. Norris (2005) cites a lack of empirical research on equating emergent development with other, more global, measures of development:

"At this point, it remains unclear to what extent, if any, there exists a systematic and predictable relationship between developmental *emergence* of morphosyntactic rules and more global notions of language ability, proficiency, or communicative competence."

Norris (2005, p. 28)

Having considered question development according to both emergence and

accuracy gains, the results of this study provide an opportunity to explore predictable relationships between the two measures. A post-hoc analysis was conducted to compare the relationship between emergence-based and accuracy-based gains. This analysis began by regrouping all learners in FTF and CGR (n = 38) into two emergence categories: developed (n=14) and not-developed (n=24). Learners were considered to have developed if their stage assignment had increased from pre to posttest. The two groups were then compared in terms of accuracy gains at higher-level stages.

As the charts in Figure 7.10 illustrate, the learners that developed (a) had lower baseline scores than the stables and (b) demonstrated stronger accuracy gains at stage 4 and 5.

Figure 7.10.1 Mean accuracy scores for stage 4 by emergence gain.

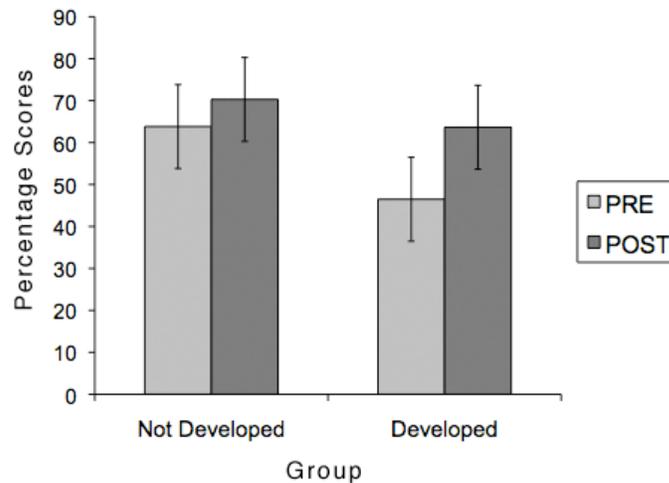
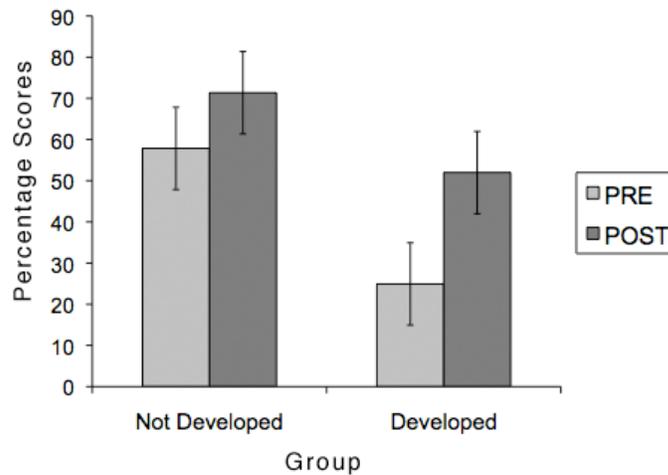


Figure 7.10.2. Mean accuracy scores for stage 5 by emergence gain.



One-way repeated measures ANOVAs on stage-specific accuracy scores revealed a significant within-subjects relationship between emergence-based gains and accuracy increase at stage 5 ( $F=6.100$ ,  $p=.018$ ).

Table 7.12. Stage-level accuracy: developed vs. not developed.

Stage Level	Within Subjects			Between Subjects		
	<i>df</i>	<i>F</i>	$\eta_2$	<i>df</i>	<i>F</i>	$\eta_2$
Stage 3	1	.540	.015	1	6.728*	.157
Stage 4	1	3.176	.081	1	4.925*	.120
Stage 5	1	6.100*	.145	1	12.909**	.264

\* $p<.05$  \*\* $p<.001$ .

The ANOVAs also revealed significant between-subjects effects wherein those learners that developed demonstrated significantly lower accuracy scores for each stage in the pretest tasks.

The results of the post-hoc analysis suggest that that emergence-based stage assignments predict broader patterns of accuracy. Learners predominantly advanced from stage 4 to 5 and demonstrated corresponding gains in stage 5 accuracy.

Additionally, the emergence-based designations of ‘developed’ and ‘not developed’ had the effect of grouping students according to their baseline level of accuracy. At all three stage levels, the learners that developed started out with significantly lower accuracy scores than those that did not demonstrate emergent development.

### **7.5. Summary**

To sum up the results of the analyses performed in this study, there appears to be a significant relationship between participation in recast-intensive conversational interaction and L2 development in ESL question formation—both in terms of emergence and accuracy. This analysis also revealed evidence that suggests a positive relationship between intensive recasting and gains in morphosyntactic accuracy – though, the FTF group failed to outperform control in terms of morphological accuracy. Finally, there was no evidence found to indicate that developmental gains are constrained by the modality of interaction to which learners were exposed—learners in written and oral experimental conditions demonstrated comparable developmental gains in the post-test tasks.

## CHAPTER VIII: DISCUSSION

### 8.1. Summary of Results

The results of this study demonstrated that recast-intensive conversational interaction facilitated developmental gains—in terms of ESL question formation and morphosyntactic accuracy—in both oral and written modalities. The mode of conversational interaction—oral (face to face) or written (computer-guided)—appeared to have no impact on the extent to which learning occurred.

#### 8.1.1. ESL question formation.

Significant developmental stage gains in ESL question formation were demonstrated by learners in both the oral and written interaction groups. These gains were evidenced both in terms emergence-based scores (based on initial productive use of higher-level stage questions) as well as accuracy-based scores (the percentage of productive uses at each developmental stage). Learners in each experimental group showed significant gains in question development in (a) emergence-based scores, as measured by a chi-square test and (b) stage-level accuracy scores, as shown by repeated-measures ANOVAs<sup>5</sup>. Taken together, these measures provide strong evidence that recast-intensive interaction promotes L2 development in ESL question formation in both oral and written communication.

#### 8.1.2. Morphosyntactic accuracy.

Significant developmental gains were also evidenced by learners in both experimental groups in terms of overall morphosyntactic accuracy. Repeated-measures ANOVAs revealed that learners in both modalities increased in syntactic accuracy—

---

<sup>5</sup> A positive correlation was also found between emergence and accuracy scores of stage-level question development.

although, only the CGR group demonstrated significant gains over control in morphological accuracy.

## **8.2. Explanation of Results**

This study was designed to comparatively explore the efficacy of recast-intensive interaction in two distinct interactional contexts: face-to-face, oral (FTF) and computer-based, written (CGR) interaction. There are three overarching questions that might be asked when interpreting the developmental gains exhibited in both interactional contexts: (1) Why did the learners in the FTF condition benefit from recast-intensive interaction? (2) Why did the learners in the CGR benefit? and (3) What contextual factors might have contributed to the gains demonstrated by the learners in the CGR condition? Before answering these questions, it is necessary to briefly consider how development was measured and how recasts were operationalized in the current study.

L2 development was measured solely in terms of gains in the oral and written post-test tasks and did not consider the role of uptake in response to implicit negative feedback (see also 3.3.5. *Treatment tasks: Oral recast provision*). In any research context, it would be a challenge to make a valid comparison of uptake in oral and written interaction. In oral conversation, a learner may produce incidental evidence of uptake by vocalizing part or all of a recast or by making a comment that indicates that the corrective feedback has been noticed. However, in meaning-focused written interaction, where production is relatively slow and effortful, learners simply do not bother to compose written equivalents to these types of utterances if the task does not explicitly require them to do so. In the current study, task advancement in the communicative written tasks was not contingent upon the learner addressing any morphosyntactic errors presented by the

corrective feedback<sup>6</sup>—unlike more form-focused, non-communicative models of written interaction, such as that of Heift’s *E-Tutor* (Heift, 2002, 2003, 2004, forthcoming), which require learners to correct their errors in order to move the task forward.

Recasts are increasingly coming to be viewed as an elastic construct that can take many different forms and may range in degree of explicitness and salience (Loewen and Philp, 2006; Sheen, 2006). The explicitness of recasts in oral interaction has been argued to be moderated by such factors as the frequency and intensity with which recasts are delivered (Oliver and Mackey, 2003), the length of the recast and the number of corrections made (Egi, 2004; Philp, 2003), the type of NNS error corrected (Carpenter et al., 2006; Mackey, Gass, & McDonough, 2000), segmentation (i.e., full/partial recasts) (Oliver, 1995; Sheen, 2004) and prosodic cues such as intonation and pitch (Loewen and Philp, 2006). Recasts in the current study were operationalized so that they would be maximally comparable in oral and written interaction. As such, recasts in both experimental conditions might be categorized as somewhat implicit in that they constituted full recasts performed without any additional corrective feedback or prosodic emphasis (Ellis, 2009).

### **8.2.1. Developmental gains in oral interaction.**

The gains demonstrated by the learners in the FTF condition following prolonged dyadic, recast-intensive interaction should come as no surprise. The task conditions in the oral treatment tasks were designed according to well-attested models from interaction research that have been widely argued to be conducive to L2 development. The current study makes few innovative claims about the effectiveness of recasts in oral interaction—

---

<sup>6</sup> There was almost no evidence that learners in the CGR group produced anything that might be considered as written uptake. Several learners were, however, observed vocalizing as they

rather, the task conditions in the oral tasks were designed to serve as a predictably beneficial baseline for comparing the effects of recast-intensive interaction in a written context. Oral SLA research has suggested that intensive recasting that consistently targets specific linguistic features can have a positive effect on L2 development (e.g., Doughty & Varela, 1998; Han, 2002; Iwashita, 2003; Long, Inagaki, & Ortega, 1998).

The developmental gains in ESL question formation exhibited by the FTF group are consistent with previous claims that recast-intensive interaction promotes emergent development according to Pienemann and Johnston's scale of developmental sequences (Mackey, 1999, 2006; Mackey & Philp, 1999; Mackey & Silver, 2005; McDonough & Mackey, 2006). In response to criticism of purely emergence-based accounts of L2 development (e.g., Norris, 2005; Pallotti, 2007) the analysis of ESL question formation in the current study attempted to look beyond emergence to also explore stage-level accuracy. The positive correlation between accuracy and emergence scores adds support to emergence-based claims by suggesting that emergence scores may predict gains in mastery of stage level question formation.

The post-test gains observed in the FTF group also provide support to claims that intensive recasting may promote L2 development in terms of morphosyntactic accuracy (Ellis & Yuan, 2004; Iwashita, 2003; Leeman, 2003; Loewen, 2005; Loewen & Philp, 2006; Lyster & Izquierdo, 2009; c.f., Lyster, 2004). Although, the fact that the FTF group failed to outperform control in terms of morphological accuracy may suggest that the syntactic reformulations presented to the learners in oral recasts may have been more salient than the morphological reformulations.

The positive developmental gains produced by oral recast-intensive interaction in this

---

interacted with the computer program, though this data was not captured or analyzed.

study may be attributed, in part, to the fact that interaction occurred in a dyadic laboratory context where feedback was consistent and opportunities to produce ESL questions were nearly ubiquitous. Interaction research set in the context of lab-based, dyadic interaction has widely found recasts to be effective at promoting L2 developmental gains (e.g., Han, 2002; Iwashita, 2003; Lyster & Izquierdo, 2009; Long, Inagaki and Ortega, 1998; Mackey & Philp, 1998; Philp, 2003). In the FTF tasks, the abundance of expert NS feedback may have made the corrective nature of the recasts more transparent to the learners (Oliver and Mackey, 2003, pp. 529-30).

### **8.2.2. Developmental gains in written interaction.**

The interactional context in which recasts were delivered in the CGR treatment tasks differed from the oral dyadic, lab-based interaction in the FTF tasks in several fundamental respects: (1) learner input was typed, not spoken (2) feedback and responses were visual, not aural and (3) the interlocutor was a conversational agent, not a human being.

The developmental gains exhibited by the CGR group suggest that recast-intensive interaction in a written, human-computer context may be as effective at promoting L2 development as in an oral, dyadic context. To date, there has been relatively little empirical evidence put forth to support such a claim. Prior to recent developments in natural language processing (NLP) and communication technologies, it was not possible to conceive of conversational interaction or recast provision in a written context. The current study represents an initial exploration of interaction in a communicative ICALL context (see also 2.1.2. *Locating discourse contexts:*

*Communicative ICALL*)— though, the results may be considered in light of previous findings obtained in non-communicative (I)CALL and communicative CMC contexts.

The positive gains in the CGR condition provide supporting evidence to claims that written, computer-based feedback may promote L2 development (e.g., Ayoun, 2001; Loewen & Erlam, 2006; Nagata, 1998; Sachs & Suh, 2007; Sagarra, 2004; Smith, 2005). The results of this study also affirm the findings of previous research that suggests that benefits obtained in written interaction transfer into learners' spoken language production (Payne & Whitney, 2002; Payne & Ross, 2005; Satar & Özdener, 2008; c.f., Abrams, 2003) as there were no significant differences in accuracy score gains on the oral assessment tasks between the written and oral groups.

The effectiveness of the feedback employed in the CGR condition (vis-à-vis the FTF condition) may have been moderated by a broad range of contextual, cognitive and affective factors. Learners in the CGR condition may have benefited from the increased time afforded by the written context for planning (and possibly editing) their output and processing feedback (as suggested by Abrams, 2003; Payne & Whitney, 2002; Payne & Ross, 2005; Smith & Sauro, 2009) as well as from the reduced social pressure accompanying the interactional context (Beauvois, 1997; Kern, 1995; Satar & Özdener, 2008). Although, countervailing arguments that suggest that this pressure may, in fact, be fundamental to the benefits of oral interaction (Bygate & Samuda, 2009; Gass, 2003).

The effectiveness of the written feedback received by learners in the CGR condition may also have been bolstered by the perceptual salience of the visual mode (Sachs & Polio, 2007). Furthermore, the corrective nature of the recasts was, perhaps, made transparent to the learners as they engaged in human-computer interaction with

consistent and intensive provision of feedback. It is unlikely that the learners might have mistaken recasts delivered by Sasha for incidental, non-corrective repetitions.

Claims about the relationship between specific contextual factors and the cognitive processes activated by written recasts remain largely speculative. However, the positive results obtained in the current study make it apparent that written, human-computer interaction may be a beneficial environment for the use of implicit negative feedback. These results have theoretical and methodological implications for recast and CALL research as well as pedagogical implications for technology-enhanced language instruction.

### **8.3. Implications**

#### **8.3.1. Implications for recast research.**

Written, computer-delivered recasts have thus far been examined in interactional contexts where the ecological validity of recasts has been strained by the discourse conditions imposed by the technologies employed (see *Ecological Validity*, below). Ortega (2008) cites the difficulty in producing comparable written and oral task conditions as a critical factor that contributes to the discordant evidence put forth by interaction research in SCMC. The remedy to this problem, Ortega argues, lies in the adoption of true comparative designs where group composition, task, topic, and context are systematically taken into account (p. 43).

Written recasts in the CGR condition were delivered in the context of a communicative ICALL environment where group composition, task and topic were comparable with the FTF condition. The two experimental conditions shared similar turn adjacency constraints—there was no opportunity for the split negotiation routines that are

typical of SCMC discourse (see below). Also, learners did not have access to a written transcript of interaction and were thus not afforded the extra opportunity for reflection and review generally provided by SCMC.

However, the discourse conditions in written and oral interaction were, by no means, isomorphic. No matter how tasks are designed, there will always be fundamental modal differences that may mediate the effect of recast provision.

### ***Working memory and written recasts.***

Learners in the CGR condition received written recasts that were displayed in Sasha's dialogue bubble for a fixed duration of four seconds and were immediately followed by a semantic response to the question asked. Because no transcript was displayed, learners in both groups had to rely solely on their working memory capacities to make the cognitive comparisons elicited by the recasts. However, it is likely that the learners in the two groups relied on separate subsystems of working memory to process the recasts they received. While it is beyond the scope of the current study to provide a nuanced model of working memory, a brief consideration of one traditional model (Baddeley & Hitch, 1974) provides a simple framework for describing how computer-centered task contexts might affect the demands placed on individual learners' working memory capacities. Baddeley & Hitch (1974) proposed a model of working memory wherein short-term memory is functionally divided into two distinct subsystems—the phonological loop (auditory information storage) and the visuo-spatial sketchpad (visual information storage). These two systems are subordinate to a central executive system that mediates selective attention to and retrieval from each system.

It is likely that learners in the CGR condition tended to rely more heavily upon the visuo-spatial sketchpad to make cognitive comparisons whereas learners in the FTF condition tended to rely on the phonological loop. Further empirical research is required to investigate explain how the reliance on these distinct subsystems may impact the effect of recasts.

***Increased planning time.***

Recast provision in the two interactional modes also differed in that learners in the CGR condition were afforded more time for planning and composition in an environment that exerted minimal social pressure. This may have provided learners in the CGR condition with greater opportunities for attending to both form and meaning (Sachs & Suh, 2007; Wong, 2001). The slower processing demands inherent in written interactive communication have widely been argued to be a boon to L2 learners in SCMC (e.g., Payne & Whitney, 2002; Payne & Ross, 2005; Smith & Sauro, 2009) and may well have played a beneficial role for the learners that interacted with Sasha.

***Learner awareness of written recasts.***

Because no direct measure of attention and awareness was employed in the current study, it is difficult to make any substantive claims about the extent to which learners noticed that they were being corrected by the written recasts they received. The only evidence of learner awareness obtained in this study was the answer to a question on the debriefing questionnaire administered at the end of the study (see Figure 8.1).

Figure 8.1. Debriefing question about noticing.

---

*I noticed that I was being corrected if I made mistakes in my questions.*

Not at all    Not much    I don't know    A little bit    A lot  
[-2]                    [-1]                    [0]                    [1]                    [2]

*Mean Responses:* FTF: .68; CGR: .63.

---

Learners in both modalities reported a moderate degree of noticing of written recasts. It has been argued that the printed text may also add to the salience of input and output in general and the noticing of non-target-like input and output in particular (Izumi, 2002; Salaberry, 2000; Smith, 2004; Smith & Sauro, 2009). This increased salience may have provided a positive benefit to learners in the written condition, though the results of the gross offline measure employed in this study do not support this claim.

### **8.3.2. Implications for CALL research.**

As CALL research has begun to mature, there is an ever-growing awareness in the field that any perceived linguistic benefits of CALL tasks and activities need to be discussed within the framework of applied linguistic theory at large (Chapelle, 2005; Ortega, 2008). Yet, because computer-guided learning is still a relatively new and highly dynamic construct in linguistic research, there are considerable challenges that remain in grounding CALL research within extant theoretical models of SLA.

One of the greatest challenges in interfacing CALL methodology and SLA theory lies in accounting for and explaining the effects of the multifarious contextual differences inherent in written and oral interaction. In written interaction, it is difficult to emulate the spontaneity and immediacy that have been argued to be essential to the efficacy of instructional techniques such as recasts. To date, written recasts have been examined in

the context of SCMC and non-communicative CALL tasks, which have been plagued by questions of ecological validity. The current study, on the other hand, explored recasts in communicative ICALL tasks that provided a discourse context similar to oral interaction—in that the learners produced spontaneous, meaning-focused output and consistently received feedback in the turn immediately following any non-targetlike utterance.

***Ecological validity of recasts in SCMC.***

In text-based SCMC, it is difficult to operationalize recast provision in a manner that corresponds to the temporal conditions inherent in oral, face-to-face interaction. The efficacy of recasts hinges upon the assumption that recasts are provided in immediate temporal and sequential response to an ill-formed utterance (Farrar, 1990; Long, 1996; Saxton, 1997, 2005). Recasts have been argued to be effective because they allow learners the opportunity to make a cognitive comparison between their ill-formed utterances and the target-like reformulations provided by an expert speaker (Ellis, 1994; Gass, 1997). In oral conversation (both dyadic and, to a lesser extent, large-group), turn-taking norms and conventions help ensure that a recast will appear in immediate juxtaposition to a non-targetlike utterance.

However, CMC discourse is not sequential in a manner comparable to oral interaction, as it is subject to what Smith (2003) referred to as *split negotiation routines*. It has widely been noted that, in SCMC, there are often multiple-turn delays between non-targetlike utterances and the responses they elicit from their interlocutor(s) (Chun et al., 2008; Herring, 2001; Lai & Zhou, 2006; Loewen & Erlam, 2006; Ortega, 2008; Smith, 2003). Given that recasts are claimed to be effective, in part, because of the

immediate juxtaposition of the learner's error and the correct reformulation, it remains an important empirical challenge to explain how the split adjacency of SCMC discourse affects the beneficial nature of written recasts (cf. Chun et al., 2008).

Written SCMC also differs from oral interaction in that written transcripts are usually displayed to the learner throughout the conversation. The presence of dynamic written transcripts allows learners opportunities for review and reflection that would be inconceivable in oral communication. When learners engage in spoken interaction, they must rely solely upon their working memory capacities to make the cognitive comparisons required to make a recast successful. Learners engaging in SCMC, on the other hand, may be likely to scan the transcript to find the trigger and response before evaluating the difference between the two written utterances. SCMC research has yet to address the extent to which learners rely on reading written transcripts versus relying on working memory and how these cognitive choices impact the effectiveness of instructional techniques such as recasts.

***Ecological validity of recasts in non-communicative CALL.***

A few studies have attempted to explore the benefits of computer-delivered recasts in the context of non-communicative CALL activities. The activities employed by Ayoun (2001, 2004) and Sagarra (2007) elicited forms of human-computer interaction that bore little resemblance to spontaneous, meaning-focused oral interaction. Learners were simply asked to complete a series of form-focused, fill-in the blank activities in which they had to accurately complete the templated model sentences. Each of these studies also introduced methodological variables that exacerbated the disconnect between oral and written recasts. Ayoun's system obscured the role of recasts as corrective

feedback by providing a ‘recast’ (the target form of the templated sentence) in response to every utterance produced by the learner—regardless of whether or not it contained an error. Sagarra’s system altered the explicitness of recasts by requiring learners to write down a numeric code to confirm that they had listened to the pre-recorded recast.

***Ecological validity of recasts in communicative ICALL.***

The current study employed an intelligent conversational agent (Sasha) to deliver recasts in a manner comparable to recast provision in spontaneous oral interaction. The recasts produced by Sasha differed from those produced by the NS interlocutors in the FTF tasks only in that they were presented in a visual, text-based format to the learners. In both oral and written conditions, learners engaged in spontaneous, meaning-focused interaction. The dyadic nature of the interaction in the treatment tasks and the rapid response time of the system (mean recast generation time: 1.07 seconds) ruled out the possibility that Sasha might produce split negotiation routines similar to those found in SCMC.

Most existing ICALL systems have been designed to handle a highly restricted range of input. In order to limit the range of input that can be expected from L2 learners, ICALL activities to date have largely consisted of (1) build-a-sentence exercises, in which responses are restricted to include only supplied word forms (e.g., Heift and Nicholson 2001), (2) cloze activities that restrict input to single words or phrases (e.g., Heift, 2004) or (3) exercises that use translation to restrict learner responses (e.g., Nagata, 2002). These types of ICALL activities more closely resemble form-focused workbook exercises (with the benefit of immediate, customized feedback) than oral communicative tasks.

The computer-based treatment tasks designed for the current study were modeled closely upon tasks that have been used extensively in oral interaction research. Sasha's expertise in interpreting L2 input, generating feedback and answering questions allowed the system to sustain meaning-focused conversational interaction and supply incidental feedback in a manner comparable to oral interaction.

In short, the current study might best be characterized as an interactionist study that employed a linguistically intelligent conversational agent to facilitate oral-like written interaction. The FTF and CGR groups in this study (1) received nearly identical instructions for each of the three experimental tasks, (2) were presented with the same picture sets and procedural task cues and (3) received recasts and semantic responses according to a similar protocol. The computational system employed in this study was innovative in that it facilitated conversational interaction that was both immediate and dynamic. This allowed learners to make immediate cognitive comparisons between their ill-formed input and the feedback in much the same manner as did the learners in face-to-face interaction. Furthermore, the restrictions on learner input in the CGR group were isomorphic with those placed on the FTF group—with the one fundamental exception that the CGR learners were forced to type their questions rather than express them verbally.

The design of this study attempted to bridge several key methodological gaps between CALL and SLA and provide an opportunity to make a valid comparison between the developmental effects of recasts in a written, computer-centered environment with those provided in an oral, dyadic context. The results indicate that, with all other task conditions being equal, the developmental benefits of recast-intensive

interaction carried out in a written computer-centered environment produced developmental gains similar to those produced in face-to-face dyadic interaction.

### **8.3.3. Pedagogical implications.**

This study provides empirical evidence that learners can receive developmental benefits from individualized, form-focused instructional intervention administered by an autonomous computational system. This finding has potential implications for the role of technology in the modern language classroom. To be clear, this study does not, in any way, suggest that a computer might be capable of effectively taking the place of a professional language teacher. Rather, it provides evidence that a well-designed computer application may serve as an effective, autonomous teaching assistant that is capable of sustaining meaning-focused conversation and providing individualized, form-focused feedback.

In both blended (technology-enhanced) and distance language courses, a reliable ICALL application might provide an effective vehicle for engaging learners in meaning-focused interaction and providing intensive, form-focused feedback. ICALL applications such as Sasha hold enormous potential to assist teachers in addressing the often disparate needs of their students. In blended language classrooms – where CALL plays a consistent, supplementary role in instruction—a well-designed ICALL system might allow a teacher to place small groups of students (e.g., heritage learners) on computers for individual instruction while she focuses on the needs of the remaining learners. An ICALL system may also be effectively used to provide form-focused instruction as homework or lab work—again, allowing the teacher to use class time for other types of language practice and instruction.

Another potential area in which ICALL might play a beneficial role in language education is as an online supplement to printed language textbooks. In recent years, it has become a mandate for language textbook authors and publishers to provide some form of online support to accompany their print materials. Online support generally comprises lists of downloadable documents and multimedia files and possibly some interactive activities and quizzes. An intelligent conversational agent such as Sasha could greatly enhance online textbook support by engaging learners in conversation about the images, text samples and audio/video clips associated with the book. Sasha could easily be trained to understand the microworlds represented in any of these types of media.

Because computer-based tasks often interact with powerful databases, there is also vast potential for storing and analyzing student production. ICALL systems are distinctly well-suited for this task in that they are capable of performing on-the-fly analysis of L2 input. The linguistic data collected in ICALL interaction might be used to inform immediate and longitudinal profiles of learner performance. In other words, ICALL systems are particularly well-suited for providing detailed information about the lexical/morphosyntactic complexity and accuracy of a learner's L2 production and pointing out where the learner's individual weaknesses may lie. If the learner interacts with the system for an extended period of time, longitudinal analysis might show how this complexity and accuracy changes over time.

To sum up, the current study presented a theoretically and methodologically grounded model for comparatively exploring the developmental outcomes of language instruction in oral and written interaction. The positive gains demonstrated by the

learners in the CGR condition suggest that written, computer-based interaction may be as productive an environment for instruction as oral interaction.

## **CHAPTER IX. LIMITATIONS AND FUTURE RESEARCH**

The results obtained from this study present compelling evidence that an intelligent conversational agent might be capable of sustaining extended meaning-focused interaction and providing corrective feedback in a manner that facilitates L2 development. However, there are a number of limitations that constrain the extent to which this claim might be generalizable within SLA research, CALL methodology and language instruction at large.

### **9.1. Generalizability of Results**

#### **9.1.1. Duration.**

The results reported in this study were based on a relatively limited amount of time that the learners engaged in recast-intensive interaction. Each learner received a total of 90 minutes of individualized instructional intervention across three sessions that occurred within the span of approximately one week. The findings of the current study also only consider developmental effects on learner performance in immediate post-tests—administered in the days that immediately followed the completion of the final treatment tasks. Although delayed post-tests were administered to each learner two weeks following the immediate post-test, the data obtained from the delayed tests has not been coded or analyzed. The developmental claims of this study are therefore limited to immediate outcomes of short-term intensive intervention. Future analysis of the delayed post-test data may provide evidence about the extent to which the immediate gains demonstrated by the learners in the experimental conditions sustain or diminish over time. Additional research, with more longitudinal designs is also needed to examine the developmental effects of long-term, sustained interaction with a conversational agent.

### **9.1.2. Target structures.**

The treatment and assessment tasks employed in this study were primarily designed to elicit ESL questions. ESL question formation is a complex linguistic construct—targetlike production of English questions is governed by a range of morphosyntactic rules that specify requirements for verb raising, auxiliary insertion and verb inflection. Because the recasts delivered in both written and oral interaction in the treatment tasks attempted to correct all learner errors—not just those specific to question formation—L2 development was assessed both in terms of emergent development according to the stages of question formation proposed by Pienemann and Johnston (1987) as well as overall morphosyntactic accuracy. However, it is theoretically and methodologically challenging to provide a fine-grained account of the relationship between accuracy and stage-level emergence (see 6.1.2. *Stage-level accuracy*). The measures employed in the current study examined this relationship in very broad terms and the results provide only limited evidence of a positive relationship between developmental gains in accuracy and emergence scores. Providing a more nuanced analysis of the relationship between morphosyntactic accuracy and stage-level question development may require a more fine-tuned coding schema that examines accuracy in terms of the specific rules that underlie developmental stages, such as that proposed by Norris (2005, p. 53-58).

### **9.1.3. Sample size.**

The generalizability of the findings from this study is also constrained by the relatively small sample size of the CGR group (n=19). The learners that participated in

this study were recently-immigrated U.S. high school students—90% of whose native language was Spanish. Although the sample size was large enough to reveal statistically significant mean developmental gains vis-à-vis control, with such a small cohort, it was difficult to empirically explore the effects of individual learner variables such as: L1, age, gender, learning styles, familiarity with computers and typing speed. Further research is required with larger and more diverse cohorts to explain how these individual learner variables may affect the extent to which different types of learners may benefit from computer-based interaction.

## **9.2. Interactional Context**

Task conditions in the written and oral treatment conditions were comparable in terms of time, content, purpose, feedback provision and task management<sup>7</sup>. However, there are myriad and fundamental differences between oral, face-to-face interaction with a human interlocutor and written interaction with a computer-based conversational agent. While the results of this study suggest that learning may occur equally in these two conditions, many empirical questions remain as to how and why learners may benefit from written, computer-based interaction with a conversational agent.

### **9.2.1. Psychological factors of human-computer interaction.**

When learners engage in conversation with a virtual interlocutor, there are a number of psychological factors that may affect the quality and quantity of interaction. It has been suggested that interaction in computer-based contexts reduces many of the social and performance pressures that may affect learners' production as they engage in face-to-face interaction (Abrams, 2003). It remains an empirical question as to how this

reduced pressure positively or negatively affects L2 development in computer-based interaction.

The quality of learner interaction with virtual interlocutors like Sasha may also be affected by the extent to which learners perceive the conversational agent to be an acceptable facsimile of a human interlocutor. In human-computer interaction research (HCI), it has widely been argued that synthesizing human-like facial expressions and gestures may help make conversational agents more believable (Bartneck & Lyons, 2007; Picard, 1997). However, computer-based conversational agents may be prone to problems of acceptability as they attempt to emulate the linguistic and paralinguistic cues of a human interlocutor. Many highly realistic robots and computer-based characters struggle with a phenomenon commonly referred to as the ‘uncanny valley’ (Mori, 1997). The ‘uncanny valley’ describes to a sense of unease and discomfort that people experience as they interact with conversational agents that attempt to emulate realistic human-like features and facial gestures.

Sasha was intentionally designed to be minimally anthropomorphic in order to minimize any unwanted effects of attempting to establish human-like acceptability. There are numerous questions that remain regarding the relationship between the realism of a conversational agent and the quality and quantity of L2 interaction. In particular, further research is required to investigate the effect of emotional responses (e.g., expressions of approval/disapproval in response to the learner’s utterances, expressions of impatience when the learner is slow to respond, etc.) on the social and performance pressures that the learner may experience during interaction.

---

<sup>7</sup> The display and updating of visual task management cues, such as guess access and sequencing chips, were consistent in the oral and written conditions. For more on task management cues, see

### **9.2.2. Written vs. oral input.**

While learners in both oral and written conditions demonstrated gains following sustained recast-intensive interaction, it is difficult to determine the precise role of modality on the learner's ability to recognize and benefit from negative feedback. The current study did not employ robust measures of attention and awareness—rather, the learners were simply asked to report the extent to which they had noticed corrective feedback in a debriefing questionnaire. Further research is required that includes either online measures such as think-aloud protocols or offline techniques such as stimulated recall to empirically explore the relationship between receptive modality and the noticing of feedback.

### **9.2.3. Written vs. oral output.**

Learners in the two experimental groups were given equivalent time and opportunity to produce ESL questions. However, there are a variety of modal factors that might have affected the amount and quality of the questions produced in each group. In the written modality, where there was reduced pressure to perform without delay, learners may have been afforded a greater amount of time for planning, producing and editing each individual utterance. The data captured in written treatment tasks in the current study included information about the amount of time between utterances, but no attempt was made to separate planning time (i.e., time spent reviewing the pictures on display and thinking of an appropriate question) from time spent composing a question. Additionally, only final, edited questions were submitted by the learners—meaning that there was no opportunity to examine strategies of self-repair that may have been exhibited in written interaction. Further research might employ keystroke capturing techniques to (1) more

---

3.2.2. *Lab setup for oral tasks.*

accurately measure the amount of time learners spend typing individual utterances and (2) provide evidence of the strategies of production and repair that learners employ in written interaction.

### **9.3. Recasts and Control**

The experimental design of the current study makes the precise role of recasts in facilitating L2 development less than clear. This is due to the fact that the control group served strictly as a testing control. In other words, learners in the control group only engaged in the oral and written tasks in the pre- and post-tests and did not participate in any treatment sessions. As a result, the control group could only serve to rule out any testing effect or gains achieved via some external factor (i.e., classroom instruction during the treatment period). In order to implicate the effect of recasts more explicitly, learners in the control group would ideally have received the same interactional opportunities as the experimental groups during the treatment sessions—without the provision of recasts.

The decision to limit control to a testing control was necessitated by scheduling constraints at the participating high school (see 3.3.2. *Experimental groups*). Because learners in control were not exposed to interaction in any treatment tasks, this study is limited to examining the developmental effects of recast + intensive interaction in each modality, since it did not isolate recasts as an independent variable. It is thus not possible to rule out the possibility that learners in the control group might have demonstrated post-test gains similar to those exhibited by learners in the experimental groups had they engaged in equivalent interaction without any feedback.

## 9.4. Limitations of Sasha

Although the ICALL system designed for sustaining written interaction in the CGR condition was capable of delivering accurate recasts and semantic responses in response to most of the open-ended L2 questions asked during the treatment sessions, there are many areas in which the performance of the system might be improved (see *Chapter V. System Analysis* for a complete discussion).

### 9.4.1. Recast provision.

Sasha accurately delivered recasts in response to approximately 65% of all non-targetlike learner questions. Though this percentage may appear low vis-à-vis an optimal delivery rate of 100%, it is worth noting that such an idealized rate is rarely, if ever, achieved by even the most attentive of native speakers in face-to-face oral interaction. Further analysis of the recorded data obtained from the oral FTF treatment tasks might help determine a more precise baseline for comparison.

Recasts are arguably the most challenging type of feedback for an autonomous conversational agent such as Sasha to generate. Unlike generating meta-linguistic feedback, recast generation requires not only that the sources and types of errors are identified, but that the target-like form of each error is established and that the ill-formed input be algorithmically reformulated so that the recast conforms with the rules specified by the target form.

Given that Sasha was designed to provide recasts in response to largely unrestricted L2 input, it was difficult to completely militate against the provision of recasts that contained non-targetlike or misleading linguistic forms (see 5.3.4. *Bad recasts*). Even though the system employed thorough morphosyntactic checks on each

potential recast before presenting it to the learner (see 4.6.4. *Recast confirmation*), 53 of the 857 recasts that Sasha generated were presented to the learners despite the fact that they contained morphosyntactic or lexical errors. It remains an empirical question what effect the delivery of error-laden recasts may have had on the learners' ability to interpret the corrective feedback delivered by Sasha.

#### **9.4.2. Semantic response provision.**

In order to sustain prolonged interaction, Sasha needed to provide accurate and reliable answers to learners' questions throughout the three 30-minute CGR task sessions. In total, Sasha was asked 2912 questions about the pictures displayed in the tasks—86% of which were responded to with a valid, task-appropriate response. Nearly one in ten questions was responded to with an answer that did not accurately represent the situation portrayed in the target picture that Sasha was meant to be looking at. Another 4% of the questions were responded to with an incomplete response (e.g., *I'm not sure.*) when the situation depicted in the target picture clearly demonstrated an opportunity for an accurate response.

The extent to which the learners accepted Sasha as a reliable partner in the information-gap tasks employed in this study depended on the ability of the system to provide consistent and accurate answers to their questions so that they could advance through the tasks. It is important to note that learners had limited knowledge of the target picture. In the matching and sequencing tasks, learners did not usually know which of the pictures on display was the target picture that Sasha's answers referred to. The learner's task was to ask questions and then guess which of the pictures on display was the target picture. In the spot-the-difference task, learners never saw the modified

version of the picture on display that Sasha was supposed to be looking at. It is therefore likely that many of the incorrect or incomplete semantic responses went unnoticed by the learners.

### **9.5. Future Instantiations of Sasha**

The current study represents the preliminary instantiation of Sasha as an intelligent language tutor. There are several concrete areas where Sasha's ability to accurately deliver feedback and semantic responses can be improved (see 5.5. *Improving System Performance for Future Use*). Recast delivery can be greatly improved by refining the syntactic rule-sets that Sasha used to qualify the structure of the learner's input and by implementing a solid strategy for word sense disambiguation. Together with a more accurate mark up of the semantic representation of the task pictures, a sophisticated disambiguation strategy would also help improve the accuracy of Sasha's semantic responses.

Because Sasha was designed to be highly reusable and scalable, there are a wide variety of future pedagogical and research interests that it might serve. In order to facilitate a broad range of future use, all of Sasha's components—including semantic markup and response, feedback generation, task interface and data coding/analysis—will be open-source licensed to encourage further use in applied and computational linguistic research and language instruction. The system is currently ready to be extended to account for a broad range of corrective feedback types, morphosyntactic features and task conditions.

## CHAPTER X: CONCLUSION

This study demonstrated that it is possible to design a computer-based system to behave like a human interlocutor—Sasha was capable of attending to both form and meaning as well as providing feedback and answers with response times comparable to oral interaction with a native speaker. Because conversational interaction and recast provision in the written and oral treatment tasks were nearly identical (apart from broad questions related to modality), it was possible to make a valid comparison of the developmental gains demonstrated by the learners in the two modes of communication. The findings of this study suggest that an autonomous conversational agent may be capable of sustaining meaning-focused interaction with incidental form-focused interventions in a manner that is as beneficial for L2 development face-to-face interaction.

### 9.1. Potential Benefits and Challenges of Computer-Based Language Instruction

Computers have been employed in language instruction for nearly forty years. While early models of CALL were highly artificial and bore little resemblance to spoken interaction, there are now ever-increasing opportunities for naturalistic language production through the use of computers and handheld devices.

Computer-based interaction provides a fertile environment for collecting a broad range of data about a learner's performance history and linguistic competency. As a learner engages in written interaction, it is possible to record and track a variety of performance indicators, such as: (1) learner output, (2) written repair moves, (3) production time/typing speed and (4) the use of instructional resources. In NLP-assisted ICALL applications, a wide range of lexical or morphosyntactic indicators may also be

recorded and tracked. Collectively, this L2 performance data may be organized into a sophisticated learner model that is capable of providing detailed evidence of learner performance and development. A learner model is a formal representation of a learner's performance history and abilities in a particular knowledge domain. In the area of language learning, learner models have largely been used to inform decisions made in the execution of artificial intelligence systems and computer-adaptive testing algorithms. However, learner models can also be used to allow learners, teachers and researchers to receive detailed evidence of linguistic performance (Bull & Kay, 2008; Dimitrova, 2003; Mitrovic & Martin, 2007; Shahrour & Bull, 2009).

The proliferation of online written communication technologies has increased the potential for effective foreign language instruction in a distance learning environment. However, given the current state of technology, distance learning foreign language courses cannot possibly provide benefits equal to those obtained through participation in a traditional language class. Yet, because it eliminates many barriers of time and place, distance learning can provide opportunities for language instruction for potential learners whose schedules or physical locations might otherwise prohibit them from enrolling in a foreign language program<sup>8</sup>. The use of distance learning might be particularly effective in supporting programs in the less-commonly taught languages, where restrictions of time and place have traditionally contributed to critically low enrollments.

Technology is also increasingly being employed as a means of augmenting traditional language instruction in blended classrooms. In the past decade, the use of

---

<sup>8</sup> For an example of current policy on distance learning in foreign language instruction, see the position statement of the National Council of State Supervisors for Languages (NCSSFL). Founded in 1960, NCSSFL is an organization of education agency personnel from all states of the

technology-based materials (e.g., Internet resources, computer-assisted instructional materials, instructional technology) has increased significantly in U.S. foreign language classrooms (Rhodes & Pufahl, 2008). In blended classroom environments, synchronous and asynchronous CMC tools allow learners to interact with one another and/or the instructor outside of the time and place constraints of the traditional classroom. Intelligent conversational agents such as Sasha may also enable learners in blended classes to engage in meaning-focused interaction and receive form-focused feedback without direct human intervention—creating conditions conducive to differentiated learning that could, heretofore, only be achieved by employing additional human teaching assistants. Unlike humans, intelligent agents are capable of simultaneously maintaining individualized dyadic interaction with large numbers of learners and do not suffer from fatigue.

In the current study, Sasha sustained a combined total of 28.5 hours of one-on-one conversational interaction with 19 beginner and intermediate ESL learners—during which, thousands of open-ended questions were posted to the system. Sasha was able to consistently maintain high levels of accuracy in the delivery of both semantic responses and morphosyntactic feedback and was as effective at promoting L2 development as an expert native speaker.

It is necessary to note that creating a conversational agent capable of autonomously maintaining meaning-focused interaction and providing form-focused feedback is not, by any means, a trivial undertaking. In order to prepare Sasha for sustaining open-ended L2 conversational interaction and delivering accurate recasts,

---

United States who have the responsibility of foreign/world language education at the state level. The position statement is available at: <http://www.ncssfl.org/papers/index.php?distancelearning>.

nearly one and a half years (approximately 2500 hours) of intensive development was required on the part of the researcher.

Technical development primarily focused on the interweaving of a wide range of open-source NLP technologies (Brill Tagger, Collins Parser, Google Suggest and Wordnet) in order to create a communicative system capable of generating accurate recasts and providing semantic responses to largely unrestricted L2 written input. A considerable amount of development time was required in order to train the embedded NLP technologies to handle non-targetlike L2 input (see 4.2. *Preparing L2 Input for Analysis* and 4.3.1. *L2 retraining of the syntactic parser*).

A significant amount of time was also invested in creating robust, reusable Web-based tools to facilitate (1) rapid, intelligent markup of the semantic representations present in visual task prompts (see Figure 4.9) and (2) multi-rater coding of oral and written learner production (see Figures 6.1 and 6.2). All of Sasha's linguistic modules and the accompanying research tools were designed to be open-source, scalable and capable of accounting for any number of task conditions and linguistic forms. There is thus enormous potential for reuse of the system (in part or whole) in future research and language instruction.

## **9.2. The Future of NLP-Driven Technologies in Language Instruction**

Computer-assisted language learning is steadily and rapidly evolving. As diverse forms of online communication find their way into in the academic and social life of L2 learners and the availability of powerful linguistic resources continues to proliferate, there are ever-increasing opportunities to use NLP tools to assist in language instruction.

A growing number of NLP tools (i.e., syntactic parsers, parts of speech taggers, morphological analyzers, multi-lingual dictionaries, translation services and various corpus tools) are becoming widely available to CALL developers—presenting them with the opportunity to bring to bear the power of NLP-driven technology on the language learning process. Traditionally, NLP-driven (i.e., ICALL) systems have taken the form of stand-alone, workbook-like applications capable of delivering customized feedback to learners as they engage in form-focused, discrete-point language activities. Several recent innovative applications of NLP technology have begun to explore ways in which ICALL can be put to use in less artificial and restrictive task conditions.

One example of how NLP technologies are being explored in more open-ended, naturalistic learning contexts are the reading- and writing-assistant tools for L2 learners of Southeast Asian languages produced by the Center for Research in Computational Linguistics<sup>9</sup>. To assist with the reading of authentic texts, the *Reader's Helper* uses bi-text corpora to provide translation and transliteration of authentic texts and employs electronic dictionary resources to enable dynamic highlighting of word classes in the text as well as automatic glossing. The *Writer's Helper* employs a variety of NLP resources to provide support for word order, segmentation and lexical control. The *Writer's Helper* also features an auto-completion tool that learners may use as they produce open-ended compositions in the target language.

Other examples of innovative uses of NLP technologies that break from the restrictive and proscriptive nature of traditional ICALL activities have emerged from the ICALL Research Group at The Ohio State University. TAGARELA, for example, is an ICALL system that has been designed to support instruction in an individualized, self-

paced curriculum for University-level Portuguese<sup>10</sup>. TAGARELA employs a range of NLP technologies to analyze learner input and L2 errors and creates sophisticated learner and instructor models that allow the system to cater to the specific needs of individual learners by generating customized feedback. A separate system developed by this group is WERTi (Working with English Real-Texts)<sup>11</sup>. WERTi allows learners to freely choose the topic of the authentic texts that they read as they engage in form-focused language practice and is capable of dynamically creating form-focused practice activities on authentic texts obtained from Reuters news feeds. WERTi's model of employing NLP technology for dynamic activity generation is compelling for several reasons. First, it gives learners control over the topical content that they work with—ensuring that the selected texts align with the learner's interests. Second, it eliminates the considerable amount of development time required by creating traditional form-focused CALL activities.

Newly emerging technologies are expanding the range of authentic language content that might be compatible with WERTi's model of dynamic activity generation. It is becoming increasingly possible to bring NLP to bear on authentic audio and video content. For example, Google has recently unveiled a new set of technologies that enable automatic captioning of authentic, multilingual audio and video<sup>12</sup>. By providing written transcripts of audio and video files, automatic captioning might allow NLP-driven tools like WERTi to provide form-focused support to learners as they engage with recorded spoken language content.

---

<sup>9</sup> See <http://sealang.net/lab/>

<sup>10</sup> See <http://prospero.ling.ohio-state.edu/tagarela/>

<sup>11</sup> See <http://prospero.ling.ohio-state.edu/werti/>

<sup>12</sup> See <http://www.nytimes.com/2009/11/20/technology/internet/20google.html>

One of the most intriguing prospects for employing NLP technology in online language use and instruction may lie in the development of “portable” NLP-driven agents. As learners engage in online communication in the target language (e.g., emailing an instructor, socially networking with native-speaking friends, etc.), future instantiations of NLP-driven agents might act as ever-present assistants that are capable of analyzing written production and providing suggestions and/or feedback to learners as they compose messages or texts for submission. In addition to providing on-demand assistance to the learner, these tools would extend the range of discourse contexts in which evidence might be captured to inform diagnostic learner models. Portable agents might take the form of stand-alone, Web-based applications that function alongside online communication tools or as embedded chatbots that function as add-ons to existing communication tools.

A new generation of online communication tools has begun to emerge that support the integration of third-party NLP-driven applications into naturalistic online communication. One such example is the *Wave* platform under development by Google. *Google Wave* facilitates online collaboration by allowing groups of users to communicate and work together with richly formatted text and multimedia. *Wave* also allows non-human expert agents (or *Wave* bots) to participate in online communication. Currently, there are several *Wave* bots have been developed to assist with cross-linguistic communication—providing lexical support and translation services<sup>13</sup>. One such bot is *Rosy Etta*—an expert agent built on top of Google’s online translation services that is

---

<sup>13</sup> See <http://wavety.com/google-wave-gadgets-and-robots/>, “Language”.

capable of spontaneously translating CMC speech in over 40 languages<sup>14</sup>. While translation services per se are not particularly useful for L2 instruction, NLP-driven agents like Sasha might be employed to act as conversational assistants that provide corrective feedback to the learner as they engage in a broad range of CMC discourse contexts.

### **9.3. The Future of CALL in Second Language Research and Instruction**

In its early years, CALL was relegated to the margins of SLA as it was widely considered to be highly experimental and artificial. Yet, as technology has advanced and research methodologies have matured, the use of CALL has gone from being a fringe form of language instruction to a staple in most modern language classrooms (Rhodes & Pufahl, 2008; Thorne & Payne, 2005).

Communication technology continues to evolve and proliferate at an ever-increasing rate. As new generations of learners enter the language classroom, they come equipped with communicative competencies and expectations that few could have imagined ten, much less, twenty years ago. Technology is increasingly presenting language learners with new vehicles for accessing authentic target-language content, utilizing intelligent language learning resources and engaging in naturalistic communication. There is thus a growing challenge for CALL and SLA research to empirically demonstrate the extent to which language development might occur in the various learning contexts engendered by these emerging technologies. As interactive written discourse pervades the social, professional and academic life of modern foreign language learners there is a mounting need for second-language researchers and teaching

---

<sup>14</sup> Note that Rosy Etta is only capable of translating each language as accurately as the underlying structures in Google Translate allow.

professionals to determine how new technologies might be optimally employed to promote language learning. It is vital that these new technologies are subject to rigorous research and validation, in which theoretical foundations and task conditions are reconcilable within broader contexts of SLA research.

In order for CALL research to further coalesce with broader models of SLA, there is a pressing need to cultivate a larger body of researchers and research tools that can help bridge the gap between traditional linguistic theory and the use of computer-based methodologies. Effective CALL development is, by nature, a cross-disciplinary endeavour—requiring expertise in applied and computational linguistics as well as instructional design. Extensible, open-source platforms like Sasha provide a productive vehicle for promoting the cross-disciplinary collaboration necessary for creating optimal online learning conditions.

The current study represents a giant methodological leap in creating conditions conducive to meaningful interaction between a second-language learner and an autonomous computer-guided system. Although there are many lingering theoretical questions regarding the extent to which modality impacts instruction, this study provides early empirical evidence that feedback-intensive written interaction with an autonomous conversational agent may be a productive vehicle for language learning.

## REFERENCES

- Abrams, Z. I. (2003). The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal*, 87 (2), 157–167.
- Adams, R. (2004). Learner-Learner Interactions: Implications for Second Language Acquisition. Unpublished doctoral dissertation, Linguistics Department, Georgetown University.
- Allwright, R., & Bailey, K. M. (1991). *Focus on the language classroom*. Cambridge: Cambridge University Press.
- Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of passé composé and imparfait. *Modern Language Journal*, 85(2), 226-243.
- Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study. *Modern Language Journal*, 88(1), 31-55.
- Baddeley, A.D., & Hitch, G.J. (1974) . Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation*, 8 (pp. 47- 89) . New York: Academic Press.
- Bailey, S. & Meurers, D. (2008). Diagnosing Meaning Errors in Short Answers to Reading Comprehension Questions. In *Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Applications. Association for Computational Linguistics*, 107-115.
- Bailin, A. (1995). AI and language learning: Theory and evaluation. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology* (pp. 327-343). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bartneck, C., & Lyons, M. J. (2007). HCI and the Face: Towards an Art of the Soluble. In J. Jacko (Ed.), *Human-Computer Interaction, Part 1, HCII2007, LNCS 4550* (pp. 20-29). Berlin: Springer.
- Beauvois, M. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals* 25(5), 455-642.
- Beauvois, M. H. (1997). Write to speak: The effects of electronic communication on the oral achievement of fourth semester French students. In J. Muyskens (Ed.), *New ways of learning and teaching: Issues in language program direction* (pp. 93–116). Boston: Heinle.

- Beauvois, M. H. (1998). Conversations in slow motion: Computer-mediated communication in the foreign language classroom. *Canadian Modern Language Review*, 54, 198–217.
- Bell L. (2003) Linguistic adaptations in spoken human-computer dialogues. Unpublished doctoral dissertation, Department of Speech, Music and Hearing, KTH, Stockholm.
- Bernstein, J., Najmi, A., & Ehsani, F. (1999). Subarashii: Encounters in Japanese spoken language education. *CALICO Journal*, 16 (3), 361-384.
- Bikel, D. (2005). Multilingual Statistical Parsing Engine. <http://www.cis.upenn.edu/~dbikel/software.html#stat-parser>.
- Blake, R. (2000). Computer mediated communication: A window on L2 Spanish Interlanguage. *Language Learning & Technology*, 4 (1), 120-136.
- Blake, R. (2005). ‘Bimodal CMC: The glue of language learning at a distance’, *CALICO Journal* 22(3), 497–511.
- Bley-Vroman, R. and C. Chaudron (1994) Elicited imitation as a measure of L2 competence. In E. Tarone, S. Gass and A. Cohen (eds.), *Research Methodology in SLA*. Hillsdale, NJ: Erlbaum. 245–61.
- Bou, B. (2005). Wordnet SQL Builder. Retrieved June 01, 2006 from <http://wnsqlbuilder.sourceforge.net/>.
- Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative - speaker interactions. *Language Learning*, 52(1), 1-42.
- Brill, E. 1992. A simple rule-based part-of-speech tagger. *In Proceeding of the Third Conference on Applied Natural Language Processing*, Trento, Italy.
- Bump, J. (1990) ‘Radical changes in class discussion using networked computers’, *Computers and the Humanities* 24: 49–65.
- Bygate, M. and Samuda, V. (2009) Creating Pressure in Task Pedagogy: The Joint Roles of Field, Purpose, and Engagement within the Interaction Approach. In A. Mackey and C. Polio (eds.), *Multiple Perspectives on Interaction*. (pp. . 90-116.) New York, NY: Routledge
- Carpenter, H., Jeon, K. S., MacGregor, D., & Mackey, A. (2006). Learners’ interpretations of recasts. *Studies in Second Language Acquisition*.
- Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, 22, 67-83.

- Cassell, J., & Tartaro, A. (2007). Intersubjectivity in Human-Agent Interaction. *Interaction Studies* 8 (3), p 391-410.
- Chapelle, C. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology*, 1, 9–43.
- Chapelle, C. (1998). Multimedia CALL: Lessons to be learned from research on instructed LA. *Language Learning & Technology*, 2(1), 22-34.
- Chapelle, C. (2001). *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research*. Cambridge, UK: Cambridge University Press.
- Chapelle, C. (2004). Technology and second language learning: expanding methods and agendas. *System* 32, 593–601.
- Chapelle, C. A. (2005). Interactionist SLA theory in CALL research. In J. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 53-64). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chun, D. (1994). Using computer networking to facilitate the acquisition of interactive competence, *System*, 22 (1), 17-31.
- Chun, L., Fei, F. and Roots, R. (2008). The Contingency of recasts and noticing. *CALICO Journal*, 26 (1), 70-90.
- Collins. M. J. (1996). A new statistical parser based on bigram lexical dependencies. In Arivind Joshi and Martha Palmer, editors, *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics* (pp. 184–191), San Francisco. Association for Computational Linguistics, Morgan Kaufmann Publishers.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163–191.
- Delmonte R. (2002). Feedback generation and linguistic knowledge in 'SLIM' automatic tutor, *ReCALL* 14 (1), 209-234.
- DeSmedt, W. H. (1995). Herr Kommissar: An ICALL Conversation Simulator for Intermediate German. In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology* (pp. 153-174). Mahwah, NJ: Lawrence Erlbaum Associates.
- Doughty, C. (1999). The cognitive underpinnings of focus on form. *University of Hawai'i Working Papers in ESL*, 18, 1–69.

- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.
- Ehsani, F., Bernstein, J. and Najmi, A. (2000). An interactive dialog system for learning Japanese. *Speech Communication*, 30, 167–177.
- Ellis, R. (1989). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Ellis, R. (2001). Investigating form-focused instruction. In R. Ellis (Ed.), *Form focused instruction and second language learning* (pp. 1–46). Oxford, UK: Blackwell.
- Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, 18(1), 4-22.
- Ellis, Rod. (2009). Corrective Feedback and Teacher Development. *L2 Journal*, 1(1), 3-18.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative classrooms. *Language Learning*, 51(2), 281-318.
- Ellis, R., & Yuan, F. (2004). The effects of careful on-line planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit language knowledge: an empirical validation study. *Applied Linguistics*, 27, 1-14.
- Farrar, M. J. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language* 17, 607-24.
- Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, 28, 90-98.
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.
- Felshin, S. (1995). The Athena Language Learning Project NLP system. In Holland, V. M., Kaplan, J. D. and Sams, M. R. (eds.), *Intelligent Language Tutors: Theory Shaping Technology* (pp. 257–272), Hove, UK: Erlbaum.
- Fitze, M. (2006). Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning and Technology*, 10(1), 67–86.

- Foster, R., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323.
- Foster, P. & Skehan, P. (1997). Modifying the task: The effects of surprise, time and planning type on task-based foreign language instruction. *Thames Valley working papers in English language teaching*, vol. 4.
- Garrett, N. (1995). ICALL and Second Language Acquisition. *Intelligent Language Tutors: Theory Shaping Technology*, edited by V.M. Holland, J. Kaplan, and M. Sams. Mahwah, NJ: Lawrence Erlbaum.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, Lawrence Erlbaum.
- Gass, S. (2003). Input and interaction. In C. J. Doughty and M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224-255). Malden, MA: Blackwell Publishing.
- Gieselmann, P. & Stenneken, P. (2006). How to talk to robots: Evidence from user studies on human-robot communication. In Fischer, K., *How people talk to computers, robots, and other artificial communication partners* (pp. 68-79). Bremen/Freiburg: Transregional Collaborative Research Center.
- Hamburger, H. (1995). Tutorial Tools for Language Learning by Two-Medium Dialogue. In V. M. Holland, J. D. Kaplan, M. R. Sams, editors, *Intelligent Language Tutors: Theory Shaping Technology*. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.
- Havranek, G. (1999). The effectiveness of corrective feedback: Preliminary results of an empirical study. *Acquisition et interaction en langue étrangère: Proceedings of the Eighth EUROSLA Conference*, 2, 189–206.
- Heift, T. (2002). Learner Control and Error Correction in ICALL: Browsers, Peekers and Adamants. *CALICO Journal*, 19 (3), 295-313.
- Heift, T. (2003). Multiple learner errors and meaningful feedback: A challenge for ICALL systems. *CALICO Journal*, 13 (in press).
- Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16 (2), 416-431.
- Heift, T. (forthcoming). Prompting in CALL: A longitudinal study of learner uptake. *Modern Language Journal*, Volume 94 (2), in press.

Heift, T., & Rimrott, A. (2008). Learner Responses to Corrective Feedback for Spelling Errors in CALL. *System*, 36(2), 196-213.

Heift, T., & Schulze, M. (2007). *Errors and Intelligence in CALL. Parsers and Pedagogues*. New York: Routledge.

Herring, S. C. (2001). Gender and power in online communication. CSI Working Paper (No. WP-01-05). Retrieved January 21, 2006 from <http://rkcsi.indiana.edu/archive/CSI/WP/WP01-05B.html>.

Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, 15, 461-493.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hulstijn, J. (2000). The use of computer technology in experimental studies of second language acquisition: a survey of some techniques and some ongoing studies. *Language Learning & Technology*, 3 (2), 32-43.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of Japanese as a foreign language. *Language Learning*, 54(2), 311-394.

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, 25, 1-36.

Izumi, S. (2002). Output, input enhancement and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541-577.

Jansen, L. M. (2000). Second language acquisition: from theory to data. *Second language research*, 16, 27-43.

Jepson, K. (2005). Conversations and negotiated interaction in text and voice chatrooms. *Language Learning & Technology*, 9(3), 79-98.

Kan, M.Y. (2003). Daemonized Collins Parser. Retrieved from <http://www.comp.nus.edu.sg/~rpnlpir/daemonCollins/>

Kaplan, J. D., & Holland, V. M. (1995). Application of Learning Principles to the Design of a Second Language Tutor. In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology* (pp. 273-287). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaplan, J. D., Sabol, M. A., Wisher, R. A., & Seidel, R. J. (1998). The Military Language Tutor (MILT) Program: An Advanced Authoring System. *Computer Assisted Language Learning*, 11(3), 265-287.

Kelm, O. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. *Foreign Language Annals*, 25, 441–454.

Kelm, O. (1996). Application of computer networking in foreign language education: Focusing on principles of second language acquisition. In M. Warschauer (Ed.), *Telecollaboration in foreign language learning: Proceedings of the Hawai'i symposium* (pp. 19–28). Honolulu, HI: University of Hawaii Second Language Teaching and Curriculum Center.

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and quality of language production. *The Modern Language Journal*, 79, 457-476.

Kitade, K. (2000). L2 learners' discourse and SLA theories in CMC: Collaborative interaction in Internet chat. *Computer Assisted Language Learning*, 13(2), 143–166.

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. *Applied Psycholinguistics*, 21, 297-320.

Kopp, S. (2006). How people talk to a virtual human - conversations from a real-world application. In Fischer, K., *How people talk to computers, robots, and other artificial communication partners* (pp. 101-112). Bremen/Freiburg: Transregional Collaborative Research Center.

Kötter, M. (2001). MOOrituri te salutant? Language Learning through MOO-Based Synchronous Exchanges between Learner Tandems. *Computer Assisted Language Learning*, 14 (3-4), 289 - 304.

Kroonenberg, N. (1995). The French connection: Public journals for high school students. In M. Warschauer (Ed.), *Virtual connections: Online activities and projects for networking language learners* (pp.27-29). Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.

Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology*, 10, 102–120.

Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E. and Rouet, J. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies*, 65, 983-991.

- Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, 25, 37-63.
- Leow, R., Egi, T., Nuevo, A. M. & Tsai, Y. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, 13(2), 93-108.
- Linnell, J. (1995). Can negotiation provide a context for learning syntax in a second language? (ERIC 393 305) *Working Papers in Educational Linguistics*, 11 (2), 83-103.
- Lochtman, K. (2000, August). *The role of negative feedback in experiential vs. analytic foreign language teaching*. Paper presented at the Conference on Instructed Second Language Learning, Brussels, Belgium.
- Loewen, S. (2003). Variation in the frequency and characteristics of incidental focus on form. *Language Teaching Research*, 7, 315 – 345.
- Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Language Learning*, 54(1) 153-187.
- Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, 19, 1–14.
- Loewen, S., & Philp, J. (2006). Recasts in the adult L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal*, 90(4).
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W.R., & Bhatia, T.J. (eds.), *Handbook of second language acquisition* (pp. 413-68). San Diego: Academic Press.
- Long, M. H. (2007). Recasts in SLA: The story so far. In Long, M. H., *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative evidence in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, 82, 357-371.
- Long, M. H., & Robinson, P. (1988). Focus on form: Theory, research and practice. In Doughty, C. J., & Williams, J. (eds.), *Focus on form in second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Lyster, R. (1998a). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48, 183-218.
- Lyster, R. (1998b). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 51-81.

- Lyster, R. (2001). Negotiation of form, recasts and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 51, 265-301.
- Lyster, R. (2002). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research*, 37, 237-253.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399-432.
- Lyster, R. & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, 59 (2), 453-498.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19, 37-66.
- Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557-587.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics* 27, 405-430
- Mackey, A., Gass, S., & McDonough, K. (2000). Do learners recognize implicit negative feedback as feedback? *Studies in Second Language Acquisition*, 22, 471-497
- Mackey, A. & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30(4), 459-477.
- Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, 53(1), 35-66.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82, 338-356.
- Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-208). Philadelphia: John Benjamins.
- Mackey, A., Polio, C. & McDonough, K. (2004). The relationship between education and teachers' incidental focus-on-form. *Language Teaching Research*, 8(3), 301-327.
- Mackey, A. & Silver, R. E. (2005). Interactional tasks and English L2 learning by

immigrant children in Singapore. *System*, 33, 239-260.

Marcus, M., Santorini, B. & Marcinkiewicz, M. (1993). Building a Large Annotated Corpus of English: The Penn Treebank, in *Computational Linguistics*, 19 (2), pp. 313—330.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27(1), 79-103.

McDonough, K., & Mackey, A. (2006) Responses to recasts: Repetitions, primed production and linguistic development. *Language Learning*, 54, 693–720.

McDonough, K. and Mackey, A. (2008). Syntactic Priming And Esl Question Development. *Studies in Second Language Acquisition*, 30, pp 31-47.

Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in second language acquisition. *Studies in Second Language Acquisition*, 3, 109-135.

Mohan, B., & Beckett, G. H. (2003). A Functional Approach to Research on Content-Based Language Learning: Recasts in Causal Explanations. *Language Learning*, 87(3), 421-432.

Mori, M. (1970). The Uncanny Valley. *Energy*, 7, 33-35.

Murray, J. H. (1995). Lessons Learned from the Athena Language Learning Project: Using Natural Language Processing, Graphics, Speech Processing, and Interactive Video for Communication-Based Language Learning. In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology* (pp. 243-256). Mahwah, NJ: Lawrence Erlbaum Associates.

Nabei, T., & Swain, M. (2002). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness*, 11(1), 43-62.

Nagata, N. (1995). An effective application of natural language processing in second language instruction. *CALICO Journal*, 13 (1), 47-67.

Nagata, N. (1997a). The effectiveness of computer-assisted metalinguistic instruction: A case study in Japanese. *Foreign Language Annals*, 30 (2), 187-199.

Nagata, N. (1997b). An experimental comparison of deductive and inductive feedback generated by a simple parser. *System*, 25 (4), 515-534.

Nagata, N. (1998). Input vs. output practice in educational software for second language acquisition. *Language Learning & Technology*, 1(2), 23-40

- Nagata, Noriko (2002). BANZAI: An Application of Natural Language Processing to Web based Language Learning. *CALICO Journal*, 19(3), 583–599.
- Nagata, N. & Swisher, M. V. (1995). A Study of Consciousness-Raising by Computer: The Effect of Metalinguistic Feedback on Second Language Learning. *Foreign Language Annals*, 28 (3), 337-347.
- Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719-758.
- Norris, J. M. (1996). *A validation study of the ACTFL Guidelines and the German Speaking Test*. Honolulu, HI: University of Hawaii, unpublished M.A. thesis.
- Norris, J. M. (2005). Using developmental sequences to estimate ability with English grammar: Preliminary design and investigation of a web-based test. *Second Language Studies*, 24(1), Fall 2005, pp. 24-128.
- Ohta, A. S. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese language classroom. In Hall, J. K. & Verplaeste, L. S. (eds.), *The construction of second and foreign language learning through classroom instruction* (pp. 47-71). Mahwah, NJ: Lawrence Erlbaum.
- Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition*, 17, 459-481.
- Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal*, 87, 519-533.
- Ortega, L. (2008). Text-based SCMC for L2 learning: Reappraising the research agenda. In A. Mackey & C. Polio (Eds.), *Multiple Perspectives on Interaction in SLA*. Mahwah, NJ: Erlbaum.
- Oxford, R. L. (1993). Intelligent Computers for Learning Languages: The View for Language Acquisition and Instructional Methodology. *Computer Assisted Language Learning*, 6(2), 173-188.
- Oxford, R. L. (1995). Linking Theories of Learning with Intelligent Computer-Assisted Language Learning (CALL). In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology* (pp. 359-369). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, 28(3), 361-382.

- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36(4), 573-595.
- Payne, J. S., & Ross, B. M. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning and Technology*, 9(3), 35–54.
- Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20(1), 7–32.
- Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.
- Pienemann, M. & Johnston, M. (1987). Factors influencing the development of language proficiency. In Nunan, D. (ed.) *Applying second language acquisition research*. National Curriculum Resource Centre, Adult Migrant Education Program, Adelaide.
- Pienemann, M., Johnston, M., & Brindley, G. (1988), Constructing an acquisition-based procedure for second language assessment, *SSLA (10)*, 217-243.
- Pellettieri, Jill. (2000). "Negotiation in cyberspace: The role of chatting in the development of grammatical competence." In M. Warschauer and R. Kern (Eds.), *Network-based language teaching: Concepts and practice*. New York: Cambridge University Press.
- Philp, J. (2003). Constraints on noticing the gap: Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25, 99-126.
- Poza, M. I. C. (2005). *The effects of asynchronous computer voice conferencing on learners' anxiety when speaking a foreign language*. Unpublished doctoral dissertation, West Virginia University, Morgantown.
- Rhodes, N. C. & Pufahl, I. (2008). *Foreign Language Teaching in U.S. Schools: Results of a National Survey*. Washington, D.C. Center for Applied Linguistics.
- Sachs, R., & Suh, B., (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197-227). Oxford: Oxford University Press.
- Sagarra, N. (2007). The effect of computer-delivered recasts and working memory on L2 development and modified output during face-to-face interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press.
- Salaberry, R. (2000). Pedagogical Design of Computer Mediated Communication Tasks: Learning Objectives and Technological Capabilities. *Modern Language Journal*, 84 (1), 28-37.

- Sauro, S. (2009). Computer-mediated corrective feedback and the development of 12 grammar. *Language Learning & Technology*, 13(1), 96-120.
- Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language*, 24, 139-161.
- Saxton, M. (2005). 'Recast' in a new light: insights for practice from typical language studies. *Child Language Teaching and Therapy*, 21(1), 23-38.
- Schwienhorst, K. (2002). Evaluating tandem language learning in the MOO : Discourse repair strategies in a bilingual Internet project. *Computer-Assisted Language Learning*, 15(2), 135-146.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 17-46. Schraw, G., & Dennison, R. S. (1994). The effect of reader purpose on interest and recall. *Journal of Reading Behavior*, 26, 1-17.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: University of Hawaii Press.
- Schmidt, R. (2001). Attention. In Robinson, P. (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition*, (pp. 237-326) Rowley, MA: Newbury House.
- Schulze, M. (2003). Grammatical Errors and Feedback: Some Theoretical Insights. *CALICO Journal*, 20 (3), 437-450.
- Schulze, M. (2008). AI in CALL: Artificially Inflated or Almost Imminent? *CALICO Journal*, 25 (3), 510-527.
- Schneiderman, B. (1992). *Designing the user interface*. Reading, MA: Addison-Wesley.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165-179.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8(3), 263-300.
- Shook, D. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning*, 5, 57-93.

- Silver, R. E. (2000). Input, output, and negotiation: Conditions for second language development. In B. Swierzbin, F. Morris, M. E. Anderson, C. A. Klee, & E. Tarone (Eds.), *Social and cognitive factors in second language acquisition: Selected proceedings of the 1999 Second Language Research Forum* (pp. 345–371). Somerville, MA: Cascadilla Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, 87, 38–58.
- Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition*, 26, 365-398.
- Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly*, 39 (1), 33-58.
- Smith, B. (2009). The Relationship between scrolling, negotiation, and self-initiated self-repair in an SCMC environment. *CALICO Journal*, 26(2), 231-245.
- Smith, B. & Sauro, S. (2009). Interruptions in chat. *Computer Assisted Language Learning*, 22(3),229-247.
- Sotillo, M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4 (1),82-119.
- Spada, N. & Lightbown, P.M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition*, 15(2), 205-224
- Spada, N. & Lightbown, P. M. (1999). Instruction, L1 influence and developmental "readiness" in second language acquisition. *Modern Language Journal*, 83(1),1-22.
- St. John, E., & Cash, D. (1995). Language learning via e-mail: Demonstrable success with German. In M. Warschauer (Ed.), *Virtual connections: Online activities and projects for networking language learners* (pp. 191-197). Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Sutcliffe, A.G. (1995). *Human-computer interface design, 2<sup>nd</sup> edition*. London, etc: MacMillan Press Ltd.
- Sykes, J. (2005). Synchronous CMC and Pragmatic Development: Effects of Oral and Written Chat. *CALICO Journal*, 22 (3), 399-431.
- Thorne, S. (2003). Artifacts and cultures-of-use in intercultural communication.

*Language Learning & Technology*, 7(2), 38–67.

Thorne, S.L. and Payne, J.S. (2005). Evolutionary trajectories, Internet-mediated expression, and language education. *CALICO Journal* 22, 371–397.

Tudini, V. (2003). Using native speakers in chat. *Language Learning & Technology*, 7(3), 141–159.

Van Patten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, 12, 287–301.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

Vinthner, T. (2002) Elicited imitation: a brief overview. *International Journal of Applied Linguistics* 12, 54-73.

Volle, L. M. (2005). Analyzing oral skills in voice e-mail and online interviews. *Language Learning & Technology*, 9(3), 146-163.

Wang, Y. (2004). Distance language learning: Interactivity and fourth-generation Internet-based videoconferencing. *CALICO Journal*, 21(2), 373–395.

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal* 13(2), 7-26.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *Modern Language Journal*, 81, 470-481.

Warschauer, M. (1998) *Electronic Literacies: Language, Culture, and Power in On-line Education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Warschauer, M. & Kern, R. (2000). *Network-based language teaching: Concepts and Practice*. Cambridge University Press.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition* 23, 345-368.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1–27.

Zacharski, R. (2003). A discourse system for conversational characters. In A. Gelbukh (Ed.), *Proceedings of the fourth international conference on intelligent text processing and computational linguistics* (pp. 492-495). Heidelberg: Springer Verlag.

## APPENDIX A.1: EXAMPLE INTERACTION WITH EMDODIED AGENTS

Embodied conversational agents interact with humans through the visual representation of some character that is capable of sustaining some form of communication with the user. Embodied agents have appeared in the form of both human and non-human interlocutors. Because speech-recognition technology remains fraught with challenges (even in handling L1 speech), conversational agents require the users to type their utterances. However, many embodied agents are capable of producing synthesized, computer-generated speech in addition to text-based responses.

Embodied agents are generally not static images. Rather, they are animated characters that are capable of emulating-- with varying degrees of effectiveness— some of the paralinguistic cues that might be exhibited in oral interaction. Movement of the agent is achieved either through videocentric (prescribed video clips) or pure computer animation.

The following examples represent the current state of the art of human-computer interaction with embodied conversational agents. All of these agents are animated in that they turn their heads and move their eyes and mouths to simulate conversational interaction and are capable of presenting responses to the user both visually and aurally. Example 1 shows a sample conversation with “Fake Kirk”. Kirk has domain knowledge about all things *Star Trek* and sometimes responds to inquiries with abstruse quotes of the venerable captain. Example 2 demonstrates a typical interaction sequence facilitated by “Joan”-- a videocentric agent that displays an elaborate set of paralinguistic cues—both facial and corporal. Example 3 shows an example of interaction with a purely animated conversational agent with domain knowledge about the company that it has been designed to represent.

## Example 1. Interaction with Artificial Solutions<sup>1</sup> speaking avatar, “Fake Kirk”



**Please identify yourself.**

**You said:** How do you feel about Khan?  
**Fake Kirk:** How? I've seen a part of myself no man should ever see.

**You say:**  Say

**Captain's Log:**

Lieutenant: What is the stardate?  
Fake Kirk: Stardate Thursday, November 19, 2009.  
Lieutenant: What do you think about Spock?  
Fake Kirk: Mr. Spock? he is the first officer of the Enterprise.  
Lieutenant: Do you have a phaser?  
Fake Kirk: A what?  
Lieutenant: Are you on the Enterprise?  
Fake Kirk: On the bridge.  
Lieutenant: where are you?  
Fake Kirk: On board the Enterprise.  
Lieutenant: are you on the bridge?  
Fake Kirk: In the Engine Room.  
Lieutenant: who is with you  
Fake Kirk: A member of my crew.  
Lieutenant: which member of your crew is with you?  
Fake Kirk: Miss Uhura.  
Lieutenant: When did you leave Earth?  
Fake Kirk: Anything else?  
Lieutenant: How long have you been in space?  
Fake Kirk: 34 years old.  
Lieutenant: How do you feel about Khan?  
Fake Kirk: How? I've seen a part of myself no man should ever see.

*Simulated by author at <http://bandore.pandorabots.com/pandora/talk?botid=f4c90421ae345a9f>, 11/20/2009.*

---

<sup>1</sup> <http://www.artificial-solutions.com/>

## Example 2. Interaction with Icogno's<sup>2</sup> speaking, videocentric avatar, "Joan"

---



**Joan:** Hi there, I'm Joan. What would you like to talk about?  
**You:** What is today's date?  
**Joan:** Tomorrow.  
**You:** Why do you keep fidgeting about?  
**Joan:** Because you ask silly questions.  
**You:** What do you mean by silly questions?  
**Joan:** Why are you talking?  
**You:** I am trying to talk to you!  
**Joan:** Really?  
**You:** What can you tell me about?  
**Joan:** It's really good and the person I'm giving it to really likes them.  
**You:** Who made you?  
**Joan:** My parents, of course.  
**You:** Who are your parents?  
**Joan:** Their names are Michael and Sue, and they live in Hastings.  
**You:** Are you married?  
**Joan:** No, I'm still single and searching.  
**You:** What kind of man are you looking for?  
**Joan:** Somebody who shares my interests and above all respects me.  
**You:** What you want to talking about?  
**Joan:** I want to talk about intelligence.

---

*Simulated by the author at <http://www.icogno.com/joan.html>, 11/20/2009.*

---

<sup>2</sup> <http://www.icogno.com>

### Example 3. Interaction with Daden<sup>3</sup>'s animated virtual receptionist, "Abi"

---



- You:** How big is the company?  
**Abi:** Sorry, I only have limited knowledge. I can only talk about Daden  
**You:** How big is Daden  
**Abi:** Daden have a core staff in the UK, and then developers spread across the globe and SL  
**You:** Where are you located?  
**Abi:** I am at the Daden offices on Daden Prime in Second Life  
**You:** Where are the Daden offices in real life?  
**Abi:** We are tenants at the Serious Games Institute in Coventry, England, and also have our main office on Daden Prime in Second Life  
**You:** What do you do?  
**Abi:** I can tell you about Daden, and about our services.  
**You:** What does Daden do?  
**Abi:** Daden help companies get into second life and other virtual worlds. We also build chatbots that companies can use as interactive guides to their web sites and virtual world locations. We offer four main services: inform, involve, integrate and interact  
**You:** Is you are in England?  
**Abi:** Say again? I only have limited knowledge. I can only talk about Daden

---

*Simulated by the author at <http://www.daden-cs1.co.uk/chatbots/livebots/abi.html>, 11/19/2009.*

---

<sup>3</sup> <http://www.daden.co.uk/>

## APPENDIX A.2: EXAMPLE INTERACTION WITH HEIFT'S *E-TUTOR*

Example 1 illustrates how the E-Tutor's Build a Sentence task handles multiple learner errors in a serial fashion as learners make multiple attempts at producing the target utterance.

### Example 1. Human-computer interaction in E-Tutor's Build a Sentence tasks

<i>Sentence Template</i>	<i>Anne / aus / Amerika / sein</i>	Error
Learner Attempt 1:	*Anne aus America sein.	L-S-M
System Response 1:	<i>Anne aus <b>America</b> sein</i> There is a spelling mistake with <b>America</b> . Suggestions: <b>Amerika</b>	
Learner Attempt 2:	*Anne aus Amerika sein.	S-M
System Response 2:	<i>Anne [...] aus Amerika <b>sein</b></i> Careful! There is a problem in word order. How about moving <b>sein</b> ? <i>See help on: <u>word order</u></i>	
Learner Attempt 3:	*Anne sein aus Amerika.	M
System Response 3:	<i>Anne <b>sein</b> aus Amerika</i> You made a mistake with the verb <b>sein</b> . You need an inflected form of the present tense here. <i>See help on: <u>verb, sein</u></i>	
Learner Attempt 4:	Anne ist aus Amerika.	-
System Response 4:	Prima!	

*Simulated by the author at <http://www.e-tutor.org>, 09/24/2009.*

In this example, there are three instances of successful learner uptake. In attempt 2, the learner successfully incorporated the lexical (L) feedback and cleared the way for the system to address syntax (S). In attempt 3, the learner successfully corrected the syntax of the statement, but failed to correct the morphological (M) error. In attempt 4, the learner applied the proper inflection to the copular verb—making the statement target-like and producing a congratulatory phrase.

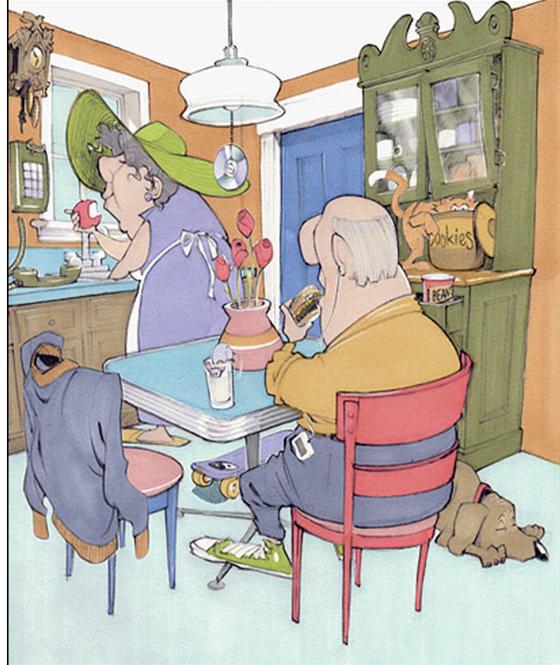
## APPENDIX A.3: EXAMPLE INTERACTION WITH SASHA

### Example 1. Sasha: Spot the Difference Task

Learner's Picture



Sasha's Picture



- Learner:** Is he reading the news?  
**Sasha:** I don't think that there is a news in this picture.  
**Learner:** \*Does the cat drinking the milk?  
**Sasha:** *Is the cat drinking the milk?*  
**Sasha:** No, it is not.  
**Learner:** Is there a jacket on the floor?  
**Sasha:** No, there is not.  
**Learner:** \*Does the men have green shoes?  
**Sasha:** *Do the men have green shoes?*  
**Sasha:** Yes, they do.  
**Learner:** Do the women have a hat?  
**Sasha:** Yes, they do.  
**Learner:** \*Is there flowers on the table?  
**Sasha:** *Are there flowers on the table?*  
**Sasha:** Yes, there are four of them.  
**Learner:** \*Is the men reading the new paper?  
**Sasha:** *Are the men reading the new paper?*  
**Sasha:** I don't think that there is a paper in this picture.  
**Learner:** What is the dog doing?  
**Sasha:** It is sleeping.  
**Learner:** \*What is the women doing?  
**Sasha:** *What are the women doing?*

**Sasha:** They are eating the apple.  
**Learner:** What is the cat doing?  
**Sasha:** I'm not sure.  
**Learner:** What is the man doing?  
**Sasha:** He is eating the sandwich.  
**Learner:** What is on the floor?  
**Sasha:** A dog.  
**Learner:** \*What the woman wearing?  
**Sasha:** *What is the woman wearing?*  
**Sasha:** A dress.  
**Learner:** \*Do the woman wearing a pant?  
**Sasha:** *Is the woman wearing pants?*  
**Sasha:** No, she is not.  
**Learner:** \*Do the man have glasses?  
**Sasha:** *Does the man have glasses?*  
**Sasha:** No, he does not.  
**Learner:** is there a cat?  
**Sasha:** Yes, there is.  
**Learner:** is the cat drinking milk?  
**Sasha:** No, it is not.  
**Learner:** Is the dog in the floor?  
**Sasha:** Yes, it is.  
**Learner:** What is the woman doing?  
**Sasha:** She is eating the apple.  
**Learner:** Where is the old man at?  
**Sasha:** He is at the table.  
**Learner:** What is the old man doing?  
**Sasha:** He is eating the sandwich.  
**Learner:** Where is the cat at?  
**Sasha:** It is in the cookie jar.  
**Learner:** \*What the dog has in his mouth?  
**Sasha:** *What does the dog have in his mouth?*  
**Sasha:** I don't think that there is a mouth in this picture.  
**Learner:** Is the old woman wearding a pant?  
**Sasha:** No, she is not.  
**Learner:** is the jaket in the floor?  
**Sasha:** I don't understand what you mean by 'jaket'  
**Learner:** Is the jacket in the floor?  
**Sasha:** No, it is not.  
**Learner:** Is a red chear there?  
**Sasha:** I don't understand what you mean by 'chear'

*Source: Learner 53. Spot the Difference Task 1, "Kitchen Scene"*

**APPENDIX B: PARTICIPANT CONSENT FORM**

**GEORGETOWN UNIVERSITY  
CONSENT TO PARTICIPATE IN RESEARCH**

**PROJECT TITLE**

**1. Implicit Corrective Feedback In Computer-Guided Interaction (IRB # 2006-169)**

**PROJECT DIRECTOR**

Ken Petersen

**PRINCIPAL INVESTIGATOR**

Ken Petersen

**TELEPHONE**

(202) 833-7522

**SPONSOR**

Georgetown University

The Georgetown University Institutional Review Board (IRB) has approved this research project. For information on your rights as a research subject, call the Institutional Review Board office at 202-687-1506.

**INTRODUCTION**

You are invited to consider participating in a research study to investigate the effects of feedback on language development among ESL learners. This form will describe the purpose and nature of the research, its possible risks and benefits, and your rights as a participant in the study. The decision to participate, or not to participate, is yours. If you decide to participate, please be sure to sign and date the last page of this form.

**WHY IS THIS RESEARCH STUDY BEING DONE?**

In this research study, we are investigating the effect that feedback has on language development among learners studying English as a second language. The study will examine the effects of feedback in both oral, face-to-face conversation as well as in a computer-mediated environment.

**HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?**

About 60 people will take part in this study.

**WHAT IS INVOLVED IN THE STUDY?**

You will be assigned to one of 3 research groups. A computer will determine your group through a process that is much like picking names out of a hat. This process is called randomization. Your chance of being in any group is 1 in 3.

The three research groups in this study are as follows:

1. **Control.** If you are assigned to this group, you will participate in three sessions that will last approximately 30-45 minutes each. In each session, you will be asked to perform three tasks: (1) A computer-based multiple-choice test in which you will need to identify grammatically correct sentences in English. (2) A computer-based writing task where you will be asked to write a series of sentences or questions. (3) An oral conversation task where you will work together with a researcher to complete a story. You will be asked to fill out a brief questionnaire at the beginning of the first session.

2. **Oral Interaction.** If you are assigned to this group, you will participate in the three sessions described above in 'Control' as well as five 30-minute sessions in which you will interact with a researcher to complete a series of language tasks (e.g., picture description, story completion, 20 questions).

2. **Computer Interaction.** If you are assigned to this group, you will participate in the three sessions described above in 'Control' as well as five 30-minute sessions in which you will engage in a series of language tasks on the computer (e.g., picture description, story completion, 20 questions). In each of these tasks, you will be asked to type in your responses by a 'virtual' conversation partner.

### HOW LONG WILL I BE IN THE STUDY?

We expect that you will be in the study for one and a half to two weeks. You will be asked to report back for a follow-up session two weeks after the last regular session. The schedule for the experiment is as follows:

---

**a. Experimental Schedule**

---

	Session 1	Sessions 2-6	Session 7	Session 8
Activities	Bio-data questionnaire	Treatment Tasks	Post-test tasks	Delayed post-test tasks (two weeks after post-test)
	Pre-test tasks			
Duration	1 hour	30 minutes	30 minutes	30 minutes
Participants	All	Groups 2 & 3 only	All	All

---

The investigators may stop the study or take you out of the study at any time they judge it is in your best interest (e.g., if you do not comply with the study plan) or for a variety of other reasons. They can do this without your consent.

You can stop participating at any time. However, if you decide to stop participating in the study, we ask you to talk to the researcher first.

### WHAT ARE THE RISKS OF THE STUDY?

There are no known risks entailed by your participation in this study.

### **ARE THERE BENEFITS TO TAKING PART IN THE STUDY?**

It is reasonable to expect that, by participating in this study, you will receive structured language practice that may benefit your knowledge of the English language.. However, we cannot guarantee that you will personally experience benefits from participating in this study, though others may benefit in the future from the information we obtain in this study. In addition, we will not provide you with either a qualitative or quantitative assessment of your language performance and you will not receive any formal credit towards your language study by participating.

### **WHO CAN PARTICIPATE IN THE STUDY?**

This study is designed for novice to intermediate adult learners of English.

### **WHAT ABOUT CONFIDENTIALITY?**

Your name will not be used when data from this study are published.

Every effort will be made to keep your research records and other personal information confidential. However, we cannot guarantee absolute confidentiality.

Individuals from the Georgetown University IRB and other Georgetown University offices may look at records related to this study, both to assure quality control and to analyze data. Your name and any material that could identify you will remain confidential except as may be required by law.

We will take the following steps to keep information about you confidential, and to protect it from unauthorized disclosure, tampering, or damage: None of your data (either personal or performance-related) will be linked to your name. Instead, you will be assigned a random ID number at the beginning of the study that we will use to identify you throughout the collection, analysis and reporting of the data.

### **WILL I BE PAID FOR PARTICIPATING?**

You will be paid a modest honorarium (\$5 per session) for your participation in this study. If you are assigned to the control group, you will receive a total of \$15 for attending the 3 sessions. If you are assigned to one of the experimental groups, you will receive a total of \$40 for attending the eight sessions. Following the completion of each session you attend, you will receive your \$5 in cash.

### **WHAT ARE MY RIGHTS AS A RESEARCH PARTICIPANT?**

Participation in this study is entirely voluntary at all times. You have the right not to participate at all or to leave the study at any time. Deciding not to participate or choosing to leave the study will not result in any penalty or loss of benefits to which you are entitled, and it will not harm your relationship with Georgetown University or any of its employees.



## **APPENDIX C: RESEARCHER MANUAL**

### **TABLE OF CONTENTS**

#### **Introduction**

1. Overview of the Study
2. List of Materials
3. Student Registration

#### **Administering Assessment Tasks**

4. Overview
5. Step-by-Step Instructions

#### **Administering Treatment Tasks**

6. Overview of Tasks
7. Giving Grammatical Feedback
8. Answering Content Questions
9. Task Management
10. Step-by-Step Instructions
11. Lab Setup Diagrams

## STUDY OVERVIEW

### *Purpose of the Study*

The overall purpose of this study is to examine the effects of intensive recasts (a form of feedback) on intermediate ESL students' ability to produce questions in English. In the design of the study, there are 3 groups of students: (1) Face-to-Face, (2) Computer-Guided and (3) Control. Students in all three groups will be given a battery of assessment tasks (as a Pre-test and Post-test) to determine their Developmental Stage in regards to English question formation. Students in the two experimental groups undergo three treatment sessions in which they have to ask countless questions about pictures they are shown. Throughout the tasks, their grammatical errors are corrected, in the form of recasts, either by a researcher or a computer program. The pedagogical effectiveness of the recasts in the two experimental modalities is then measured by developmental gains from pre-test to post-test in the two experimental groups. The purpose of the control group is to rule out testing effects in explaining any apparent language gain found in either experimental group. The purpose of the delayed post-test (given to the 2 experimental groups) is to test whether any gains demonstrated in the immediate post-test hold up over time.

### *Registering Students*

All students begin the study by adding their name and email address to a sign-up sheet that is distributed throughout the ESL classes. Each student on the list is then randomly assigned to a group and given a student id number. Students are then given a paper consent form to read and sign and receive an email that explains their role in the study, points them to the online bio-data questionnaire and tells them their schedule.

### *Data Collection Schedule*

A total of approximately 60 students will be participating in this study. Each student is randomly assigned either to one of the two experimental groups or to the control group. The number of sessions required for each student varies according to which group she is randomly assigned (as shown in Table 1.1).

**Table 1: Experimental Schedule**

Experimental Group	Pre-Test	Sess. 1	Sess. 2	Sess. 3	Post-Test	Delayed Post-Test
Face-to-Face (n=20)	30 min.	30 min.	30 min.	30 min.	30 min.	30 min.
Computer-Guided (n=20)	30 min.	30 min.	30 min.	30 min.	30 min.	30 min.
Control (n=20)	30 min.				30 min.	

### ***Assessment Tasks***

Before and after the three experimental sessions, the students in all groups submit to a variety of written and oral assessment tasks that test their ability to form questions. Each 30-minute testing session requires approximately 15 minutes on a computer and approximately 15 minutes interacting with a researcher.

### ***Face-to-Face Tasks***

Tasks in the Face-to-Face group are carried out at a table in one-on-one student/researcher pairs. This means that each of the 60 sessions in this group (3 sessions x 20 students) has to be scheduled with a native speaker of English.

### ***Computer-Guided Tasks***

Tasks in the Computer-Guided group are carried out by placing each student at a computer to interact with a virtual researcher. Since no sustained, one-on-one, human interaction is required in this group, the students can be scheduled to complete the experimental session in one large group, provided that there are computers available.

## **MATERIALS**

<b>Contents of Researcher Packet:</b>
<ol style="list-style-type: none"><li>1. Researcher Manual</li><li>2. Researcher Task Booklet (Answer Keys)</li><li>3. Researcher Data Binder</li><li>4. Student Assessment Task Booklet (Picture Sets)</li><li>5. Student Treatment Task Booklet (Picture Sets)</li><li>6. CD containing PDFs of all files above</li></ol>

**Contents of Researcher Common Resource Box:**

1. Stopwatches
2. Dictaphones
3. Guess Access Chips (YN/WH)
4. Difference Count Chips (1-10)
5. Sequence Count Chips (1-6)
6. Cardboard Page Covers
7. Spare Batteries
8. Sticky Tack
9. No. 2 Pencils
10. Pencil Sharpener
11. Blank CDs

**STUDENT REGISTRATION*****1. Announcement of the Study***

In the first 5-10 minutes of a chosen class period in each participating ESL class, the teacher reads a short announcement inviting the students to participate in the study. The signup list is then passed around the class for interested students to sign up.

***2. ID/Group Assignment***

Once the signup sheets are complete, each student on the list is randomly assigned to one of the three groups and given a student id.

**Random Group Assignment**

For each student, a die is rolled with three values: Ctrl, F2F, CG. Whatever value is facing up, determines the group assignment for each student on the list.

**Computer Login/Password**

The computer login and password for each student adheres to the following convention:

[ first initial + last initial + underscore + studentID ]

Name: John Smith, studentID: 11  
Login: js\_11  
Password: 11

Name: Mary Sue Jones, studentID: 21  
Login: mj\_21  
Password: 21

### **3. *Introductory Email***

Once ID/Group Assignment is complete, each student is sent an email informing them of their ID number, group assignment, their schedule and information about receiving payment. The email will also include a link to a bio-data questionnaire that the students are required to complete before arriving at the first session.

### **4. *Consent Form***

In the first 5-10 minutes of a chosen class period, the teacher distributes the consent form for the study and asks the students to read, sign and return the forms. If students have any questions or concerns about their participation in the study, they should contact Ken Petersen (kpeter@actr.org).

## **OVERVIEW OF ASSESSMENT TASKS**

There are three testing sessions in this study:

1. Pre-Test (Prior to first treatment session)
2. Immediate Post-Test (Immediately following last treatment session)
3. Delayed Post-Test (Many days after last treatment session)

Each testing session will last approximately 30 minutes (15 minutes oral, 15 minutes written) and will consist of the following assessment tasks:

### **1. *Oral Imitation Test***

***Mode: Face-to-Face***

***Time Allotted: 8 minutes***

Students are presented with a series of 10 individual pictures by the researcher. With each picture, there is a question asked aloud by the researcher that the student has to answer. After the student has given an answer, she must repeat the question she heard back to the researcher.

## 2. Oral Role Play Task

*Mode: Face-to-Face*

*Time Allotted: 7 minutes*

Students are presented with a brief scenario in which they are required to play a role and ask the researcher questions. For example, the students might be asked to imagine themselves as detectives and listen to the background details of a made-up crime. They are then asked to produce as many questions as they can that might help solve the crime.

## 3. Written Picture Dialogues

*Mode: Computer-based*

*Time Allotted: 8 minutes*

Students are presented with a series of 10 pictures by the computer. Their task is to complete a question-answer dialogue for each picture they see. With each picture, there is a text field for the student to type in a question that would be appropriate for the answer that is written out below.

## 4. Written Free Response

*Mode: Computer-based*

*Time Allotted: 7 minutes*

Students are given an open-ended, free response task to read a short passage and then ask 10 questions about what they have read.

### *Scheduling*

The most important thing to remember when scheduling the assessment tasks for students is that each 30-minute testing session consists of two separate 15-minute sessions. One is administered in an oral, face-to-face modality, the other is handled by a computer. Each researcher can test 4 students per hour as is shown in Table 2.

*Table 2: Assessment Task Schedule: Sample Hour*

	Face-to-Face		Computer	
	Start	End	Start	End
Student 1	12:00	12:15	12:15	12:30
Student 2	12:15	12:30	12:00	12:15
Student 3	12:30	12:45	12:45	1:00
Student 4	12:45	1:00	12:30	12:45

Notice that each 30-minute block of time is filled by testing 2 students simultaneously.

### *Managing the Schedule*

The procedure for completing the schedule shown in Table 4.1 is as follows:

1. Make sure that the "Testing in Progress" sign is hung on the lab door.
2. At 12:00, students 1 & 2 report at the door of the lab for 30 minutes of testing.
3. If a session is in progress when they arrive, they must wait outside the door until the previous pair of students is dismissed.
4. Both students enter their names, the date and time on the Student Sign-In Sheet.
5. Student 1 is lead to the research table and asked to sit and wait.
6. Student 2 is lead to a computer and is logged in to the online assessment tasks.
7. Once Student 2 is underway on the computer, the researcher administers the oral assessment tasks to Student 1.
8. When the fifteen-minute task window has expired, the students switch seats and receive the second half of the assessment tasks.
9. At the end of the 30 minutes, students are released and reminded of the time of their next appointment.
10. Students 3 & 4 are admitted to the lab.

## **ASSESSMENT TASKS: STEP-BY-STEP PROTOCOLS**

<b>Before Arriving For Data Collection</b>
--

1. For each scheduled student, open the Researcher Data Binder to a blank 'Assessment Task' data sheet.
2. Complete the first five fields on each sheet:
  - ❑ **StudentID:** Look at your schedule for the day.
  - ❑ **Time of Written Test:** The time (EST) that the written session began.
  - ❑ **Time of Oral Test:** The time (EST) that the oral session began.
  - ❑ **Session Type:** Pre, Post or Delayed.
  - ❑ **DictaphoneID:** Look at the sticker on your dictaphone.
3. Make sure that your dictaphone has been backed-up to the computer and wiped clean.

## ***Oral Imitation Task***

### ***Required Materials***

1. Researcher Data Binder à 'Assessment Task' data sheet
2. Student Assessment Task Booklet
3. Stopwatch
4. Dictaphone
5. Small cardboard cover
6. Large cardboard cover
7. 3 No. 2 pencils

### ***Instructions***

1. Open the Researcher Data Binder to the 'Assessment Task' data sheet you have prepared for this 15-minute testing session.
2. Fill in the Start Time field for the *Oral Imitation Test* with the number displayed on the dictaphone (e.g., 01:23:46 ).
3. Open the Student Assessment Task Booklet to the initial picture of the set designated for this testing session and place the small cardboard cover in the booklet as a bookmark.
4. Close the booklet and place it in front of the student.
5. Open the Researcher Task Booklet to the appropriate key, cover it with the large cardboard cover and place it on the table in front of you (next to your data binder).
6. Ask the student to sit down at the table.
7. Greet the student, then ask: *Are you ready to begin?*
8. Read the following instructions:

*We are going to do two different tasks today, totaling 10 minutes. The first task is a question imitation task. I am going to show you a series of pictures in this book. I will ask you a question about each picture, like, 'What is the man doing?'. You will then answer my question, like, 'He is playing football'. I will then ask you to repeat my question back to me, where you would say, 'What is the man doing?'. We will have 5 minutes for this task. Are you ready to begin?*

9. Press record on the dictaphone and identify the session aloud, directly into the dictaphone:  
*e.g., 'Learner 5...Pre-Test... Imitation'*
10. Place the dictaphone on the table.
11. Start the timer (countdown from 8 minutes).
12. Open the Student Assessment Task Booklet to the picture you have bookmarked.
13. Read question 1 from the key in your Researcher Task Booklet.

14. Allow the student to answer the question.
15. Say: *Now, please repeat the question I asked you.*
16. Turn the page to the next picture and repeat the steps above until either time runs out or you complete all ten pictures in the set.
17. When time has run out:
  1. Announce that the task is complete.
  2. Stop the dictaphone.
  3. Fill in the End Time field for the *Oral Imitation Test* with the number displayed on the dictaphone.
  4. Fill in the # of Questions field for the *Oral Imitation Test* with the total number of questions asked by the student in the task.
  5. Move on to Free Response Test.

### ***Oral Role Play Tasks***

#### ***Required Materials***

1. Researcher Data Binder à "Assessment Task' data sheet
2. Stopwatch
3. Dictaphone
4. 3 No. 2 pencils
- 5.

#### ***Instructions***

1. Open the Researcher Data Binder to the 'Assessment Task' data sheet you have prepared for this 15-minute testing session.
2. Fill in the Start Time field for the *Oral Role Play Test* with the number displayed on the dictaphone (e.g., 01:23:46 ).
3. Open the Researcher Task Booklet to the Oral Role Play Test for this testing session and place it on the table in front of you.
4. Read the following instructions:

*In this task, we are going to do a little role-play where you will need to ask me as many questions as you can. I will describe the role you will play and suggest some questions that you might ask and then you will begin asking me questions.*

*We will have 5 minutes for this task. Are you ready to begin?*

5. Press record on the dictaphone and identify the session aloud, directly into the dictaphone: e.g., 'Learner 5...Pre-Test... Role Play'
6. Place the dictaphone on the table.

7. Start the timer (countdown from 7 minutes).
8. Read the scenario for this task from your Researcher Task Booklet.
9. Answer the student's questions and assist her in asking questions. Do not model questions, but provide suggestions like, 'Try asking me about the old man.'
10. When time has run out:
  1. Announce that the task is complete.
  2. Stop the dictaphone.
  3. Fill in the End Time field for the *Oral Role Play Test* with the number displayed on the dictaphone.
  4. Fill in the # of Questions field for the *Oral Role Play Test* with the total number of questions asked by the student in the task.
  5. Announce to both students that the session is finished.
  6. Confirm the time and date of the next appointment with each student.
  7. Release the students.

### ***Administering Written Tests***

The administration of the written tests is performed by the computer. The researcher should do the following to get the student started.

1. Open an Internet browser and go to the online assessment page.
2. Log in the student: [ first initial + last initial + underscore + studentID ]
3. Seat the student at the computer.
4. Tell the student that they will have 15 minutes to take two tests.

Instruct the student to click "Begin S

## **OVERVIEW OF TREATMENT TASKS**

Each learner in the Face-to-Face group participates in three 30-minute treatment sessions. Each session consists of three distinct tasks, with duration of 10 minutes each. The three tasks are as follows:

### **Picture Matching**

#### *a. Overview*

In the picture matching tasks, the learner is presented with a series of picture sets containing four pictures. With each set of pictures presented to the learner, one of the pictures has been selected as the target picture that the researcher is looking at. The task of the learner is to ask questions in order to determine what the target picture is. Before she can guess which picture is the target picture, she needs to ask at least one Yes/No question and one WH- question. Once she has met this requirement, she may point to the picture and ask, "Is it this picture?". If the correct picture has been identified, the researcher advances the learner to the next set of pictures.

### *b. Guess Access Chips*

At the beginning of each task session, the researcher explains that, in order to guess which picture is the target picture, the learner must first ask at least one Yes/No question and one WH- question. The researcher shows the learner a pair of poker chips (YN/WH) that are both black on one side and white on the other and explains that the learner may not try to guess the target picture until both chips have been turned over so that the white sides of the chips are facing up. At the onset of the task, the chips are then laid in the center of the table with the black sides face-up and the researcher prompts the student to start asking questions. Once the learner has asked one Y/N and one WH question, the corresponding chip is turned over to expose the white side. When both chips are white-side-up, the learner is eligible to guess which picture is the target. A correct guess advances the task to the next picture set. An incorrect guess resets the guess access chips to black.

### *c. Advancing Through Picture Sets: Matching*

Each learner begins the first treatment session at picture matching set 1. Within the 10-minute time period, the learner advances through the picture matching sets by correctly guessing the target picture. When the 10-minute time limit has been exceeded, the researcher notes the current picture matching set number on the task management sheet and proceeds to the Spot the Differences task. Note: When each learner begins the matching portion of subsequent sessions, the researcher begins by presenting the learner with the last unresolved picture set for each learner.

## **Spot the Differences**

### *a. Overview*

In the spot the differences tasks, the learner is presented with a single picture containing a complex scene. The learner is told that the researcher is looking at a very similar picture, containing a number of distinct differences from the picture being viewed by the learner. The task of the learner is to ask questions about the picture she sees in order to determine what the specific differences between the two pictures are. The learner receives a point for each difference she identifies. The identification of a difference is determined by the researcher by comparing the two pictures.

For example, if the learner's picture contains a parrot in a cage and the researcher's picture contains a parrot sitting on a man's shoulder, and the learner asks either: "Is there a parrot in a cage?" or "Where is the parrot?" or "What is in the cage?", then a difference has been found.

### *b. Difference Count Chips*

At the beginning of each task session, the researcher places the chip marked '0' in the center of the table. Each time a difference is found, the researcher incrementally updates the *Difference Count Chip*. After the first difference is found, the researcher replaces the '0' chip with the chip marked '1'. When the second difference is found, the '1' chip is replaced with the chip marked '2' and so on until the learner has identified 10 differences.

### *c. Advancing Through Picture Sets*

There are three Spot the Differences picture sets, numbered 1,2 & 3, which correspond with the number of the session for the current learner.

## **Picture Sequencing**

### *a. Overview*

In the Picture Sequencing Tasks, the learner is presented with a series of six pictures that represent a short story, displayed in scrambled order. The task of the learner is to ask questions to determine the correct order of the six pictures. She does this by identifying one picture at a time – beginning with the first picture and moving on to the second, third, etc. Once the learner has identified the correct order of the sequenced story, the researcher advances the task to the next sequence.

### *b. Guess Access Chips*

As with the *Picture Matching* task, the learner is required to ask at least one Yes/No question and one WH- question before guessing which picture is next in the sequence. See *Picture Matching*, above, for details on managing guess access chips.

### *b. Indicating Sequencing Count*

Once the learner is eligible to guess the next picture in the sequence,, she may point to the picture and ask, "Is it this picture 1?". If the correct picture has been identified, the researcher places the adhesive sequence chip marked '1' at the bottom right of the picture and asks the learner to move on to identifying the second picture in the sequence. If the guess is incorrect, the guess access chips are reset to black and the learner needs to ask at least one more Y/N and one more WH question before being allowed to venture another guess.

### *c. Advancing Through Picture Sets*

Once the correct order of a sequencing set is determined, the researcher advances the task to the next sequencing picture set. If the 10-minute time limit expires before the correct sequence is found, the learner will begin on the current sequence at the outset of the next session. In this case, when the task resumes in the subsequent session, the researcher will replace the sequencing chips on the pictures that have already been identified.

## PROVIDING FEEDBACK

Throughout the Face-to-Face treatment tasks, the researcher carefully monitors the student questions and provides a recast for every grammatically ill-formed question. Please adhere to the following guidelines when providing recasts to the students throughout the tasks.

### *What to correct*

Only questions that contain grammatical (syntactic and morphological) errors are responded to with a recast. In Table 7. 1, we see a student question containing a simple syntactic error. The researcher repeats the question back to the student, with the word order corrected.

**Table 7. 1: Sample Recast: Syntactic Error**

<i>Student:</i>	'What the men are doing?'
<i>Researcher:</i>	'What are the men doing?'

In Table 7. 2, we see a student question containing a simple morphological error. The researcher repeats the question back to the student with the verb inflection corrected.

**Table 7. 2: Sample Recast: Morphological Error**

<i>Student:</i>	'What is the men doing?'
<i>Researcher:</i>	'What are the men doing?'

In Table 7. 3, we see a student question containing multiple morpho-syntactic errors. The researcher repeats the question back to the student with all errors corrected.

**Table 7. 3: Sample Recast: Multiple Errors**

<i>Student:</i>	'What the men they is do?'
<i>Researcher:</i>	'What are the men doing?'

As Table 7. 3 illustrates, a recast can be used to address multiple errors simultaneously. The important thing to remember is that recasts should only be provided in response to grammatical errors. The next section presents a number of ways that recasts should not be used.

### *What NOT to correct*

#### **a. Lexical Errors**

If the student produces a grammatically well-formed question, but one or more of the words used in the question do not correspond semantically with the picture in question, do not provide a recast.

#### **b. Conceptual Errors**

If the student produces a grammatically well-formed question, but seems to have missed the point of the picture in question, do not provide a recast—perhaps offer an explanation if they continue to struggle.

#### **c. Fragments**

Do not fill in any missing words or structure for fragmented questions. Wait until the student has definitively completed the question. If the question is fragmented or severely mangled, ask the student to try again.

### *Timing of Recasts*

Recasts are always given IMMEDIATELY following an ill formed question asked by the student. If an error is detected, first provide the recast, then proceed to answer the question or advance the task.

There should be no explicit pause before the recast, but please wait until the student's turn is fully complete.

A brief (1-2 seconds) pause should be provided between the recast and the answer to the semantic content of the question.

### *Prosodics of Recasts*

All recasts should end in a rising inflection appropriate for the question asked by the student. This rising intonation is meant to indicate that the statement is an interrogative. It is not intended as a means of drawing attention to the fact that the learner made an error. In fact, no special pitch or stress should be used to indicate the corrections that have been made in the student's question. Similarly, no facial

expressions or other paralinguistic cues should ever be used to signal that an error has occurred.

## **ANSWERING STUDENT QUESTIONS**

Throughout the Face-to-Face treatment tasks, the researcher provides answers to the questions that the students ask about the pictures that they see. Please adhere to the following guidelines when providing answers to the students' questions throughout the tasks.

### ***What to answer***

Only answers that are clearly evident from looking at the pictures should be given to the students. Only one discrete answer should be given for any question asked by the student. Once the answer is given, the student is prompted to ask another question.

#### *Answering Broad, Open-Ended Questions*

Students are likely to ask many open-ended questions that could be answered with a variety of responses. For example, the student may ask, "What is the man doing?". The man in question may be doing a number of different things. The researcher should simply pick the most obvious thing the man is doing, answer the student and prompt for a new question.

***Table 8.1: Handling open-ended questions***

<i>Student:</i>	'What is the man doing?'
<i>Researcher:</i>	'He is sitting in a chair. Ok, ask me another.'

### ***What NOT to answer***

#### ***Conjecture Questions***

Answers that require any degree of conjecture should be avoided. For example, the researcher should not answer questions about what the people in the pictures are thinking, why something is happening, what the researcher thinks/feels about something, etc.

**Table 8.2: Cutting off conjecture questions**

<i>Student:</i>	'Why is the boy happy?'
<i>Researcher:</i>	'I can't really tell you why something is happening. I just see the picture.'
<i>Student:</i>	'What does the boy want to do?'
<i>Researcher:</i>	'I don't know what he is thinking. I only see what he is doing.'
<i>Student:</i>	'What do you think the man wants?'
<i>Researcher:</i>	'I don't know. Please ask questions about what is happening in the picture, not about what I think.'

### ***Illegitimate Questions***

Some overly vague questions should not be considered legitimate questions and do not merit an answer.

**Table 8.3: Cutting off illegitimate questions**

<i>Student:</i>	'What is happening in this picture?'
<i>Researcher:</i>	'A whole lot of things. Try asking me a specific question.'
<i>Student:</i>	'What else is he doing?'
<i>Researcher:</i>	'You need to ask me a new question.'

### *Repeat Questions*

If you find a student asking the same question over and over again in the course of the tasks, point out to the student that she needs to stop asking this question and try asking some other questions.

## **TASK MANAGEMENT**

Throughout the treatment tasks, the researcher has the following roles and duties to fulfill:

### **1. Maintain Researcher Data Binder (Primary Data!)**

Included in the researcher packet is a binder with numerous sets of blank data sheets that will be filled out throughout the administration of the tasks. At the end of data collection, the data binder must be submitted in the following format:

1. All data fields should be completely filled in with a No. 2 pencil for each task, for each student.
2. Irregularities should be thoroughly documented.
3. The three task sheets (Match, Seq, Spot the Diff) for each 30-minute student session should be paper clipped together in the binder.

### **2. Audio Recording/Archiving (Primary Data!)**

The conversational interaction for each task session must be recorded on one of the dictaphones provided. Please adhere to the following daily protocol for recording and archiving the audio transcripts.

#### *Pre-Session – Checkout Dictaphone*

11. At the beginning of any day of data collection, select a dictaphone to use throughout the day.
12. Sign the dictaphone out on the Dictaphone Log located in the Researcher Materials Box.
13. Fill in the dictaphone ID number on all data sheets that you anticipate using on the current day.
- 14. Make sure that there is no data stored on the selected dictaphone and that the counter is reset to 0:00. If there is lingering data on a dictaphone, either contact the previous user (check the Dictaphone Log) or archive the data yourself on the Audio Archive Workstation. Do not erase data that you are not POSITIVE has been archived!**

### *In Session – Record Each Task & Note Start/End Times*

- ❑ Fill-in the start time on the appropriate task data sheet.
- ❑ Press record at the beginning of each 10-minute task.
- ❑ Provide verbal header for each task.
- ❑ Press stop at the end of each 10-minute task.
- ❑ Fill-in the end time on the appropriate task data sheet.

### **3. Set Up the Research Area**

*See: Researcher Manual [11]: Physical Task Layouts.*

### **4. Procedural Discourse Management**

The researcher controls the procedural discourse of the data collection in the following ways:

#### *Initiate/Closeout Tasks*

The researcher greets students, invites them to the table and makes sure that they are ready to begin the tasks and informs students when the task is finished.

#### *Task Advancing Questions*

In order to advance through the Matching and Sequencing tasks, the student asks procedural questions that potentially move the task forward (e.g., "Is this the picture?"). Before answering, the researcher makes sure that the Guess Access Chips are both showing white. If not, he informs the student that she needs to ask either a Yes/No or WH question before being allowed to make a guess. If the guess is both licensed and correct, the task advances.

### **5. Task Advancement**

#### *Matching*

When the student guesses which picture in the student set matches the key in the Researcher Task Booklet, the task advances to the next picture set.

#### *Spot the Differences*

There are no Guess Access Chips or explicit task advancing questions in these tasks. Instead, the Difference Count Chip is updated whenever the student asks a question that reveals a difference between the two pictures. For example, Picture A (student) portrays a mouse eating cookies and Picture B (researcher) portrays him drinking

milk. If the student asks, "Is the mouse eating cookies?", the researcher says, "No, he's not". I think you found a/another difference!' and updates the chip before prompting for another question.

### *Sequencing*

When the student guesses which picture in the student set matches the target key in the sequence, a Sequencing Chip is placed on the selected picture in the Student Task Booklet. Once all of the pictures in a sequence have been identified, the task advances to the next picture set.

## 6. Conversational Discourse Management

In all of the tasks, the conversation about the pictures should loop continuously through the following sequence. The researcher creates the context for this sequence by prompting the student to ask a question.

**Table 9.1: Prototypical Sequence of Conversational Turns**

<i>Student:</i>	Asks question.	'What the boy he is eating?'
<i>Researcher:</i>	Provides recast*	'What is the boy eating?'
<i>Researcher:</i>	Answers question	'He is eating a sandwich.'
<i>Researcher:</i>	Prompt for New Question	'OK, now ask me another question.'

(\* if error detected)

## TREATMENT TASKS: STEP-BY-STEP PROTOCOLS

### Before Arriving For Data Collection

1. For each scheduled student, open the Researcher Data Binder to a blank set of data sheets (Matching, Sequencing and Spot the Difference).
2. Complete the first five fields on each sheet:
  - ❑ **StudentID:** Look at your schedule for the day.
  - ❑ **Date:** The date of the data collection session
  - ❑ **Session #:** The number of the session for this learner (1, 2 or 3)
  - ❑ **DictaphoneID:** Look at the sticker on your dictaphone
  - ❑ **Initial Picture Set:** For session 1, this value is set at 1. For subsequent sessions, look at the value of Final Picture Set on the data sheet from the previous session for this learner and copy the value here.
  - ❑ **Sequence Index (Sequencing Only):** For session 1, this value is set at 1. For subsequent sessions, look at the value of Sequence Index on the data sheet from the previous session for this learner and copy the value here.

3. Make sure that your dictaphone has been backed-up to the computer and wiped clean.

## Picture Matching Task

### *Required Materials*

1. Researcher Data Binder → 'Picture Matching Task' data sheet
2. Student Treatment Task Booklet
3. Guess Access Chip [YN/WH]
4. Stopwatch
5. Dictaphone
6. Small cardboard cover
7. Large cardboard cover
8. 3 No. 2 pencils

### *Initiating a Task Session for a Student*

1. Place the Matching Task Data Sheet for the current learner on the table in front of you.
2. Fill in the Start Time field on the Matching Task Data Sheet with the number displayed on the dictaphone (e.g., 01:23:46 ).
3. Open the Student Treatment Task Booklet to the initial matching set and place the small cardboard cover in the booklet as a bookmark.
4. Close the booklet and place it in front of the student.
5. Open the Researcher Task Booklet to the appropriate key, cover it with a large cardboard cover and place it on the table in front of you (next to your data sheet).
6. Place the black and white Guess Access Chips in the middle of the table, black side up.
7. Ask the student to sit down at the table.
8. Greet the student, then ask: *Are you ready to begin?*
9. Read the following instructions:

*We are going to do three different tasks today, each lasting ten minutes. The first task is a picture matching task. You will see set of four pictures in the booklet in front of you. I will be looking at one of those pictures in my booklet. Your job is to ask me questions about the pictures you see to guess which one I am looking at.*

*To guess a picture, just point to it and ask "Is this the picture?".*

[Point to the Guess Access Chips]

These chips tell you when you are allowed to guess which picture I am looking at. You may only guess when the white sides of both chips are facing up.

*[Show the white sides of the chips and then flip back to black]*

*To make the chips white, you need to ask at least one Yes or No question and one question beginning with Who, What, Where, etc.*

*I will now start by showing you the first set of four pictures and you may begin asking me questions about the pictures to help you guess which picture I am looking at.*

*Are you ready to begin the task?*

*[Answer any lingering questions]*

OK. We will have 10 minutes on this task.

10. Press record on the dictaphone and identify the session as shown in this example:

*'Learner 5...Session 1... Matching'*

11. Place the dictaphone on the table.

12. Open the Student Task Booklet to the bookmarked set of pictures.

13. Start the timer.

14. Say: *Begin asking me questions.*

### ***Managing the Conversational Task***

10. Provide recasts in response to grammatical errors. (See: [7]: *Giving Feedback*)

11. Provide answers to the content of student questions. (See: [8]: *Answering Content Questions*)

12. Keep Guess Access Chips Updated:

1. Any time the student asks a Yes/No question, if the Y/N chip is black, flip it to white.

2. Any time the student asks a WH- question, if the WH chip is black, flip it to white.

3. Following any guess (right or wrong), flip both chips to black.

4. Advance Picture Sets:

1. If the student guesses the picture correctly, go on to the next picture set.

2. If a guess is incorrect, remain on the current picture set.

3. Keep a Running total of Student Questions and Recasts Provided

Use the spaces provided on the data sheet to tick off how many questions

the student asked during the 10 minute task and how many times you provided a recast.

### ***Closing Out the Task***

1. When the time has expired, say: *The task is now complete.*
2. Stop the dictaphone.
3. Fill in the End Time field on the Matching Task Data Sheet with the number displayed on the dictaphone.
4. Tally the number of questions asked by the student and the number of recasts given and fill in the appropriate fields at the top of the Matching Task Data Sheet.
5. Fill in the Final Picture Set field on the Matching Task Data Sheet with the current picture set. If the student guessed the picture before you could advance to the next picture set, put a star next to the number. Otherwise, simply write down the number of the current picture set.
6. Note any irregularities that may have occurred during the session. (e.g., session was interrupted, student was consistently off task, etc) .

## **Spot the Difference Task**

### ***Required Materials***

1. Researcher Data Binder → 'Spot the Differences Task' data sheet
2. Student Treatment Task Booklet
3. Difference Count Chips [numbered, 0-10]
4. Stopwatch
5. Dictaphone
6. Small cardboard cover
7. Large cardboard cover
8. No. 2 pencil

### ***Initiating a Task Session for a Student***

- Place the Matching Task Data Sheet for the current learner on the table in front of you.
- Fill in the Start Time field on the Matching Task Data Sheet with the number displayed on the dictaphone (e.g., 01:23:46 ).
- Open the Student Treatment Task Booklet to the spot the difference picture set that corresponds with the session number (1, 2 or 3) and place the small cardboard cover in the booklet as a bookmark.
- Close the booklet and place it in front of the student.
- Open the Researcher Task Booklet to the spot the difference picture set

that corresponds with the session number (1,2 or 3), cover it with a large cardboard cover and place it on the table in front of you.

□ Read the following instructions:

*Our next 10-minute task is a 'Spot the Differences' task. In this task, there are two pictures—one that you can see [ place hand on Student booklet ] and one that only I can see [ place hand on Researcher booklet ]. The two pictures are very similar to each other, but there are at least ten differences between the two pictures. Your job is to ask me questions about my picture to locate those differences.*

*Each time you ask a question that points to one of the differences, you will get another point. I'm going to start you off with 0 points.*

[ Place Difference Count Chip [0] in the middle of the table. ]

*When I show you the picture in this book, just start asking me questions about the things you see in the picture and I will look to see if the same thing is happening in my picture.*

*You will only be allowed to ask five Yes/No questions. After that, you will need to use WH- question words, like, What? Where? Who? How many? I will tell you when you have used up all of your Yes/No questions.*

*Are you ready to begin the task?*

[Answer any lingering questions]

OK. We will have 10 minutes on this task.

1. Press record on the dictaphone and identify the session as shown in this example:  
*'Learner 5...Session 1... Spot Difference'*
2. Place the dictaphone on the table.
3. Open the Student Task Booklet to the bookmarked picture.
4. Start the timer.
5. Say: *OK, Begin asking me questions.*

### ***Managing the Conversational Task***

1. Provide recasts in response to grammatical errors. (See: [7]: *Giving Feedback*)
2. Provide answers to the content of student questions. (See: [8]: *Answering Content Questions*)
3. Keep Difference Count Chips Updated:

1. Any time the student asks a question that highlights a difference between the 2 pictures:
  - Answer the question
  - Say something like: *Ok, you found another one.*
  - Update the Difference Count Chip
2. Any time the student asks a question that does not highlight a difference:
  - Answer the question
  - Say something like: *That is happening in both pictures.*
4. Advance Picture Sets:
  1. If the student finds 10 differences before the time expires, move on to the next picture.
2. Keep a Running total of Student Questions and Recasts Provided.
  1. Use the spaces provided on the data sheet to tick off how many questions the student asked, how many times you provided a recast.
  2. Keep a Running total of Yes/No Questions asked.
  3. Only allow five Y/N questions per picture set. Announce to the student when the 5<sup>th</sup> Y/N question has been asked.

### ***Closing Out the Task***

1. Say: *The task is now complete.*
2. Stop the dictaphone.
3. Fill in the End Time field on the Spot the Difference Task Data Sheet with the number displayed on the dictaphone.
4. Tally the number of questions asked by the student and the number of recasts given and fill in the appropriate fields at the top of the Spot the Difference Task Data Sheet.
5. Note any irregularities that may have occurred during the session. (e.g., session was interrupted, student was consistently off task, etc) .

<b>Picture Sequencing Task</b>
--------------------------------

### ***Required Materials***

1. Researcher Data Binder → 'Picture Sequencing Task' data sheet
2. Student Treatment Task Booklet
3. Guess Access Chip [YN/WH]
4. Sequence Index Chips [numbered, 1-10]
5. Stopwatch
6. Dictaphone
7. Small cardboard cover
8. Large cardboard cover
9. No. 2 pencil

### *Initiating a Task Session for a Student*

1. Place the Picture Sequencing Task Data Sheet for the current learner on the table in front of you.
2. Fill in the Start Time field on the Picture Sequencing Task Data Sheet with the number displayed on the dictaphone (e.g., 01:23:46 ).
3. Open the Student Treatment Task Booklet to the initial sequencing set and place the small cardboard cover in the booklet as a bookmark. If the picture set has been started in a previous session, place the number of sequence index chips on the page indicated in the Sequence Index field at the top of the data sheet.
4. Close the booklet and place it in front of the student.
5. Open the Researcher Task Booklet to the appropriate key, cover it with a large cardboard cover and place it on the table in front of you (next to your data sheet).
6. Place the black and white Guess Access Chips in the middle of the table, black side up.
7. Greet the student, then ask: *Are you ready to begin?*
8. Read the following instructions:

*The last task for the day is a picture sequencing task. In this task, you will see a series of six pictures that represent a short story. The order of the pictures, however, is scrambled. Your job is to figure out the correct sequence of the pictures by asking me questions about the individual pictures. You will begin by trying to identify the first picture in the story. Once you have found the first picture, you will move on to the second, third picture, etc. When you have identified the entire sequence, we will move on to a new story.*

*To guess the next picture in the sequence, just point to it and ask "Is this picture 1?".*

*As in the matching task, you may not guess the next picture until you have asked at least one Yes/No and one WH question, as indicated by these chips.*

[Point to the Guess Access Chips]

*I will now start by showing you the pictures that make up the first story. Begin by asking me questions about what you think the first picture is.*

*Are you ready to begin the task?*

[Answer any lingering questions]

OK. We will have 10 minutes on this task.

9. Press record on the dictaphone and identify the session as shown in this example:

*'Learner 5...Session 1... Sequencing'*

10. Place the dictaphone on the table.
11. Open the Student Task Booklet to the bookmarked set of pictures.
12. Start the timer.
13. Say: *Begin trying to find the first picture.*

### ***Managing the Conversational Task***

1. Provide recasts in response to grammatical errors. (See: [7]: *Giving Feedback*)
2. Provide answers to the content of student questions. (See: [8]: *Answering Content Questions*)
3. Keep Guess Access Chips Updated:
  - a. Any time the student asks a Yes/No question, if the Y/N chip is black, flip it to white.
  - b. Any time the student asks a WH- question, if the WH chip is black, flip it to white.
  - c. Following any guess (right or wrong), flip both chips to black.
4. Keep Sequence Index Chips Updated:
  - a. When the student correctly identifies the next picture in the sequence, stick the appropriate adhesive sequence index chip in the bottom-right hand corner of the selected picture.
5. Advance Picture Sets:
  - a. Once the student guesses the first five pictures in the sequence correctly (picture six requires no additional guess. as it is the only picture left), go on to the next picture set. Note that picture number six will always be answered by default.
6. Keep a Running total of Student Questions and Recasts Provided
  - a. Use the spaces provided on the data sheet to tick off how many questions the student asked during the 10 minute task and how many times you provided a recast.

### ***Closing Out the Task***

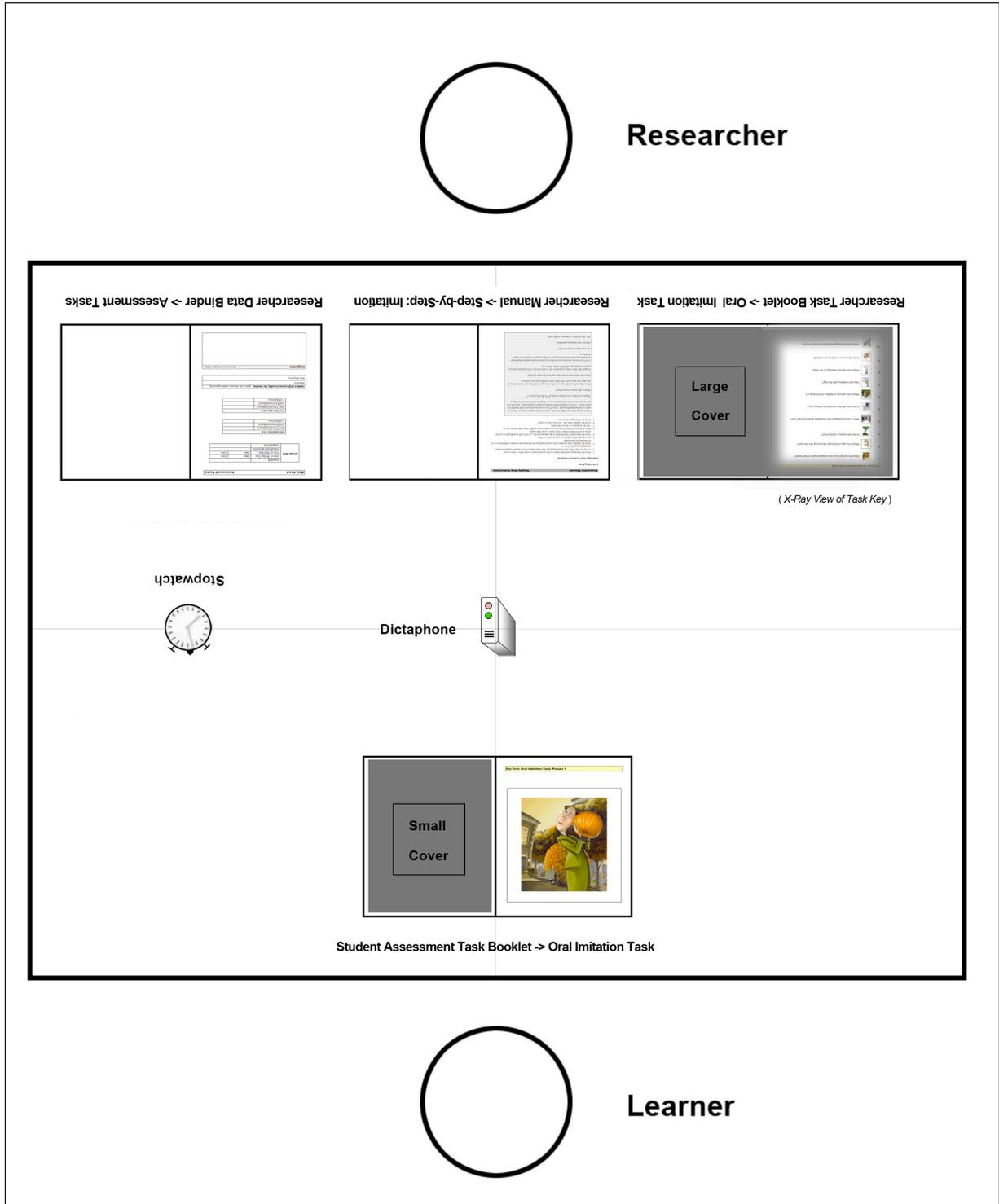
1. Say: *The task is now complete.*
2. Stop the dictaphone.
3. Fill in the End Time field on the Sequencing Task Data Sheet with the number displayed on the dictaphone.
4. Tally the number of questions asked by the student and the number of recasts given and fill in the appropriate fields at the top of the Sequencing Task Data Sheet.
5. Fill in the Final Picture Set field on the Sequencing Task Data Sheet with the current picture set.
6. Fill in the Sequence Index field on the Sequencing Task Data Sheet with

the number of sequence index chips displayed.

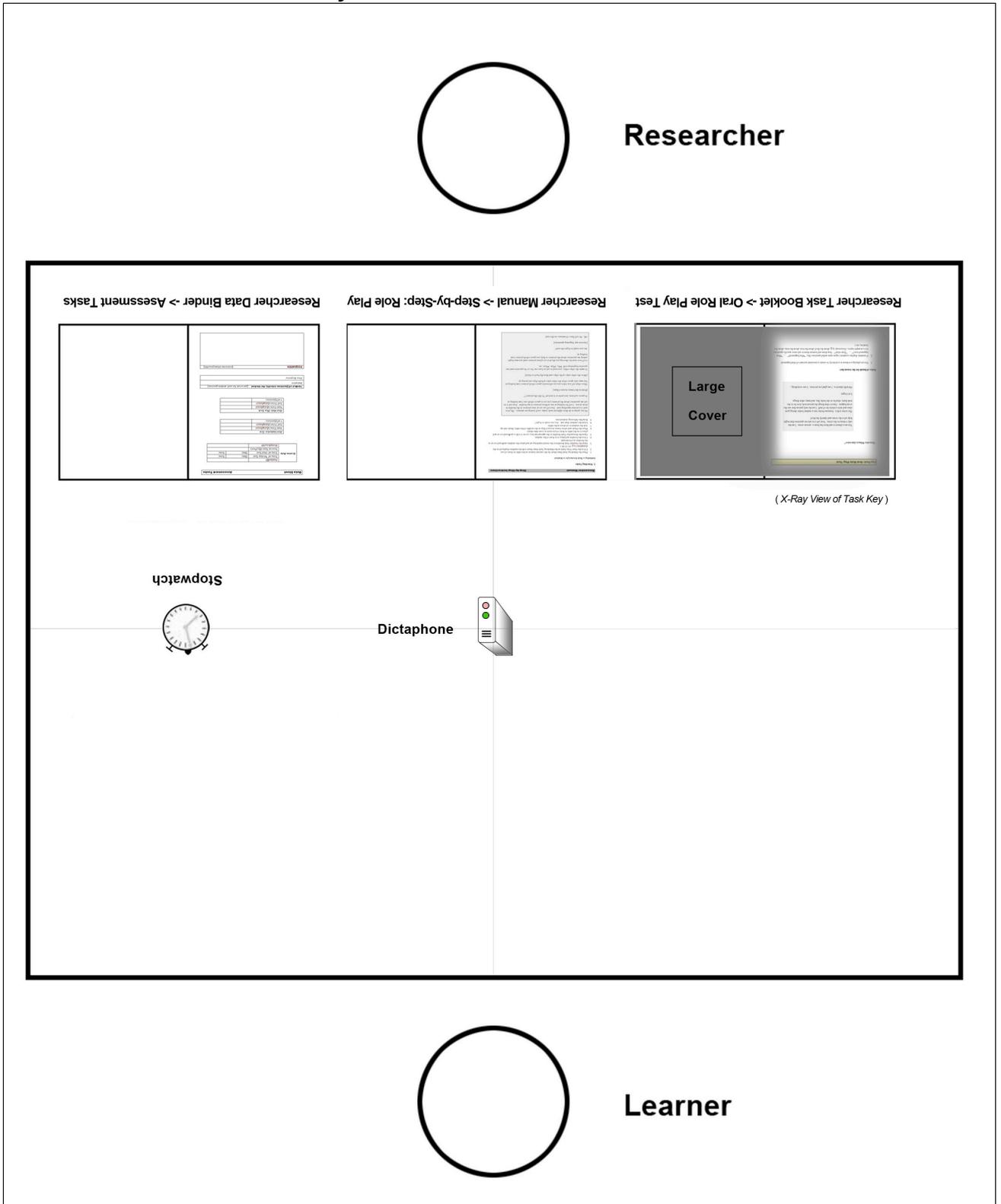
7. Note any irregularities that may have occurred during the session. (e.g., session was interrupted; student was consistently off task, etc)

# FACE-TO-FACE TASK LAYOUTS

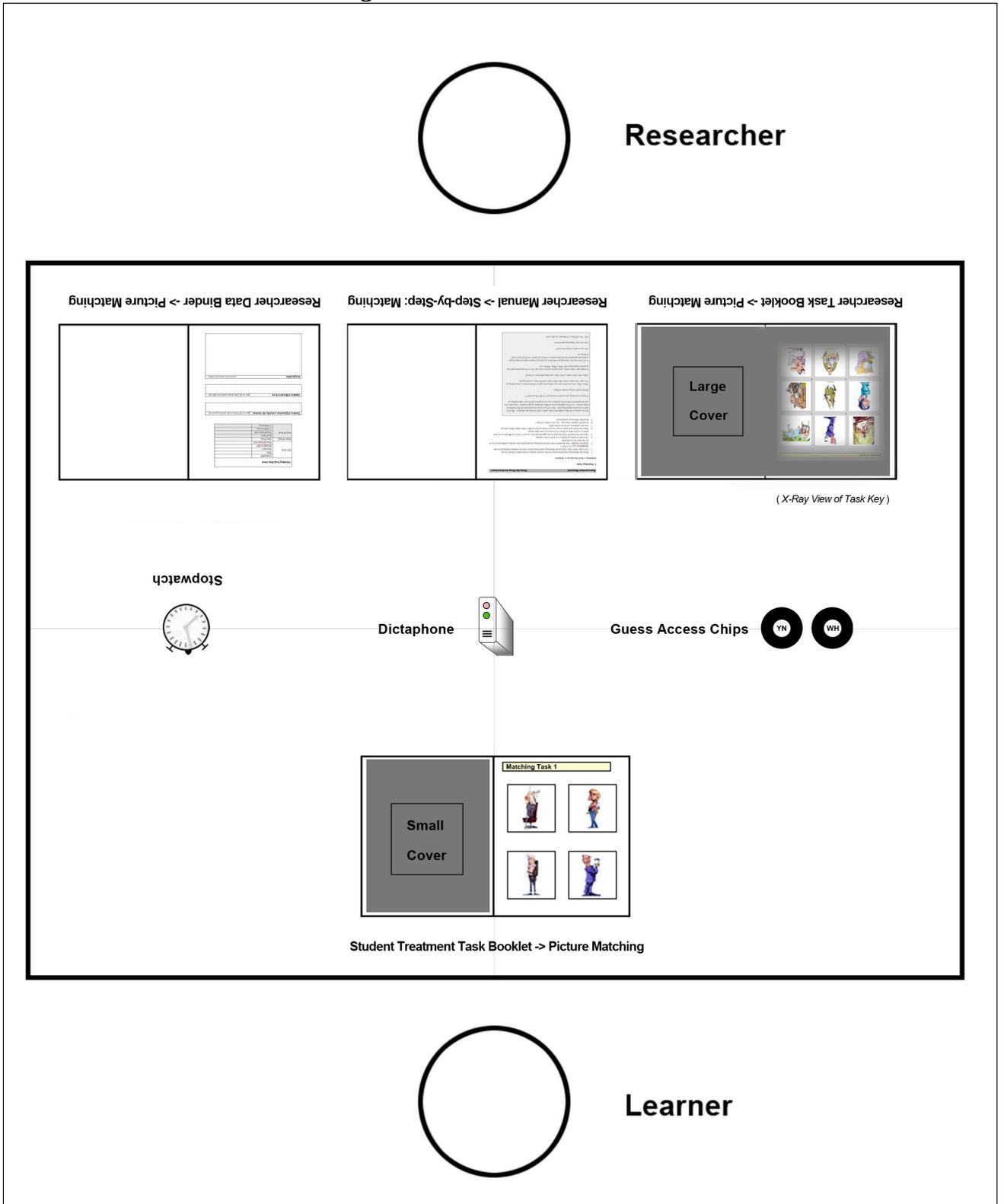
## Assessment Task: Oral Imitation



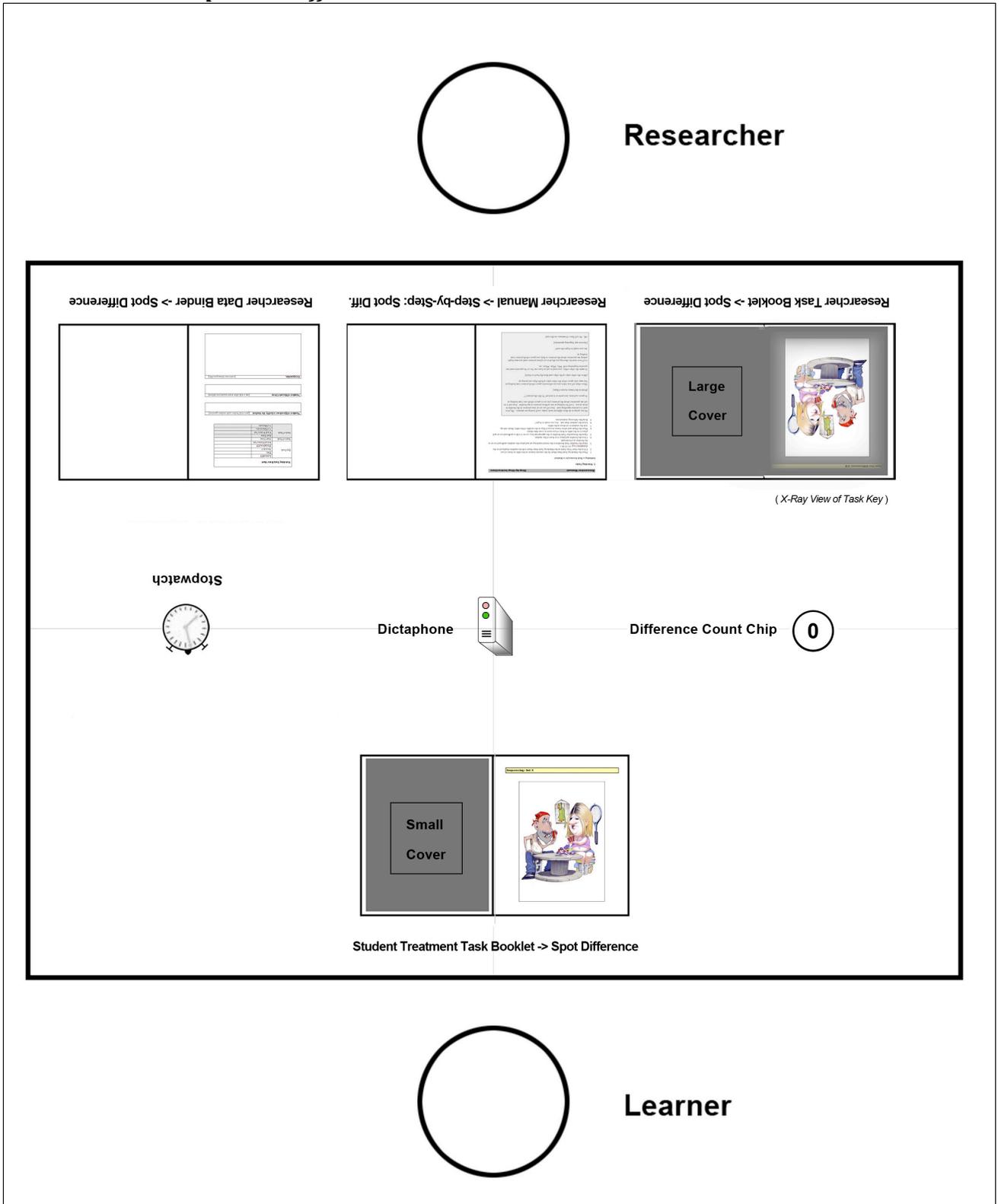
**Assessment Task: Oral Role Play**



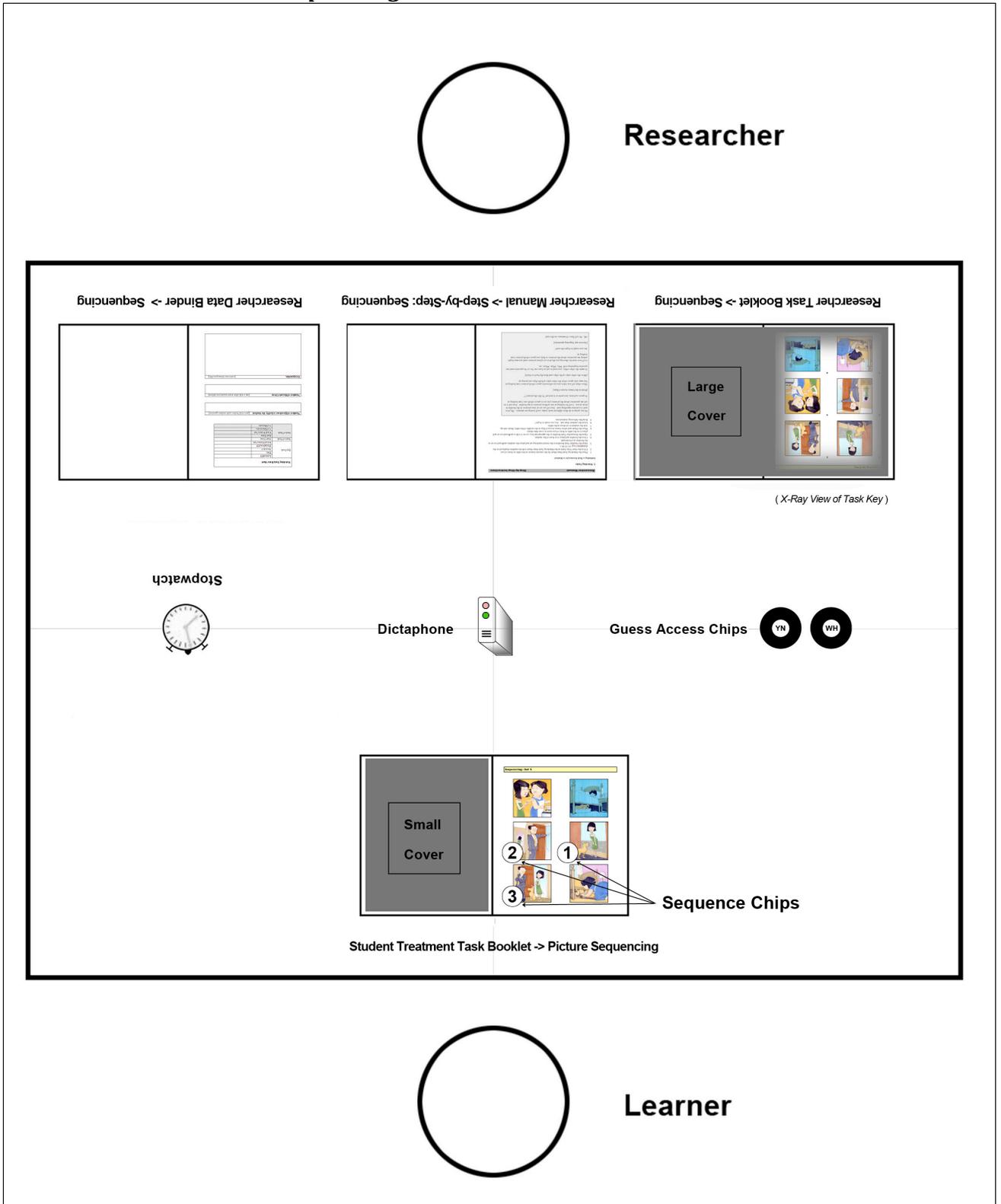
**Treatment Task: Picture Matching**



**Treatment Task: Spot the Difference**



**Treatment Task: Picture Sequencing**



## APPENDIX D: EXAMPLES FROM RESEARCHER TASK BOOKLET

### A. Assessment Tasks

#### 1. Key for Imitation Task

##### Pre-Test: Oral Imitation Task: Key

1.



Does the woman have an orange pumpkin in her hands?

2.



What is the guy in the crazy hat pouring into the bowl?

3.



Is the man relaxing under a tree?

4.



Who is running between the mountains behind the two men?

5.



Is the man with the microphone a happy guy?

6.



What is the woman in the blue shirt looking at?

7.



How tall is the boy with the dog?

8.



What is the woman wearing on her head?

9.



Is the old woman on the bench smiling?

## 2. Script for Role Play Task

### Pre-Test: Oral Role Play Test

#### *"Detective/Witness Interview"*

*Read To Student:*

Are you ready? In this task, you are a detective and there has been a serious crime. I am the only witness to the crime. Your job is to ask me questions that might help solve the crime and identify the thief.

The crime is this: Someone broke into a student locker during gym class and stole something. I am the only person that saw the event happen. I know what things the person took, how he or she took it, what he or she looks like, and many other things.

Let's begin our interview:

Oh hello detective. I am glad you are here. I saw everything.

*Here is what you saw:*

About an hour ago, a 3-foot tall man, dressed as a clown, came into the locker room. He got up on the bench and broke the lock off with a hammer. He then took a purple bag out of the locker and ran away.

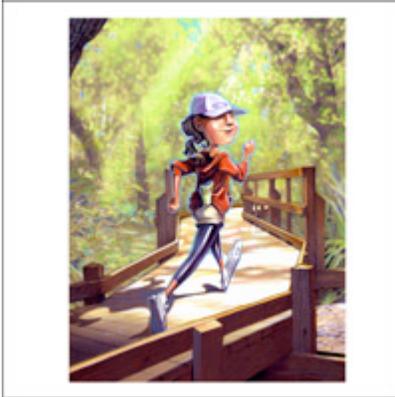
*Rules of thumb for the researcher:*

1. Use your imagination to fill in any unspecified details that the student might ask about. As you do this task over and over, you should be building a consistent account of what you witnessed.
2. If students displays a pattern vague open-ended questions like, "What happened?" ... "What happened next?" ... "Then what?" Stop them and instruct them to ask more specific questions. Give example topics, if necessary (e.g. about the thief, about the loot, about the time, about the location, etc.)

## B. Treatment Tasks

### 1. Key for Matching Task

#### Matching: Keys (1)



1



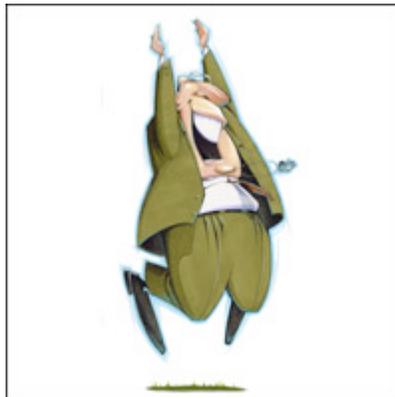
2



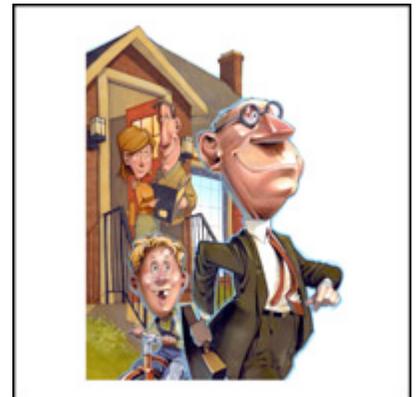
3



4



5



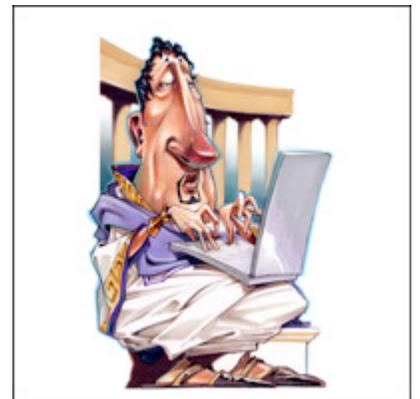
6



7



8



9

2. Alternate Picture for Spot the Difference Task

**Spot the Differences: 1-A**



3. Correct Order for Sequencing Task

Sequencing: Set 3: Key

1



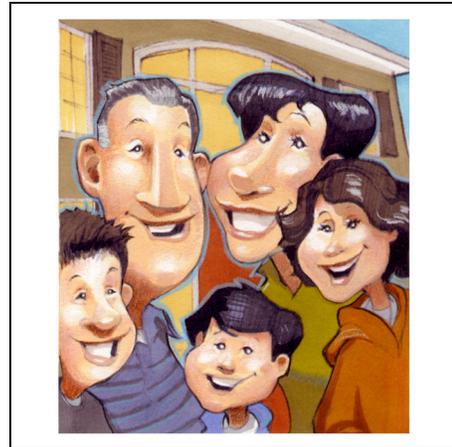
2



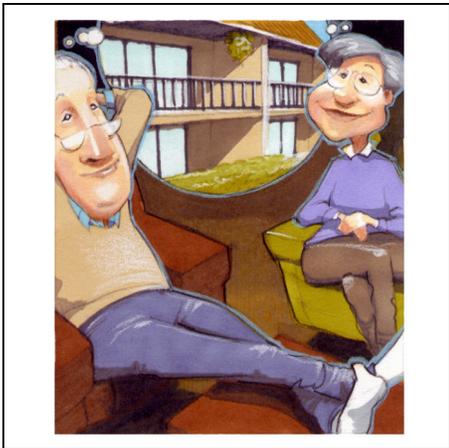
3



4



5



6



APPENDIX E: EXAMPLES FROM STUDENT TASK BOOKLET

1. Assessment Tasks

*a. Imitation Task*

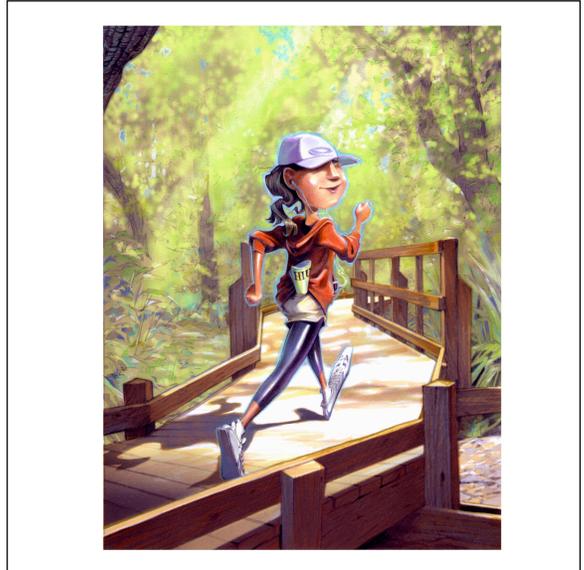
**Pre-Test: Oral Imitation Task: Picture 1**



## 2. Treatment Tasks

### a. Matching Task

#### Picture Matching: Set 1



*b. Spot the Difference Task*

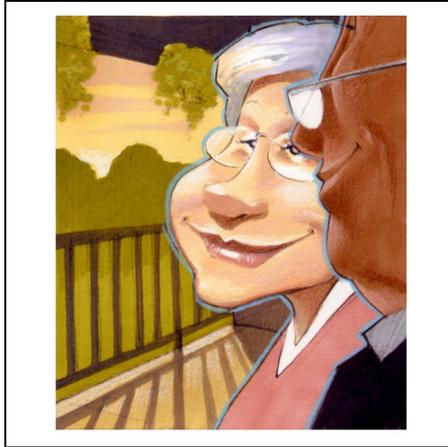
**Spot the Differences: 1-B**



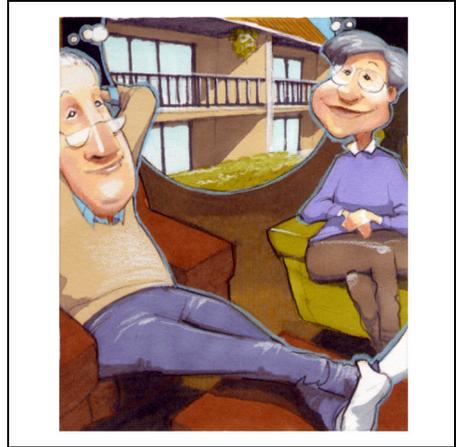
c. Picture Sequencing Task

Sequencing: Set 3: Key

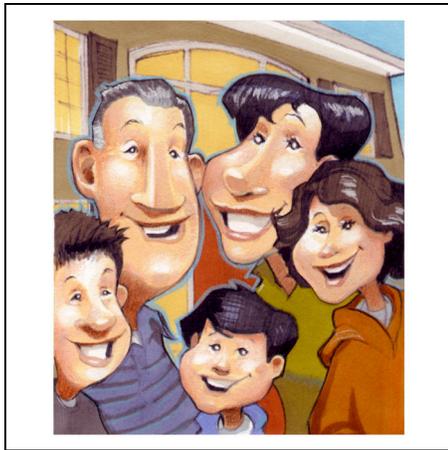
1



2



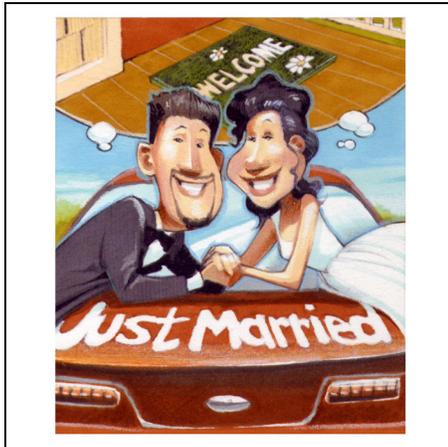
3



4



5



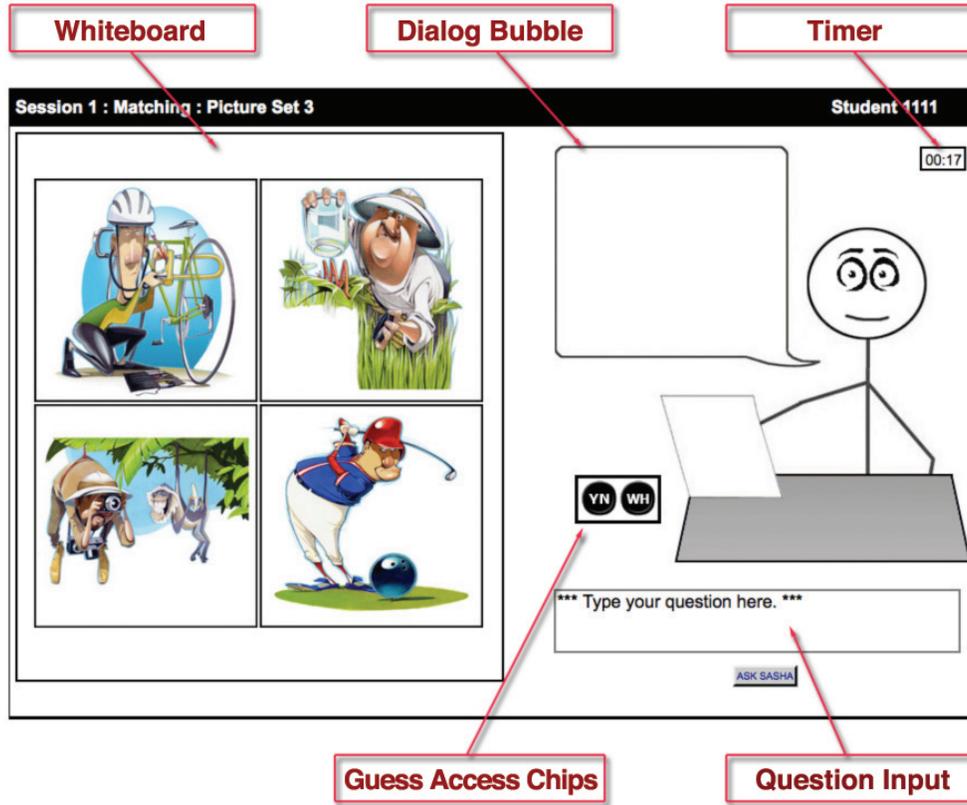
6



All illustrations used in this study were created by a professional illustrator. Bruce MacPherson's work has appeared in publications such as the New York Times, the Washington Post and the Wall Street Journal. He has also illustrated two popular children's books – *Josefina Javelina: A Hairy Tale* and *Thank You, Aunt Tallulah!* Samples of his work can be found at: <http://www.brucemacpherson.biz/>.

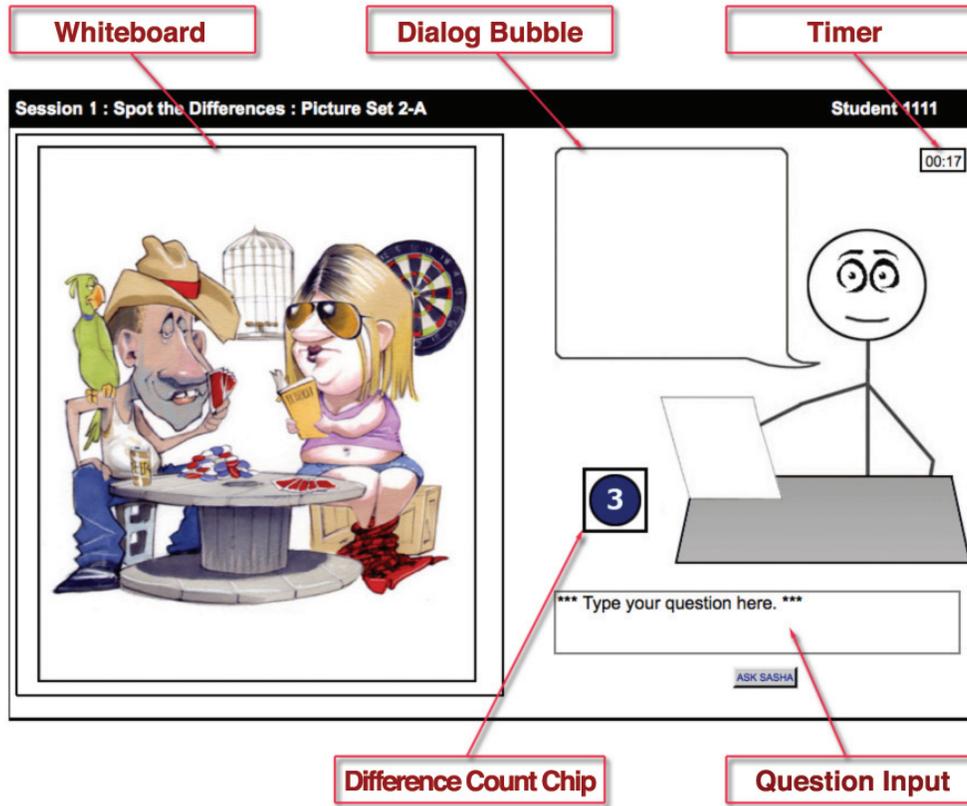
## APPENDIX F: SASHA TREATMENT TASK EXAMPLES

### 1. Picture Matching Task



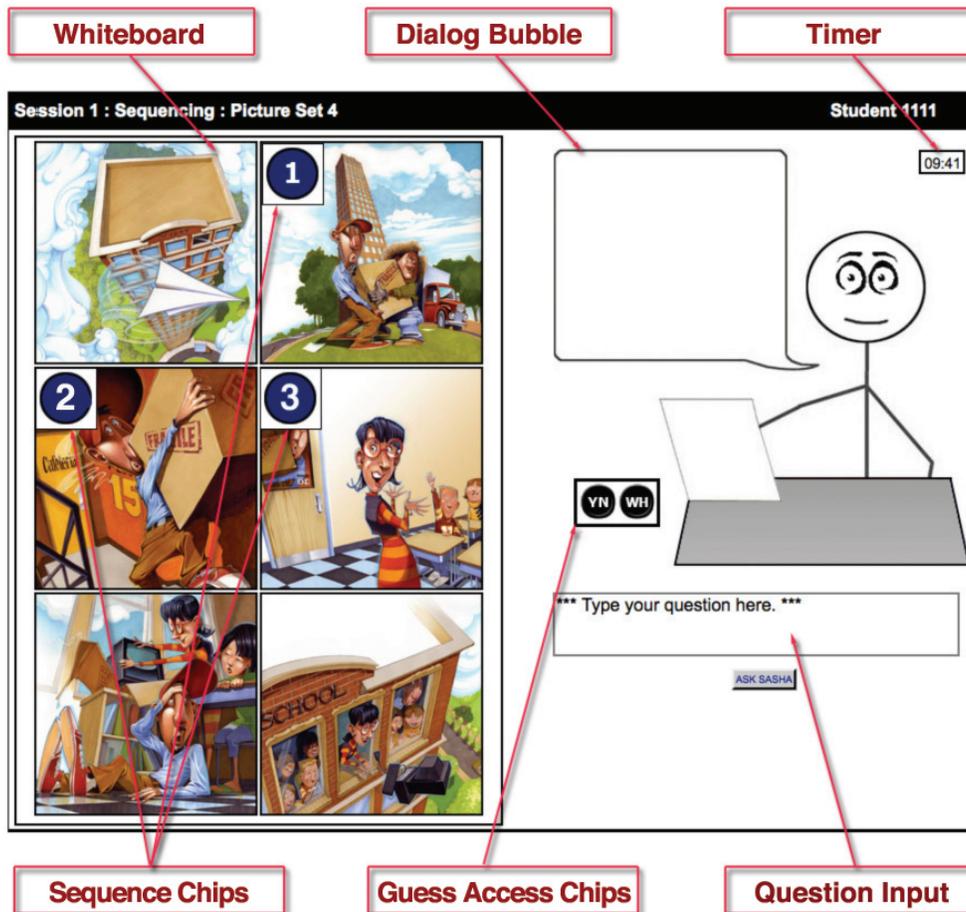
Interface Components	
Whiteboard	Display of picture sets.
Dialog bubble	Sasha's feedback and responses.
Timer	Time remaining in task.
Guess Access Chips	Yes/no & Wh- question requisite indicators for picture guessing.
Question Input	Text area for learner questions.

## 2. Spot the Difference Task



Interface Components	
Whiteboard	Display of picture sets.
Dialog bubble	Sasha's feedback and responses.
Timer	Time remaining in task.
Difference Count Chip	Number of differences identified by the learner.
Question Input	Text area for learner questions.

### 3. Picture Sequencing Task



Interface Components	
Whiteboard	Display of picture sets.
Dialog bubble	Sasha's feedback and responses.
Timer	Time remaining in task.
Sequence Chips	Indicates correctly identified pictures in the sequence.
Guess Access Chips	Yes/no & Wh- question requisite indicators for picture guessing.
Question Input	Text area for learner questions.

## APPENDIX G: DEBRIEFING QUESTIONNAIRE

### English Question Study Exit Questionnaire

1. Name \_\_\_\_\_

2. Grade:  9<sup>th</sup>  10<sup>th</sup>  11<sup>th</sup>  12<sup>th</sup>

3. What country are you from? \_\_\_\_\_

4. How many years have you been in America? \_\_\_\_\_

5. How many years have you been studying English? \_\_\_\_\_

***Please select an answer for each of the following statements:***

† 6. I am comfortable speaking English with native speakers.

Not at all       Not much       I don't know       A little bit       A lot

† 7. I can write well in English.

Not at all       Not much       I don't know       A little bit       A lot

† 8. I think the tasks I did for this study were helpful for my English.

Not at all       Not much       I don't know       A little bit       A lot

† 9. I enjoyed coming to the sessions.

Not at all       Not much       I don't know       A little bit       A lot

† 10. I noticed that I was being corrected if I made a mistake in my questions.

Not at all       Not much       I don't know       A little bit       A lot

‡ 11. I am good at typing English on a computer.

Not at all       Not much       I don't know       A little bit       A lot

‡ 12. I liked talking with Sasha.

Not at all       Not much       I don't know       A little bit       A lot

‡ 13. I think Sasha is pretty smart.

Not at all       Not much       I don't know       A little bit       A lot

†      FTF & CGR only

‡      CGR only

## APPENDIX H. SASHA'S RULE SET FOR ESL QUESTIONS

The following is the method from Sasha's grammar module (written in object-oriented Perl 5). The method uses a series of regular expression shorthand variables to generate a rule set that describes the range of target-like and non-targetlike ESL questions. Learners' L2 input is compared to the regular expression strings stored in the rule set to determine what stage level the question is, whether or not it is a targetlike utterance and (if non-target) to trigger a proscribed set of morpho-syntactic repairs.

Example rule:

```
# {*Who going (to) eat(ing) (fish)?} -> Who is going to eat fish?
'ruleKey' => 1000222,
'regExp' => $self->addSlashes("($WHNP|$WHDT) ?($VPBE|$ISNT|$DONT|NOT)? $VPF ($VPB|$VPG)?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
```

This rule accounts for L2 attempts at future ('going to') WH questions. This rule will match on any questions such as:

2. \*Who/what going to eat?
3. \*Who/what going to eat fish for dinner?
4. \*Who/what going to eating?
5. \*Who/what going to eating fish with a fork?
6. Who is going to eat fish?

Note that this rule will catch a broad range of non-targetlike attempts at future constructions, yet it will also match on targetlike questions. The rules have been ordered according to the developmental stage level that they represent. A single question may match on more than one rule and subsequent rule assignments overwrite any previous assignments. Since the strictness of the rules increases with stage level, targetlike matches on higher stage level rules override any earlier assignments. The question in Example 5 will first be assigned as a potential Stage 3 question, but will move to Stage 5 when it later matches on the targetlike rule for a Stage 4 question using the future tense. None of the other examples will meet these requirements and will be classified as Stage 3 questions.

```

sub initializeGrammar() {
my $self = shift;
##### REGEX MATCHING VARIABLES #####

# ##### GLOBALS #####
my $word = '\w+';
my $notBe = '(\w{3`cm`}ldolgo)';
my $notDoBe = '(\w{3`cm`}lgo)';
my $stative = '(like|love|hate|enjoy|loathe|detest|dig)*';
my $stativeINF = '(want|need|forget)*';
# ##### /GLOBALS #####

# ##### VP/AUX/MOD VARIABLES #####
my $VB = 'VB(ZIPID)?'; # any simple verb
my $VPINF = 'VP[0,to,TO]'; # infinitive phrase marker
my $VPH = "VP[1,$stative(sled)?,($VBI|VBG)]"; # stative ('helper') verb (allow for VBG in cases where
# 'wanting' has been corrected)
my $VPHI = "VP[1,$stativeINF(sled)?,($VBI|VBG)]"; # stative ('helper') verb (allow for VBG in cases
# where 'wanting' has been corrected)
my $VPH_INF = "VP[1,$stative(sled)?,($VBI|VBG)] $VPINF"; # stative ('helper') verb
my $VPHI_INF = "VP[1,$stativeINF(sled)?,($VBI|VBG)] $VPINF"; # stative ('helper') verb + TO
my $VPHI_NOINF = "VP[1,$stativeINF(sled)?,($VBI|VBG)]"; # [MAL] stative ('helper') verb - TO
my $VPHI_G_INF = "VP[1,$stativeINF"."ing"."",VBG] $VPINF"; # [MAL] stative ('helper') verb + ing + TO
my $VPHI_G_NOINF = "VP[1,$stativeINF"."ing"."",VBG]"; # [MAL] stative ('helper') verb + ing + TO
my $VPH_ALL = "($VPHI|VPH_INF|VPHI_INF|VPHI_NOINF|VPHI_G_INF|VPHI_G_NOINF)*"; # any helper verb
my $VPBE = "VP[(0|1)+,be,$VB]"; # be as either aux or main
my $VPBE_INF = "VP[(0|1)+,be,$VB]"; # be as infinitive 'be' -> 'VB'
my $VPDO = "VP[(0|1)+,do,$VB]"; # be as either aux or main
my $VPCOP = "VP[1,be,VB(ZIPID)+]"; # copular not 'be' {i.e. not label: VB}
my $VPCOP_MAL = "(VP[(0|1)+,$word,MD])|VP[(0|1)+,do,$VB]";
my $VPF = "VP[0,going,VBG] $VPINF"; # future tense
my $AUX = "VP[0,$word,$VB]"; # any aux
my $AUXBE = "VP[0,be,$VB]"; # aux be
my $AUXDO = "VP[0,do,$VB]"; # aux do
my $AUXSPECIAL = "VP[(0|1)+,(gets?|getting|becomes?|becoming)+,($VBI|VBG)+]"; # get/become as aux verb
my $VPHAVE = "VP[(0|1)+,have,$VB]"; # aux do
my $GOT = "VP[1,got,$VB]"; # main verb got (w/have)
my $MD = "VP[(0|1)+,$word,MD]"; # any modal
my $DONT = "$AUXDO NOT"; # don't, doesn't, didn't
my $ISNT = "$VPBE NOT"; # isn't, aren't, wasn't, weren't
my $MDNT = "$MD NOT"; # can't, won't, wouldn't
my $VP = "(VPINF)? ?VP[1,$word,$word]"; # any main verb
my $VPN = "VP[1,$word,VPN]"; # participial form
my $VPB = "(VP[1,$notBe,$VB]|$AUXSPECIAL $VPN)+"; # any simple verb phrase
my $VPB_NODO = "(VP[1,$notDoBe,$VB]|$AUXSPECIAL $VPN)+"; # any negative verb phrase
my $VPG = "($AUXSPECIAL )?VP[1,$word,VBG]+"; # any progressive verb
# ##### /VP/AUX/MOD VARIABLES #####

# ##### QUESTION WORDS ##### #
my $WH = "WH[$word,$word]";
my $HOW = "WH[how,$word]";
my $WHAT = "WH[what,$word]";
my $WHICH = "WH[which,WDT]";
my $WHDT = "WH[$word,WDT]";
my $WHNP = "WH[$word,WP]";
my $WHADVP = "WH[$word,WRB]";
# ##### /QUESTION WORDS ##### #

my $NP = '(NP( PP)?( CC NP( PP)?)?)'; # series of conjoined NPs (with possible PP children)
my $VP_ARGS = "(((($NP)|(ADJPIADVPIPPIS)) ?)+"; # possible POS arguments to main verbs

##### /REGEX MATCHING VARIABLES #####
##### RULE SETS FOR ESL QUESTIONS #####

$self->rules (
# ##### STAGE 2 ##### #

```

```

##### Stage 2 [Mal] ##### #
{ # { *The man are eats (fish)? *The man gets angry? } -> Does the man eat fish?
  # IMPORTANT --> This will catch copulars -- Stage 2 rule 10002 teases them out
  'ruleKey' => 10001,
  'regExp' => $self->addSlashes("$NP ?($VPBEI$ISNTI$DONTINOT)? $VPB ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man eats (fish)? *The man gets angry? } -> Does the man eat fish?
  # IMPORTANT --> This will catch copulars -- Stage 2 rule 10002 teases them out
  'ruleKey' => 100011,
  'regExp' => $self->addSlashes("$NP ?(NOT)? ($VPBI$VPN) ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man is a clown/angry?/The man is on the bus } -> Is the man a clown?
  'ruleKey' => 10002,
  'regExp' => $self->addSlashes("$NP ($VPCOPI$ISNTINOT) ($NP ?($VP_ARGS)?IADJPIADVPIPP)"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man likes to eat (fish)? } -> Does the man like to eat fish?
  'ruleKey' => 10003,
  'regExp' => $self->addSlashes("$NP ?($DONT)? $VPH_ALL $VPB ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man is eating (fish)? *The man is getting (angry)? } -> Is the man eating fish?
  'ruleKey' => 10004,
  'regExp' => $self->addSlashes("$NP ?($VPBEI$ISNTINOT)? $VPG ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man is going to eat (fish)? } -> Is the man going to eat fish?
  'ruleKey' => 100041,
  'regExp' => $self->addSlashes("$NP ($VPBEI$ISNTINOT) $VPF $VPB ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man can/will eat (fish)? } <- Can the man eat fish?
  'ruleKey' => 100042,
  'regExp' => $self->addSlashes("$NP ($MDI$MDNT) $VPB ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *There is a man (eating (fish))? } <- Is there a man eating fish?
  'ruleKey' => 10005,
  'regExp' => $self->addSlashes("EX ($VPBEI$ISNTINOT) $NP ?($VPG ?($NP)?)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { *The man is getting involved? } <- $AUXSPECIAL
  'ruleKey' => 1000051,
  'regExp' => $self->addSlashes("$NP ($VPBEI$ISNTINOT) $AUXSPECIAL $VPN ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
{ # { * (Is) there is a man with some money? } <- Is there a man with some money?
  'ruleKey' => 1000052, #
  'regExp' => $self->addSlashes("?( $VPBEI$ISNTINOT)? EX ($VPBEI$ISNTINOT) $NP ?($VP_ARGS)?"),
  'stageLevel' => 2,
  'isTarget' => 0
},
##### /Stage 2 [Mal] ##### #
##### /STAGE 2 ##### #
##### STAGE 3 ##### #
##### General [MAL] ##### #
{ # { *Eats he fish? } -> Does he eat fish?

```

```

        # IMPORTANT--> 'Is he an artist?' will be matched-- it is overwritten
        # by the target rule in Stage 4 (rule #)
        'ruleKey' => 10006,
        'regExp' => $self->addSlashes("$VPB $NP ?($VP_ARGS)?"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*Eating he fish?} -> Does he eat fish?
        # IMPORTANT--> 'Is he an artist?' will be matched-- it is overwritten
        # by the target rule in Stage 4 (rule #)
        'ruleKey' => 1000666,
        'regExp' => $self->addSlashes("$VPG $NP ?($VP_ARGS)?"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*Has he a fish?} -> Does he have a fish?
        'ruleKey' => 1000061,
        'regExp' => $self->addSlashes("$VPHAVE $NP $NP"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*'6:25' on his clock? He fat? The boy a farmer?} -> Does he have a fish?
        'ruleKey' => 1000062,
        'regExp' => $self->addSlashes("$NP ?(NOT)? ?($VP_ARGS)?"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    # ##### /General [MAL] ##### #
    # ##### WH @ end [MAL] ##### #

    { # {*He (wants to) lives where?} -> Where does (he want) to live?
        'ruleKey' => 10007,
        'regExp' => $self->addSlashes("$NP ?($VPH_ALL)? $VPB $WH"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*He will live where?} -> Where will he live?
        'ruleKey' => 10008,
        'regExp' => $self->addSlashes("$NP $MD $VPB $WH"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*He is going to eat (fish) where?} -> Where is he going to eat fish?
        'ruleKey' => 10009,
        'regExp' => $self->addSlashes("$NP $VPBE $VPF $VPB ?($NP)? $WH"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*He is (eating) (fish) where?} -> Where is he eating fish?
        'ruleKey' => 10010,
        'regExp' => $self->addSlashes("$NP ?($VPBE)? ?($VPG)? ?($NP)? $WH"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    # ##### /WH @ end [MAL] ##### #
    # ##### Have + SV0? [MAL] ##### #
    { # {*Has the man eat (fish)?} -> Has the man eaten fish?
        'ruleKey' => 10011,
        'regExp' => $self->addSlashes("$VPHAVE $NP $VPB ?($VP_ARGS)?"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*Has the man (is) eating (fish)?} -> Is the man eating fish?
        'ruleKey' => 10012,
        'regExp' => $self->addSlashes("$VPHAVE $NP ?($VPBE)? $VPG ?($VP_ARGS)?"),
        'stageLevel' => 3,
        'isTarget' => 0
    },
    { # {*Has the man going to eat (fish)?} -> Is the man going to eat fish?

```

```

'ruleKey' => 10013,
'regExp' => $self->addSlashes("$VPHAVE $NP ?($VPBE)? $VPF $VPB ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
# ##### /Have + SVO? [MAL] ##### #
# ##### Do + SVO? [Mal] ##### #
{
  # {*Does the man (is) eating (fish)?} -> Is the man eating fish?
  'ruleKey' => 10014,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT) $NP ?($VPBE)? $VPG ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*Does the man is eat (fish)?} Does the man eat fish?
  'ruleKey' => 10015,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT) $NP $VPBE $VPB ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*(Does|Is) eats the man (fish)?} Does the man eat fish?
  'ruleKey' => 1000151,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT!$ISNT!$AUXBE) $VPB $NP ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*(Does|Is) eating the man (fish)?} Is the man eating fish?
  'ruleKey' => 1000152,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT!$ISNT!$AUXBE) $VPG $NP ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*Does the man (is) going to eat (fish)?} -> Is the man going to eat fish?
  'ruleKey' => 10016,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT) $NP ?($VPBE)? $VPF $VPB ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*Does/Is) the man (is) want(ing) (to) eat(ing) (fish)?} -> Does the man want to
  # eat fish?
  # IMPORTANT --> This will catch "Does the man want to eat fish?"
  # It will later be reassigned a target match in Stage 4
  'ruleKey' => 10017,
  'regExp' => $self->addSlashes("($AUX!$DONT!$ISNT!NOT) $NP ?($VPBE)? ($VPH_ALL)* ($VPB!$VPG)
  ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*Does the man (is) (a clown|angry)+?} -> Is the man a clown?
  'ruleKey' => 10018,
  'regExp' => $self->addSlashes("($AUXDO!$VPCOP_MAL!$DONT) $NP ?($VPCOP)? ?($VP_ARGS)?"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # {*Does there (is) a chair?} -> Is there a chair?
  'ruleKey' => 10019,
  'regExp' => $self->addSlashes("($VPDO!$DONT!NOT) EX ?($VPCOP)? $NP"),
  'stageLevel' => 3,
  'isTarget' => 0
},
{
  # { Do they (are) involved with the plan?}
  'ruleKey' => 1000191,
  'regExp' => $self->addSlashes("$VPDO $NP ?($VPBE)? ADJP ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 0
},
# ##### /Do + SVO? [Mal] ##### #
# ##### Do + SVO? [Target] ##### #
{
  # { Does he eat (fish)? Does he get angry? }
  'ruleKey' => 1,

```

```

'regExp' => $self->addSlashes("($AUXDO!$DONTINOT) $NP $VPB ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 1
},
{
    # {Does he want to eat (fish)?}
    'ruleKey' => 2,
    'regExp' => $self->addSlashes("($AUXDO!$DONTINOT) $NP ($VPH_INF!$VPHI_INF) $VPB
        ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 1
},
{
    # {Does he like eating (fish)?}
    'ruleKey' => 3,
    'regExp' => $self->addSlashes("($AUXDO!$DONTINOT) $NP $VPH $VPG ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 1
},
},
# ##### /Do + SVO? [Target] ##### #
# ##### Wh + (Aux/MD) + SVO? [Mal] ##### #
{
    # {*Why (is/does) the man (is) eats (fish)?} -> Why does the man eat fish?
    # IMPORTANT --> This will catch "Why does the man eat fish?"
    # It will later be reassigned a target match in Stage 5
    'ruleKey' => 10020,
    'regExp' => $self->addSlashes("$WH ?($AUX)? $NP ?($VPBE!$ISNTINOT)? $VPB ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*Why (is/does) the man doesn't eats (fish)?} -> Why doesn't the man eat fish?
    # IMPORTANT --> This will catch "Why does the man eat fish?"
    # It will later be reassigned a target match in Stage 5
    'ruleKey' => 1000201,
    'regExp' => $self->addSlashes("$WH ?($AUX)? $NP ($DONT!$ISNTINOT) $VPB ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*What color (is/does) the lady (is) wearing? Why she is wearing blue?} -> WH BE NP Main
    'ruleKey' => 1000063,
    'regExp' => $self->addSlashes("$WHDT NP ?($AUX)? $NP ?($VPBE)? $VPG ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*What color shirt (is/does) the lady (is) wearing? Why she is wearing blue?} -> WH BE
    NP Main
    'ruleKey' => 10000631,
    'regExp' => $self->addSlashes("$WHDT NP NP ?($AUX)? $NP ?($VPBE)? $VPG ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*What color wears the lady?} -> What color does the lady wear? What book does the man
    give him?
    'ruleKey' => 1000064,
    'regExp' => $self->addSlashes("$WHDT NP $VPB ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*What color shirt wears the lady?}
    'ruleKey' => 10000641,
    'regExp' => $self->addSlashes("$WHDT NP NP $VPB ?($VP_ARGS)?"),
    'stageLevel' => 3,
    'isTarget' => 0
},
{
    # {*Why (is/does) the man (is) eating (fish)?} -> Why is the man eating fish?
    # IMPORTANT --> This will catch "Why is the man eating fish?"
    # It will later be reassigned a target match in Stage 5
    # Also will catch -> 'what color the person wearing' -- reassigned by 1000063
    'ruleKey' => 10021, #WH[why,WRB] NP VP[1,eating,VBG]
    'regExp' => $self->addSlashes("($WHADVP!$WHNP) ?($AUX!$ISNT!$DONTINOT)? $NP
        ?($VPBE!$ISNT!$DONTINOT)? $VPG ?($VP_ARGS)?"),
    'stageLevel' => 3,

```

```

'isTarget' => 0
},
{ # {*Why (is/does) the man is (angry/a clown/beside the bus?) -> Why is the man angry?
'ruleKey' => 1000211,
'regExp' => $self->addSlashes("$WHADVP ?($AUX|$ISNT|$DONT|NOT)? $NP ($VPBE|$ISNT)
?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Why (is/does) the man (is) want(ing) (to) eat(ing) (fish)?} -> Why does the man want
to eat fish?
# IMPORTANT --> This will catch "Does the man want to eat fish?/Does the man like
eating?"
# These will later be reassigned target matches in Stage 5
'ruleKey' => 10022,
'regExp' => $self->addSlashes("$WH ?($AUX|$ISNT|$DONT|NOT)? $NP ?($VPBE|$ISNT)? $VPH_ALL
($VPBI$VPG) ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Who going (to) eat(ing) (fish)?} -> Who is going to eat fish?
'ruleKey' => 1000222,
'regExp' => $self->addSlashes("$WHNP|$WHDT) ?($VPBE|$ISNT|$DONT|NOT)? $VPF ($VPBI$VPG)
?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Who is eats his fish(with a fork)?} -> Where does the man with the fish eat?
'ruleKey' => 1000223, #WH[who] VP[0,be,VBZ] VP[1,wants,VBZ] VP[0,to,TO] VP[1,eating,VBG] NP
# WH[what] VP[1,doing,VBG] NP
'regExp' => $self->addSlashes("$WHNP|$WHDT) ?($AUX|$VPBE|$ISNT|$DONT|NOT)? $VPB $NP
?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*What doing the man (with the fish)?} -> What is the man with the fish doing?
'ruleKey' => 1000224,
'regExp' => $self->addSlashes("$WHNP|$WHDT) $VPG $NP ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where (is/does) he will/can eat(ing) (fish)?} -> Where will he eat fish?
'ruleKey' => 10023, #WH[where,WRB] VP[0,can,MD] VP[1,eating,VBG] NP
'regExp' => $self->addSlashes("$WH ?($AUX)? $NP ($MD|$MDNT) ($VPBI$VPG) ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where can eat(ing) the man?} -> Why can the man eat quickly?
'ruleKey' => 1000231, #WH[where,WRB] VP[0,can,MD] VP[1,eating,VBG] NP
'regExp' => $self->addSlashes("$WHADVP ($MD|$MDNT) ($VPBI$VPG) ?($VP_ARGS)? $NP
?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Why (is/does) the man is going to eat (fish)?} -> Why is the man going to eat fish?
'ruleKey' => 10024,
'regExp' => $self->addSlashes("$WH ?($AUX|$ISNT|$DONT|NOT)? $NP ?($VPBE|$ISNT)? $VPF $VPB
?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where (has/is/does) there is a man?} -> Where is there a man?
'ruleKey' => 10025,
'regExp' => $self->addSlashes("$WH ?($AUX|$ISNT|$DONT|NOT)? EX $VPCOP $NP"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*What (is) there is on the bus? Why (is) there is man on the bus?} -> What is there on

```

the bus? Why is there a man on the bus?

```
'ruleKey' => 1000251,
'regExp' => $self->addSlashes("$WH ?($VPBE)? EX $VPBE ?($NP)? ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Why (is|does|would) the man (is) angry?} -> Why is the man angry?
# IMPORTANT --> This will catch "Why is the man angry?"
# Will be reassigned in Stage 4 (rule 22)
'ruleKey' => 10026,
'regExp' => $self->addSlashes("$WH ?($VPCOP_MAL $MD $AUXDO $VPBE $ISNT $DONT $NOT)? $NP
?($VPBE $ISNT $NOT)? ADJP"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where should he eaten (the fish)?} -> Where should he eat the fish?
'ruleKey' => 10027,
'regExp' => $self->addSlashes("$WH $MD $NP $VPN ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where can he (is) eat(ing) (the fish)?} -> Where can he eat the fish?
'ruleKey' => 1000270,
'regExp' => $self->addSlashes("$WHADVP $MD $NP ?($VPBE)? ($VPB|$VPG) ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where has he eating (the fish)?} -> Where is he eating the fish?
# Where is he eating the fish? Is caught later by target rule 9
'ruleKey' => 1000271,
'regExp' => $self->addSlashes("$WHADVP $AUX $NP $VPG ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Where wants the man to eat (the fish)?} -> Where is he eating the fish?
'ruleKey' => 1000272,
'regExp' => $self->addSlashes("$WHADVP ($VPH|$VPHI) $NP $VPINF $VPB ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
# ##### /Wh + (Aux/MD) + SV0? [Mal] ##### #
# ##### Be + SV0? [Mal] ##### #
{ #Is the man (is) eat (fish|at the table)?} -> Does the man eat (fish)?
'ruleKey' => 10028,
'regExp' => $self->addSlashes("($VPBE $ISNT $NOT) $NP ?($VPBE)? $VPB ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ #Is the man (is) eating (fish|at the table)?} -> Is the man eating (fish)?
'ruleKey' => 10029,
'regExp' => $self->addSlashes("($VPBE $ISNT $NOT) $NP ?($VPBE)? $VPG ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ #Is they are friends|on the bus|angry?} -> Are they friends?
'ruleKey' => 1000291,
'regExp' => $self->addSlashes("($VPBE $VPCOP_MAL $ISNT $NOT) $NP $VPBE ?($NP|$PP|$ADJP)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ #{*Is the man is going to eat(ing) (fish)?} # Is the man going to eat fish?
'ruleKey' => 10030,
'regExp' => $self->addSlashes("($VPBE $ISNT $NOT) $NP $VPBE $VPF ($VPB|$VPG) ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Is the man can eat(ing) (fish)?} -> Can the man eat fish
'ruleKey' => 10031, # NOTE: Verb is optional to account for gerund-like assignment (Collins)
of $VBG -- ideosyncratic structure, so little worry of overlap
```

```

'regExp' => $self->addSlashes("($VPBE|ISNT|NOT) $NP $MD ?($VPB|VPG)? ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Is the man (gets|seems)* angry? } -> Does the man get angry? Does the man eat quickly?
'ruleKey' => 10032,
'regExp' => $self->addSlashes("($VPBE|ISNT|NOT) $NP $VPB (ADJPIADVP)"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Is the man is angry? } -> Is the man angry?
'ruleKey' => 10033,
'regExp' => $self->addSlashes("($VPBE|ISNT|NOT) $NP $VPCOP ADJP"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Is there is a man?} # -> Is there a man? needs more training!!
'ruleKey' => 10034,
'regExp' => $self->addSlashes("($VPBE|ISNT|NOT) EX $VPCOP $NP"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*Will he eating (fish)?} -> Will he eat fish?
'ruleKey' => 10035,
'regExp' => $self->addSlashes("($MD|$MDNT) $NP $VPG ?($VP_ARGS)?"),
'stageLevel' => 3,
'isTarget' => 0
},
{ # {*what color his hair? What time in your clock} -> what color is his hair?
'ruleKey' => 10036,
'regExp' => $self->addSlashes("$WHDT NP (NP|PP)"),
'stageLevel' => 3,
'isTarget' => 0
},
# ##### /Be + SV0? [Mal] ##### #
# ##### /STAGE 3 ##### #

# ##### STAGE 4 ##### #
# ##### ($WH)? + copula + S [Target] ##### #
{ # {Is he a policeman/angry/beside the glass (of water)?}
'ruleKey' => 5,
'regExp' => $self->addSlashes("($VPBE|ISNT) $NP ((($NP ?($VP_ARGS)?|ADJPIADVPIPP) ?(PP
NP)?)"),
'stageLevel' => 4,
'isTarget' => 1
},
{ # { Is there a dog on the bus?}
'ruleKey' => 6,
'regExp' => $self->addSlashes("($VPBE|ISNT) EX $NP ?(PPI$VPG ?($NP)?)"),
'stageLevel' => 4,
'isTarget' => 1
},
{ # { What is this? How tall is the man?}
'ruleKey' => 7,
'regExp' => $self->addSlashes("$WH $VPBE ($NP ?(ADJP)?|PP)"),
'stageLevel' => 4,
'isTarget' => 1
},
{ # { Who is angry? Why is he angry?}
'ruleKey' => 8,
'regExp' => $self->addSlashes("$WH ($VPBE|ISNT) ?($VP_ARGS)?"),
'stageLevel' => 4,
'isTarget' => 1
},
{ # { How hard is English to understand?/How tall is the clock in the picture?}
'ruleKey' => 9,
'regExp' => $self->addSlashes("$HOW (ADJPIADVPIRB) $VPBE $NP ?($VPINF $VPB)? ?($VP_ARGS)?"),
'stageLevel' => 4,
'isTarget' => 1
}

```

```

},
{ # { How many flowers are in the vase?}
  'ruleKey' => 10,
  'regExp' => $self->addSlashes("$HOW ?(ADJPIADVP)? $NP $VPCOP (EX ?($VP_ARGS)?|$VP_ARGS)"),
  'stageLevel' => 4,
  'isTarget' => 1
},
# ##### /($WH)? + copula + S [Target] ##### #
# ##### Aux/modal + SV [Target] ##### #
{ # { Has he eaten (a fish)?}
  'ruleKey' => 11,
  'regExp' => $self->addSlashes("$VPHAVE $NP ($VPNI$GOT) ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Are they (get) involved with the plan?}
  'ruleKey' => 11,
  'regExp' => $self->addSlashes("$VPBE $NP $VPN ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Do they get involved with the plan?}
  'ruleKey' => 12,
  'regExp' => $self->addSlashes("$VPDO $NP $AUXSPECIAL $VPN ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Is he eating (fish)? Is the man getting angry?}
  'ruleKey' => 13, #VP[0,be,VBP] NP VP[0,get,VBP] VP[1,involved,VBN]
  'regExp' => $self->addSlashes("$VPBE $NP $VPG ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Is he going to eat (fish)?}
  'ruleKey' => 14,
  'regExp' => $self->addSlashes("$VPBE $NP $VPF $VPB ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Will he eat (fish)?}
  'ruleKey' => 15,
  'regExp' => $self->addSlashes("$MD $NP $VPB ?($VP_ARGS)?"),
  'stageLevel' => 4,
  'isTarget' => 1
},
{ # { Are there any other people?}
  'ruleKey' => 16,
  'regExp' => $self->addSlashes("$VPCOP EX ($NPIADVPIADJP)"),
  'stageLevel' => 4,
  'isTarget' => 1
},
# ##### /Aux/modal + SV [Target] ##### #
# ##### /STAGE 4 ##### #
# ##### STAGE 5 ##### #

# NOTE: There is some spill over between rules 4 & 5. Collins parser is somewhat
# unpredictable in how it assigns NP (gerund) vs. VPG. For the purposes of this application,
# this is an acceptable lack of granularity. Further work (parser training?) may be
# needed to tease these apart.
# Test data: Why does the man like eating fish? vs. Why does the man like eating?

{ # {Where does the man (want to) eat (fish)? }
  'ruleKey' => 17,
  'regExp' => $self->addSlashes("$WH ($AUXDO|$DONT) $NP ?($VPH_INF|$VPHI_INF)? $VPB
    ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},

```

```

{ # {Why does the man like eating (fish)? }
  'ruleKey' => 18,
  'regExp' => $self->addSlashes("$WH ($AUXDO!$DONT) SNP $VPH $VPG ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # {How do you think it is suitable for him?}
  'ruleKey' => 19,
  'regExp' => $self->addSlashes("$WH ($AUXDO!$DONT) SNP $VPB S"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{
  # {Do you think it is suitable for him?}
  'ruleKey' => 4,
  'regExp' => $self->addSlashes("($AUXDO!$DONT!NOT) SNP $VPB S"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # {Why is he eating (fish)?}
  'ruleKey' => 20,
  'regExp' => $self->addSlashes("$WH ($VPBE!$ISNT) SNP $VPG ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{
  # {*Where is he going to eat fish?}
  'ruleKey' => 21,
  'regExp' => $self->addSlashes("$WH ($AUXBE!$ISNT) SNP $VPF $VPB ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # {Where will the man eat (fish)?}
  'ruleKey' => 22,
  'regExp' => $self->addSlashes("$WH ($MD!$MDNT) SNP ($VPB!$VPBE_INF) ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Where is there a man (eating (fish))?}
  'ruleKey' => 23,
  'regExp' => $self->addSlashes("$WH ($VPBE!$ISNT) EX SNP ?($VPG ?($VP_ARGS)?)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Why is there a man on the bus? What is there in the drawer?}
  'ruleKey' => 24,
  'regExp' => $self->addSlashes("$WH ($VPBE!$ISNT) EX ?($SNP)? PP"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Who (ateleats) the fish?}
  'ruleKey' => 25,
  'regExp' => $self->addSlashes("($WHNP!$WHDT) ?($DONT)? $VPB ?($NPIADJPIADVP!PP)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Who is (going to eatleating) the fish?}
  'ruleKey' => 26,
  'regExp' => $self->addSlashes("($WHNP!$WHDT) ($AUXBE!$ISNT) ($VPG!$VPF $VPB) ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Who likes eating fish?}
  'ruleKey' => 261,
  'regExp' => $self->addSlashes("($WHNP!$WHDT) ?($DONT)? $VPH $VPG ?($VP_ARGS)?"),
  'stageLevel' => 5,
  'isTarget' => 1
},
{ # { Who (wants/likes to/will) eat the fish?}

```

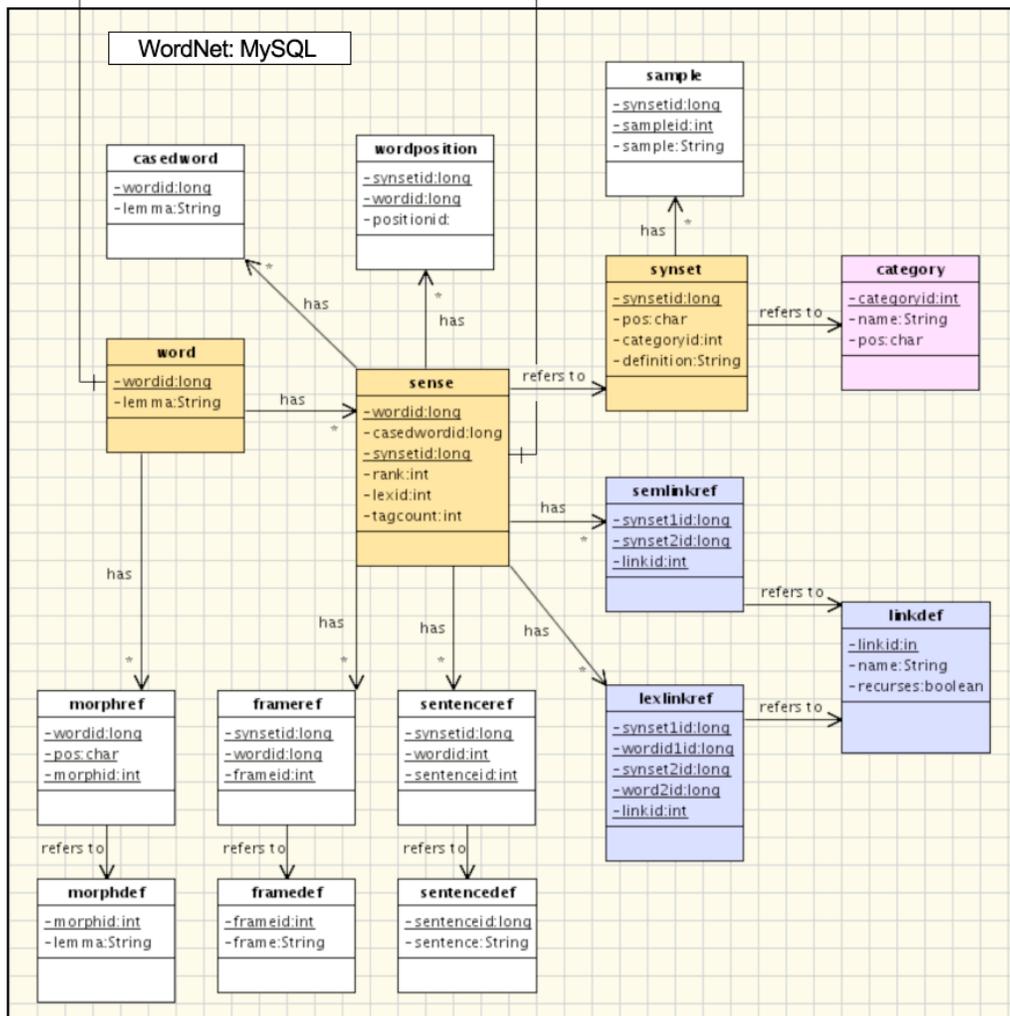
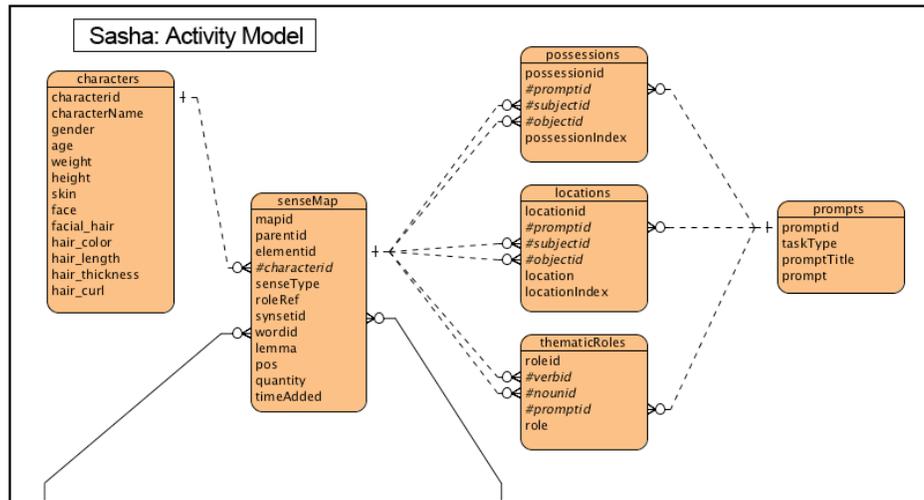
```

'ruleKey' => 27,
'regExp' => $self->addSlashes("$WHNP|$WHDT) ?($DONT)? ($VPHI_INF|$VPH_INF|$MD) $VPB
?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 1
},
{ # { What has he eaten?}
'ruleKey' => 28,
'regExp' => $self->addSlashes("$WH $VPHAVE $NP ($VPB|$VPN) ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 1
},
{ # { What color is his hair?}
'ruleKey' => 29,
'regExp' => $self->addSlashes("$WHDT NP $VPBE $VP_ARGS"),
'stageLevel' => 5,
'isTarget' => 1
},
{ # { What fish is the man eating/going to eat? What color shirt is he wearing?}
'ruleKey' => 30,
'regExp' => $self->addSlashes("$WHDT NP ?(NP)? ($VPBE|$ISNT) $NP ?(NOT)? ($VPG|$VPF $VPB)",
'stageLevel' => 5,
'isTarget' => 1
},
{ # { What fish does/can/will the man eat? What color shirt does he wear?}
'ruleKey' => 31,
'regExp' => $self->addSlashes("$WHDT NP ?(NP)? ($VPDO|$DONT|$MD) $NP $VPB"),
'stageLevel' => 5,
'isTarget' => 1
},
{ # { Which animal eats the cookies?}
'ruleKey' => 32,
'regExp' => $self->addSlashes("$WHDT NP ?($DONT)? $VPB ?($NP)? ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 1
},
{ # { *Where eats the man?} -> Where does the man eat?
'ruleKey' => 10037,
'regExp' => $self->addSlashes("$WHADVP ?($VPH_ALL)? $VPB_NODO ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 0
},
{ # { *Where is eats the man?} -> Where does the man eat?
'ruleKey' => 10038,
'regExp' => $self->addSlashes("$WHADVP $AUX $VPB_NODO ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 0
},
{ # { *Where is|does eating the man?} -> Where is the man eating?
'ruleKey' => 10039,
'regExp' => $self->addSlashes("$WHADVP $AUX $VPG ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 0
},
{ # { *Where eating the man?} -> Where is the man eating?
'ruleKey' => 10040,
'regExp' => $self->addSlashes("$WHADVP $VPG ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 0
},
{ # { *Where likes eating the man?} -> Where does the man like eating?
'ruleKey' => 10041,
'regExp' => $self->addSlashes("$WHADVP $VPB $VPG ?($VP_ARGS)?"),
'stageLevel' => 5,
'isTarget' => 0
},
{ # { How many flowers there (are) in the vase?}
'ruleKey' => 10042, # WH[how,WRB] ADJP NP PP EX VP[1,be,VBP]
'regExp' => $self->addSlashes("$WHDT|$WHADVP) (ADJPI|ADVP) $NP ?(EX ?($VPBE))?")

```

```
                                ?($VP_ARGS)?"),
    'stageLevel' => 4,
    'isTarget' => 0
  },
  # ##### /STAGE 5 ##### #
);
##### /RULE SETS FOR ESL QUESTIONS #####
}
```

## APPENDIX I: ACTIVITY MODEL DATABASE STRUCTURE



Source: <http://wmsqlbuilder.sourceforge.net/images/tables-wordnet.png>