

Generativism and Emergentism: Evidence from Second Language Acquisition Studies of  
Poverty of the Stimulus Phenomena

Laurel B. Preston

A dissertation

submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Julia R. Herschensohn, Co-Chair

Toshiyuki Ogihara, Co-Chair

Karen T. Zagona

Program Authorized to Offer Degree:

Linguistics

©Copyright 2015  
Laurel B. Preston

University of Washington

Abstract

Generativism and Emergentism: Evidence from Second Language Acquisition Studies of  
Poverty of the Stimulus Phenomena

Laurel B. Preston

Co-Chairs of the Supervisory Committee:

Professor Julia R. Herschensohn

Linguistics

Associate Professor Toshiyuki Ogihara

Linguistics

This work applies evidence from second language acquisition studies to the question of the existence of an inborn, language-specific system that constrains the set of possible human grammars. Three sets of studies are examined and critiqued. The sets were selected so that the studies within each of the sets are comparable on the basis of the target phenomenon investigated and the experimental techniques, but contrast on the basis of their chosen theoretical frameworks. The target phenomena are the Overt Pronoun Constraint, scrambling, and quantifier scope ambiguities. Each set of studies includes at least one study conducted within the generativist approach, advocating for L2

learner access to an innate language-specific endowment (often called Universal Grammar) that is essential to the acquisition of grammar in the presence of impoverished linguistic input. Also, each set of studies contains at least one study conducted within an emergentist approach, advocating that domain-general learning mechanisms applied to rich and complex linguistic input suffice to explain L2 learner attainment. All of the target phenomena investigated in these studies are constructions that have been argued in the generative literature to pose poverty-of-the-stimulus (POS) problems for L2 learners. That is, the linguistic input typically available to learners in the form of second language input, instruction, and through their first language, has been argued to underdetermine the acquisition of the construction. The poverty of the stimulus is considered a cornerstone argument for the existence of UG.

Based on my review and critique of the selected studies, I find that their aggregate empirical results support the generativist approach. However, I do not rest this claim on a POS argument; I argue that the premise of insufficiency of input has not been well enough established for the POS argument to be applied. Instead, I argue that the evidence from these studies is more consistent with generative predictions for sudden, step-wise acquisition of grammatical contrasts, possibly on the basis of limited input, rather than with emergentist predictions for gradual, incremental acquisition on a necessarily substantial amount of input.

# Table of Contents

List of Figures .....	v
List of Tables .....	vi
Chapter 1: Introduction .....	1
1.1 Purpose and Overview .....	1
1.2 Summary of Chapters .....	4
1.3 Generativism and Emergentism Compared on Key Issues.....	6
1.3.1 Structure of the Mind .....	7
1.3.2 The Role of Input .....	9
1.3.3 The Generativist View Applied to Acquisition .....	10
1.4 Emergentism .....	14
1.4.1 Structure of the Mind .....	15
1.4.2 The Role of Input .....	16
1.4.3 The Emergentist Point of View on Language Acquisition .....	18
1.5 Conclusion .....	20
Chapter 2: Poverty of the Stimulus (POS).....	21
2.1 Overview.....	21
2.2 The Argument from the Poverty of the Stimulus: Basic Statement for L1A ...	22
2.3 The Generativist View on the POS Argument.....	24
2.3.1 The Generativist View on the Key Assumptions.....	24
2.3.2 The POS Argument for L2A.....	43
2.3.3 Confirming the POS.....	47
2.4 The Emergentist View on the POS Argument.....	48

2.4.1	The Emergentist View on the Key Assumptions .....	48
2.4.2	Disconfirming the POS .....	54
2.5	Conclusion .....	55
Chapter 3: Methodologies for Assessing and Modeling Acquisition .....		57
3.1	Overview .....	57
3.2	Grammaticality Judgment Tests (GJs).....	58
3.2.1	Rationale for GJ Tests.....	58
3.2.2	Characteristics of GJ Tests.....	59
3.2.3	Evaluation of GJs.....	69
3.2.4	Issues Particular to L2A Research .....	75
3.3	Truth-Value Judgment Tasks (TVJs).....	76
3.3.1	Rationale .....	76
3.3.2	Characteristics of TVJs.....	77
3.3.3	Evaluation of TVJs .....	83
3.3.4	Issues Particular to L2A Research .....	85
3.4	Artificial Language Paradigms .....	86
3.4.1	Rationale .....	86
3.4.2	Characteristics of Artificial Language Paradigms .....	86
3.4.3	Evaluation of Artificial Language Paradigms.....	87
3.4.4	Issues Particular to L2A Research .....	88
3.5	Connectionist-Type Computer Simulation Models .....	88
3.5.1	Rationale .....	88
3.5.2	Characteristics of Connectionist Models .....	89

3.5.3	Evaluation of Connectionist Models.....	90
3.5.4	Issues Particular to L2A Research .....	91
Chapter 4:	Literature Review: Empirical Evidence from L2A Studies .....	93
4.1	Overview.....	93
4.2	Overt Pronoun Constraint Studies .....	94
4.2.1	Kanno (1997, 1998) .....	97
4.2.2	Pérez-Leroux and Glass (1999) .....	106
4.2.3	Sheen (2000).....	109
4.3	Scrambling Studies .....	117
4.3.1	Hopp 2005.....	121
4.3.2	Williams and Kuribara (2008) .....	131
4.4	Scope Studies.....	144
4.4.1	Marsden (2009).....	146
4.4.2	O’Grady, Lee and Kwak (2009) .....	158
4.4.3	O’Grady (2013a).....	166
Chapter 5:	Critical Evaluation .....	170
5.1	Overview.....	170
5.2	Micro-Level Evaluation: Empirical Issues .....	171
5.2.1	Subject Characteristics.....	171
5.2.2	Task Design .....	177
5.3	Theoretical Issues .....	183
5.3.1	Generative Studies .....	183
5.3.2	Emergentist Studies .....	199

Chapter 6: Revisiting Poverty of the Stimulus: Discussion and Conclusions .....	220
6.1 Overview.....	220
6.2 Difficulties with Implementing the POS Argument .....	220
6.2.1 Determining Acquisition.....	221
6.2.2 Determining (In)Sufficiency of Input .....	228
6.3 Summarizing the Empirical Results of these L2 Studies: More Support for Generativism than Emergentism.....	235
6.4 Reflections .....	243
6.4.1 Common Ground .....	243
6.4.2 Other Experimental Tools Reconsidered .....	246
6.5 Conclusion .....	250
Bibliography .....	253

## LIST OF FIGURES

Figure 3.1: Context Presented as an Abstract Diagram .....	84
Figure 6.1: Subset-Superset Relationship Between L1 and L2: All Studies .....	229
Figure 6.2: Learning Path, Generative Prediction .....	238
Figure 6.3: Learning Path, Emergentist Prediction.....	238

## LIST OF TABLES

Table 4.1:	Possible Antecedents for Embedded Subject Pronouns in Japanese .....	98
Table 4.2:	Native Speaker Control Group (n=20), Percentage of Acceptances in Each Test Condition .....	100
Table 4.3:	L2 Learner Group (n=28), Percentage of Acceptances in Each Test Condition.....	101
Table 4.4:	Responses to Translation Task Coded by Pronoun Type and Story Type.....	108
Table 4.5:	Characteristics of Learner Groups by Proficiency.....	127
Table 4.6:	Percentage Acceptance of Sentence Type in Infinitival Paradigm, All Groups .....	129
Table 4.7:	Construction Types Included in Experimental Tasks .....	137
Table 4.8:	Acceptance Rates (%) by Native Speaker Group in Subject-Wide Scope Contexts.....	153
Table 4.9:	Acceptance Rates (%) by Native Speaker Group in Object-Wide Scope Contexts.....	153
Table 4.10:	Acceptance Rates (%) by Learner Group in Subject-Wide Scope Contexts.....	154
Table 4.11:	Acceptance Rates (%) by Learner Group in Object-Wide Scope Contexts.....	155

Table 4.12:	Individual Consistency (number of individuals in each category) of Performance on Canonical ‘every N’ Sentence Type, Object-Wide Scope Reading.....	157
Table 4.13:	Percentage of ‘True’ Responses by Subject Group .....	165
Table 5.1:	Empirical Assessment—Subjects .....	171
Table 5.2:	Mean Scores on Cloze Test Proficiency Assessment by Learner Groups .....	176
Table 5.3:	Adequacy Assessment—Subject Characteristics.....	177
Table 5.4:	Adequacy Assessment—Task Design Characteristics.....	181
Table 5.5:	Adequacy Assessment—Empirical Factors .....	182
Table 5.6:	POS Arguments Offered in Each Study.....	185
Table 5.7:	Adequacy Assessment, Generative Studies—Theoretical Factors, Part 1 .....	189
Table 5.8:	Adequacy Assessment, Generative Studies—Theoretical Factors.....	199
Table 5.9:	Adequacy Assessment, Emergentist Studies—Theoretical Factors, Part 1 .....	207
Table 5.10:	Adequacy Assessment, Emergentist Studies—Theoretical Factors.....	218
Table 5.11:	Adequacy Assessment—Empirical and Theoretical Factors .....	219
Table 6.1:	Native Speakers—Percentage Acceptance of Test Constructions .....	223
Table 6.2:	Which Studies Support Which Predictions.....	242

## **Acknowledgements**

I would like to express my sincere thanks to my committee, who have patiently nurtured my growth as a researcher. I thank Toshiyuki Ogihara for his support and encouragement while I wrestled with the challenges of semantic theory. I thank Karen Zagona for her insightful comments and discussions that helped restore clarity out of confusion and vagueness. Julia Herschensohn deserves special thanks for her expert guidance and for generously sharing her time and expertise throughout the course of my writing.

I'm grateful for help and support from many people within the University of Washington community and beyond. Particular thanks go to Jeffrey Kim, Doug Wulf, David Goss-Grubbs, Naoko Komoto and Alli Germain for key conversations that advanced my work. I appreciate Kyoko Sano for her interest and conversations. Many thanks go to Joyce Parvi and Mike Furr for constantly smoothing the way. I wish I could express my thanks in person to Laurie Poulson, who was a valued friend and colleague and is greatly missed.

The support of my friends and family was vital. I thank Paula Rowland for her friendship and her mentoring on matters personal and professional over the years. I thank Mary Gewolb for her sustained encouragement and support, including practical advice. I am also grateful to my book group friends who have been constantly supportive, and patient with my lack of preparation.

Finally, I thank my family for making room in their lives for this project. Their love and support was essential to my progress. Conrad's confidence in me gave me courage when I was feeling uncertain. Hillary served as a willing sounding board many

times, helping me to get past the roadblocks. Jason's interest and curiosity restored my sense of excitement when I was feeling burdened. Mónica's skilled coaching and insightful questions raised the bar on my presentations. I am deeply grateful to Steve for so many things, from conversations about abstract issues to assistance on concrete details of formatting, so I will simply say that this dissertation would not have been possible without him.

## **Dedication**

To my family, especially Steve.

# CHAPTER 1: INTRODUCTION

## 1.1 Purpose and Overview

This work applies evidence from adult second language acquisition studies of poverty-of-the-stimulus phenomena to the question of the existence of an inborn, language-specific system that constrains the set of possible human grammars. Which approach better accounts for the evidence: generativism or emergentism?

Generativism is a theoretical framework whose adherents hold that language acquisition can only be explained by positing a language-specific innate endowment often called Universal Grammar (UG), which takes the form of categories, representations, and grammatical operations of some kind (Chomsky 1959; Pinker 1984; Anderson and Lightfoot 2002; Berwick et al. 2011; and many others). There are several standard arguments for innateness with respect to child first language acquisition (L1A), including but not limited to: speed of acquisition, the existence of a critical period, convergence on a common grammar, language universals, and poverty (insufficiency) of input data (Sampson 2005). Many people consider the argument from the poverty of the stimulus (POS) to be the strongest argument for the existence of UG. Briefly, the POS argument states that if it can be shown that the linguistic input that a language learner receives underdetermines the grammar that he or she acquires, then the successful learner must be relying on innately supplied grammatical information. More controversially, some generativists hold on the basis of POS argumentation that UG also plays a role in adult second language acquisition (L2A) (see for instance Dekydsprotter 2001; White 2003a; Schwartz and Sprouse 2000, 2013; and many others).

Emergentism is a theoretical framework (less unified than generativism) whose adherents hold that both L1A and L2A take place using general-domain cognitive capabilities such as categorization, statistical analysis, pattern recognition, abstraction, and generalization (Elman et al. 1996; Ellis 1998, 2012; Tomasello 2003; Sampson 2005; O'Grady 2008, 2013; and many others). Most emergentists reject the idea of poverty of the stimulus, instead arguing (for instance Ellis 2012, p.198) that linguistic input is sufficiently rich and complex, determining the acquired language skills through characteristics such as form, frequency, function, and the interactions between these. O'Grady differs from the majority of emergentists in holding that poverty of the stimulus exists, but is overcome by general processing constraints rather than by something like UG.

Consideration of evidence from adult second language acquisition studies of poverty-of-the-stimulus phenomena will advance the investigation into innateness for two reasons. First, POS arguments are considered to be very strong. If emergentist models of acquisition are successful on these often complex and subtle phenomena, they pose a serious challenge to generativism and the linguistic nativist point of view. It is only relatively recently that emergentist-type studies have begun to consider accounting for acquisition of POS-type phenomena, enabling a comparison between the two approaches on this basis. Second, adult learners are different from child learners. Unlike children, adults have fully mature cognition, as well as well-developed pragmatic skills. If adult learners are not successful at acquiring complex phenomena such as scope ambiguities, it is not because they are not capable of perceiving the ambiguities or computing pragmatic

inferences. With those confounds out of the way, the role of input in learner performance can possibly be better characterized.

This dissertation examines three sets of L2A studies. The studies were selected for comparability. Each set of studies focuses on a common target phenomenon (the Overt Pronoun Constraint, scrambling, and quantifier scope ambiguities), where each of these phenomena has been argued to pose a POS problem for adult second language learners. Each set contains at least one study arguing that learners rely on UG, and one study arguing that learners rely on general-domain learning strategies. After providing a review and critique of each individual study, I argue that in the aggregate, the body of evidence offered supports the existence of UG. This conclusion is not based on a POS argument. I argue that while the POS argument is logically sound, POS arguments are difficult to apply in practice because it is difficult to establish the premises of the argument. Instead, my conclusion rests on the empirical evidence for the learners' developmental path. The evidence conforms more to the generative prediction for sudden, step-wise acquisition of grammatical contrasts, possibly on the basis of limited input, than to the emergentist prediction for incremental acquisition on the basis of a necessarily substantial amount of input.

Generativism and emergentism will be described in more detail in Sections 1.3 and 1.4 below, with particular attention to their respective positions on the existence of UG, the role of linguistic input, and the process of L1 and L2 acquisition. To be clear, this dissertation is not a full-fledged comparison between or referendum on generativism versus emergentism. A comprehensive comparison of the two approaches would require a comparison of their empirical coverage, their compatibility with evidence from human

processing, and their explanatory power regarding why grammars are the way they are and why acquisition is the way it is. This work compares the generative and emergentist approaches only with respect to the two issues mentioned above: UG and input (in)sufficiency.

## **1.2 Summary of Chapters**

The remainder of Chapter 1 will provide some brief background on key characteristics of the generativist and emergentist theoretical frameworks. The relevant characteristics will include their respective views on the language learning capacity, the process of acquisition including the role of input, and consequent predictions for a learner's developmental path.

Chapter 2 will begin by summarizing the logic of the POS argument for L1A (Chomsky 1965; Schwartz 1987; Baker and McCarthy 1981; Pullum and Scholz 2002; among others). I will then consider various ways that linguistic input has been argued to be impoverished. Several criteria will be treated in detail and illustrated with specific examples offered in the literature. This discussion will bring up considerations of the differences between input available to L1 learners versus L2 learners, leading to a restatement of the POS argument, this time for L2A (Cook 2003; White 2003a).

Chapter 2 continues by reporting the emergentist views on the POS argument and on the sufficiency of linguistic input to learners. After considering how to go about disconfirming the poverty of the stimulus, Chapter 2 concludes with a short summary.

Chapter 3 will describe and consider the merits of the experimental tasks and methodologies that are used in the studies under consideration in this dissertation: grammaticality judgment tasks, truth value judgment tasks, artificial language paradigms,

and computer simulation modeling. For each task, we will consider its usefulness and challenges, as well as issues that are specifically relevant to L2 investigations. The goal of this chapter is to provide an understanding of the nature and limitations of these tasks, as a basis for evaluating the learner performance data reported in the studies.

Chapter 4 will survey a selected group of L2A studies of POS phenomena. The chapter is organized into three sets of studies. Each set includes multiple studies of L2 acquisition of a particular phenomenon (the Overt Pronoun Constraint, scrambling, and scope ambiguity), and contains at least one study done within each of the competing frameworks (generativist and emergentist). In order to preview the critical evaluation in Chapter 5, the survey will highlight the nature of the phenomena under investigation, the study methodology, and the explanations offered for the observed learner performance.

Chapter 5 will provide a critical evaluation of the studies reviewed in Chapter 4, in light of the theoretical and methodological issues raised in Chapters 2 and 3, respectively. For each theoretical framework, the studies will be critiqued on empirical issues of subject characteristics and task design, and on theoretical issues such as the appropriateness of the conclusions to the data. This detailed, micro-level discussion in Chapter 5 will form the basis for a macro-level discussion of broader themes in Chapter 6.

Chapter 6 will conclude the dissertation with a critique of POS-style argumentation. I argue that although the POS argument is logically sound, in practice, establishing the premise of stimulus poverty is too difficult, hence the argument's conclusions cannot be drawn. Despite abandoning the POS argument, Chapter 6 will argue that the nativist position is better supported by the empirical evidence presented in

these studies than the emergentist position. This conclusion rests on the closer conformity of the empirical evidence to the generativist predictions than to the emergentist predictions for the shape of a learner's developmental path.

### **1.3 Generativism and Emergentism Compared on Key Issues**

At the most general level, the questions that motivate this dissertation are: How is the brain structured? How does the brain interact with the environment? These questions have been considered since at least the time of Aristotle and Plato, and still are examined in various fields of inquiry such as philosophy (Descartes 1637/1978; Locke 1690/1964; Hume 1739/1978), child development (Piaget 1923; Skinner 1957), and animal development (von Frisch 1950, 1974 for bees; Brenowitz 2004 for songbirds; Goodall 1986 for chimpanzees; among many others). The corresponding particular questions as formulated within the field of linguistics can be expressed as: Is there an innate language faculty? What is the role of linguistic input in the acquisition of language? Two current approaches in the linguistics literature, generativism and emergentism, address these questions (and others, but our attention is on these). Their respective positions are outlined in this section and the next. Their views on the process of language acquisition will also be summarized.

Generativism views language as a combinatorial system that generates an infinite set of possible sentences from a finite set of elements.<sup>1</sup> The generativist approach has its modern origin in the early works of Chomsky (1957, 1965). Although there are approaches that can be characterized as generative without being in the tradition of these

---

<sup>1</sup> Due to the variety of existing proposals, here, I am intentionally vague with respect to what these elements are exactly.

works (Pollard and Sag 1994, Head-driven Phrase Structure Grammar is one example), I will concentrate on the Chomskyan generativist tradition.

### **1.3.1. Structure of the Mind**

The generative view of the mind/brain is one of an entity with innate, possibly rich structure, taking the form of independent modules. Vision, audition, and face recognition are oft-cited examples of mental modules. And of course, in this framework, the language faculty is posited to be another. However, the claim is *not* that everything that we know about language is innate. The claim is that there is a core, invariant, language-specific faculty. Chomsky (1975, p.29) calls it ‘Universal Grammar’:

“Let us define “universal grammar” (UG) as the system of principles, conditions, and rules that are elements or properties of all human languages not merely by accident but by necessity – of course, I mean biological, not logical necessity. Thus UG can be taken as expressing “the essence of human language.” UG will be invariant among humans. UG will specify what language learning must achieve, if it takes place successfully.”

The characterization of the content of UG is an empirical question of primary importance to the generative endeavor. Proposals for the richness of content of the UG component have varied over the decades. Under previous theoretical programs, UG was richly structured with detailed sets of principles and parameter setting options that served to highly constrain the possible set of human grammars. However, under the more recent Minimalist Program (Chomsky 1995 and others), the trend is toward simplicity and reduction of the number of mechanisms that are necessary to postulate. These differences aside, the point is that under the generativist view, we have a mental grammar. That is, we have mental representations of components of language and possible structures of sentences.

Gregg (2005, p.840-841) distinguishes Chomsky-modularity and Fodor-modularity. Chomsky-modularity is simply the idea of the separateness of the contents of the language module: principles such as Binding Theory or c-command are unique to the language module. Fodor-modularity, on the other hand, is a more developed theory of mind (see Fodor 1983). According to Fodor, modules are domain-specific and autonomous. That is, each module (various ones for perception, also language) recognizes only the input that is relevant to its own computations, with such calculations being performed wholly inside the module without interaction with other modules. Modules interact indirectly through a central processor of some sort that can accept and use information from the different modules. In Fodor's (1983) view, the language module is only concerned with the well-formedness of sentences, and not with the processing of meaning. He considers certain things such as quantifier scope and binding of reflexives to be syntax, and hence included in the module. Other semantic factors such as reference and co-reference are out. On this view, knowledge of grammar is only one system that contributes to use of grammar. Extra-grammatical factors such as pragmatic inference and the incorporation of context into an interpretation would have to be coordinated through the central processor.

The modularity at issue here is cognitive, not anatomical. A cognitive language module as conceived here is not necessarily physically contiguous within the brain, and no such claim is made within the generative framework. Fodor argues for a fixed neural architecture, but does not necessarily make claim as to the exact shape of that architecture. In fact, see Progovac (2014) and Chesi and Moro (2014) for reviews of evidence of multiple locations of syntactic processing activity in the human brain.

### 1.3.2 The Role of Input

As Chomsky (2005/1980, p.35) puts it, “the stimulus does not contain the elements that constitute our knowledge”. For an example of the separation of grammatical knowledge and output, Newmeyer (2003, 2005) presents the following evidence. Consider the following question, and the possible and impossible answers (2005, p.145):

(1) Who does John<sub>i</sub> want to shave?

- a. Himself<sub>i</sub>
- b. Him<sub>j</sub>
- c. Me
- d. \*Myself
- e. \*Him<sub>i</sub>

The elliptical answers must conform to the argument structure that would be required by the appropriate full sentences (had the full sentences been uttered).

Newmeyer argues that examples of these types offer evidence that full argument structure is represented in the mind/brain even when it is not apparent in the utterance. So we see that surface forms do not necessarily resemble the abstract, underlying regularities that constitute grammatical knowledge. We will see further examples of this in Chapter 2, and of generative arguments to the effect that this disconnect between the linguistic input and the acquired grammar constitutes a poverty of the stimulus.

The role of input is not seen as providing direct evidence on which sentences are grammatical and ungrammatical. Instead, the linguistic input has an indirect role, providing the basis for grammatical analysis of crucial structure (Meisel 2011, p.52) rather than for simple observation and storage of surface forms. Chomsky (1988, p.35)

states: "...very limited input data suffice for the mind/brain to provide a rich and complex language...". On the generativist view, the input triggers but does not directly shape the grammar. Although it seems that the extreme UG versions of the past have given way to a greater role for input recently, in a strict triggering approach there is no role for input statistics such as frequency in the acquisition of grammatical rules.<sup>2</sup> In summary, on the generative view, a learner requires the right kind of input, and not necessarily a substantial amount of it. The studies considered in Chapter 4 below will provide evidence on these points.

### **1.3.3 The Generativist View Applied to Acquisition**

#### ***1.3.3.1 Child L1 Acquisition***

On the generativist view, child L1 acquisition (L1A) is an unconscious process of growth rather than a project in which the child is actively, intentionally engaged. Chomsky (1988, p.134) puts it this way: "Language learning is not really something that the child does; it is something that happens to the child placed in an appropriate environment...". The process of acquisition is pre-programmed under the guidance of UG. Children search the input for evidence for or against innately known options in the grammar. According to Yang (2004, p.455), "UG instructs the learner" with respect to what is important to notice in the input. The child's ultimately attained linguistic competence is claimed to not be sensitive to the quality and quantity of the input received, except in cases of extreme deprivation.

Research in the generative framework encompasses different proposals for the mechanisms of L1A. A classic approach going back to Chomsky (1965) is that children

---

<sup>2</sup> But see Newmeyer (2005) for discussion of evidence that processing constraints do affect grammars.

compare different grammars against the input, constantly testing and evaluating hypothesized grammars, accepting or rejecting them until the adult-like grammar is obtained. A more recent approach (Roeper (2000); Yang (2004); among others) is one where competing potential grammars are assigned probabilities. Proposals for formulating these hypothesized grammars include parameter-setting, where the parametric options are defined by UG (White 2003; Meisel 2011; many others) or by selecting features from a predefined set made available by UG (Lardiere 2008, 2009). Under any of these proposed mechanisms, UG is viewed as providing principles that serve to constrain the hypothesis space and therefore constrain all grammars. When children come to ‘know’ a language, they know the principles that delimit that language and therefore they know not only what is grammatical, but also what is ungrammatical according to the grammar that they have obtained.

### ***1.3.3.2 Adult L2 Acquisition***

In contrast, the availability and role of UG in adult L2A is a matter of controversy within the generativist research community. There are three possible positions to take on this issue: that UG is fully available to L2 learners (Schwartz 1987; Dekydsprotter, Sprouse and Anderson 1997; White 2003; Herschensohn 2009; Lardiere 2009; Schwartz and Sprouse 2013; among others); that UG is partially available to L2 learners (Bley-Vroman 1990; Hawkins and Chan 1997; Meisel 2011; among others); or that UG is completely inaccessible to L2 learners (Clahsen and Muysken 1986). Each position determines a different characterization of the process of adult second language acquisition, as described below.

Before turning to that discussion, it is important to note two of the well-known differences between L1A and L2A (Bley-Vroman 1990).<sup>3</sup> The former is achieved uniformly well, with child learners (in a particular community) converging on essentially the same grammar. Second language acquisition is the opposite of each of these: learners show variable attainment and multiple adult L2 learners may attain grammars that are measurably different from each other. These two properties are often termed variability and non-convergence, respectively. Any theory of L2A needs to address these differences.

With that in mind, let's review the predictions of the three positions on UG access in second language acquisition with regard to the L2 path and endstate. First, if L2 learners have full access to UG, then the L2 acquisition process is predicted to be *similar* to the L1 acquisition process in that UG guides and constrains learners' hypotheses about the target grammar. Learner grammars at all stages are predicted to conform to universal principles. More than that, all options provided by UG should be available to learners, even those options that are not instantiated in the native language. But such a scenario on its own would guarantee reliably successful acquisition of convergent grammars, just as in L1A. In order to account for the differences between L1A and L2A, proponents of full access suggest a variety of factors that may interfere with what would otherwise be an L1-like acquisition path. For instance, Schwartz (1987) suggests that input to L2 learners may be misleading in that it may contain ungrammatical sentences produced by other L2 learners. She also proposes (1987, p.171) a model of the mind whereby auditory and

---

<sup>3</sup> Additional differences include the importance of instruction, the role of affective factors (Bley-Vroman 1990), voluntary versus mandatory progress, and whether or not attained grammars are complete (Herschensohn 2000). A full discussion of the differences between L1A and L2A is beyond the scope of this chapter.

visual input in the second language go to a 'central processor' before going to the language module. The central processor has the opportunity to filter or block input from going to the language module at all, thus interfering with the input being available to interact with UG. Although the full access approach requires such types of emendations<sup>4</sup>, in its basic form, full access accounts for acquisition of constraints in the L2 that do not exist in the L1.

Under the partial access or no access hypotheses, second language acquisition is predicted to be substantially *different* from first language acquisition. Considering the partial access point of view, we first note that there are many ways to make the idea of partial access more precise (see for example Hawkins and Chan 1997; Hawkins and Hattori 2006; among others). A classic account (Bley-Vroman 1990) holds that adults have only indirect access to UG through their L1 grammar. Under this proposal, adult second language acquisition is not driven by UG. Instead, adults rely on their fully mature cognition and their domain-general learning strategies, consciously employed, in the attempt to learn the target language. Variability and non-convergence are thus expected. Only the part of UG instantiated in the L1 will constrain the learner grammars to conformity with linguistic principles. This proposal is part of the Fundamental Difference Hypothesis (FDH) which states that L1A and L2A are different in qualitatively significant ways. The FDH (Bley-Vroman 1990) was first offered at a time when UG was viewed as richly endowed with principles and with parameters, motivating the view of L1 acquisition as dependent on language-specific cognitive systems. Ensuing advances in theoretical syntax have evolved the view of UG from a richly detailed

---

<sup>4</sup> For now, I am putting aside discussion of possible effects of the additional differences noted in footnote 3, and also passing over a detailed discussion of L1 transfer.

structure to one whose contents are quite general, thus raising the possibility that UG is a minimal set of principles that is instantiated identically in all languages (Hale 1996). If the innate language module is invariant across languages, that is, not parameterized in any way, then there would be no such thing as partial access, only full access. Each learner's L1 would provide the full UG faculty.

Finally, with regard to the hypothesis that L2 learners have no access to UG, in this case learners must depend entirely on general cognitive and learning skills. Meisel (2011, p.94) notes that this strong form of the 'no access to UG' proposal is not widely held. Early proponents Clahsen and Muysken (1986) revised their view (1989) to acknowledge a role for UG through the L1. Such a revised proposal is similar to the partial access proposal discussed above.

The generative studies under consideration in Chapters 4 through 6 all adopt a full access hypothesis.

## **1.4 Emergentism**

The term 'emergentism' as used here is an umbrella term for a number of approaches that are distinguished by the mechanics of their accounts of language acquisition, but share in common the view that language is constructed as an accumulation of statistical analyses performed on the entire history of a language learner's experience, rather than from an inborn system of grammatical rules. The modern emergentist approach could be said to have its linguistics origin in the early Parallel Distributed Processing models, of which a classic example is Rumelhart and McClelland's (1986) connectionist model of child acquisition of past tense morphology. The connectionist modeling approach, described in more detail in Chapter 3 and exemplified in a study in Chapter 4, is noted for its

deliberate attempt to model possible physiological processes in the brain.<sup>5</sup> For the purpose of this section, emergentism encompasses this and other approaches that posit different mechanisms by which statistical language entities are created: connectionism (Elman et al. 1996; Christiansen and Chater 2001), cue-based models (Bates and MacWhinney 1987 for the Competition Model; MacWhinney 2012 for the Unified Model), and dynamic learning systems (see contributions in Verspoor et al. 2011). Without further exploring the details of these different approaches, we will summarize the ideas they share in common regarding the three points of the structure of the mind, the role of input, and the process of acquisition.

#### **1.4.1 Structure of the Mind**

In the emergentist framework, the mind is not necessarily considered to be a *tabula rasa*. To be clear, emergentism is not an anti-nativist point of view. (Note that innateness and domain-specificity are in principle independent. For instance, Gregg (2005, p.841) mentions honorifics as an example of (language) domain-specific knowledge that is not innate.) Since computer simulations are so frequently employed in this framework, they provide a good illustration of the unavoidability of postulating some innate capabilities. Plunkett (1998, p.103) points out: “Of necessity, all the [computer simulation] models make specific assumptions about the nature of the learning environment and the nature of the network architecture that is brought to bear on the problem. These assumptions constitute the theoretical stance taken by the authors in characterising the trade-off between computational machinery and environmental resources. Note that this trade-off

---

<sup>5</sup> See Medler 1998 for a useful review of connectionist history and methods. It is worth noting that connectionist models have been applied to diverse areas such as engineering and weather, not just linguistics.

is never conceived as a tabula rasa approach to learning. Specific computational processes are always defined in order to demonstrate how learning gets off the ground.” When a computer simulation model is constructed (as it must be) with a certain number of individual units, with certain connectivity patterns between units, and with specific algorithms for operating on the input data, all of these decisions constitute an assumption about what is available to the learner in an initial state.

That being said, the emergentist view rejects the existence of an innate *language-specific* faculty. O’Grady (2008, p.448) offers what he calls the ‘emergentist thesis for language’: “The phenomena of language are best explained by reference to more basic non-linguistic (i.e., ‘non-grammatical’) factors and their interaction—physiology, perception, processing, working memory, pragmatics, social interaction, properties of the input, the learning mechanisms, and so on.” Under this approach, language learning ‘emerges’ out of the interactions between various non-linguistic factors.

#### **1.4.2 The Role of Input**

In contrast to the generativist emphasis on abstract rules, emergentists focus on particular constructions and language in use. The learner has access to surface forms, and applies general learning mechanisms to the steadily evolving and increasing ‘database’ of their individual experience. Concrete experience of the input is necessary and sufficient to the learning. Holme (2013, p.605) states: “Language acquisition, then, is not the abstraction of hidden structures from input but the learning of the forms that the learner actually encounters.” Input is determinative, not simply a ‘trigger’; put in the terms of computer simulation models, repetition of input results in strengthened network connections which

essentially constitute the knowledge of language. In this sense, linguistic input is directly formative of language acquisition.

On this view, frequency is of high importance. This claim has been supported for L1A with empirical studies, including Brooks and Tomasello (1999), who found correct production of passive utterances using nonce verbs by children of the age of 3 to 3 ½, after extra exposure. Normally, passive is acquired much later (up to 2 years later).

‘Token’ frequency refers to how often an item appears in the input. ‘Type’ frequency is defined as “the number of distinct lexical items that can be substituted in a given slot in a construction” (Ellis 2012, p.198). For instance, the English past tense *-ed* has a very high type frequency, whereas the irregular past tense *went* has a very high token frequency.

Bybee and Thompson (2000) show that both types of frequency are relevant to acquisition. Since people learn via experience and interaction with their environment, we need lots of input. Ellis (2002, p.167) observes: “The enormity of the lexical pool, the range of frequencies from 60,000 per million down to 1 per million and below, and the wide range of different linguistic constructions, when considered from the point of view of sampling theory, makes it clear that the necessary representative experience for fluency must be vast indeed.”

In summary, on the emergentist view, the surface forms of the linguistic input contain all the information necessary for language acquisition. All the learner needs is sufficient quantity so that the relevant statistical analyses will converge to correctly predict what is grammatical.

### **1.4.3 The Emergentist Point of View on Language Acquisition**

#### ***1.4.3.1 Child L1 Acquisition***

Similarly to what we saw in the generativist program, the emergentist program has not settled on one particular account of language acquisition. However, several ideas that are held in common within the framework will be described here. First, reliance on general cognitive and social skills entails that learning language is fundamentally like learning anything else (music, chess, etc): a process of pattern analysis and distributional analysis on the one hand, and a social process of imitation and intention-reading on the other. On this usage-based approach, learning occurs in a piecemeal and incremental fashion, and there are many possible developmental paths. Ultimately, the learner's knowledge comprises "...the learner's entire collection of memories of previously experienced utterances" (Ellis and Larsen-Freeman 2006, p.565). Note that such a 'collection' will be unique to each learner; therefore variability between learners (even child learners) is expected. Convergence is not guaranteed in principle, but could be explained on the basis of factors such as universal processing constraints and transmission of social and cultural conventions. Unlike in the generativist program, knowledge of ungrammaticality is not a primary concern.

#### ***1.4.2.2 Adult L2 Acquisition***

As we will see in Chapter 4, the investigation of L2A within the emergentist framework is in its early stage. The relative scarceness of empirical studies (compared to those done in the generative framework so far) does not prevent us from considering the predictions that the program makes. First, the same basic domain-general learning mechanisms are available to adults as to children. MacWhinney (2008, 2012) proposes a "Unified

Competition Model' which identifies several 'risk factors' and 'support factors' that are present in children as well as adults, but which respectively hinder or facilitate the learning process. The difference between child and adult learning is attributed to a different balance and interaction among these factors.

Although common learning mechanisms are posited, the differences between the child L1 initial state and the adult L2 initial state are predicted to affect the developmental paths for each type of learner. Whereas an infant enters the world with some small experience of language in the womb (de Boysson-Bardies 1999), an adult L2 learner has a fully-established language instantiated in his or her neural architecture. This first language provides entrenchment of pre-set habits that may need to be overcome (see Tomasello 2003), but as Myles (2013) points out, the L1 also provides resources and strategies for communication if a needed L2 construction has not been learned yet. Adult learners also have fully mature cognitive capabilities (for instance, longer attention span, better working memory, mature theory of mind). Further, adults have more mature social skills and a far better understanding of social meaning.

In summary, the emergentist program makes the following predictions with respect to L2A. First, variability among learners is expected on the basis of variation in input, both linguistic and non-linguistic. Convergence is unexpected, on the same basis. As to the question of ultimate attainment, this approach would seem to predict that learners should in principle be able to attain native-like behavior. Continued input should be able to eventually overcome whatever barriers to learning may exist.

## **1.5 Conclusion**

The purpose of this work is to apply evidence from adult L2A studies of complex phenomena to the question of the existence of UG. This chapter set out a few key proposals in the generativist and emergentist programs with respect to the structure of the mind, the role of input in acquisition, the processes of acquisition and the consequent predictions for ultimate attainment and developmental paths. In the coming chapters we will explore the ways in which these theoretical frameworks investigate and justify their claims. We begin this investigation in the next chapter by examining the POS argument.

## CHAPTER 2: POVERTY OF THE STIMULUS (POS)

### 2.1 Overview

In Chapter 1 we saw that the general idea that experience underdetermines knowledge has ancient roots, and that the particular question of the gap between linguistic experience and linguistic knowledge has been of serious interest to generative linguists since the mid-twentieth century. According to Thomas (2002, p.52), it was not until 1980 that the phrase ‘the poverty of the stimulus’ was introduced by Chomsky (1980, p.34) as a moniker for this gap, the postulated existence of which is a key assumption in what is now considered the cornerstone argument for linguistic nativism, ‘the argument from the poverty of the stimulus’.<sup>6,7</sup> Section 2.2 will provide a general statement of the POS argument for L1A (Pullum and Scholz 2002). Section 2.3 will explore the generativist view of the key assumptions regarding learning and input, considering both L1 and L2. This discussion will lead to presentation of a revised statement of the POS argument, adapted for L2A (based on Cook 2003). The section will conclude with a discussion of how to go about confirming the poverty of the stimulus. Section 2.4 will discuss the emergentist views on the POS argument and on the content and sufficiency of linguistic input to learners. Section 2.4 will conclude with a look at how to go about disconfirming the poverty of the stimulus. Section 2.5 concludes the chapter with a brief summary.

---

<sup>6</sup> Singling out POS is a matter of broad consensus, however, there are some researchers who disagree. In particular, Sampson 2004 considers the claim of the existence of language universals to be a more compelling argument than POS for innate linguistic knowledge (see his Chapter 5).

<sup>7</sup>In the literature, we find several additional terms, sometimes conflated, not always appropriately so: ‘the projection problem’ (Peters 1972); ‘the logical problem of language acquisition’ (Baker and McCarthy 1981); ‘Plato’s problem’ (Chomsky 1986); and ‘the deductive gap’ (Baker 1979), to name a few. In fact, only the last is synonymous with the basic idea of stimulus poverty. The first three more accurately refer to the general question: how is it that learners acquire grammars for which there is little or no evidence in the input? See Thomas (2002) for a discussion of the origins and nuances of the various terms.

## 2.2 The Argument from the Poverty of the Stimulus: Basic Statement for L1A

We will use the formulation provided in Pullum and Scholz (2002, p.18), because the logic is clearly laid out and the argument is given in general terms rather than tied to a particular phenomenon or a particular description of the input. By ‘data-driven learning’, Pullum and Scholz (henceforth P&S) mean some (unspecified) process in which the learner applies domain-general capabilities and learning algorithms to the linguistic input. By ‘innately-primed learning’, they mean that the learner makes use of innate, language-specific faculties (again, the detailed nature of which are not specified). Here is their presentation:

(1) Pullum and Scholz (2002) statement of the POS argument

“a. Human infants learn their first languages either by data-driven learning or by innately-primed learning. [Disjunctive premise; by assumption.] (here and throughout, brackets in original)

b. If human infants acquire their first languages via data-driven learning, then they can never learn anything for which they lack crucial evidence. [By definition of data-driven learning.]

c. But infants do in fact learn things for which they lack crucial evidence. [Empirical premise.]

d. Thus human infants do not learn their first languages by means of data-driven learning. [From (b) and (c), modus tollens.]

e. Conclusion: human infants learn their first languages by means of innately-primed learning. [From (a) and (d), disjunctive syllogism.]”

To put it less formally, the essential POS argument states that if the input to the learner lacks evidence that is crucial for identifying the target grammar, then successful language acquisition must rely on prior innate, language-specific knowledge. The premise is that learners go beyond the data to which they are exposed, farther than any domain-general learning mechanisms such as imitation, analogy, statistical analysis, or induction could take them. In what ways do learners ‘go beyond’ the input? Specific claims for these include, among others, the production and comprehension of novel sentences; understanding of entailment relations; ambiguity, and sometimes even a lack of ambiguity that might be unexpected, given the input. We will see proposals for specific examples of ‘going beyond the input’ in Section 2.3 below where we examine the possible deficiencies of the input, and also in Chapter 4 where we review existing POS studies.

A few comments are in order. First, the argument as presented is specifically phrased for L1 acquisition, and in Section 2.3 we will examine whether or not the argument needs modification for L2A. Second, in part a., P&S set up a stark binary choice between data-driven learning (DDL) and innately-primed learning for first languages. In a footnote, they say that this is possibly (probably?) a false dichotomy. As mentioned in Chapter 1 above, generativists do not deny a role for input or for general-domain learning mechanisms in the language acquisition process as a whole. The choice is not ‘DDL *or* innately-primed language-specific learning?’, but ‘DDL alone *or* DDL along with innately-primed language-specific learning?’. As Berwick et al. (2011, p.1210) put it, “The point of a POS argument is not to replace appeals to “learning” with appeals to “innate principles” of Universal Grammar (UG). The goal is to identify

phenomena that reveal [innate, domain-specific] contributions to linguistic knowledge, in a way that helps characterize those contributions.” The idealized argument as given by P&S abstracts away from the detailed hypotheses about grammars and about learning mechanisms. For the purpose of illustrating the general argument, this level of generality for the proposed faculties and learning mechanisms is appropriate. However, we will see in Chapters 4 and 5 below that once we are considering particular elements of grammar, the specifics of hypotheses about grammatical structure and about learning mechanisms are crucial to the success of the argument.

We will examine the POS argument in more detail in the next two sections, first from the generativist perspective (Section 2.3) and then from the emergentist perspective (Section 2.4).

## **2.3 The Generativist View on the POS Argument**

### **2.3.1 The Generativist View on the Key Assumptions**

The POS argument rests on two premises, both of which are encoded in part c. of the argument as stated by P&S, given in (1) above. The first premise is that learning has taken place. The second is that the linguistic input is too impoverished to be able to account for the learning: input underdetermines the acquired language. These premises need to be established in order to draw the conclusion. In this section, we inspect these assumptions and how they can be established in more detail so that we will be prepared to evaluate the quality of the POS arguments offered in the L2A studies surveyed in Chapter 4. We will investigate each assumption from the point of view of L1 acquisition and L2 acquisition, and we will see that the assumptions may hold differently in the two different processes.

### ***2.3.1.1 Learning***

The first assumption that we will consider is that learning has taken place. For L1, on first thought this assumption simply seems obvious and uncontroversial. Human infants grow to become native speakers of the languages to which they are exposed. The child attains a stable state in which he or she has acquired “a grammar that generates *all* and *only* the sound/meaning pairs of the local language, modulo limitations on vocabulary” (Crain and Pietroski 2001, p.144, emphasis original).

If the assumption about successful learning is perhaps obvious for L1 acquisition, its appropriateness for L2A is less so. How should we define ‘success’? If we were to require ‘native-like performance’, we might be setting an impossible bar. But even as we take the view (following Lardiere 2008, 2013; Slabakova 2006, 2008; see discussion in Section 1.3 above) that L2ers in principle may attain native-like performance in at least some respects, we must acknowledge the body of L2 studies that provide evidence against such attainment. Several studies (Coppieters 1987; Birdsong 1992; Abrahamsson and Hyltenstam 2009; among others) show that L2 learners who appear native-like on some measures can nevertheless be distinguished from native speakers. And processing studies reveal that non-native processing is slower than native processing, even when performance on behavioral tasks is comparable (see, for instance, White and Genesee 1996).

In fact, we cannot say for sure what it is that L2-ers acquire. Learner behavior that appears similar or nearly identical to native behavior does not guarantee the acquisition of the target grammar. White and Genesee (1996) show that L2ers may attain a grammar that is systematically different from the target in subtle ways. Meisel (2011, p.174)

argues, “...in my view, [many studies] have at best shown that some L2 learners are able to use certain constructions which can be argued to depend on UG-related grammatical knowledge in a fashion similar to native speakers of the target language in question.” Because of these issues, it is difficult to define ‘successful learning’ in an L2 setting.

On the other hand, Bley-Vroman’s (1983) *comparative fallacy* argues against comparing L2 learner grammars to native speaker grammars, suggesting instead that L2 learner grammars be examined on their own. White (2003b, p.25) points out that L2 interlanguages<sup>8</sup> may differ from target grammars yet still show evidence of being constrained by UG principles and parameters. For instance, Hawkins and Chan (1997) and White (1992) argue for an alternative, UG-compatible analysis of L2 learner interlanguages that seem to contain subadjacency violations. In other words, L2 performance that is not completely target-like is not necessarily indicative of an absence of UG in L2A.

On this view, we can recruit POS arguments for L2A without requiring that the target grammar has been acquired (Herschensohn, p.c.). The basic requirement for the POS argument is that learners acquire structures and grammaticality contrasts that cannot be induced from the input. Grimshaw and Rosen (1990, p.191) have argued the same standard for assessing child L1 acquisition (here, speaking specifically about binding theory): “If subjects treat the two conditions differently, they are detecting a difference in the acceptability of the sentences in the two conditions, and we can reason from this that must have knowledge of the binding theory”. So as long as L2 learners trend toward native-like performance, demonstrating a relative contrast between grammatical and

---

<sup>8</sup> I adopt this term from Selinker (1972) to indicate systematic, non-native grammars

ungrammatical constructions, then the ‘learning’ assumption of the POS argument is satisfied. For an example of the relative-contrast-type approach, see Dekydsprotter et al. (2001) on the acquisition of asymmetries in scope interpretations. In Chapter 3, we will examine the experimental methodologies in use for investigating and arguing for such learning. But for now we turn to the second key assumption that underlies the POS argument.

### ***2.3.1.2 Stimulus Poverty: Deficiencies in the Input***

In this section we explore the idea of stimulus poverty, that is, the idea that the primary linguistic data (PLD) are impoverished to the extent that learners are unable to identify the target grammar based on the input. Several specific deficiencies have been proposed: for one thing, natural language input contains performance errors; for another, the PLD is finite while our language capacity is infinite; further, the PLD might not contain negative data in the form of corrections; and finally, the positive examples in the PLD may be deficient in numerous ways. For instance, certain utterance types may be rare or complex, others under-informative, and some patterns of utterances have been argued to be misleading. Some of these deficiencies are quantitative and some are qualitative. What follows is a brief discussion of each from a generativist perspective. For each type of deficiency, the discussion will make note of any differences between the characteristics of L1 and L2 input. Information in this section will form the basis for an evaluation of POS arguments in L2 studies, to be presented in Chapters 4 and 5.

#### **DEGENERACY**

Many performance errors occur in the course of natural language discourse. Speakers hesitate, sometimes (often) producing non-words such as *aaah* or *mmmm*. Utterances can contain slips-of-the-tongue, be incomplete, or even ill-formed. Although child-directed

speech is rather free of errors (Snow and Ferguson 1977), adult L2 learners, especially in immersion settings, certainly hear a full spectrum of such noisy data. Schwartz (1987, p.200ff) suggests that L2 learners in both instructed and immersion settings may receive aberrant data from an additional source: other L2 learners. To the extent that learners converse with fellow language students or with non-native speakers in other settings, their input may contain interlanguage data that does not conform to the target language. From a generativist point of view all learners, L2 learners in particular, must attune to the reliable data and ignore the variable, possibly misleading degenerate data, the existence of which is uncontroversial.

#### **FINITENESS**

By virtue of our finite brains and finite existence in time, it is also uncontroversial that the input learners receive comprises a finite subset of the infinitely possible sentences of a language. This is true for L1A and L2A. Should we consider finiteness to be a severe deficiency? Cowie (2010, p.17) points out that having a finite sample is ‘endemic to any kind of empirical inquiry’; in any endeavor, we can only make a finite number of observations. A capacity for induction or generalization (comparable perhaps to those used for other sciences) allows language users to ‘go beyond’ the finite set of examples. L1 and L2 learners alike develop the capacity to produce and comprehend infinitely many novel sentences. So, the challenge is not in the leap from finite to infinite. Rather, the challenge is that any finite set of observations is in principle compatible with an infinite number of hypotheses. An easy illustration of this principle involves visualizing a set of points plotted equidistantly along the x axis of a graph, all at the same height. The points can be connected with a straight line, or they can be connected by an infinite number of sinusoidal curves of fixed period but varying amplitudes (see Piatelli-

Palmarini 1980, p.259-260 for discussion). There is no information contained in the points themselves that determines which curve fits the data best (where the choice of ‘curve’ includes the straight line). In order to induce the correct hypothesis, additional information is required. A classic linguistic example concerns the formation of yes/no questions in English. On the assumption that the question in (2) below was formed by auxiliary fronting on the declarative sentence *The dog is hungry*, a learner might in principle need to consider several different hypotheses for the grammar that generates the question. Sample hypotheses are given in (3a) and (3b) (question and hypotheses taken from Lasnik and Uriagereka (2002, p.147)):

- (2) Is the dog hungry?
- (3) (a) Front the first auxiliary.
- (b) Front the auxiliary in the matrix Infl.

A learner confronted with example (4) would be able to rule out hypothesis (3a):

- (4) Is the dog that is in the corner hungry?

as (4) does not result from fronting the first auxiliary in the corresponding declarative (5):

- (5) The dog that is in the corner is hungry.

Now we have a new set of finite data, namely (2) and (4). Lasnik and Uriagereka (2002, p.149) offer three more possible hypotheses, each compatible with the new data set. The proposed hypotheses are listed in (6):

- (6) (a) Front the first auxiliary (that comes after an intonation change).
- (b) Front the first auxiliary (that comes after the first complete constituent).
- (c) Front the first auxiliary (that comes after the first semantic unit you parsed).

Enter another piece of disconfirming data (Lasnik and Uriagereka 2002, p.149):

(7) Will those who are coming and those who are not coming raise their hands?

On the basis of the hypotheses in (6), one would expect the second occurrence of *are* to be fronted, rather than *will*. At this point we now have a new, finite data set consisting of (2), (4), and (7). Other hypotheses that generate these three examples could be proposed.

So we see that a finite set of data is always consistent with multiple hypotheses.

The learner's task is to identify the correct hypothesis.

### **NEGATIVE EVIDENCE**

If learners are capable of generalizing from data that underdetermine a target grammar, how is it that learners discern the correct boundaries of grammaticality? How do L1 learners uniformly acquire the same language, and how do both L1 and L2 learners come to know that a particular item or structure does *not* occur in the target grammar? To take an example from L1, it is well-known that child learners do overgeneralize.

Overgeneralization has been documented in several areas, including (Bowerman 1988, p.79-81): dative alternations (*I said her no*); lexical causatives (*Don't giggle me*); passives (*Mommy will get lightnined*); locative alternations (*...I spilled it of orange juice*); and *un*-prefixation (*Uncapture me!*). Another famous example is the use of regular past tense morphology on irregular verbs (*goed* instead of *went*). A learner's production of such constructions indicates that he or she has hypothesized a grammar that is 'larger' (that is, more permissive) than the target grammar. No amount of additional positive evidence will contradict this hypothesis. Additional grammatical examples will support the 'smaller' grammar, but will not disconfirm the errors. Yet somehow children eventually retreat from these overgeneralizations, and some L2-ers are able to retreat eventually.

It has been argued that negative evidence could play a crucial role in the acquisition of constraints, but that the primary linguistic data (PLD) are deficient in this respect: negative evidence is not available to the learner. Let us briefly explore the types, existence and usability of negative evidence.

Direct negative evidence consists solely of direct corrections in response to particular ungrammatical utterances. Such evidence has not been documented to be systematically offered to children, and anecdotal reports of attempts to offer direct correction show the children being impervious to it. Of course, instructed L2 learners do receive negative input. There is a body of studies which suggests that both L1 and L2 learners do in fact receive linguistic input that *functions* as corrective evidence.<sup>9</sup> An array of feedback that assists learners in identifying ungrammaticalities in their production has been shown to be available, and possibly effective. Various techniques for feedback include repetition, recasts (reformulations), prompts, elicitation, requests for clarification, and (for L2 learners) metalinguistic explanations. To understand the potential utility of linguistic input that is not specifically corrective, consider reformulations.

Reformulations are offered immediately after an utterance that contains an error; reformulations maintain the intended meaning of the original utterance, but offer an alternate form. The contrast between the two forms calls attention to the locus of the error, and provides a grammatical alternative. Although in one sense a reformulation is a positive exemplar, Choiunard and Clark (2003, p.666) argue that recasts “fit the classic definition” of negative evidence in that they provide a direct contrast to the learner’s utterance, thus signaling that the learner’s utterance may have been ungrammatical. Note

---

<sup>9</sup> One example is discussed here. The interested reader is referred to reviews in Cowie 2010, p.45-46 for L1 and Leowen 2012 and Carroll 2001, Chapter 8 for L2.

that Chouinard and Clark are forced to say “may have been” ungrammatical, because it is the case that for many meanings, alternate grammatical forms exist for expressing it.

Chouinard and Clark found that recasts were offered to children even after grammatical utterances, but at much lower frequency than after error productions. Still, in this way, there is a difference between negative evidence strictly defined, in which it is clearly stated that an error was produced, and the types of feedback that are being argued to act as corrective evidence but are not so explicit. Loewen (2012, p.27) suggests that “it may be better to view explicitness as a continuum rather than a dichotomy”.

L1 learners receive this potentially corrective input from parents and caregivers; Chouinard and Clark (2003) show that younger children receive more frequent feedback (in the form of recasts) than older children do. L2 learners receive feedback primarily in the classroom, if they are instructed. Loewen (2012, p.29), citing Mackey (1999) notes that an L2 learners’ ability to use feedback may depend on proficiency: “..with studies suggesting that learners need to be developmentally ready to benefit from feedback”. Loewen (2012) and Carroll (2001) both remark that research on the question of whether or not feedback is effective shows mixed results, however, they feel that the balance of evidence is in favor of its usability and use.

From the point of view of the POS argument, it is not enough that some learners receive appropriate and helpful negative evidence. If negative evidence is crucial to the learning process, then a PLD that is sufficient for learning will offer appropriate negative evidence to all learners in sufficient quantity at the right times. Contrary to that requirement, Loewen (2003) found that the quantity of negative feedback offered to L2 students was inconsistent across classrooms in the same school, and even across students

within the same classroom. I am not aware of any studies that attempt to quantify the amount of negative evidence that an immersion learner might receive. Intuitively, one would think that classroom instruction would provide considerably more feedback than naturalistic settings. Thus, current information suggests that negative evidence cannot be guaranteed to be a part of every L2 learner's experience.

We will consider one final alternative to direct negative evidence as a source of constraints on overgeneralization, before concluding this discussion on the taxonomy of negative evidence, and that is: a principles approach. The section above on finiteness of the PLD touched on the fact that any discussion of the adequacy of input needs to take into consideration what specific learning mechanisms are available to the learner. For instance, induction alone on the basis of finite data is not sufficient to determine a single correct hypothesis for a grammar, as there would be no mechanism to retreat from overgeneralizations. But what if other learning principles are at work along with induction? For example, a 'uniqueness principle' (Pinker 1984 and others) could act to impose a 'one form-one meaning' limit. An L1 or L2 learner could then correct an overgeneralization on the basis of only positive data by noticing that the conventional form (i.e. that used by adults in the case of L1, or used by native speakers in the case of L2) is different from the novel form in their own utterances. For example, a child who utters *telled* but notices *told* in adult speech will retreat to the adult form, correcting the overgeneralization. This particular proposed principle has a number of problems, including some confusion as to why the overgeneralization would have occurred in the first place, given such a principle and prior correct production of the word *told*. In addition, overgeneralization of regular past tense persists in children's production for

several months, unexpected under this approach (Crain and Pietroski 2001, p.145). Finally, while there may be unique verb forms, when it comes to the acquisition of syntactic constructions there are multiple forms for expressing the same or similar meaning. Unless the L1 or L2 learner can identify a single unique form to acquire, a ‘uniqueness principle’ is not helpful.

A second learning principle that has been offered as a compensatory constraining mechanism in the absence of negative data is the Subset Principle (Berwick 1985). White (1989, Chapter 6) discusses the Subset Principle from the point of view of L2. Under the Subset Principle, grammars are ordered in a subset/superset relationship. Given two grammars that meet the Subset Condition, the following predictions hold for L2 learners:

- when the L2 is a subset of the L1, the learner should posit the more restrictive grammar and therefore is predicted to acquire the L2 rather than to transfer his/her L1 superset grammar;
- when the L1 is a subset of the L2, the learner should be able acquire the superset L2 grammar on the basis of positive evidence.

White (1989), Inagaki (2006) and others have investigated these claims. Evidence to date is against the former claim, but mostly in favor of the latter. Several studies cited in White 1989 show learners overgeneralizing their superset L1 behavior to the subset L2 rather than positing the more restrictive grammar based on the input. Inagaki (2002) reviewed several studies which demonstrated successful acquisition of a superset L2. However, Inagaki (2006) studied manner-of-motion verbs with PPs, which in English can have directional or locational readings (e.g. *John swam under the bridge*), but in Japanese have only the locational reading. He found that Japanese learners of English had trouble

acquiring the directional reading, counter to the above prediction. Inagaki (2006) argues that in order to successfully acquire a superset L2, the positive evidence must be frequent and clear.<sup>10</sup>

### **POSITIVE EVIDENCE**

Positive evidence is the final category of stimulus poverty that we will discuss. Positive evidence is simply the set of examples of grammatical sentences that are part of the target language, and that learners hear.<sup>11,12</sup> Generativists argue that positive data is not completely up to the task of identifying the target grammar, in three important ways: examples of some phenomena are too rare or complex to be useful; strings of sounds or words can be uninformative as to their structure; and patterns of sentences can be misleading, with the potential to create incorrect expectations about grammaticality. We will briefly consider each deficiency in turn.

Possibly the most-cited example of rarity in L1 input is relevant to the acquisition of structure dependency in yes/no question formation in English, an example of which was given above (see p.29). The claim is that learners need examples of auxiliary inversion over complex NPs in order to arrive at the correct structure-dependent hypothesis for these types of questions. The rarity of such examples has been debated (see Chomsky in Piatelli-Palmarini 1980, and elsewhere; Pullum and Scholz 2002; Legate and Yang 2002; Sampson 2004; among many others). Many sources allude in a general way to the rarity of positive examples such as *Is the dog that is in the corner hungry?*, but Legate and Yang (2002, p.157) present quantitative evidence. Examining

---

<sup>10</sup> these descriptive terms are not quantified, but need to be, for this prediction to be tested.

<sup>11</sup> I am putting aside written input for now. However, availability of and reliance on written input is a significant difference between L1A and instructed L2A, and for that reason deserves consideration.

<sup>12</sup> For a different definition of positive evidence that involves parsing in addition to the speech signal, see Carroll (2001, p.18ff)

the Nina corpus in the CHILDES corpus<sup>13</sup>, they find 46,499 sentences spoken by adults, 20,651 of which were questions. Of the questions, none (zero!) were of the form of interest, that is, auxiliary inversion over a complex NP.

Several factors combine to weaken the force of an insufficiency argument based on rarity of occurrence. First, even though Legate and Yang's zero count sounds definitive, Schwartz and Sprouse (2013, p.151) point out that "The mere absence of tokens of a sentence type from any given corpus can never serve as documentation that such sentences are "unavailable to the learner"". Second, there is a question about what sentence types serve as relevant evidence. Pursuing the particular example of subject auxiliary inversion further, Pullum and Scholz (2002, p.42) argue that sentences exhibiting non-subject *wh*- movement such as *When will the man who is in charge \_\_\_ be back?* also provide relevant evidence for structure dependency and therefore yes/no question formation, thus arguing that useful positive evidence does not have to be direct. Yang (2002, p.100) concurs that such *wh*- questions assist in ruling out structure-independent hypotheses. If you accept in general that similar-but-not-exact data can contribute to the acquisition of a given phenomenon<sup>14</sup>, then the question of rarity must be established on a different (and larger) set of examples.<sup>15</sup> Finally, even when an argument from rarity can be established, it may not be persuasive in that without a clear understanding of the amount of input that would be *sufficient*, it is difficult to declare any non-zero amount of exposure to be *insufficient*.<sup>16</sup> Under a generative-type 'triggering'

---

<sup>13</sup> Child Language Data Exchange System

<sup>14</sup> We will see in Section 2.4 that this is a controversial claim and it must be evaluated in the context of what learning mechanisms are assumed.

<sup>15</sup> Pullum & Scholz (2002, p.42) find several examples of structural-dependence auxiliary fronting in the first 500 interrogatives in the Wall Street Journal 1987-89 corpus (Linguistic Data Consortium 1993).

<sup>16</sup> See Legate and Yang (2002, p.155-159) for an attempt to calculate the required exposure to a certain construction whose acquisition is being investigated by measuring the frequency of occurrence of a

view of the role of input, it is not necessarily the case that a lot of data will be required for acquisition. Chomsky (1988, p.35) states (for L1A) that “...very limited data suffice for the mind/brain to provide a rich and complex language...”.

When evaluating the rarity of input for a given phenomenon in L2A instead of L1A, we must also consider the fact that the learner’s L1 may instantiate a target construction. This in effect renders the ‘input’ to be rich with respect to that construction, even if a learner is exposed to relatively few positive examples. So we can see that rarity of positive input, even when it exists, might not be a severe deficiency in L1 or L2.

A clearer example of impoverished data lies in the under-informative nature of the data. Clark and Lappin (2011, p.56) put it this way: “An important feature of the PLD is that it is “raw data””. For instance, the stream of speech does not contain specific information about grammatical categories, about constituency, or morphosyntactic representations in general. Under the generative assumption that learners are acquiring a set of features, representations and constraints on representation, this lack of information presents challenges to the learner. For the L1 learner, the challenge is severe. For the L2 learner, as we saw earlier, the native language provides a source of information, albeit sometimes in conflict with the L2. But the L2 learner will still be challenged by phenomena that exist in the L2 and are not instantiated in the L1, in the same way that an L1 learner is challenged to discover the correct set of possible structures.

A significant example of the under-informative nature of the positive data is structural ambiguity. Consider the following example (Berwick et al. 2011, p.1212):

(8) The senator called the donor from Texas.

---

different construction that has already been shown to be acquired. The assumption that required frequencies would be the same across constructions seems problematic to me.

Example (8) could be structured to be interpreted in at least two ways. Perhaps the senator called the donor who was from Texas (9), or the senator called the donor while he (the senator) was in Texas himself (10):

(9) The senator [called [the [donor [from Texas]]]]

(10) The senator [[called [the donor]] [from Texas]]

Although ambiguity is rampant in natural language, Berwick et al. point out that ambiguity can be surprisingly constrained. Consider the following examples (Berwick et al. 2011, p.1211):

(11) The goose is ready to eat

(12) Darcy is eager to please

(13) Darcy is easy to please

While the goose in (11) could be the subject or the object of eating, Darcy in (12) can only be the subject (the one who is eager to do the pleasing), and Darcy in (13) can only be the one who is easily pleased by others.

This empirical fact that ambiguity exists in some sentences and not others contributes to the severity of the under-informative nature of the positive evidence (for discussion, see Schwartz and Sprouse 2013, section 7.2.3). We see complete variety in the possibilities for surface variation in relation to possible meanings. Examples (12) and (13) show us form-meaning pairs that are in 1-to-1 correspondence; each form has one possible meaning. Example (8) is a 1-form-to-many meanings match-up. In (14) below, we see a many-forms-to-one-meaning relationship (example taken from Schwartz and Sprouse (2013, p.144)):

(14) Joe looked up the number

(15) Joe looked the number up

The two forms of (14) and (15) are minimally different, each containing the same words and differing only because of the particle shift. Pairs (or more) of sentences with larger surface differences can still correspond to the same meaning, as seen in (16) and (17) (from Fodor and Crowther (2002, p.134)):

(16) Two houses are owned by Susan

(17) Susan has two houses

In principle, the presence of both uniqueness and variability in form-meaning mappings hampers a learner's ability to form correct generalizations. The availability of multiple structures leaves the potential hypothesis space wide open. In the previous section on negative evidence, we discussed the fact that if a uniqueness principle did exist, it could be applied to the correction of overgeneralizations. An interesting point for the consideration of positive evidence here is that there are many overgeneralizations that would be logically possible based on the positive data alone, but have not been documented to be made by learners. Lightfoot (1998, 2005) gives the following examples:

(18) Tim's happy

(19) Kim's happier than Tim is

where he argues that (18) should lead the learner to expect to be able to reduce the second *is* in (19). But learners come to know that *Kim's happier than Tim's* is ungrammatical without committing that overgeneralization.<sup>17</sup>

---

<sup>17</sup> Pullum and Scholz (2002) propose that the presence or absence of stress clues the learner in to the correct generalization, thereby explaining the lack of overgeneralization. In response, Fodor and Crowther (2002) argue that stress is still a linguistic clue, therefore the basic argument for domain-specific prior knowledge is only strengthened by the P&S analysis.

In summary, we have seen some examples in which the positive data are under-informative. First, the positive data contain ambiguous sentences whose syntactic structure is not uniquely identified.<sup>18</sup> Second, examples such as (11) through (13) do not signal their ambiguity or lack thereof, raising the question of how a learner correctly comes to know that a sentence is unambiguous. Third, surface alternations such as in (14) and (15) do not contain information regarding the limits of possible alternations, raising the question of how learners decide when an alternate form is not possible. Yet native speakers do arrive at such conclusions correctly, and we will see in Chapter 4 that some L2 learners demonstrate correct intuitions as well, in the L2.

For discussion of our third and final category of poverty of the positive data, we turn to Schwartz and Sprouse (2013, p.152ff) and their exposition of what they call ‘the bankruptcy of the stimulus’. These are cases where, on the basis of positive evidence alone, patterns of sentences could lead learners to incorrect predictions about grammaticality. One of the examples that they offer is the phenomenon of intact movement versus remnant movement in German.<sup>19</sup> They argue that remnant topicalization and remnant scrambling pose a POS problem for L1 learners and for L2 learners whose L1 does not instantiate scrambling (English, for instance). The linguistic facts involving word order alternations in German are intricate, so we will not present a full exposition here. Our focus at the moment is on Schwartz and Sprouse’s argument for the bankruptcy of the stimulus. The interested reader is referred to Chapter 4 below for

---

<sup>18</sup> I have concentrated on syntactic examples here, however, analogous examples could be made in the phonological domain.

<sup>19</sup> This has been studied by Schrieber and Sprouse (1998), and Hopp (2002, 2005) for English learners of German.

data and to the studies cited in fn 19, as well as to Müller (1996, 1998) for more detailed discussion of the claims about the restrictions on remnant movement in German.

Schwartz and Sprouse summarize the German facts as follows: German permits intact movement such as topicalization (fronting of an XP) and scrambling (leftward movement of Direct Objects (DO)). In addition, under certain circumstances German permits movement of an XP from which some other movement has already applied. This is known as remnant movement. Citing Müller (1996), Schwartz and Sprouse (2013, p.154) say that the circumstances under which remnant movement is permitted are when the two different instances of movement are of two different types. You can topicalize a remnant after scrambling, but you cannot scramble a remnant after scrambling has already taken place. Schwartz and Sprouse (2013, p.156) summarize these facts:

- |      |   |   |
|------|---|---|
| (20) | Topicalization of intact phrase         | ✓ |
|      | Remnant topicalization after scrambling | ✓ |
|      | Scrambling of intact phrase             | ✓ |
|      | Remnant scrambling after scrambling     | x |

They argue that learners (both L1, and L2 whose L1 does not instantiate scrambling) would find the positive evidence in the PLD misleading by the following reasoning: On the basis of examples of intact movement and remnant topicalization after scrambling, all of which are permitted, a learner would incorrectly expect remnant scrambling after scrambling to also be permitted.

The success of this argument depends on the correct choice of relevant linguistic facts and their analysis. Again we see the question of whether or not similar-but-not-exact data should be considered as evidence for a grammaticality of a particular construction.

Schwartz and Sprouse base their bankruptcy argument on positive evidence from a subset of logically possible intact and remnant movements. Because three of the four constructions considered are attested in the input, it seems intuitively reasonable to suggest that a learner would expect that the fourth would also be possible. A different argument could be made on the basis of a different assumption about what constitutes relevant data. Consider example (21) below. It shows a hypothetical table that a learner could construct based on a larger set of logically possible remnant movements than that used in (20), above. The scenario in (21) assumes that a restriction on multiple movements of the same type is in effect (shown by the ‘no’ entries on the diagonal). On the possibility that some of the instances of multiple movement of different types could be unattested for independent reasons, all of the table entries where the second movement is of a different type than the first are represented with an uncommitted ‘x’ rather than a ‘yes’ or ‘no’.

(21) Remnant movement:

	<b>Topicalization after...</b>	<b>Scrambling after...</b>	<b><i>wh</i>- after...</b>
<b>Topicalization</b>	no	x	x
<b>Scrambling</b>	x	no	x
<b><i>wh</i>-movement</b>	x	x	no

Under the scenario in (21), it is less reasonable than in (20) to think that a learner would be misled by the positive evidence. Even if all of the x’s in (21) were ‘yes’, a learner would have to generalize from multiple movements of different types to multiple movements of the same type. So we see that the bankruptcy argument relies on a notion of similarity that needs to be carefully justified.

So, we leave this idea of bankruptcy of the stimulus, where patterns of sentences lead to incorrect generalizations, with a provisional feeling. Such bankruptcy may exist but the patterns must be carefully chosen to reflect the correct set of relevant facts. This is a task that may not be easy, especially for intricate patterns; it is also a task that is potentially dependent on the particular analysis of the linguistic facts.

In this section, we have examined four proposed ways in which the primary linguistic data (PLD) is argued to be impoverished: degeneracy, finiteness, the lack of negative evidence, and deficiencies in the positive evidence. This discussion has brought out a few differences between input that is available to L2 learners versus L1 learners. Accordingly, the next section will reconsider the statement of the POS argument.

### **2.3.2 The POS Argument for L2A**

Section 2.2 presented the version of the POS argument as stated in Pullum and Scholz (2002), repeated here for convenience. You will notice that the argument is specifically stated in terms of L1A:

(1) Pullum and Scholz (2002) statement of the POS argument

“a. Human infants learn their first languages either by data-driven learning or by innately-primed learning. [Disjunctive premise; by assumption.] (here and throughout, brackets in original)

b. If human infants acquire their first languages via data-driven learning, then they can never learn anything for which they lack crucial evidence. [By definition of data-driven learning.]

c. But infants do in fact learn things for which they lack crucial evidence.  
[Empirical premise.]

d. Thus human infants do not learn their first languages by means of data-driven learning. [From (b) and (c), modus tollens.]

e. Conclusion: human infants learn their first languages by means of innately-primed learning. [From (a) and (d), disjunctive syllogism.]” (brackets original)

In this section, we will modify the statement of the argument to make it more appropriate to L2A, taking into account the differences between L1A and L2A as discussed in Section 2.3.1.

Recall that the crucial assumptions on which the POS argument rests are those of successful learning and stimulus poverty. Section 2.3.1 revealed differences between L1 and L2 acquisition with respect to both of these assumptions. First, with respect to learning, the L1 learner achieves native competence (represented by ceiling performance on linguistic tasks of comprehension and production), but the L2 learner does not. As discussed in Section 2.3.1.1, for the purpose of the POS argument for L2, learning is considered successful if a learner demonstrates a trend toward native-like performance. Second, with respect to the linguistic input, L2 learners have additional resources available to them. The most important of these is the L1. It is an empirical question whether the L1 facilitates or impedes L2A. However, for the purpose of the POS argument, the L1 is an advantage in the sense that the L2 learner has the benefit of the knowledge of the structures and features instantiated in the L1. This knowledge is in principle available to help ‘fill the gap’ between the PLD and the target grammar by supplying information about possible structures. The other major element of additional input available to the L2 learner over the L1 learner is in the category of negative evidence, and complementary to the negative evidence is the mature cognition of the

adult learner. As we saw earlier, instructed L2 learners in particular have access to correction, general feedback, and metalinguistic explanations. In these ways, the L2 input is richer than the L1 input.

Recall from Section 2.3.1 that sufficiency of input must be considered in the context of specific assumptions with respect to the learning mechanisms available to the learner. Here again, L2 learners have additional resources in the form of mature cognitive abilities (including specifically linguistic abilities) and a rich conceptual system that can facilitate the creation of representations (see Carroll 2001, p.207ff). Carroll (2001) as well as Meisel (2011, p.170-174) propose that the additional cognitive resources brought to bear on the L2 acquisition problem act to compensate for the already reduced deficiencies in the input relative to L1 input data. Carroll and Meisel separately argue that there is no ‘logical problem of L2 language acquisition’, because the L2 learner is neither ‘linguistically uninformed’ nor ‘cognitively uninformed’<sup>20</sup>; the L2 learner is in principle capable of inducing a grammar from the linguistic input in combination with the already instantiated L1. If this is true, then we cannot construct a POS argument for L2.

However, there are reasons to think that we can, that is, that the L2 input in conjunction with the available learning mechanisms may still be insufficient to successful acquisition. The learner’s L1 may not be of assistance in cases where the L2 instantiates a property not instantiated in the L1. Immersion learners may not get much negative evidence, and we have seen that instructed learners cannot rely on receiving a consistent amount of negative evidence. Even though adults are capable of requesting help as well as self-reflecting on the differences between their performance and the input, it is not

---

<sup>20</sup> The terms *linguistically uninformed* and *cognitively uninformed* are from Carroll.

guaranteed that they will do so. (The POS assumptions concern what every learner *must* have available in the input, not what some learners *might* have available in the input.) Finally, in terms of the cognitive advantages of adulthood, there are different views on how fundamental and abrupt the differences are. Carroll (2011, p.243) characterizes the difference between child L1 learners and adult L2 learners as ‘profound’. For another view, Herschensohn (2009, p.283) argues that “the difference between adult and child language learning is one of gradient degree.”<sup>21</sup> On the possibility that L2 learning mechanisms are not profoundly different from L1 learning mechanisms, but rather on a continuum, and L2 input is relatively but not sufficiently enriched, the POS argument can be made for L2.

We conclude this section with a summary in the form of a restatement of the POS argument, adapted to L2A (based on Cook 2003, p.202):

(22) The POS argument for L2A:

- a. If L2 learners can be shown to trend toward reproducing the linguistic behavior of native speakers, and
- b. If the linguistic input typically available to L2 learners in the form of natural language, instructed language, negative feedback, and the L1 does not contain information sufficient for acquisition on the basis of general-domain learning mechanisms, then
- c. The L2 learner must rely on innate, language-specific principles and/or constraints.

---

<sup>21</sup> See MacWhinney (2012)’s Unified Competition Model for support for similarity between L1 and L2 learning mechanisms.

### 2.3.3 Confirming the POS

In preparation for evaluating POS studies in future chapters, it will be useful to consider what it takes to confirm a POS-type argument. It is worth noting that POS arguments are non-demonstrative, that is, the truth of the premises does not guarantee the conclusion. A new proposal for a different learning mechanism or a discovery of a new way to extract information encoded in the PLD, or a different theory of grammar and what is acquired can serve to obviate the need for a proposed innate, language-specific capacity.<sup>22</sup> But as Crain and Pietroski (2001, p.151) put it, "...one can hardly object to the strategy of providing several converging (non-demonstrative) arguments for an empirical thesis."

That being said, the primary task in supporting a POS argument is to establish the premises: successful learning and stimulus poverty. We can conduct investigations to establish that learners have acquired certain particular linguistic phenomena; Chapter 3 will discuss relevant experimental techniques. Discussion earlier in this chapter has considered that for L2 acquisition, we are challenged to be thoughtful in how we measure success. But showing that the input is too impoverished to support acquisition is harder. According to Meisel (2011, p.173): "...it is difficult if not impossible to provide positive evidence for the non-existence of a fact or a process."

Schwartz and Sprouse (2013) concur, to a certain extent. As we saw in Section 2.3.1.2, they argue that one cannot for instance take a lack of a particular type of example in a corpus as evidence for non-existence of that example type in the learner's entire input. From this point of view, POS arguments for phenomena that are rare are 'soft'

---

<sup>22</sup> Indeed, we have seen an evolution of grammatical theory that affects the nature of the proposals for UG, and have similarly seen an evolution in the Fundamental Difference Hypothesis (Bley-Vroman 1989, 2009) as a result of theoretical and empirical discoveries.

(their term). Schwartz and Sprouse consider that phenomena of ungrammaticality and ambiguity provide stronger cases for L2 stimulus poverty, as long as the L1 grammar does not already instantiate the necessary constraints.

In Chapter 4 we will examine the ways in which several target phenomena are argued to be POS phenomena. In Chapter 5 we will evaluate those arguments, concluding that they are not entirely successful.

## **2.4 The Emergentist View on the POS Argument**

### **2.4.1 The Emergentist View on the Key Assumptions**

The emergentist framework takes the point of view that language acquisition is a data-driven, bottom-up process. Emergentists therefore dismiss POS claims, instead maintaining that the PLD is much more rich and informative than generativists give credit for. This section discusses the emergentist position in more detail as it relates to the premises of the POS argument concerning learning and stimulus poverty. The discussion reiterates a few of the general points about the emergentist approach presented in Chapter 1, however, more details are recruited in order to show how emergentism works towards refuting the POS argument specifically.

#### ***2.4.1.1 Learning***

Recall the learning assumption as stated in Pullum and Scholz's version of the POS argument: "Human infants learn their first languages either by data-driven learning or by innately-primed learning". By utilizing the word 'languages', P&S avoid a theoretical commitment to either the generative or emergentist point of view, which differ in their view of what is acquired. As mentioned in Section 2.3.1, generativists take it that

the learner acquires a stable ‘grammar’, a system of linguistic principles or constraints, which generates all and only the grammatical sentences in the language.

In contrast, emergentists do not agree that learners acquire a grammar in the same sense. The nature and result of the learner’s induction problem are not the same as those proposed in the generativist approach. Gregg (2003, p.107-108) gives this summary statement: “The underlying competence of a language learner from this [i.e., emergentist] perspective consists of an ability to do distributional analyses and an ability to remember the products of the analyses, competences which of course extend to nonlinguistic domains as well.” Knowing a language is a matter of knowing statistics; principles are not rules in the generative sense but merely regularities. These statements apply to both first language learners and second language learners. O’Grady (2013a, p.254) goes so far as to call language acquisition ‘an illusion’: “...at least in the case of syntax, there is no such thing as language acquisition, if we take that term to refer to a developmental process devoted specifically to the construction or growth of linguistic knowledge.” O’Grady proposes that what looks like language proficiency is simply the successful automatization of processing routines, where the increasing efficiency of those routines in the face of repeated exposure to linguistic input produces the appearance of grammatical generalizations.<sup>23</sup> What a learner knows is constantly evolving in response to continually received input.

What are the implications of this point of view for satisfying the assumption of learning as part of the POS argument? In a simple sense, emergentists agree that learning takes place, therefore, the assumption is satisfied. The emergentist approach does not put

---

<sup>23</sup> O’Grady’s proposal will be examined in more detail in Chapter 4.

a dent in the POS argument by dismantling the learning assumption. However, the argument will play out slightly differently because of the different view of what it is that learners come to know. How one measures success in learning depends on what one thinks is being learned. Emergentist proposals may need to rely on processing evidence in addition to performance tasks to fully establish their claims. For instance, for L2A claims, the concept of ‘native-like’ or ‘trending towards native’ will encompass similarities in processing as well as in behavior on tasks such as grammaticality judgment tests and truth value judgment tasks.

#### ***2.4.1.2 Stimulus Poverty: Does it Exist?***

Emergentists maintain that the overall input is sufficient for acquisition in conjunction with only domain-general learning mechanisms and domain-general prior knowledge. The following discussion presents the emergentist view on the nature of L1 and L2 input, and their arguments as to why no existing deficiencies are so severe that they cannot be overcome without innate, language-specific tools. The section will conclude with some remarks about the special circumstances of L2A compared to L1A.

#### **DEGENERACY**

The process of bottom-up language acquisition begins with individual exemplars, some of which, as we have noted earlier, are degenerate; however, the process of acquisition is a process of extracting regularities from the total data. From this point of view, low frequency ungrammatical utterances are ultimately not determinative. One might ask how it is that (possibly equally) low frequency grammatical utterances are attended to and acquired when ungrammatical utterances are not. It is important to keep in mind that emergentists propose a number of factors for statistical tracking, not simply raw frequency, some of which are recency, redundancy with other cues, reliability

(contingency of mapping), saliency, and prototypicality of meaning. There are likely to be interactions among these factors as well. These factors may account for the differential acquisition of various low-frequency phenomena.

### **FINITENESS**

As we saw under the generativist approach, the leap from finite data to the ability to produce and comprehend infinitely many sentences can be made because of domain-general learning mechanisms such as induction and analogy. The bigger challenge from finite data rests on the need to identify the correct target grammar based on data that are compatible with multiple possible grammars. As we will see below, non-nativists believe that that challenge can be met by the positive data.

Emergentist models do, however, require a certain amount of data in order to extract the necessary generalizations. Ellis (2012, p.203) says, “Learners’ cognitive systems have to be allowed sufficient exposure to allow Saussure’s somewhat mysterious process by which “thought-sound’ evolves divisions, and language to take place with its linguistic units in between these two amorphous masses.” So finiteness is not a problem per se, but the language sample cannot be too small.

### **NEGATIVE EVIDENCE**

Emergentists would not dispute any of the information about the existence (or not) of direct negative evidence that was presented in Section 2.3.1.2. When it exists, as in instructed L2A, direct negative evidence can serve to provide constraints on the learning space, thus increasing efficiency and decreasing computational resource requirements.

Under this approach, direct negative evidence is welcome but not in principle necessary.

Positive evidence in combination with certain learning mechanisms can provide constraints on overgeneralizations.

## **POSITIVE EVIDENCE**

Our review of positive evidence in Section 2.3.1.2 centered on three proposed deficiencies: rarity, unformativeness, and bankruptcy. As we saw in the quote from Ellis (2012) above, extreme rarity could be a problem for the emergentist approach. There may be some compensatory benefit to multiple converging cues such as those mentioned in the section on degeneracy. But in general, statistical models need a certain amount of data in order to establish baseline frequencies, etc.

With regard to the lack of structural information in the speech stream, this is less of a concern. First, in the emergentist view, learners are not acquiring structure, rather, form-meaning pairings. Structural information per se is not relevant. The claim is that the necessary units (words, clauses) can be recognized through pattern searches. The acquisition of ambiguity proceeds differently simply because what is being acquired is not forms, but form-meaning pairs. So somewhat by definition, ambiguity must be resolved in order for learning to take place. Thus we see that the differences in the views of what is acquired account for the differences in the views of what constitutes sufficiency in the PLD.

The final deficiency that was considered under the generative approach was ‘bankruptcy of the stimulus’. Recall that this term is from Schwartz and Sprouse 2013, and it refers to patterns of sentences in the positive data that, under certain possible generalization procedures, create an incorrect expectation of grammaticality of sentences that turn out to be unattested (because ungrammatical). This situation can be overcome with certain stochastic modeling techniques. Larsen-Freeman (2012, p.77) explains: “Its absence (of positive evidence, that is) allows learners to decrement the probability of a

relationship in the target language that they would have otherwise expected.” The non-occurrence of certain sentences in the positive data functions as negative evidence.

Tomasello (2003, p.178ff) discusses two other factors involving positive data that serve to constrain syntactic phenomena: entrenchment and preemption. Tomasello’s discussion concerns verbs specifically. Entrenchment refers to the solidification of the use of a verb in high-frequency constructions, and the reluctance of a learner to extend the verb to other constructions that (in their experience) are unattested. Preemption refers to a learner’s tendency to adopt a construction that they have heard, even if it is unexpected (to them). Tomasello gives the example *He made the rabbit disappear* in contrast to *He disappeared the rabbit*. The idea is that having heard the first sentence, the second is preemptively eliminated from their hypotheses. Thus we see that there are several ways that positive evidence can obviate the need for negative evidence.

An important difference between the emergentist and the generative view of positive data concerns the applicability of similar but not exact sentences as evidence for a given construction. Recall that Pullum and Scholz (2002) proposed that *wh*- questions could provide evidence for structure dependency in yes/no question formation. Clark and Lappin (2011, p.40) argue differently, pointing out that:

“An inductive learner does not know whether the principles for forming polar interrogatives also apply to constructing *wh*-questions. As far as such a learner is concerned, it could be possible for *wh*-questions to be formed through fronting the auxiliary in the main clause, while polar interrogatives are generated by fronting the first auxiliary. In order for a DDL (data-driven learning) learner to extract the correct rule...it needs a prior notion of similarity between the two constructions.”

This then is one way in which the PDL would be considered less rich (for a given construction) under the emergentist approach than under the generative approach.

Regarding the differences between L1A and L2A input, emergentists would concur with all of the concrete examples discussed in Section 2.3 such as the fact that instructed L2 learners receive more direct negative feedback, and that the L2 input is enriched by the existing L1. There are three differences that are particularly salient to the emergentist approach. First, the cognitive and social maturity of adult learners is particularly helpful from the standpoint that emergentists value the role of meaning in acquisition. Adults bring more knowledge of the world and more understanding of social meaning to the task, enriching their input relative to child L1 learners. This is an advantage. The second and third factors are disadvantages. Adults receive less support than children, who generally have enthusiastic and attentive caregivers interacting with them. Perhaps most importantly, the adult brain is already committed to the processing routines of the first language, therefore less malleable (MacWhinney 2012, p.211) and less able to use the input it receives (Ellis and Larsen-Freeman 2006, p.571), since the data and statistics compiled for the L1 may bias the L2 statistical analyses, at least early on.

In conclusion, we have seen that the generative and emergentist approaches disagree in principle on what constitutes learner competence and on the extent to which learner input is sufficient for the development of such competence. The bottom line is that in a bottom-up approach to language acquisition, there is no poverty of the stimulus.

#### **2.4.2 Disconfirming the POS**

What must an emergentist do in order to concretely undermine the POS argument? The answer is to build models of acquisition that demonstrate that data-driven learning (DDL) is possible without recourse to language-specific prior assumptions. Fodor and Crowther

(2002, p.142) say that we must “be vigilant for any scraps of unlearned information that may have crept into the learning program.”

What does a successful model look like? First, success on novel examples (beyond the training set) and in particular the ability to recognize ungrammaticality is a crucial part of language acquisition that must be able to be replicated. Second, a successful model will not overgenerate where learners do not overgenerate.

## **2.5 Conclusion**

Chapter 2 provided a general statement and discussion of the POS argument. The logic of the argument is solid and uncontroversial, however, its underlying assumptions concerning learning and data deficiency can be difficult to establish. Regarding learning, for L1A, of course, it is not difficult to demonstrate that (cognitively normal) learners are ultimately successful; however, in the case of L2A, which is our main area of interest, learner success may require a nuanced definition and careful measurement. Chapter 3 will review the main experimental methodologies used to confirm or disconfirm successful L2 acquisition.

The assumption of data deficiency is highly controversial outside the generative community. In addition, statements regarding sufficiency or insufficiency of linguistic input must be made in concert with proposals regarding grammatical theory, knowledge, and available learning mechanisms. The generativists aim to show that the POS argument stands; their challenge is to substantiate the premise that the data available to all learners does not contain sufficient information to enable acquisition on the basis of domain-general learning mechanisms. Generativists must establish that something doesn't exist; this is a formidable task. The emergentists aim to show that the POS argument fails by

disconfirming the assumption that the data are impoverished. To do this, they need to construct a model that demonstrates success given only naturalistic input to domain-general learning mechanisms, without positing language-specific priors. Again, a formidable task. After reviewing the experimental techniques in Chapter 3, Chapter 4 will consider generativist and emergentist L2A studies that support and dispute, respectively, the POS assumptions and the POS argument.

## **CHAPTER 3: METHODOLOGIES FOR ASSESSING AND MODELING ACQUISITION**

### **3.1 Overview**

The sheer number and variety of experimental tasks that have been used to assess and model acquisition precludes presentation of a comprehensive inventory here. Doughty (2003, p.299-302) lists over 50 different ways to measure L2 ability; other reviews of L2 assessment provide additional examples (see for instance Chaudron 1983, 2003; Norris and Ortega 2012). Westermann, Ruh and Plunkett (2009) review connectionist (modeling) approaches to language acquisition.

This chapter will focus on the experimental tasks and methodologies that are used in the studies that will be under review and critique in Chapters 4 through 6. These include grammaticality judgment tasks (GJs), truth-value judgment tasks (TVJs), the use of artificial languages, and connectionist computer simulation models. These methods have in common that they are highly controlled and thus well-suited to the assessment and modeling of learner knowledge of rare and/or complex linguistic phenomena. In addition to providing a rationale and description of each method, the chapter will consider methodological advantages and disadvantages, including (for the judgment tasks) issues around the validity and reliability of the evidence provided by that method. The points discussed in this chapter will provide an understanding of the appropriateness and limitations of these tasks, thus providing a basis for evaluating the acquisition claims made in the studies reviewed and critiqued in Chapters 4 through 6. Most of the discussion will be theory-neutral, as both generative and emergentist approaches make use of these experimental techniques. However, where useful, I will mention any theory-particular issues. Most of the discussion will be relevant to both L1A and L2A, drawing

examples from both. However, particular issues for L2A will be highlighted for each task.

### **3.2 Grammaticality Judgment Tests (GJs)**

The essence of a GJ test involves a simple discrimination task, in which subjects are asked to decide whether or not a sentence is grammatical. Although the term ‘grammaticality judgment’ is pervasive in the literature, it is important to recognize that language users provide *acceptability* judgments rather than *grammaticality* judgments. Technically, *grammaticality* refers to whether or not a sentence is generated by a grammar. Hence, grammaticality is referenced to a particular theory, and the judgment of grammaticality is not available to a naïve language user. Language users report whether or not they find a sentence to be *acceptable*. Acceptability can be based on the user’s internal grammar, but may also involve extra-grammatical factors such as comprehensibility, and ease of interpretation. These factors will be discussed in more detail in Section 3.2.3.

In this work the term ‘grammaticality judgment’ will be used, in part because it is in common use in most (but not all) of the literature, and also because quite often the researcher who is employing such a test is searching for information regarding the subjects’ internal grammar, albeit indirectly.

#### **3.2.1 Rationale for GJ Tests**

The use of GJ tests is well-established in native speaker and L2A studies. There are several rationales for their use (see Schütze 1996; Gass and Selinker 2001; Norris and Ortega 2012; among others). First, native speaker intuitions and L2 learner intuitions provide information on competence, that is, the properties of a language user’s grammar.

When a subject in an experiment accepts a sentence, experimenters infer that that sentence could be possibly generated by the subject's grammar.<sup>24</sup> Similarly, when a subject does not accept a sentence, the experimenter infers that that sentence is not generated by his or her grammar; this ability to expose the limits of grammaticality provides a second rationale for the GJ technique. Third, the highly controlled experimental conditions force the subject to confront the target construction, thus guaranteeing the collection of relevant data (especially concerning rare constructions) in a way that production tasks cannot. Finally, GJs can easily present multiple stimuli to multiple subjects, assisting in creating more robust sample sizes for statistical analysis.

### **3.2.2 Characteristics of GJ Tests**

Grammaticality judgment tests come in a number of guises. The most basic version involves a simple discrimination task, in which subjects are asked to decide whether or not a sentence is grammatical. More information can be obtained by requesting the subject to do additional tasks such as locating errors in the sentences, correcting errors, and providing grammatical explanations of errors. The mode of presentation of the test sentences can be aural or written. A variety of factors enter in to the design of the test instrument. In the remainder of this section we will consider possible designs for stimuli, for the types of judgments elicited, and other procedural considerations such as instructions. For more exhaustive treatments, the interested reader can consult several sources including Schütze (1996), Birdsong (1989), and Featherston (2007a,b).

---

<sup>24</sup> In order for this inference to be valid, other performance factors that could affect the judgment must be ruled out. See Section 3.2.3 below for further discussion.

### ***3.2.2.1 Stimulus Design***

There are a number of important factors in the design and presentation of stimuli in GJs. Test sentences should be controlled for length, complexity of structure, complexity of vocabulary, high frequency words, and location of error within the sentence. Sentences can be presented in random order with filler sentences as distractor items, or they can be presented in an organized sequence which draws attention to the target construction. In the former approach, subjects are less likely to identify the target of study and therefore less likely to use conscious strategies in arriving at their judgments. In the latter approach, subjects are more likely to be aware of the target. Although this design may promote the very disadvantage that the random order presentation avoids, it has the advantage of increasing the likelihood that the target structure is in fact the basis for the judgment given. We will return to this point in Section 3.2.3 where we discuss the validity of judgment data.

Another important element in the design of stimuli is the presence or absence of context preceding the test sentence. Contextual material can be related to the test sentences either semantically or structurally. Schütze (1996, p.150-160) discusses issues around the potential effects of context on judgments. He points out that a sentence presented out of context can seem less acceptable than when the same sentence is situated in a semantic context. Crain and Steedman (1985) note that if semantic context is not provided, the experiment is less controlled than if context is provided. To the extent that subjects may invent contexts for sentences as part of their judgment decision process, the invented contexts will obviously differ for each subject. They recommend (p.338) that the experimental design “unambiguously establish a known set of assumptions under which

the subject will approach the target sentence.” Providing a common context for all subjects may contribute to reducing inter-subject variability.

With regard to structural contexts, Warner and Glass (1987) found that a structurally related, similar preceding sentence can influence subjects to a correct grammaticality judgment of a test sentence; and, a structurally related but dissimilar preceding sentence can lead subjects toward an incorrect judgment. They studied native speaker judgments of ‘garden path’ sentences, which are well-formed but seem bad until a temporary ambiguity is resolved, for example a verb that can be either transitive or intransitive (strike) may be parsed with the following NP as object when it’s actually the subject of the second clause. The garden-path, ambiguous test sentences were preceded by similar but unambiguous sentences, as illustrated in examples (1) and (2):

(1) Positive bias

- (a) Before [the man sleeps] [the cat eats].
- (b) When [the boys strike] [the dog kills].

(2) Negative bias

- (a) If [the girl pets the cat] [she sings].
- (b) When [the boys strike] [the dog kills].

(example sentences taken from Warner and Glass 1987, p.720)

In example (1), the (a) sentence prompts a correct parse for the (b) sentence (i.e. biclausal with intransitive verb in first clause). In example (2), the (a) sentence prompts a misleading parse for the (b) sentence (i.e. biclausal with transitive verb in first clause). Warner and Glass found that within subjects, the garden-path sentence (b) was rated grammatical on average 87% of the time when preceded by a positive biasing context,

and only 65% of the time when preceded by a negative biasing context.<sup>25</sup> So we have seen evidence that both semantic and structural contexts can influence judgments.

What if no context is provided? Schütze (1996, p.155) notes that in the absence of added contextual material, the neighboring stimuli surrounding any particular test sentence serve as a kind of context in that they provide a source of contrasts. Bever (1970) proposed that in the context of better sentences, a marginal sentence would be judged as worse and in the context of worse sentences, a marginal sentence would be judged as better. Subsequent studies (Greenbaum 1976; Snow 1975; Nagata 1992) confirm this prediction, whether or not the surrounding sentences are related to the target sentence. It is not clear that these effects are of great concern, based on an experiment by Cowart (1994), who compared GJ tests where filler items were all highly acceptable to GJ tests where filler items were two thirds highly acceptable and one third of low acceptability. Although the test sentences in the mixed-acceptability filler scenario were rated overall higher than the test sentences in the consistent-acceptability filler test, the relative ratings among the test sentences were the same across experiments. Cowart (1997, p.52) recommends including filler items of various acceptabilities.

As a final point with respect to context, if a test sentence contains anaphora or other discourse-dependent forms, then a suitable context must be provided in order for a subject to arrive at a judgment. We will see in Section 3.3 that there are other methodologies besides GJs for assessing acquisition of such things as anaphora and null subjects.

---

<sup>25</sup> A discussion of why native speakers are not 100% accurate on this task is beyond the scope of this chapter. However, we will return to this point in Chapter 4, where we see another example of native speaker uncertainty on a GJ task.

### ***3.2.2.2 Judgment reports***

We turn now from consideration of test item design to considerations of options on how judgments are reported. Three main methods for reporting judgments include binary choice, an n-point scale, and magnitude estimation. A binary choice judgment is the simplest form of response; the subject is asked to state categorically whether the test sentence is or is not grammatical. Some researchers argue against the use of a binary choice on the basis that acceptability is a continuum rather than a stark dichotomy.

Newmeyer (2007, p.398) puts it this way: “We all know – and have always known – that judgments about acceptability are highly gradient. Hence any test for acceptability that forces an informant to choose between ‘fully acceptable’ and ‘fully unacceptable’ as the only two choices is dead on the ground.” Hopp (2005, p.48) points out that gradient acceptability is particularly evident in judgments on optional constructions such as scrambling.

However, Bader and Häussler (2010, p.276) point out that a binary choice study can in fact produce a picture of gradience because judgment data are averaged across subjects and across sentences. The resulting mean percentages of what are individually binary judgments range over all values from 0% to 100%. For instance, one structure could be categorized as grammatical 75% of the time, and another structure categorized as grammatical 90% of the time. So a binary choice test is not completely uninformative with respect to gradience. Cowart (1997, p.68) argues that as the number of subjects increases, finer-grained distinctions in grammaticality can be revealed.

One problem with a binary choice test design is that there may be undetectable inter-subject variation in responses. That is, two different subjects could have the same

reaction to a given sentence, yet one categorizes the sentence as grammatical and the other categorizes it as ungrammatical. Featherston (2007a, p.285) makes the analogy to color perception: “If we show informants the colour grey, and ask them to categorize it as black or white, they will show much variation.” As mentioned above, in the aggregate, the group data can reveal gradience as lighter grays will more likely be categorized as white and darker grays as black. However, for grays that are closer to the middle than the extremes, judgments will vary.

It is worth noting that the notion of gradient acceptability does not require a view of grammar as gradient or fuzzy. Weskott (2011, p.254) argues based on phonological evidence that gradient judgments do not preclude categorical mental representations: “McMurray and colleagues (2002) showed that participants are sensitive to gradient effects in phoneme detection, depending on the task they are instructed to perform. Nonetheless, no one would doubt that phoneme perception in everyday speech comprehension constitutes a paradigm case of categorical, that is, non-gradient, perception.” Theoretical frameworks such as Optimality Theory (Prince and Smolensky 1993; McCarthy and Prince 1993) do subscribe to the idea of gradient competence; however, it is an open empirical question whether grammars (in the generative sense) are gradient or not.<sup>26</sup>

A second method for reporting judgments, an interval scale task, facilitates the placement of judgments on a continuum. Subjects are given a discrete scale, typically anywhere from 3 to 7 points, and are asked to choose the point that best represents their perception of the acceptability of the sentence. For instance, subjects could be asked to

---

<sup>26</sup> The interested reader is referred to Aarts (2007).

rate sentences on a scale of 1 to 7, with 1 being ‘impossible’ and 7 being ‘fully acceptable’. In practice, a 3-point scale (good, middling, bad) is not recommended, as it is difficult to interpret. Subjects might choose the middle option either because they perceive a sentence to be of middling acceptability or because they are unsure of their judgment. In principle, acceptability and certainty are independent (Sorace 1996, p.397) and the potential to combine both types of response in one data point is problematic.

Scales with a higher number of defined points offer more resolution but it can still be difficult to interpret the middle ratings. The points on the scale are designed to represent equal spacing across the continuum of acceptability, that is, the psychological distance between 1 and 2 should be the same as the psychological distance between 4 and 5.<sup>27</sup> However it is difficult if not impossible to guarantee that subjects will adhere to such regularity in assigning their judgments (Sprouse 2011, p.276); and in fact there may be some evidence that intervals are not treated as uniform (Stevens and Galanter 1957, cited in Cowart 1997, p.71). Sorace (1996, p.398) suggests clear labels for each scale point to assist subjects in being consistent in their ratings. In contrast, Cowart (1997, p.70-72) advises that only the endpoints be defined, on the basis that subjects may have an even harder time maintaining consistent distances between descriptions such as ‘better’, ‘doubtful’, or ‘worse’ than if they were free to simply choose from numerically defined points between two extremes.

Although interval scales provide the advantage of higher resolution, it is possible that subjects might perceive more levels of grammaticality than the scale provides. Thus interval scales are still somewhat vulnerable (as with binary choices above) to creating

---

<sup>27</sup> As an analogy, consider a thermometer. 10 degrees Fahrenheit represents the same size interval between 40 and 50 degrees as it does between 70 and 80 degrees.



compared acceptability judgment tests on word order variation in German<sup>28</sup> across three experimental paradigms (ME, seven-point interval scale, and binary choice), finding no difference in informativity. Similarly, Bader and Häussler (2010) compare MEs to binary choice GJs in studies of German word order variation<sup>29</sup>, again finding similar results using both judgment response methods. Further arguments against the privileged utility of MEs are based on the question of whether or not subjects are capable of the “modest degree of mathematical sophistication” (Cowart 1997, p.74) required for valid use of the ratio method. Narens (1996) proposes that ME judgments should have the property of commutativity, that is, the order in which sentences are compared and rated should not matter. Sprouse (2011) tested native speakers of English on various syntactic structures such as left branch extraction, double center embeddings, and object relative clauses, among others. He concludes that his experimental results do not display the property of commutativity and therefore he disputes the idea that subjects can make ratio judgments on syntactic stimuli.

We have considered ME judgment tasks here because they have been used in both L1 and L2 studies (for L2, see Sorace 1992, 2000 and others), and some have argued that MEs represent the ‘gold standard’ in experimental protocol. However, MEs are not the method of choice in any of the studies that we will consider in Chapter 4, therefore, we will leave further investigation of the pros and cons of this methodology to future work.

### ***3.2.2.3 Other Protocol Design Issues***

---

<sup>28</sup> Weskott and Fanselow examine scrambling of arguments of a ditransitive verb.

<sup>29</sup> Bader and Häussler examine the order of subject and object in embedded clauses, verb cluster linearization (verb-modal), and verb argument alternations (subject-dative object-accusative object).

We conclude our discussion of GJ tests with attention to a few additional aspects of experimental design: instructions, additional elicitation, and timing.

Adequate instructions are critical to the validity of judgment data. It is particularly important to clarify the idea of acceptability to the naïve subject, who might otherwise rely on a variety of criteria in developing a judgment. For instance, subjects may consider comprehensibility, positive or negative reaction to content, perceived prescriptive errors, punctuation, or even truthfulness. Schütze (1996, p.132) discusses an amusing example from Hill (1961) in which subjects who had rejected the sentence *I never heard a green horse smoke a dozen oranges* reversed their judgments “once it was pointed out to them that the sentences was true”.

In the interest of verifying the actual basis for judgments, some experimental protocols include the elicitation of additional information such as asking the subject to locate the error(s) in the test sentences that were rated unacceptable or less acceptable, to share their thoughts about each sentence (‘think-aloud task’), or to provide explanations of possible rules that they invoked in their judgment process or perceived to be at work in the test sentences. Schütze (1996, p.190) feels that the inclusion of such elicitation measures is crucial because of the fact that it is so crucial to be able to know that the judgments were based on the target structure. However he cautions that in order to mask the target structure from the subject (a protocol that he recommends), one must elicit some corresponding information about the good sentences, so as to maintain a balance between the amount of work the subject is asked to do for each item. Asking subjects for their reasoning has the possible disadvantage of creating confounds in their thinking and in prompting the introduction of extra-grammatical factors into the decision process,

which in a way is the very thing that should be avoided. An unanalyzed reaction is considered to be a better reflection of underlying competence.

To promote obtaining subjects' initial, intuitive reaction, a Speeded Grammaticality Judgment (SGJ) can be set up. In an offline setting the best the experimenter can do is to instruct the subjects to work quickly, or to provide an overall time limit for the entire task. In an online setting, time limits can be strictly enforced through computer presentation of stimuli on a set schedule, and a set time-out for registering judgments (on the order of seconds). Comparisons between timed and untimed GJs might reveal effects of processing (if subjects perform differently under the different experimental procedures).

Having looked at a range of characteristics that can be manipulated in GJ study design, we will now consider advantages and disadvantages of the general method.

### **3.2.3 Evaluation of GJs**

The following discussion of advantages and disadvantages of GJs applies to assessing both native speaker judgments and L2 learner judgments, and issues will be illustrated with examples from L1A and L2A studies. Issues particular to assessing L2 learners will be discussed in Section 3.2.4 below.

#### ***3.2.3.1 Advantages of GJs***

No matter what the particular design of the test, the grammaticality judgment approach has a number of strengths and weaknesses, as discussed in Featherston (2007), Cowart (1997), Schütze (1996), and Sorace (1996), among others. A few of the advantages of GJs were mentioned in the rationale section (3.2.1 above). Judgment tasks improve on production tasks in two ways: first, by being uniquely qualified to identify what is *not*

possible in a grammar; and second, by being able to guarantee to collect data on rare and complex phenomena. Although in principle a production task could elicit such phenomena, it is far more efficient to force the subject to confront the construction and react to it. A third advantage of the test methodology lies in its potential to abstract away from some non-grammatical factors which, in other situations, could affect the subject's response. For instance, the presentation of arbitrary and isolated test sentences eliminates any unintended and unwelcome introduction of experimenter bias such as would be possible in an interview situation.

Yet another advantage of a judgment task, especially if augmented with the collection of think-aloud or timed response data, is that it has the potential to identify subtle differences among subjects whose performance appears similar. Coppieters (1987) compared a group of near-native speakers to native speakers and found a much higher degree of variability among the near-natives on a variety of structures. Birdsong (1992), repeating the Coppieters study but using more precise subject selection criteria and improvements to the test structure, found a much closer correspondence between natives and near-natives. Although the findings are contradictory, both studies exemplify that learner/native differences can be studied using judgment tests. (For a full critique of Coppieters (1987), see Birdsong (1992)).

Fifth and finally, as an argument in support of grammaticality judgment test methodology, there are studies that show good correlation between judgment data and production data, indicating that language users may be accessing the same grammar for judgment as for production. As evidence, Chaudron (2003, p.805) cites several studies showing consistency between GJs and a variety of other tasks, including: Ortega (2000)

comparing GJs with Oral Proficiency Interviews; White et al. (1997) comparing GJs with a paragraph-story task; and Duffield and White (1999) comparing GJs with a sentence-matching task.

### ***3.2.3.2 Criticisms of GJs***

On the other hand, numerous criticisms have been directed at the use of judgment tests and judgment data. Some of the criticisms can be deflected with careful test design, but a few seem fairly insurmountable, at least until we have a better understanding of the brain and language processing. Three major concerns are accuracy, validity, and reliability of the judgments.

One of the more difficult objections to overcome is the fact that the researcher cannot guarantee the accuracy of the reported judgment. Only the subject knows whether or not their reported judgment matches their internal judgment. A judgment involves a mental state which cannot be checked independently.<sup>30</sup> For instance, it is possible that a subject could overrule an intuition on the basis of a test-taking strategy; perhaps he or she is concerned with balancing the number of ‘grammatical’ versus ‘ungrammatical’ judgments, or perhaps they report a definite judgment when in fact they are unsure. A basic discrimination test with no follow-up is especially unable to detect this discrepancy.

A second concern is the validity or soundness of reported judgments. Judgments can be invalid for the question at hand in cases where the subject does not base the judgment on the target structure. In a study conducted by Birdsong, Johnson, McMinn and Ingmundson (1985), as described in Birdsong (1989), the researchers attempted to direct the subjects’ attention to particular syntactic problems by explicitly instructing the

---

<sup>30</sup> Sorace (1996, p.380) makes an analogy to other sensations. For instance, self-reported pain levels.

subjects to base their judgments on *missing words*, *words out of place*, or *too many words*. L2 learners of French at varying levels of proficiency were asked to judge the grammaticality of sentences containing relativizers. The target structures corresponding to the instructions were relativizer deletion, resumptive pronouns and stranding. A follow-up test revealed that subjects did not necessarily attend to the syntax, but in fact attended to the morphology of the relativizer. So in this case, even with the help of instructions, valid data were not obtained.

Even assuming we could get accurate, valid responses, judgment data has the potential to be unreliable. Ellis (1991) tested L2 learners twice, one week apart. The second test comprised a subset of sentences from the first test. On the second test, 7 out of 8 subjects changed their judgment on at least one out of the ten test sentences. Problems of intra-subject inconsistencies have been documented to occur even within a single testing session. Snyder (2000) found that with repeated exposure over the course of a GJ test, subjects increasingly perceived certain types of ungrammatical sentences as grammatical. On the other hand, Rebuschat (2013, p.612) suggests the possibility that subjects may learn during the test, becoming more accurate as they get more exposure to constructions.<sup>31</sup> In both cases, intra-subject performance has the potential to change over the course of the test instrument, resulting in unreliable data. Inter-subject inconsistencies could arise as a result of differences in metalinguistic abilities and/or processing differences such as in working memory capabilities, as will be discussed in more detail below.

---

<sup>31</sup> Rebuschat cites Rohrmeier et al. (2011) as an example of learning during the test, however, the object of study was melodic structure, not linguistic structure.

Grammaticality judgments are potentially unreliable due to the fact that they are vulnerable to manipulation through interference from extra-grammatical factors. There are many extra-grammatical factors involved in the judgment process; it is a decision process that involves cooperation with (or interference from) performance issues as well as from other cognitive systems not specific to language. To the extent that the judgment process is conscious and/or analytical, a judgment is a reflection of some combination of performance and cognition as well as the competence, which is the intended target of study.

In the earlier discussion of accuracy we have already seen one performance issue: test-taking strategies. Other performance constraints can include physical factors such as fatigue, as well as mental factors such as world-knowledge, prescriptive knowledge and general metalinguistic knowledge. Intuitively one would expect prescriptive knowledge to influence a subject's judgments, and indeed, studies that collect data on reasons for judgments do have subjects report the use of criteria such as punctuation, or "what I learned in elementary school about correct grammar" (cited in Schütze 1996, p.162). As Sorace (1996, p.385) puts it: "It is difficult to tell whether subjects reveal what they think or what they think they should think." In Chapter 4, we will see an additional example where subjects apply a rule in the formation of their judgments, rather than relying on unconscious knowledge.

The effects of metalinguistic knowledge in the judgment process are more difficult to pin down. On the one hand, it has been shown that lack of metalinguistic knowledge does not necessarily correlate to the ability to make judgments. As described in Birdsong (1989), Scribner and Cole (1981) studied an illiterate, 'metalinguistically

impoverished' population in coastal Liberia who were able to accurately discriminate deviant sentences in their native language, but not able to explain how to correct them. And on the other hand, when subjects have sophisticated metalinguistic skills, these skills may participate in the judgment decision process. Birdsong (1989) discusses research suggesting that metalinguistic skills vary among individuals and can be augmented by activities such as exposure to other languages, exposure to language games, and being read to (see his Chapter 1, also Chapter 2, p.51-54). By these criteria, all of the L2 learner subjects in the studies presented in Chapter 4 below have highly developed metalinguistic skills.

With regard to cognition, as shown in Herschensohn (2007, Chapter 3), evidence from exceptional circumstances such as Down Syndrome, Williams Syndrome and specific language impairment argues for the dissociation of language from other cognitive systems. However, even if language *knowledge* is modular, language *performance* certainly involves interaction with more general cognitive functions. For instance, short-term memory has been shown to play a role in semantic integration, comprehension, and possibly parsing, although the role of short-term memory in parsing has become controversial (Harley 2001, p.398-402).

We have seen that there is a potential for a myriad of factors to affect judgments and therefore to affect the validity and reliability of judgment data. Featherston (2007b, p.403) likens a grammaticality judgment to an onion. His recommendation: "I therefore see Grammaticality as something that we work towards, slowly discarding irrelevant information, our current best efforts at any one time being known as Well-formedness."

### 3.2.4 Issues Particular to L2A Research

The discussion so far applies to GJ tasks in general, whether administered to native or nonnative speaker subjects. However, L2 research must grapple with additional challenges due to the fact that L2 grammars and learners are different from native grammars and learners in important ways. First, it is necessary to keep in mind that an L2 learner's judgments are characterizing the L2 learner's grammar (i.e. an interlanguage), not the target native grammar of the second language, and L2 grammars are different from L1 grammars. Herschensohn (2007) presents an accumulation of evidence regarding L2 grammars.<sup>32</sup> L2 grammars are transitional, incomplete or divergent from native grammars, and indeterminate (lacking a consistent analysis for some phenomena). Even a grammar of a learner who might be categorized as near-native is potentially incomplete. As discussed in Sorace (1996), incomplete grammars will yield random or inconsistent responses to target structures that are missing, whereas divergent grammars will yield responses that are consistent but different from a native grammar.

Second, adult L2 learners are different from child L1 learners. L2 learners must learn all aspects of the target language at once, and they draw on a variety of cognitive strategies and learning styles as they progress in their learning. L2 learning is to some extent a conscious process, and learning styles may affect judgments (see discussion in Birdsong 1989, p.86).

For these reasons, judgment test results from L2 learners differ from those of natives in several regards. For one thing, L2 learner judgments show more variability

---

<sup>32</sup> Note that Herschensohn and Sorace take a generativist perspective, and recall that emergentists differ on their conception of the content of a grammar. However, the characteristics of L2 grammars described here are empirically demonstrated, therefore need to be taken into account in either approach.

than native judgments. We have already seen two studies documenting this variability: Ellis (1991) and Coppieters (1987). In the Coppieters study, the learners were judged by their peers to be highly proficient, and yet their performance did not match the consistency of native performance. Based on work by several researchers cited in Birdsong (1989), another difference between natives and L2 learners is that natives are more confident in their judgments. In several studies where subjects were given the opportunity to express their judgment in the form of a rating, natives tended to use the extremes of the scale, while the judgments from beginning L2 learners “typically hover around neutrality” (Birdsong 1989, 68.)

This discussion has shown that GJs are quite useful but must be designed well, so that learners’ judgments have the best possibility of reflecting their underlying knowledge.

### **3.3 Truth-Value Judgment Tasks (TVJs)**

Truth-value judgments are also discrimination tasks and share many characteristics with GJ tasks. However, TVJs are judgments of truth, not grammaticality. All of the test sentences to which subjects are asked to respond in a TVJ task are grammatical. The goal of a TVJ is to ascertain whether or not various interpretations of grammatical sentences are available in the subject’s grammar. TVJs have been used to study a variety of complex phenomena such as the interpretation of reflexives and of quantifier scope, examples of which will be seen below.

#### **3.3.1 Rationale**

The TVJ is a comprehension task that was developed to ascertain children’s comprehension of complex phenomena such as anaphora and quantifier scope (Crain and

McKee 1985, Crain and Thornton 1998). The procedure, which will be described in Section 3.3.2, has been adapted for adult L2 learners as well (see White et al. 1997; Dekydsprotter et al. 2001; Slabakova 2003; O’Grady 2013a; among others). The rationale for TVJ tasks parallels that for GJ tasks in several respects. First, TVJ tasks provide a window on subjects’ competence. When a test sentence is accepted in a context that determines a particular interpretation, experimenters infer that that interpretation for that test sentence is available in the subject’s grammar, and if a test sentence is rejected, that that interpretation is not available.<sup>33</sup> The task thus has the potential to assess both what is possible and what is not possible in a subject’s grammar. Second, TVJ tasks are controlled and can therefore guarantee that data will be gathered on a particular linguistic construction of interest. Third, although TVJs can be more complicated to set up than GJs, these tasks can still be offered to a potentially large number of subjects, thus improving the robustness of data collection and statistical analysis compared to individual introspection.<sup>34</sup> The TVJ task is well-suited to testing for ambiguity and constraints on ambiguity.

### **3.3.2 Characteristics of TVJs**

In a TVJ task, subjects are presented with a series of contexts paired with test sentences. The context determines a particular interpretation of the test sentence which follows it. The presentation can be in the form of a story, a video, or (commonly, in the case of child subjects) a puppet show. Following the presentation of the context, the learner is asked a yes/no question or asked to respond to a True/False-type statement that makes use of the target construction. In order to answer correctly, the learner must understand the context

---

<sup>33</sup> But see Section 3.3.3 for caveats on interpreting experimental results.

<sup>34</sup> But see Phillips (2008) for a defense of ‘armchair linguistics’.

and whether or not the test sentence is appropriate in that context. Since the subjects' responses are categorical, the task is not metalinguistic; there is no need for the subject to engage in a conscious consideration of grammar or parsing, for instance. The subject/respondent need only decide if the test statement corresponds with the facts as given in the preceding context.

Since the task was originally designed to reach children in an entertaining way, a puppet show is a commonly used presentation method (see for example Crain and Thornton 1998, Demirdache and Lungu 2008). A puppet show involves both linguistic and nonlinguistic input as the context story is acted out with dialogue. The test sentence is presented verbally by one of the puppets. This method has a few disadvantages, including the trouble and expense of doing live presentations. Perhaps more importantly, the procedure may introduce bias. Having the puppet present the test sentence can bias the subject towards a 'true' response, on the natural assumption that speakers (even puppets) intend to speak truly (Crain and Thornton 1998, p.53). Also, since the test sentence is spoken after the context is presented, in order to answer accurately, the subject must be able to fully recall the details of the context without having an opportunity to review it. However, a live presentation is sometimes advantageous for representing simultaneous events, if temporal sequence is important to the target construction. Another advantage of live presentation is that it facilitates follow-up questions by the experimenters to the subjects who can then offer information on the bases for their judgment decisions.

Other methods of context presentation include providing the subject with a picture or series of pictures (Bott and Radó 2007), or a written story (Dekydsprotter et al. 2001),

or both (O'Grady 2013a). Pictures and stories both have an advantage compared to live presentations in providing completely consistent presentations to all subjects. Another advantage is that the subject is able to review the context rather than having to remember it while considering the test sentence. This availability facilitates the use of more complex contexts in the form of more detailed pictures or wordier stories.

Comparing the picture context method to the story context method, White et al. (1997) found that the use of a story context yielded better information about which readings of ambiguous sentences were possible in the grammars of early L2 adult learners of English.<sup>35</sup> In English, the reflexive in sentence (5) can take either the subject (Mr. Brown) or the object (Mr. Green) as antecedent:

(5) Mr. Brown sold Mr. Green a picture of himself.

However, the subject antecedent reading is preferred. White et al. found that learners accepted the subject antecedent reading equally well across both the picture context and the story context test conditions. However, learners showed significantly lower accuracy on object antecedent contexts, accepting object antecedents at a far lower percentage. They argue that the story task is better at discovering licit interpretations even when those interpretations are dispreferred. White et al. attribute this difference to the fact that people who are taking the picture-matching test can (and often do) read the sentence before looking at the picture, in which case they access the preferred reading first and then have difficulty matching the sentence to the given (dispreferred) context.

---

<sup>35</sup> White et al. tested L1 Japanese and L1 Canadian francophones. Their control group comprised native speakers of Canadian English.

In (6) below, you see an example of a complete test item (context and test sentence) from a TVJ task in a study of adult L2 acquisition of the relative scope of quantifiers and negation (O’Grady, Lee and Kwak 2009, p.82):

(6) Context:

Tom is at his uncle’s repair shop. Tom’s uncle is about to go out for lunch. He asks Tom to fix three radios and three computers before he returns. Tom promises to do so.

Tom fixes the three radios early. Then, Tom examines the first computer. But, he can’t fix it. He decides to wait until his uncle comes back. Then, Tom looks at the second computer. There is something wrong with the sound, but he can’t fix it.

Finally, Tom comes to the third computer. There is something wrong with the screen.

Screens are very hard to fix. But, Tom manages to fix it.

Test sentence: Tom didn’t fix all the computers.

The English test sentence has two possible interpretations. One is a “full set” interpretation and one is a “partitioned set” interpretation, as follows:

(7) (a) Full set interpretation (*all* takes wide scope)

All of the computers are such that Tom did not fix them.

(Number of computers fixed: 0)

(7) (b) Partitioned set interpretation (*not* takes wide scope)

It is not the case that Tom fixed all the computers.

(Number of computers fixed: 0,1 or 2)

The context given in (6) above favors the partitioned set interpretation, since Tom did manage to fix 1 of the 3 computers. An alternative context could support the full set

interpretation by substituting the text in (8) for the corresponding final paragraph in the context in (6):

(8) Finally, Tom comes to the third computer. There is something wrong with the screen. He thinks that he can fix it quickly. However, after Tom works on it for a while, he gives up.

We would expect a native English speaker or a successful acquirer of L2 English to accept the test sentence in both contexts. However, if an individual's grammar (native or interlanguage) did not permit (say) the partitioned set interpretation, then we would expect the individual to accept the test sentence in context (8) and reject it in context (6). We will discuss this relative scope study from which this example is taken in more detail in Chapter 4 below.

There are several considerations in designing TVJ materials. As noted by Schmitt and Miller (2012, p.38) the contexts throughout the whole task need to be of similar sizes, controlled for vocabulary with respect to frequency and proficiency level, and controlled for complexity. Similarly, test sentences should differ minimally in these respects. As discussed for GJs in Section 3.2.2, filler items and control sentences should be included in the overall task. Some training items prior to the experimental task help ensure that subjects understand and can correctly follow the test procedures. Crain and Thornton (1998, Chapter 38) and Bott and Schlotterbeck (2012, p.19) discuss the usefulness of control items in investigating sentences that are ambiguous in the target language. If subjects have very low acceptance rates on one particular reading that is available in the target grammar, it can be difficult to decide if that particular interpretation is available but highly dispreferred, or truly unavailable in a subject's grammar. When control items

are included, the experimenter can compare acceptance rates on controls and on test items. If there is a statistically significant difference between performance on test items versus control items, then the experiment may provide evidence for or against the existence of an interpretation in the subjects' grammars.

Crain and Thornton (1998) formalize the ideas that the experimental contexts should be clear and felicitous in two design principles: the condition of falsification and the condition of plausible dissent.<sup>36</sup> The condition of falsification applies in an investigation of constraints such as binding principles. For example, consider the following scenario:

(9) Context: In a video, Big Bird and Ernie are standing next to each other.

Big Bird is patting his own head.

(10) Test sentence: Big Bird is patting him.

The test sentence (10) is false in the context of (9) in adult English grammar. Principle B of binding theory requires that the pronoun *him* must not be bound by the local antecedent *Big Bird*. Notice that the context supports the falsity of the test sentence as well as the truth of the negation of the test sentence (that is, *It is not the case that Big Bird is patting him* is true). Crain and Thornton argue that for maximum clarity for the experimental subject, contexts should make negative judgments available by clearly making the negation of the test sentence true in the given story.

---

<sup>36</sup> Drozd (2004) criticizes the conditions of falsification and plausible dissent as being general conditions on discourse and therefore unnecessary. I do not see the harm in codifying conditions so that care is taken in the proper design of experiments.

The condition of plausible dissent requires that the felicity conditions for a test sentence be met. Consider the following scenario based on a discussion in Crain (2012, p.170-173):

- (11) Context: Sam and Ted went out to lunch. Sam ordered chowder, and Ted ordered sushi and pasta.
- (12) Test sentence: Ted ordered sushi or pasta for lunch.

The test sentence (12) is true under a logical interpretation of *or* in adult English grammar. However, pragmatically it is odd to make the statement (12) after the lunch has already taken place, when everyone knows that the stronger statement using the conjunction *and* is true (*Ted ordered sushi and pasta*). The context is not felicitous for the utterance of (12), therefore, subjects may get confused and their response might not be a true reflection of their underlying grammar.

### **3.3.3 Evaluation of TVJs**

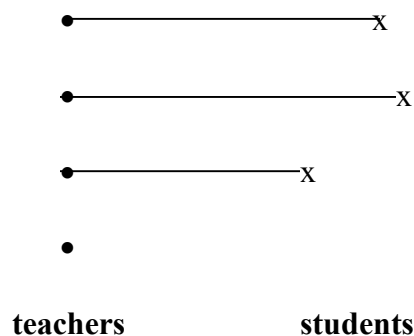
#### ***3.3.3.1 Advantages of TVJs***

TVJs share all of the advantages of GJs. Because they are highly controlled, the experiment is guaranteed to obtain information about the target structure. TVJs are designed to evaluate both what is and what is not possible in the subject's grammar. TVJs are better at obtaining information on licit but dispreferred readings compared to multiple choice selection tasks or production tasks. In a TVJ, only one test sentence is presented for each context. Putting aside the issue with picture contexts discussed above, with well-designed story contexts an interpretation that is normally dispreferred can be made available to the subject because there is no competition from other possible readings. In a multiple-choice selection task where more than one possible interpretation is offered at

the same time, the preferred reading could overwhelm the other possibilities in the subject's mind. Therefore the subject would be less likely to ever choose a response that indicates that they have the dispreferred reading available in their grammar.

### 3.3.3.2 Disadvantages of TVJs

Truth-value judgment tasks are vulnerable to the same criticisms in terms of accuracy, validity, and reliability as grammaticality judgment tests are. We must rely on the honesty of the test-taker in reporting accurate judgments. If a subject has any difficulty understanding the context as presented, then we have concerns for the validity of the judgments. Studying scope ambiguities, Bott and Radó (2007) compare three different types of disambiguating contexts: one linguistic and two visual. They find that providing a simplified visual context in the form of abstract diagrams yields reliable and valid results compared to more naturalistic-type pictures. Figure 3.1 below shows a sample abstract diagram that could provide one possible context for a sentence such as *Exactly one teacher praised each of these three students.*



**Figure 3.1: Context Presented as an Abstract Diagram**

Bott and Schlotterbeck (2012) found a high number of false rejections by their subjects in an experiment on scope ambiguities. An investigation of the test items found that atypical scenarios promoted errors in judgment. Figure 3.1 is atypical in the sense that there is an

additional teacher with no corresponding student. Bott and Schlotterbeck argue that a mismatch in the number of elements in the two sets induces an increased cognitive load relative to a context in which teachers and students are in one-to-one correspondence, and that an increased cognitive load gives rise to errors. So we see that the validity of the task is dependent on context descriptions, if complexities in the context can interfere with judgments.

In terms of task reliability across and within subjects, first, it seems plausible that TVJs would be vulnerable to the same concerns as GJs with regard to subjects gaining mastery of the target construction as they go through the test. Second, with regard to the stability of subjects' performance over time, I am not aware of any studies that can speak to that issue. Finally, we have already seen that judgments can suffer interference based on confusion or bias, to the detriment of task reliability.

### **3.3.4 Issues Particular to L2A Research**

As mentioned in the section on GJ tasks, L2 learner grammars are incomplete, and L2 learners' judgments are more variable than native speaker judgments, and these facts must be kept in mind when interpreting L2 performance on any judgment tasks.

Although originally designed for experiments with children, TVJ tasks can be adapted for adult subjects in several ways. The mode of presentation can be written stories (assuming that the adults are literate). The content of the stories can be chosen to be appropriate and interesting to adults. Finally, the number of test items can be larger because adults generally do not fatigue as quickly as children. Given that it is crucial that subjects understand the context, story contexts can be presented in the L1 (but test sentences are always presented in the L2). The TVJ task is actually more suitable for

adults than for children from the point of view that adults on average have better working memory skills, better attention spans, more world knowledge, and are better at making inferences than children. These skills should combine to contribute to optimal performance.

### **3.4 Artificial Language Paradigms**

#### **3.4.1 Rationale**

Reber (1967) pioneered the use of artificial grammars to investigate implicit learning, that is, the unconscious induction of rules from input. An artificial grammar consists of a set of strings of letters or geometric symbols, where the strings are generated by a finite set of rules. Such grammars are unlike natural languages in that they are simpler and have no semantics. Artificial languages, on the other hand, are constructed with words rather than letter sequences, and rules that approximate natural language syntax.

#### **3.4.2 Characteristics of Artificial Language Paradigms**

The artificial language learning paradigm consists of two phases: a learning phase, and a testing phase. During the learning phase, subjects are exposed to a set of input data under an ‘incidental’ learning condition. That is, the subjects are not told anything about the crucial facts of the training stimuli, nor are they told that they will be tested. For example, subjects may be asked about the semantic plausibility of the constructions presented to them, taking their attention away from any conscious consideration of syntactic patterns. After the highly controlled learning phase is completed, subjects are tested (often with a GJ test, but other tests are possible) on novel sentences in order to ascertain whether or not any syntactic learning took place.

### 3.4.3 Evaluation of Artificial Language Paradigms

The use of an artificial target language in language acquisition studies confers several advantages.<sup>37</sup>To the extent that an artificial language is constructed, the experimenter has control over the complexity of the phenomena presented to the subjects as well as the frequency of exposure that the subjects experience in the learning phase. Importantly for a POS study, an artificial language is guaranteed to be new. Participants will not have encountered the language before, will not have received any former instruction, and the instruction that they are given in the experimental setting can be completely controlled.

On the other hand, Williams and Rebuschat (2011, p.239) point out some challenges. Certain experiments use a *totally* artificial language, that is, one with invented lexical items. If the lexical items are not assigned referents and sentences do not have meanings, then meaning is excluded from the artificial language and the learning process may not be representative of natural language learning. If on the other hand the lexical items are assigned referents, then the experiment is usually limited to a small system that is manageable for the learners. But there is such a thing as too small, for instance when there are not enough lexical items available to be able to present the learner with a robust number of novel lexical items and sentences in the test phase. The use of a semi-artificial language can overcome these problems. Lexical items from the subjects' native language can be used in a language whose syntactic rules are defined by the experimenter. For instance, Williams and Kuribara (2008) use English lexical items with Japanese syntax (including case inflection). For subjects who are native speakers of English, such a language lightens the processing load (relative to nonce words) and allows both the

---

<sup>37</sup> For more detailed discussion, see Winter and Reber (1991), Williams and Rebuschat (2011), and Rebuschat (2013), among others.

learning and testing phases of the experiment to present a sufficiently large number of comprehensible sentences. These semi-artificial systems also have the advantage of better resembling natural languages, in that they have semantic meaning.

Another challenge in the use of artificial languages is that although officially the experiment is divided into learning and testing phases, it is possible that learners will continue to learn during the testing phase. Experimenters can check for continued learning by checking for increased accuracy of performance on test items presented later rather than earlier on the test instrument. If learning appears to be continuing, then any inferences based on the amount or quality of input to the learner will need to take the additional input into account.

Finally, when a grammaticality judgment task is used for the assessment of learning, the task is subject to all of the issues around accuracy, reliability, and validity that were discussed above in Section 3.2.

#### **3.4.4 Issues Particular to L2A Research**

Regarding issues particular to L2A research, it is fair to say that artificial language experiments are *always* L2A experiments. The artificial language paradigm has been popular in L2A research for investigating questions about pedagogy (Norris and Ortega 2000) and to investigate implicit (unconscious) versus explicit (conscious) learning (Reber 1967; R. Ellis 2005; and others).

### **3.5 Connectionist-Type Computer Simulation Models**

#### **3.5.1 Rationale**

Connectionist-type models are of interest to acquisition researchers in part because of the perception that such network-type models are biologically plausible models of the brain

(Elman et al. 1996; Clark and Lappin 2011). On this view, modeling provides a window into cognition. The connectionist framework lends itself well to the investigation of processes and developmental patterns through implementation of machine learning algorithms.

### **3.5.2 Characteristics of Connectionist Models**

Connectionist models share a common set of components: units, connectivity, and learning rules. Every model posits a set of units (nodes) with defined activation levels. The units exist in layers: an input layer, an output layer, and a ‘hidden’ layer that enables the model to transform the input representations as needed, according to a specified learning procedure. Individual units are connected (or not) in a specified architecture. Learning procedures are specified by algorithms (rules) for manipulating the inputs; such rules are in the form of adjustments to the strengths of the connections between units. The model learns on the basis of every input it receives; the model embodies an adaptive, dynamic system.

The study by Williams and Kuribara (2008) which will be considered in Chapter 4 uses a particular architecture known as a simple recurrent network (SRN) (Elman 1990; Elman et al.1996). This architecture provides an additional layer of ‘context’ units that store the activation levels of the hidden units for an additional iteration of the model. These activation levels are then fed back into the hidden units, in a way ‘recycling’ the information from previous steps in the model. Effectively, the context layer makes the model sensitive to the whole history of an input sentence. Elman’s (1990) model was successful in recognizing simple sentences and in categorizing words into linguistic categories such as noun, verb, inanimate, animate, human and nonhuman.

### 3.5.3 Evaluation of Connectionist Models

Every model instantiates a myriad of design decisions. The number of nodes in each layer, the particular learning algorithms, the initial state of the weights on the connections, the type and amount of input presented to the model, all must be specified. Rather than giving a detailed critique of a number of models, in this section I will mention advantages and disadvantages of the methodology in general. In Chapter 5, I will offer specific comments on the model proposed in Williams and Kuribara (2008).

One advantage lies in the precision that is imposed on the experiment as a result of the design decisions that need to be made in the construction of the model. The researcher is forced to make very specific assumptions regarding the initial state of the model, the input the model receives, and the learning mechanisms that the model brings to bear on whatever acquisition problem is posed to it. These decisions constitute a complete theoretical position. Another advantage of conducting a simulation is the experimenter's degree of control over the manipulation of variables. For instance, the frequency and timing of linguistic input can be rigidly specified, something that can only be attempted in naturalistic settings. Variables can be modified independently, facilitating investigation of particular effects and interactions. For example, frequency, quantity, and timing of input can be varied in a number of different simulations, holding the learning algorithms constant (Li 2009, p.641). These types of manipulations are very difficult to achieve empirically.

Regarding disadvantages to connectionist models, again, for the purpose of this section, my comments will be at a very general level, with specifics reserved until Chapter 5 for the specific connectionist model considered in Chapter 4. A major caveat at

this point in the evolution of the connectionist approach is that we cannot substantiate the theory's claims to neural plausibility. We do not at this point have a good enough understanding of the actual neural mechanisms of language acquisition and use or indeed cognition in general to be able to confidently make a correspondence between the particular architectures instantiated in a given model and network structures in the human brain. This mismatch in itself does not discount the method. Early connectionist models (Rumelhart and McClelland (1986); Thomas (1997); N. Ellis and Schmidt (1997); among others) investigated relatively simple phenomena such as word recognition, past tense morphology and plural morphology, often on a small number of lexical items. More recent models investigate more complicated phenomena and larger input sets, even including natural corpora (see for instance Li and Farkas (2002), investigating bilingual child lexical development with 500 words from each of two languages). So we see that two current challenges for the connectionist approach could be overcome in time.

#### **3.5.4 Issues Particular to L2A Research**

For the investigation of L2A, connectionist models need to pay particular attention to the initial state of the model. The initial state should reflect the instantiation of a language in the form of connection weights that reflect the kinds of biases and expectations that would come with prior knowledge of a language. Because of the above-mentioned ability to manipulate variables, connectionist models can be very useful in simulating things like early or late age of first exposure to an L2, and duration of exposure.

In this chapter we have considered key experimental methodologies for investigating learner acquisition of subtle and complex linguistic phenomena. We will

use this information in the next few chapters as we review and critique three sets of studies that make use of these judgment tasks.

## **CHAPTER 4: LITERATURE REVIEW: EMPIRICAL EVIDENCE FROM L2A STUDIES**

### **4.1 Overview**

Chapter 4 launches the empirical investigation of the theoretical issues introduced in chapters 1 and 2. This chapter surveys three sets of existing L2A studies. Each set of studies centers on a particular linguistic phenomenon that has been argued to represent a poverty of the stimulus problem: the Overt Pronoun Constraint (OPC), scrambling, and quantifier scope interactions. Each set contains studies conducted from a generativist approach and from an emergentist approach. These studies have been chosen for examination for two reasons: POS and adult L2A. If POS phenomena truly exist, and if they are truly acquired, then POS studies pose the strongest case for the existence of innate language-specific faculties (White 2003; Schwartz and Sprouse 2013). And since POS phenomena such as scope interpretations are complex, it is helpful to study adults whose cognitive faculties are mature. If it can be shown that adult L2 learners rely on innate knowledge (UG), then the case for UG in L1A is strengthened. On the other hand, if L2A is shown to be accomplished on the basis of domain-general faculties, then the case for UG in L1A is neither advanced nor contradicted.

Note that from a big-picture point of view, the Full Access version of the generative approach and the emergentist approach make the same basic prediction, that is, that in principle the ultimate attainment of an L2 grammar is possible. The approaches differ on the mechanisms in play in the acquisition process, and for this reason they differ in their predictions with respect to the developmental path of acquisition (Meisel 2011; Ellis 2002, 2003; among others). As discussed in Chapter 2, under the generative approach a learner will make a sudden leap towards native-like behavior, possibly

(though not necessarily) on the basis of a small amount of input. Under the emergentist approach, a learner will demonstrate incremental progress toward the target language, requiring a sufficiently large amount of input in order to instantiate the new patterns. We will keep these predictions in mind as we examine the empirical evidence.

The theoretical rationale, methods, results and conclusions will be summarized for each study in preparation for the evaluative discussion in Chapters 5 and 6. Particular attention will go to three areas. First, with regard to the phenomenon itself, the discussion will note the basis on which the phenomenon was classified as POS (by generativists) or the basis on which the input is argued to provide the necessary information to enable acquisition (by emergentists). Second, the discussion will highlight the study methodology including the choice of learners in terms of proficiency levels as well as the choice of task. After presenting the empirical results, the authors' interpretations of the data they present will be given. Chapter 5 below will in some cases offer a different interpretation of the same results.

## **4.2 Overt Pronoun Constraint Studies**

This chapter begins with a review of a set of L2A investigations of the Overt Pronoun Constraint (OPC). These studies comprise a classic example of POS argumentation. In particular, the generative studies have been cited as providing strong evidence for the role of UG in second language acquisition (see for instance commentary in White 2003a,b). After providing a brief statement of the OPC, this section will review three studies in the generative framework and one emergentist response.

The OPC (Montalbetti 1983) applies to null subject languages.<sup>38</sup> Whereas non-null subject languages require overt subjects, null subject languages permit overt subjects but also permit null subjects. The OPC is concerned with the interpretation of embedded null and overt pronominal subjects. The statement in (1) below is from Kanno (1997, p.267):

(1) Overt Pronoun Constraint (OPC)

In languages that permit null arguments, an overt pronominal must not have a quantified NP as antecedent. The prohibition on antecedents includes *wh*- words, as will be illustrated shortly.

The sentences in (2) and (3) below illustrate the role of the OPC in the contrast between null subject languages and non-null subject languages. English is a non-null subject language. In English, a (necessarily overt) pronominal subject of an embedded clause can take either a quantified or a referential antecedent. In (2a) under the given co-indexation, *he* receives a bound variable interpretation, that is, *he* does not refer to a particular individual. Situations that make (2a) true can include situations in which many people think of themselves as winners. In (2b), *he* is co-referential with *Tom* (examples from Kanno 1998, p.1127):

---

<sup>38</sup> Kanno (1997, fn3) gives examples from Spanish, Korean, and Chinese in addition to the Japanese examples which are the main focus of her study. She suggests, as do Pérez-Leroux & Glass (1999, p.227) that the OPC may be a linguistic universal. However, Sheen (2000, p.814) objects to characterizing the OPC as being present in all null subject languages. He points out the small number of languages in which the OPC has been documented and provides some anecdotal evidence questioning its existence in several null subject languages such as Finnish, Hungarian, and Turkish, among others. Marsden (2002, fn2) points out that the OPC has been argued to be subsumed under more general properties of UG (see for instance Noguchi 1997). In any case, the evidence for OPC constraints on the interpretation of embedded overt subjects in Japanese and Spanish does not seem to be in dispute. Since these are the languages under investigation in the studies reviewed in this section, the question of the universality of the OPC is not critical to the POS arguments presented in these studies.

(2) a. English overt pronominal subject with quantified antecedent:

Who<sub>i</sub> thinks he<sub>i</sub> is the winner?

b. English overt pronominal subject with referential antecedent:

Tom<sub>i</sub> says that he<sub>i</sub> is the winner.

In contrast, in a null subject language such as Japanese where the OPC is active, a referential antecedent is permitted for an overt pronominal subject but a quantified antecedent is not. Overt pronominal subjects in embedded clauses cannot receive a bound variable interpretation. This contrast is illustrated in (3) below. The examples in (3) are based on Kanno 1997, p.266, 267 respectively:

(3) a. Japanese overt pronominal subject with quantified antecedent:

\*Dare<sub>i</sub> ga [s kare<sub>i</sub> ga sore o mita to] itta no

Who Nom he Nom that Acc saw that said Q

‘Who<sub>i</sub> said that he<sub>i</sub> saw that?’

b. Japanese overt pronominal subject with referential antecedent:

Tanaka-san<sub>i</sub> wa [s kare<sub>i</sub> ga sore o mita to] itta

Tanaka-Mr. Top he Nom that Acc saw that said

‘Mr. Tanaka<sub>i</sub> said that he<sub>i</sub> saw that.’

Note that the OPC does not restrict the possible antecedents for null subjects. Null subjects can have either quantified or referential antecedents, as illustrated in (4) below. In (4a), the null pronoun in the embedded clause can have a bound variable interpretation. In (4b), the null embedded subject has a matrix subject referring antecedent. Examples are based on Kanno (1998, p.1129, 1127 respectively) and *e* indicates a phonetically empty pronoun, translated here as *he*:

(4) a. Japanese null pronominal subject with quantified antecedent:

Dare<sub>i</sub> ga [<sub>S</sub> kyoo e<sub>i</sub> osoko-naru] to itteiru no

Who<sub>i</sub> Nom today (he<sub>i</sub>) late-become that says Q

‘Who says that he is going to be late today?’

b. Japanese null pronominal subject with referential antecedent:

Taroo<sub>i</sub> wa [<sub>S</sub> e<sub>i</sub> kurasu de itiban da] to omotteiru

Taroo Top (he<sub>i</sub>) class in best is that thinks

‘Taroo<sub>i</sub> thinks that (*he<sub>i</sub>*) is the best in the class.’

The examples so far involve intra-sentential antecedents. However, the embedded overt pronouns in (2a) and (3a) and the embedded null pronoun in (4a) can possibly take extra-sentential referring antecedents, that is, the pronouns can co-refer with an entity established in previous discourse instead of taking (or trying to take, in the case of (3a)) their matrix subjects as antecedents. (3a) is not ungrammatical on the surface, rather, the interpretation indicated by the given co-indexation is not available. Section 4.2.1 will show how these contrasts between overt subjects in null subject and non-null subject languages will be important to POS argumentation.

#### **4.2.1 Kanno (1997, 1998)**

This section discusses Kanno’s 1997 study in some detail. Because her 1998 study shares many characteristics with respect to goals and methods, the discussion of Kanno (1998) will be abbreviated.

##### **RATIONALE**

The theoretical goal for Kanno’s 1997 study is to provide evidence on the question of whether or not UG is active in second language acquisition. Kanno (1997, p.267-268) argues that the OPC provides a clear poverty of the stimulus problem for L2 learners of a

null subject language when their L1 does not allow for null arguments, thus providing a good testing ground for UG availability. She targets L1 English learners of L2 Japanese.

For her POS arguments, consider Table 4.1 below which summarizes the Japanese facts that L2 learners need to acquire:

**Table 4.1: Possible Antecedents for Embedded Subject Pronouns in Japanese**

	<b>Null Pronoun</b>	<b>Overt Pronoun</b>
<b>Referring Antecedent</b>	yes	yes
<b>Quantified Antecedent</b>	yes	<i>no</i>

The bottom right cell of the table represents the OPC. Kanno argues that the facts in this table and their contrast to the English facts support that the OPC is a poverty of the stimulus problem. From a learner’s point of view, positive evidence in the L2 will not eliminate the possibility of an overt pronoun having a bound variable interpretation.<sup>39</sup>

The fact that null pronouns can have both bound variable and referring antecedents could be expected to lead the learner to hypothesizing that overt pronouns can also take both types of antecedent. This situation corresponds to Schwartz and Sprouse’s (2013) ‘bankruptcy of the input’ scenario, as illustrated in example (20) of Chapter 2, above. In addition to the L2 facts, the learners under study have L1 knowledge that quantified antecedents are permitted for overt pronouns in the L1. The L1 does not instantiate a constraint on the bound variable interpretation.

---

<sup>39</sup> See Chapter 2 Section 2.3.1.2 for the discussion of deficiencies in positive input.

To complete her POS argument, on the basis of examining several textbooks and instructional materials, Kanno finds that L2 instruction does not address the OPC. To summarize, Kanno argues that acquisition of the Overt Pronoun Constraint constitutes a poverty of the stimulus problem on the basis of under-informative positive evidence, the lack of the constraint in the L1, and a lack of instruction in the L2 classroom. Her research questions and predictions are:

- Do L1 English learners of L2 Japanese *accept* quantified antecedents (that is, bound variable interpretations) for overt embedded pronominal subjects?  
If yes, then learners are showing L1-like behavior.
- Or, do these learners *reject* quantified antecedents for overt embedded pronominal subjects?  
If yes, then they are showing target-like behavior and therefore may have acquired the OPC.

#### **EXPERIMENTAL METHOD**

Kanno's subjects consisted of a control group of native speakers of Japanese (n=20) and a learner group of native speakers of English who were enrolled in the fourth semester of Japanese language courses at the University of Hawai'i (n=28). She did not assess learner proficiency experimentally. With respect to language background, the learners had never lived in Japan nor had they ever lived with a native speaker of Japanese.

The experimental task consisted of a coreference judgment task in the form of a written questionnaire. Subjects were presented with four sets of biclausal sentences, representing the following four patterns:

- embedded overt pronoun with matrix subject quantified antecedent
- embedded null pronoun with matrix subject quantified antecedent

- embedded overt pronoun with matrix subject referring antecedent
- embedded null pronoun with matrix subject referring antecedent

The first of these is the crucial indicator of acquisition of the OPC. The test sentences were presented without any introductory context. Learners were asked about the interpretation of the embedded pronoun. When a test sentence contained a quantified antecedent (similar to (3a) above), the answer choices were (a) ‘same as *dare*’ or (b) ‘another person’. When the test sentence contained a referring antecedent (similar to (3b) above) such as ‘Tanaka-san’, the choices were ‘Tanaka’ or ‘someone other than Tanaka’. In both cases, answer (a) corresponds to choosing the matrix subject as antecedent, and answer (b) corresponds to choosing an extra-sentential antecedent. Learners were instructed to indicate both answers if they felt that both answers could be considered appropriate.

## RESULTS

Kanno reported results by group and by individual. The group results for the native control group are summarized in Table 4.2, and the group results for the L2 learners are summarized in Table 4.3. The tables report the percentage of total responses in each test condition reported. For instance, native speakers accepted quantified antecedents for overt pronouns only 2% of the time.

**Table 4.2: Native Speaker Control Group (n=20), Percentage of Acceptances in Each Test Condition**

	<b>Embedded Null Subject Pronoun</b>	<b>Embedded Overt Subject Pronoun</b>
<b>Referring Antecedent</b>	100%	47%
<b>Quantified Antecedent</b>	83%	*2%

Note: \*indicates reading was predicted to be disallowed. Data from Kanno (1997, p.272).

Native speakers accepted both referring and quantified antecedents for null pronouns, albeit they seemed to prefer referring antecedents. Overt pronouns are permitted to take referring matrix subject antecedents as well. Crucially for the OPC, native speakers rarely permitted a quantified antecedent for an overt embedded pronominal subject.

**Table 4.3: L2 Learner Group (n=28), Percentage of Acceptances in Each Test Condition**

	<b>Embedded Null Subject Pronoun</b>	<b>Embedded Overt Subject Pronoun</b>
<b>Referring Antecedent</b>	81.5%	42.0%
<b>Quantified Antecedent</b>	78.5%	*13.0%

Note: \*indicates reading was predicted to be disallowed. Data from Kanno (1997, p.273).

Again, both referring and quantified antecedents were accepted for null subject pronouns, and overt subjects were also permitted to take referring antecedents but at a lower rate. Quantified antecedents were accepted for overt subject pronouns only 13% of the time. Kanno reports (1997, p.273) that the learner performance is not significantly different from the native speaker performance in any test condition. Both native speakers and L2 learners differentiate between null and overt pronouns with respect to accepting a quantified antecedent. Kanno reports (1997, p.274) that the difference is statistically significant.

Kanno also analyzed the results for each individual. Native speaker results were consistent across all individuals. Eighteen native speakers rejected a quantified antecedent for an overt pronoun 5 times out of 5, and two rejected the same 4 times out of

5. Learner results showed more variability. Fifteen of the subjects rejected a quantified antecedent for an overt pronoun 5 times out of 5. Seven more subjects rejected the same at least 4 times out of 5, and four more subjects rejected the quantified antecedent for the overt pronoun either 2 or 3 times out of 5.<sup>40</sup> In summary, all native speakers and 86% of L2 learners rejected quantified antecedents for overt pronouns at least 4 times out of 5.

## **CONCLUSIONS**

Kanno argues that her results are consistent with learners having acquired the Overt Pronoun Constraint. At the group level, learners rejected quantified antecedents for overt pronouns at a rate that was not statistically significant from native speaker results. At the individual level, although the learner group showed more variability than the native speaker group, 86% of the learners were consistent in this rejection, where consistency was defined as at least 4 rejections out of 5 test items. On the basis that the OPC represents a poverty of the stimulus problem, and that learners appeared to acquire the constraint, Kanno concludes that her results support that UG is active in L2 acquisition.

Turning to Kanno (1998), we find many similarities but some important differences.

## **RATIONALE**

The purpose of Kanno's 1998 study was to explore L2 learner variability in development and in ultimate attainment. The research is based on the assumption that L2 learners have full access to UG at all times. The official research question for this study asked whether L2 learners have "longitudinal consistency", that is, whether they access UG at all times during their developmental path. Kanno performed a "limited-scope" longitudinal study,

---

<sup>40</sup> Data from two subjects was not included in this tally because those learners did not demonstrate a contrast between null and overt pronouns.

testing L2 learners at two points in time. For the purpose of her study, the research questions were:

- Do L2 learners show consistent performance across time?
- Do L2 learners show consistent performance across individuals?

The studies are of “limited-scope” because they use access to one UG principle as a proxy for access to all of UG. These studies are interesting for our purposes because the principle she focuses on is the Overt Pronoun Constraint. Her subjects are once again L1 English learners of L2 Japanese. To investigate L2 learner access to UG, Kanno (1998) assesses learner performance with respect to the OPC. She provides the same arguments for the OPC as a poverty of the stimulus problem and therefore an appropriate test of learner access to UG. For our purposes, therefore, her research questions are identical to those in the 1997 study:

- Do L1 English learners of L2 Japanese accept quantified antecedent for overt embedded pronominal subjects?
- Or, do these learners reject quantified antecedents for overt embedded pronominal subjects?

If the first question is answered affirmatively, then learners are showing L1-like behavior.

If on the other hand the answer to the second question is yes, then the learners are performing in a native-like manner and may have acquired the OPC.

#### **EXPERIMENTAL METHOD**

Kanno (1998) reran the same experiment from the 1997 study, with two modifications.

First, she has different numbers of subjects in her native control and L2 learner groups:

12 and 29. Second, she presented three rather than four sets of five test sentences to the

learners: null pronouns with quantified antecedents, overt pronouns with quantified

antecedents, and overt pronouns with referring antecedents. She did not offer test sentences that contained an embedded null subject pronoun and a referring antecedent. The test procedures were the same as in the 1997 study (described above), and in fact the test sentences themselves were the same. The same groups of individuals were tested once and then again after an interval of 12 weeks.

## **RESULTS**

Once again, Kanno presented group results and individual results for the two groups, but this time she has two sets of results for each group. She reported (1998, p.1130) that native speakers in Session 1 distinguished between null and overt pronouns with respect to their ability to take quantified antecedents: quantified antecedents were permitted for null subjects 85% of the time but only less than 2% of the time for overt pronouns. This difference is statistically significant. In Session 2, native speakers permitted quantified antecedents for null pronouns 75% of the time, and for overt pronouns, never (0%). Kanno characterizes the native speaker group results (p.1131) as “consistent and categorical”. The OPC is active and stable across time in native grammars.

The group results for L2 learners in Session 1 showed them also making a statistically significant distinction between null and overt subjects. Quantified antecedents were permitted for null pronouns 73% of the time, but for overt pronouns only 29% of the time. On the other hand, referential antecedents were permitted for both null and overt pronouns with no strong preference for either. The results in Session 2 were similar, with quantified antecedents accepted at 72% and 34% for null and overt pronouns, respectively. Results for acceptance of referential antecedents again showed no strong preference between null and overt pronouns. Overall, learner performance seems similar to native speaker performance, however, Kanno does not report whether or not

learner performance and native speaker performance are statistically significantly different.

With regard to individual results, the native speakers were consistent across individuals in both sessions. However, the analysis of individual L2 learner results showed considerable variability. Once again Kanno adopts rejection of a quantified antecedent for an overt pronoun at least 4 out of 5 times as the standard for acquisition of the OPC. On this basis, 18 learners in Session 1 and 15 learners in Session 2 showed mastery of the constraint. If the standard for rejection is relaxed to 3 rejections out of 5 (still a majority of correct rejections), then 21 out of 29 individuals in Session 1 and 22 out of 29 individuals in Session 2 showed knowledge of the OPC. However, a detailed breakdown showed that only 9 individuals showed consistent mastery of the OPC across both sessions. Nine individuals showed mastery only in the first session, 6 individuals showed mastery only in the second session, and 5 did not achieve mastery in either session.

Further, Kanno analyzed the individual learner data for consistency across sessions with respect to all three types of test sentences. She defined ‘non-uniform’ (inconsistent) performance as a difference of at least two responses. For test sentences featuring quantified antecedents and overt pronouns, by this definition 13 individuals performed inconsistently, and 16 performed consistently. For the other two sets of sentences, more individuals (21 out of 29) performed consistently. Kanno comments (p.1137) that these data support a “selective” variability in learner acquisition since the most variability occurred on the test sentence that is crucial to the OPC. If learner

performance variability were due to general factors such as attention or confusion, one would expect the variability to be more evenly distributed.

## **CONCLUSIONS**

Kanno's first finding is that as a group, L2 learners showed acquisition of the OPC. Her second finding was that as individuals, L2 learners showed considerable variation. She concludes based on the first finding that UG is active in L2 acquisition. She argues based on the second finding that UG is variably accessible across the L2 developmental path.

### **4.2.2 Pérez-Leroux and Glass (1999)**

We now turn to the third and final generative study under consideration in this section. Pérez-Leroux and Glass (1999) (henceforth P-L&G) investigated L2 acquisition of the OPC as part of a set of experiments with a broader theoretical goal. We will focus on the OPC study.

## **RATIONALE**

The background facts on OPC have been established in the previous discussion. P-L&G are studying L1 English learners of L2 Spanish rather than Japanese, but the same facts and contrasts between the L1 and L2 hold. Therefore the particular research questions of interest are once again whether L2 learners (whose L1 does not permit null subjects) accept or reject quantified antecedents for overt embedded subject pronouns.

## **EXPERIMENTAL METHOD**

Because P-L&G are interested in the developmental path of L2 learners, their subjects are three groups of L1 English speakers learning L2 Spanish at varying levels of proficiency. An independent measure of proficiency was not provided. The elementary learners (n=39) were enrolled in their fourth semester of Spanish language study at a large US

university. The intermediate proficiency learners (n=21) were enrolled in fourth-year advanced Spanish language and Spanish literature classes. The highest proficiency group included instructors and graduate teaching assistants each of whom had at least 7 years of Spanish language ‘experience’ either in the form of instruction or immersion or both. A native speaker control group (n=20) completed the subject population.

The subjects were given a controlled production task in which they were asked to translate a sentence from English to Spanish. For each test item, the subjects were given a 2-sentence ‘story’ to read in English. The story provided an unambiguous interpretation for the pronoun that appeared in the target sentence. The story either established a referent or described a situation in which people talked about themselves. The latter story type serves to provide a bound variable interpretation for the pronoun. P-L&G predict that native speakers and learners who have acquired the OPC will produce only null pronouns in the bound variable context, and either null or overt pronouns in the referential context. The test consisted of 8 stories total, 4 of each type. An example of each story type is given in (5) and (6) below, followed by a target sentence for translation:

(5) Referential story (P-L&G 1999, p.232)

In the O.J. Simpson trial, it is clear that the press has a negative bias against the defendant in their reporting. Some journalist said that he was a wife-beater.

- To translate: ‘But no journalist said that he is guilty’.
- Target translation:

Ningún periodista dijo que él era culpable.

no journalist        said that he was guilty.

(6) Bound-variable story (P-L&G 1999, p.233)

The court charged that some journalists had been in contact with the jurors. Several of them were questioned by the judge.

- To translate: ‘No journalist admitted that he had talked to the jurors.’
- Target translation:

Ningún periodista admitió que le había hablado a los jurados.

no journalist admitted that  $\emptyset$  to-them-had spoken to the jurors.

## RESULTS

P-L&G state that the task design was effective at eliciting responses containing the desired target structure. However, they report that some responses had to be eliminated from the results for a variety of reasons, including sentences produced with a lexical NP as the embedded subject, monoclausal sentences, or incomplete sentences. The results are given as a percentage of null or overt embedded pronouns produced in translations following bound variable stories and referential stories. Table 4.4 below summarizes a subset of the results. Percentages do not always add up to 100 because some responses were discarded.

**Table 4.4: Responses to Translation Task Coded by Pronoun Type and Story Type**

	Bound Variable Stories		Referential Stories	
	Null	*Overt	Null	Overt
<b>Native speakers</b>	85.0%	13.7%	31.3%	67.5%
<b>Elementary</b>	57.7%	34%	21.2%	67.9%
<b>Intermediate</b>	73.8%	26.2%	35.7%	59.5%
<b>Advanced</b>	93.1	0%	58.3%	31.9%

Note: \*indicates reading was predicted to be disallowed. Data taken from P-L&G (1999, p.234)

At the group level, native speakers and learners at every proficiency level distinguish between null and overt pronouns, using null pronouns more than overt in bound variable contexts at a statistically significant level. This behavior is consistent with the OPC. Use of null pronouns appears to increase with proficiency; however, P-L&G (1999, p.240) find no significant differences between learners and native speakers at any proficiency level based on a proportional difference score.<sup>41</sup>

Individual results were not presented.

## **CONCLUSIONS**

P-L&G conclude that their results support the acquisition of the OPC at all proficiency levels, even at the elementary level. Similarly to Kanno, they argue that the finding that early learners are successful with respect to the OPC provides evidence for an active role of UG in L2A.

### **4.2.3 Sheen (2000)**

The final study under consideration in this section was done in a non-nativist framework. Sheen (2000) proposes a non-UG account of L2 acquisition of the Overt Pronoun Constraint. He supports this theoretical goal in two ways: by providing arguments and evidence against the UG account given in Kanno (1998), and by providing arguments and evidence that input provides relevant evidence, and in combination with general problem solving skills is sufficient for successful acquisition. Sheen's overall research question asks whether or not UG is necessary to account for L2 acquisition. Sheen (2000) presents several experiments, each of which supplies particular research questions whose answers bear on the overarching question of the existence of UG. The experiments will be

---

<sup>41</sup> They calculate a proportional difference rather than using the absolute differences in order to compensate for the different number of responses from each subject.

presented separately below, during the course of the presentation of Sheen's non-UG proposal.

First we will review the arguments and evidence that Sheen presents to weaken Kanno's (1998) UG account.<sup>42</sup> Sheen (2000, p.814) argues that the OPC is not a linguistic universal. He suggests that a claim of universality should not be made until a much larger sample (ideally, all) of the world's potentially thousands of null subject languages have been verified to have an active OPC. He conducts an informal survey of linguist native speakers of eight null subject languages, including Spanish and Chinese among others, and found that none were willing to categorically rule out the possibility of a quantified antecedent for an overt embedded subject pronoun. Based on these two arguments he suggests that it is premature to claim the OPC as a linguistic universal and therefore part of UG.

Sheen presents two experiments that he argues cast doubt on a UG account of OPC L2 acquisition. One experiment seeks to verify the reasons behind learner judgments and the other investigates the performance of L1 speakers of a null subject language learning a null subject L2 language.

### **Experiment 1: Understanding learner judgments**

#### **RATIONALE**

Sheen (2000, p.799) points out that in principle, different learners can report similar judgments for different reasons (for discussion, see Section 3.2.3.2 above). The fact that learners perform similarly to native speakers does not guarantee that they have acquired

---

<sup>42</sup> Sheen (2000) comments exclusively on Kanno (1998). To the extent that Kanno (1998) builds on Kanno (1997), his comments cover both studies. However, following his lead I will refer exclusively to Kanno (1998) in this review.

the proposed target phenomenon. Sheen's research question for this experiment can be formulated as follows:

- Do L2 learners who perform correctly on OPC test sentences actually use the OPC to arrive at their judgments?

#### **EXPERIMENTAL METHOD**

The subjects were 33 native speakers of French (a non-null subject language) all of whom were students at a university in Quebec, Canada. The subjects were learning Spanish but there is no mention of their proficiency level or language experience.

The task was modeled after Kanno's experiment, however, the entire task consisted of the crucial type of sentence for determining OPC compliance: biclausal sentences (we are not told how many) with quantified antecedents and overt embedded subject pronouns. A sample test sentence is given in (7) below:

- (7) Alguien dijo que él tenía que coger el tren  
'Somebody said he had to catch a train'

The students were asked about the possible interpretations for the embedded subject and were given the same response choices as in Kanno's studies: the pronoun could receive a bound variable interpretation by taking the quantified matrix subject as antecedent; the pronoun could refer to someone else in the discourse; or both. The crucial difference between Sheen's procedure and Kanno's procedure was that Sheen asked the students to provide the reasons for their answers.

#### **RESULTS**

According to the text, Sheen reports the percentage of students who give each response type, however, in his Table 4 (p.808) he compares his percentages to Kanno's results which were given as a percent of total responses. If he only gave one test sentence to each

subject, then the percentages are equal, however, that seems implausible. Putting that aside, the crucial data point from his experimental task is that 42.5% of the students did not permit the bound variable reading and instead chose (b) ‘another person’ in the discourse as the coreferent for the pronoun. This choice is in compliance with the OPC. It is important to note that Sheen reports (p.808) that the majority of the students (57.5%) responded that both the bound variable and the referential antecedents were possible, indicating that the majority had not yet learned the restriction on the bound variable interpretation for overt subject pronouns.

When asked to explain their choice, 71% of the students whose responses were in compliance with the OPC said that they were ‘obeying a rule’ that overt pronouns refer to an antecedent from outside the matrix clause. The other 30% of the students in compliance with the OPC plus the students who permitted both interpretations said that they had a ‘feeling’ about it.

## **CONCLUSIONS**

Sheen concludes that learners who show evidence of OPC acquisition have not necessarily acquired a UG constraint. Rather, learners may be using completely different means to arrive at judgments that appear to trend toward native behavior.

Sheen presents a second experiment designed to provide evidence for or against a UG account of OPC acquisition.

### **Experiment 2: UG availability through the L1**

#### **RATIONALE**

Sheen argues (p.808) that if the OPC is universally available in null subject languages, then L2 learners whose L1 is a null subject language should have the OPC available to

them through their L1. Having the OPC available should facilitate acquisition. He proposes the following:

- Do native speakers of another null subject language perform as well as or better than native speakers of non-null subject languages on the OPC, ‘all things being equal’?

If yes, the evidence supports a UG account in which learners have access to UG principles including the OPC. If no, the evidence does not support that learners have access to UG.

#### **EXPERIMENTAL METHOD**

The subjects were Korean L1 graduate students who were attending Tottori University in Japan. No independent proficiency measure was given. The task was ‘a test similar to Kanno’s’ (p.808). No further information is given.

#### **RESULTS**

On the crucial OPC sentences that offer a quantified matrix subject and an overt embedded subject pronoun, the Korean graduate students permitted both a bound variable interpretation and a referential interpretation 73% of the time. They permitted only the referential interpretation 24% of the time. This contrasts with Kanno’s English L1 subjects in Session 1 of her 1998 experiment, whose performance at a group level was not significantly different from native speaker performance. Kanno’s subjects permitted only the referential antecedent 71% of the time, and permitted both referential and bound variable interpretations 4% of the time. These results are quite opposite the Korean learner results.

## CONCLUSIONS

Based on the hypothesis, the worse performance of L2 learners whose L1 is a null subject language compared to the performance of L2 learners whose L1 is a non-null subject language would indicate that the OPC is not available and active in L2 acquisition.

However, Sheen chooses to interpret these results as problematic for the claim that Korean is a null subject language (p.809). If Korean is a non-null subject language, then Korean L1 learners of L2 Japanese would not be expected to do any better than English L1 learners of Japanese. On the basis of information from linguist native speakers of Korean, Sheen suggests that Korean permits both types of antecedent, quantified and referential, for embedded subject pronouns. He thus concludes that the Korean learners in this experiment are behaving in an L1-like manner.

We conclude by discussing the arguments and evidence that Sheen presents with respect to the sufficiency of L1 transfer, L2 input and general cognitive learning skills for the acquisition of the OPC. Experiment 3 considers L1 transfer.

### **Experiment 3: L1 transfer**

#### RATIONALE

It is generally accepted that L2 learners rely on relevant L1 characteristics, at least early on. Kanno (1998) rejected the idea of L1 English influence on her subjects' L2 performance on the basis of a supplemental experiment in which she investigated English native speaker intuitions on English sentences that corresponded to the Japanese test sentences. She found that English native speakers preferred intra-sentential (that is, matrix subject) antecedents for an embedded subject pronoun *he* 85% of the time whether the matrix subject was quantified or referential. She concluded that English L1 learners of L2 Japanese were not transferring L1 behavior to the L2. Sheen found that his own

intuitions did not match Kanno's native speakers' intuitions, and was motivated to rerun Kanno's experiment for verification of the results.

#### **EXPERIMENTAL METHOD**

There were two native English speaker subject groups: American (n=28) and British (n=15). The British were expatriate residents of a compound in Saudi Arabia. The experiment was conducted via internet discussion lists. The task was 'similar to Kanno's' (p.804). No further information was given.<sup>43</sup>

#### **RESULTS**

For the crucial OPC test sentences with a quantified matrix subject and an overt embedded pronoun, the American and the British native speakers accepted either an extra-sentential antecedent or both the intra-sentential (bound variable) and extra-sentential (referential) interpretations 96% and 100% of the time, respectively. For test sentences with a referential matrix subject and an overt embedded pronoun, the American and the British native speakers accepted either an intra-sentential or both intra- and extra-sentential antecedents 100% of the time.

#### **CONCLUSIONS**

Sheen interprets his experimental results (p.805) as showing "a clear consistent systematicity between quantifier (type 2) and referential (type 3) antecedents". In particular, he claims that the data show that when a sentence has a quantified matrix subject, native speakers prefer an extra-sentential antecedent for an embedded pronoun, and when a sentence has a referential matrix subject, native speakers prefer an intra-

---

<sup>43</sup> A third subject group, Canadian French high school students, was also recruited. Their results will not be discussed here.

sentential antecedent.<sup>44</sup> He thus concludes that Kanno's English learners of Japanese were likely relying on L1 intuitions when forming their L2 judgments.

Having settled on L1 transfer as a positive factor in L1 English speaker acquisition of L2 Japanese, Sheen investigates the possible contribution to L2 acquisition from the learners' L2 input. On the basis of examination of L2 textbooks and instructional material, Kanno (1997, 1998) had noted that her L2 learner subjects had not had any exposure to biclausal sentences containing embedded subject pronouns. Kanno therefore dismissed L2 input as a factor in her subjects' L2 acquisition. Kanno (1998, p.1134) noted that in the instructional materials at the instructional level where her subjects were enrolled, the overt pronoun *kare* (the only pronoun used in Kanno's tests) only occurred in monoclausal sentences and always had an extra-sentential antecedent.

Sheen (p.803) argues that the monoclausal evidence is in fact relevant to the OPC. In particular, he suggests that learners can induce the following simple '*kare* rule' based on such sentences: *kare* ('*he*') always has an extra-sentential antecedent. He argues that the *kare* rule provides an alternate account of the lack of a bound variable interpretation for the biclausal sentences in Kanno's tests, because the bound variable interpretation would only come about if the (intra-sentential) quantified matrix subject served as the antecedent for the embedded overt pronoun. By rejecting the intra-sentential antecedent, the learners appear to have acquired the OPC.

In summary, Sheen proposed a non-UG account of L2 learner performance that appears to conform to the OPC. Sheen disputed the existence of the OPC as a linguistic universal. He proposed that the English L1 learners in Kanno's study relied on L1

---

<sup>44</sup> The validity of this conclusion will be discussed in Chapter 5.

transfer and induction from L2 instructional materials to provide the guidance for their experimental judgments. He offered experimental evidence that L2 learners do not use the OPC in arriving at their experimental judgments.

The merits and weaknesses of the studies presented in this section will be discussed in Chapter 5.

### **4.3 Scrambling Studies**

This section reviews two studies: a generative account and an emergentist account of second language acquisition of scrambling. Scrambling is the optional reordering of elements in a sentence. The option to scramble is available in some languages but not all, and languages that allow scrambling can vary with respect to the properties of the permitted scrambling operations. For instance, as we will see, German permits local scrambling but Japanese permits both local and long distance scrambling. After reviewing relevant data we will turn to the studies: Hopp (2005) for the generative approach, and Williams and Kuribara (2008) (hereafter W&K) for the emergentist approach. The three languages involved in these two studies are English, Japanese, and German. Hopp (2005) studies English L1 and Japanese L1 learners of L2 German. W&K study English L1 learners of a semi-artificial language in which English lexis is combined with Japanese syntax. Basic word order<sup>45</sup> facts for each language are given below.

The canonical word order in English is SVO<sup>46</sup> (see example (8a) below). Some different constituent orders are found in questions, imperatives, topicalizations and a few other constructions. However, scrambling is not permitted. Verbal arguments cannot be

---

<sup>45</sup> I will use ‘word order’ and ‘constituent order’ interchangeably.

<sup>46</sup> Here and throughout, S= subject, O= object, V= verb, I = indirect object

reordered, and adverbs may not appear between the verb and its object, as shown in example (8b-c) below. This holds for embedded clauses (8d) as well as main clauses.

- (8) a. John gave Mary the book reluctantly.  
 b. \*John gave the book Mary reluctantly.  
 c. \*John gave reluctantly Mary the book.  
 d. \* I think that the book John gave Mary reluctantly.

For Japanese, the canonical word order is SIOV. The order of the arguments can be rearranged, however, the verb is always final in both matrix and embedded clauses. Examples (9a-c) are from Hopp (2005, p.37).<sup>47</sup> (9a) gives the canonical order. (9b) shows the object scrambled across the indirect object, and (9c) shows the object scrambled across the subject. Example (9d) is from W&K (2008, p.523), where an indirect object is scrambled across a subject:<sup>48,49</sup>

- (9) a. John-ga Mary-ni sono hon-o watasita. SIOV  
 John-NOM Mary-DAT that book-ACC gave  
 b. John-ga [sono hon-o]<sub>1</sub> Mary-ni t<sub>1</sub> watasita. SOIV  
 John-NOM that book-ACC Mary-DAT gave  
 c. [sono hon-o]<sub>1</sub> John-ga Mary-ni t<sub>1</sub> watasita. OSIV  
 that book-ACC John-NOM Mary-DAT gave  
 d. Ano resutorani-de [<sub>IP</sub> John-ga t<sub>i</sub> piza-o tabe-ta]. ISOV  
 that restaurant-in John-nom pizza-acc eat-past

<sup>47</sup> Hopp uses brackets to indicate the scrambled element.

<sup>48</sup> Hopp adopts a movement analysis of scrambling. W&K do not. I will remain neutral on the syntactic analysis of scrambling for now. However, I retain the indexing and the ‘trace’ notation in the examples in order to make the original position and the moved position obvious.

<sup>49</sup> Nemoto (1995, fn 1) comments that Japanese examples lacking a topic phrase in the matrix clause can sound unnatural and suggests appending *koto* ‘fact’ to the end of example sentences as a remedy. Neither Hopp nor W&K mention this issue nor do any of their examples contain the word *koto*.

Such clause-internal scrambling can be called ‘short scrambling’.<sup>50</sup> Japanese also permits long-distance scrambling where the object scrambles out of an embedded clause to the sentence-initial position of a matrix clause. Example (10a) gives the canonical order and (10b) gives the long-distance scrambled order (data from W&K 2008, p.525):

- (10) a. Mearii-ga [Jon-ga ano resutoran-de  
 Mary-NOM John- NOM that restaurant-in  
 pizza-o tabe-ta to] omot-ta.  
 pizza-ACC eat-PAST think-PAST S [ SIOV ] V

‘Mary thought that John ate pizza in that restaurant.’

- b. Piza<sub>1</sub>-o Mearii-ga [Jon-ga ano resutoran-de t<sub>1</sub>  
 pizza-ACC Mary- NOM John- NOM that restaurant-in  
 tabe-ta to] omot-ta.  
 eat- PAST that think- PAST O S [SIV ] V

Similarly to Japanese, the canonical constituent order in German is considered to be SIOV. But German differs from Japanese in that while embedded clauses are verb-

---

<sup>50</sup> Hopp (2007, p. 98) distinguishes two types of clause-internal scrambling, that which crosses the subject and that which does not. He labels the former ‘medium scrambling’. In this work we will follow W&K 2008 in calling all clause-internal scrambling ‘short’.



‘I think that Martin watched a film about France yesterday.’

Examples (8) – (11) represent very basic facts within a far more complicated total picture. Scrambling can interact with information structure, definiteness, case-marking, and intonation. Additional operations such as *wh*- movement or topicalization can affect the possibilities for scrambling in a given sentence. These considerations are relevant to the experiments in Hopp 2005 and W&K 2008 and will be mentioned in the appropriate subsections below.

### 4.3.1 Hopp 2005

#### RATIONALE

The overarching theoretical goal of Hopp (2005) is to investigate the role of UG in L2A, if any. Hopp adopts the general methodology recommended by Schwartz and Sprouse (2000)<sup>51</sup>, that is, to investigate the L2 acquisition of POS phenomena. Although the particular POS phenomenon under study here has to do with scrambling, Hopp does not argue that the “intact” scrambling of simple, complete phrases exemplified in the data above constitutes a POS phenomenon in German. As Schwartz and Sprouse (2013, p.156) put it, “There is no doubt that tokens of intact topicalization and scrambling of non-complex XPs are aplenty in the language surrounding learners.” Instead, Hopp focuses on *constraints* on scrambling, which are more complex and rarer in the input. Specifically, Hopp considers what is known as remnant movement, including remnant scrambling and remnant topicalization. In the interest of comparability to the W&K

---

<sup>51</sup> ...and reinforced more recently in Schwartz & Sprouse (2013)

(2008) study of scrambling, this summary of Hopp’s experiment will concentrate on his investigation of remnant scrambling, examples of which are presented below.<sup>52</sup>

The concept of remnant movement was introduced briefly in Chapter 2 above. Remnant movement is a two-part operation in which an element is first moved out of a complex phrase, leaving a ‘remnant’, and then the remnant phrase moves across the previously moved element. Example (12a) shows a canonically ordered sentence, and (12b) shows the final product of ‘remnant topicalization (over scrambling)’<sup>53</sup>.

- (12) a. Peter hat schon [den Wagen zu reparieren] versucht.  
 Peter has already the car to repair tried  
 “Peter has already tried to repair the car.”
- b. [t<sub>1</sub> zu reparieren]<sub>2</sub> hat Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht.  
 to repair has Peter the car already tried

In (12b), ‘t<sub>2</sub>’ marks the original location of the complex infinitival phrase *den Wagen zu reparieren*. The first movement (subscripted ‘1’) is of the object DP (*den Wagen*), scrambling over (to the left of) the adverb *schon*. Then the remnant phrase [t<sub>1</sub> zu reparieren] topicalizes over (to the left of) the scrambled DP, all the way to the sentence-initial position.

---

<sup>52</sup> Hopp’s complete test paradigm includes: for an infinitival phrase, (a) scrambling; (b) topicalization; (c) remnant topicalization over scrambling; (d) long remnant topicalization over scrambling; \*(e1) remnant scrambling over scrambling (over an Adverb); \*(e2) remnant scrambling over scrambling (over the Subject); \*(f) remnant topicalization over topicalization, and for a DP, all of the above excluding (e2). This discussion will concentrate on (a), (c), and (e1). Further details will be given in the discussion of the test materials below.

<sup>53</sup> The examples in (12) are drawn from Hopp (2005) and from Schwartz & Sprouse (2013)

Remnant movement is only licit if the two instances of movement are of two different types (Müller 1996, 1998).<sup>54</sup> Thus in contrast to (12b), the two instances of scrambling in example (13) result in an ungrammatical sentence:

- (13) \*Ich glaube, dass [t<sub>1</sub> zu reparieren]<sub>2</sub> Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht hat.  
I think that to repair Peter the car already tried has  
'I think that Peter has already tried to repair the car.'

Hopp argues that this constraint on scrambling poses a strong POS challenge, both to child native speaker learners and adult L2 learners of German. He offers several arguments. First, Hopp (2005, p.40) points out that optional phenomena in general are challenging, because learners must figure out the range of variation, whether or not there are any restrictions on the variation, and if so, what those restrictions are. With regard to acquiring restrictions on scrambling, the situation is particularly difficult because (p.42) "there is no discourse context that requires a scrambled or topicalized sentence to be used instead of the base order." The point is, if a certain non-canonical constituent order does not appear in the input, a learner cannot infer that that order is ungrammatical rather than simply dispreferred.

On that note, Hopp presents a second POS argument from frequency data. He cites (2005, p.42) corpus studies<sup>55</sup> of spoken as well as written German that show that scrambling of complex NPs and remnant movement are "highly infrequent". For instance, Bornkessel et al. (p.B23-B24) report: "According to the 'W-Pub' corpus archive

---

<sup>54</sup> Müller (1996, 1998) proposes a general constraint called Unambiguous Domination, pertaining to combinations of several types of movement, including scrambling, topicalization, and *wh*-movement. As noted in fn 6 above, Hopp (2005) tests learners on a variety of remnant movement combinations, both grammatical and ungrammatical.

<sup>55</sup> Hoberg (1981), Schlesewsky et al. (2000) and Bornkessel et al. (2002).

(Mannheimer Institut für deutsche Sprache), in transitive sentences, the combination of *dass* ('that') + *der* ('the<sub>NOMINATIVE</sub>') occurs approximately eight times more often than the combination of *dass* ('that') + *den* ('the<sub>ACCUSATIVE</sub>') and the combination of *dass* ('that') + *dem* ('the<sub>DATIVE</sub>'), the latter two occurring with approximately the same frequency."<sup>56</sup>

Hopp argues that the complete range of movement options may not be adequately represented in the input for the learner to identify the entire paradigm of grammatical and ungrammatical orders.

Third, Hopp points out that even if the full set of options is available, the input could be misleading if the learner focuses on the relative positions of specific elements, a possibility that Hopp argues is quite plausible. For instance, the same order of elements obtains in the grammatical (12b) above versus the ungrammatical (13): [*zu reparieren*] precedes *Peter* which precedes [*den Wagen*]. He therefore argues that surface order is not sufficient to distinguish grammaticality from ungrammaticality. Fourth, Hopp consulted textbooks and instructors to verify that remnant scrambling is not taught or directly corrected in German language classes.

Having argued for the insufficiency of input to learners, Hopp completes his POS arguments by pointing out that for English L1 learners, scrambling is not instantiated in the L1 grammar and therefore neither the L1 nor the L2 input could be responsible for successful acquisition. In contrast to English, Japanese does instantiate scrambling (as demonstrated above), as well as remnant scrambling.<sup>57</sup> Japanese L1 learners of German

---

<sup>56</sup> Note that this statement does not specifically refer to complex XPs or remnant movement, nor does relative frequency data on its own provide information on the basic frequency of the noncanonical orders. See Chapter 5 below for further comments.

<sup>57</sup> For arguments and evidence see Tsujioka (2001).

would therefore have their L1 as a possible resource for the acquisition of scrambling and constraint on scrambling in L2 German.

To summarize, Hopp argues that remnant scrambling presents a severe POS challenge for English L1 learners of L2 German. In order to investigate the broad question of the role of UG in L2A, Hopp's experiment asks the following specific research questions and makes the associated predictions given below:

- Do L1 English and L1 Japanese learners of L2 German *accept* basic scrambling of complete phrases?
- Do L1 English and L1 Japanese learners of L2 German *accept* remnant topicalization across a scrambled phrase?
- Do L1 English and L1 Japanese learners of L2 German *reject* remnant scrambling across an already scrambled phrase?

If the answer to each of the above questions is yes, then the learners are behaving in a target-like manner, recognizing what is possible and what is impossible with respect to scrambling in German. For English L1 subjects, under the assumption of Full Access to UG, learners at sufficiently high proficiency are expected to demonstrate target-like behavior. If on the contrary L2 learners do not have access to UG, then English L1 learners are predicted to have difficulty with these constructions. Because Japanese L1 learners of German have scrambling and similar scrambling constraints instantiated in their L1, they are predicted to be able to perform in a target-like manner.

#### **EXPERIMENTAL METHOD**

Hopp's subject control group consisted of 26 native speakers of German who were residents of Germany at the time of the test. Beyond that residency information, we do not have any demographic data or language history on the native speakers. All of the 13

Japanese L1 subjects as well as 10 of the 26 English L1 subjects were also residents in Germany at the time of the test. The remaining 16 English L1 subjects were final-year students of German at universities in Britain. That being said, every nonnative learner had resided in Germany at some time. Lengths of residence for native English subjects varied from .5 years to 33 years, and for Japanese subjects ranged from 1 to 25 years.

The learner groups were partitioned into proficiency levels based on their performance on a 40-item cloze test. English L1 learners were rated high intermediate (n= 7, mean score 19.3), advanced (n= 13, mean score 28.7), or very advanced (n= 6, mean score 39.2). Japanese L1 learners were rated high intermediate (n= 8, mean score 18.8) or advanced (n= 5, mean score 30.4). Since no Japanese learners scored higher than 34 on the cloze test, there was no ‘very advanced’ group of Japanese L1 learners. The mean score for the native speaker controls was 37.2.

Table 4.5 summarizes the characteristics of the learner groups by proficiency. In general, subjects with longer Length of Exposure (LoE), longer Length of Residence (LoR), and lower Age of Onset (AoO) scored higher on the proficiency exam.<sup>58</sup> Note that the AoO for all of the English groups is considerably lower than either of the Japanese groups. The English High-Intermediate (Hi-Int) group has a longer average LoE but a shorter average LoR than the corresponding Japanese Hi-Int group. The English Advanced group has shorter LoE and LoR than the Japanese advanced group.

---

<sup>58</sup> Hopp does not define exposure, however, exposure seems to include instruction, perhaps in the home country or elsewhere, whereas residence is specifically residence in Germany.

**Table 4.5: Characteristics of Learner Groups by Proficiency**

	<b>English Hi-Int</b>	<b>English Adv</b>	<b>English Very Adv</b>	<b>Japanese Hi-Int</b>	<b>Japanese Adv</b>
<b>Length of Exposure</b>	12.5	16.7	29.2	6.4	22
<b>Length of Residence</b>	3.3	6.2	15.2	4.8	17.2
<b>Age of Onset</b>	15.6	12.5	13.3	24.6	23.0
<b>Age</b>	29.7	27.8	15.7 <sup>59</sup>	31.9	45.4

Note: Arithmetic mean in years. Data from Hopp (2005, p.51).

The experimental task consisted of a grammaticality judgment test (GJ). The complete suite of test item types included basic intact scrambling and topicalization as well as a variety of remnant movement combinations. The test items of highest interest to this discussion involve grammatical intact scrambling and grammatical and ungrammatical remnant movement involving scrambling. Samples of these constructions have been presented above, and are repeated here for convenience:

(11d) Intact scrambling

Ich glaube, dass [den Wagen zu reparieren]<sub>1</sub> Peter

I think that the car to repair Peter

schon t<sub>1</sub> versucht hat.

already tried has

‘I think that Peter already tried to repair the car.’

---

<sup>59</sup> 15.7 is the number reported in Table 1 on page 51 of Hopp (2005). However, this number must be incorrect because the range of ages given for the English very advanced group is 27–60 years of age.

(12b) Remnant topicalization across scrambled phrase

[t<sub>1</sub> zu reparieren]<sub>2</sub> hat Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht.

to repair has Peter the car already tried

(13) Remnant scrambling across scrambled phrase

\*Ich glaube, dass [t<sub>1</sub> zu reparieren]<sub>2</sub> Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht hat.

I think that to repair Peter the car already tried has

‘I think that Peter has already tried to repair the car.’

These examples represent part of the ‘infinitival paradigm’. In addition, Hopp tested a parallel ‘DP paradigm’ (represented in (11e) above), where the scrambled phrase is a DP such as *einen Film über Frankreich* (a film about France). There were 74 items total on the GJ test. Thirty-nine items were target items and 35 items were fillers. Altogether, 40 of the items were grammatical and 34 ungrammatical. Pilot studies and previous research (Schreiber and Sprouse 1998, among others) showed that the written presentation of isolated sentences can yield unstable judgments and low acceptance rates. Accordingly, the GJ test presented each test item bimodally and embedded in contexts. After reading/hearing the context, subjects were asked to respond to the question “Is this sentence possible in this context?”, and to rate each test sentence on the following scale: ‘-2’ = ‘not possible’; ‘-1’, ‘+1’, ‘+2’ = ‘possible’; ‘x’ = ‘don’t know’. The test was timed automatically. Subjects read/heard the test sentence twice with a 3-second pause in between, and were then given 11 seconds to make a judgment before the CD moved on to

the next track. Without being specific, Hopp (2005, p.49) says that “Extensive instructions and practice items preceded the task.”

## RESULTS

The difference between acceptance rates for grammatical versus ungrammatical scrambling sentences is statistically significant for all learner groups. Across all proficiencies and both L1s, all learner groups reliably accept grammatical sentences and reject ungrammatical ones. There was no statistically significant difference between any individual learner group and the native controls on any sentence type in the infinitival paradigm. Specific results for infinitival paradigm are given in Table 4.6 below. Results for the DP paradigm, not shown here, are consistent with the infinitival paradigm results.

**Table 4.6: Percentage Acceptance of Sentence Type in Infinitival Paradigm, All Groups**

	<b>English Hi-Int (n=7)</b>	<b>English Adv (n=13)</b>	<b>English Very Adv (n=6)</b>	<b>Japanese Hi-Int (n=8)</b>	<b>Japanese Adv (n=5)</b>	<b>Native Controls (n=26)</b>
<b>Scrambling (complete phrase)</b>	61.9	64.1	66.7	91.7	73.3	80.5
<b>Remnant top over scr across finite clause boundary</b>	57.1	38.5	27.3	43.8	50	59.6
<b>*Remnant scr over scr (over Adverb)</b>	4.8	10.3	11.8	0	6.7	12.8
<b>*Remnant scr over scr (over Subject)</b>	23.8	7.7	5.6	16.7	26.7	24.4

Note: \*indicates sentence was predicted to be ungrammatical. Data from Hopp (2005, p.53-54).

With respect to group differences between the two different L1s, there were no qualitative differences. Learners from both L1s distinguished between licit and illicit scrambling. There was however a quantitative difference between the English L1 Hi-Int group and the Japanese L1 Hi-Int. group: the former accepted grammatical scrambling at a lower absolute rate than did the latter. The quantitative difference between the Advanced English L1 group and the Advanced Japanese L1 group was smaller than that between the less proficient groups.

Hopp also analyzed the results at the individual level. Every subject in both L1 groups demonstrated a grammaticality contrast between licit scrambling of complete phrases and illicit remnant scrambling. This pattern holds over the total test items but also holds separately for the infinitival paradigm and the DP paradigm. Similarly, all subjects reliably distinguish between licit and illicit remnant scrambling.

## CONCLUSIONS

Hopp (2005) concludes that English L1 learners of L2 German are going beyond the input and accessing fully available domain-specific UG in the acquisition process.<sup>60</sup> He argues that L1 and L2 knowledge states are similar and argues against a critical period for L2 language acquisition. He cites the learners' consistent ability to distinguish grammatical from ungrammatical instances of various types of scrambling as evidence for his position. Learners perform consistently across L1s, across proficiencies, and even across individuals. Divergences are quantitative rather than qualitative.

---

<sup>60</sup> For more specifics on how Hopp characterizes UG with respect to scrambling, see Hopp (2005, p. 37-38). Briefly, Hopp adopts a movement analysis of scrambling, with movement triggered by a strong uninterpretable feature [scr] (Oka 1996, Sauerland 1999). English L1 learners must activate this feature, not available in their L1, in order to license scrambling. On the other hand, the constraints on remnant movement (Müller 1996, 1998) are proposed to be *universal*, therefore available in English, where they apply to topicalization and *wh*-movement.

Hopp argues that these results are unexpected on a L1-transfer and/or L2-input model of acquisition. If learners were relying on their L1, one would expect to see a difference between the performance of English L1 learners and Japanese L1 learners, based on the fact that they have different initial states and differential access to existing scrambling capabilities. If learners were relying on L2 input, one would expect to see individual differences across subjects based on differences in experience. One would also expect to see differences across proficiencies, with learners less successful at lower proficiency. Both of these effects would be expected to be relatively strong given the relative infrequency of examples of scrambling in the input.

We now turn to a study of scrambling that considers the emergentist approach.

#### **4.3.2 Williams and Kuribara (2008)**

W&K (2008) conduct a set of two experiments investigating early-stage acquisition of scrambling: one in a generative approach, and one in an emergentist-type approach. In the first experiment, described in Section 4.3.2.1 below, they investigate the possible role of UG in adult L2 acquisition in a manner that is similar in some ways to Hopp's (2005) study. In the second experiment they investigate the possible role of statistical input in L2A. In this second study, summarized in Section 4.3.2.2 below, W&K build a computer simulation model which they train and test on the same data that was given to their human subjects in the generative study. For both studies, the target language is a semi-artificial language which W&K call 'Japlish'<sup>61</sup>. Japlish uses English lexical items and Japanese syntax, specifically, Japanese word order and case marking. W&K draw conclusions on the nature of L2A based on comparisons between their adult L2 learners

---

<sup>61</sup> Section 3.4 above gives a brief discussion of artificial language paradigms.

and child L1 learners studied elsewhere, and between their adult L2 learners and their computer simulation performance. As with all the experiments presented here in Chapter 4, comments will be reserved until Chapter 5 below.

By using Japlish, W&K can examine the acquisition of syntax at its earliest stage, essentially at first exposure. The following examples illustrate Japlish. The case marker *-ga* indicates the Subject, *-o* indicates the Object, and *-ni* indicates the Indirect object, if present. Examples (14a-b) show simple and complex canonical word order, respectively, based on Japanese syntax. Examples (15a-b) illustrate some instances of scrambling in Japlish, again, based on what is allowed in Japanese. These examples are taken from Table 2, p.530 of W&K. The correspondence to Japanese natural language can be seen by comparing these examples to the Japanese data in (9) and (10), given at the beginning of Section 4.2 above. The ‘t’ indicates the canonical position of the scrambled element. The index number on an element indicates which element is scrambled.

- (14) a. Simple canonical: S I O V  
Student-ga dog-ni what-o offered?
- b. Complex canonical: S (Adv) [S O V] V  
John-ga angrily Mary-ga that ring-o lost that said
- (15) a. Short scrambling: O<sub>1</sub> S (I) t<sub>1</sub> V  
That sandwich-o John-ga ate.
- b. Long scrambling: O<sub>1</sub> S [S I t<sub>1</sub> V] V  
What-o Mary-ga professor-ga students-ni taught that said?

For the most part, I will follow W&K in discussing the Japlish data in terms of its word order (SIOV, etc) rather than with the use of full lexical items. With this background in place, we turn to the experiments.

#### ***4.3.2.1 Experiment 1: A Role for UG?***

##### **RATIONALE**

As mentioned, in the first experiment W&K investigate the possible role of UG in adult L2A. Adopting the position that UG is active in child first language acquisition, W&K then ask if L2A is similar to L1A (Schwartz and Sprouse 1996, among others) or fundamentally different from it (Bley-Vroman 1990, among others). They characterize the L1 acquisition of scrambling as a POS phenomenon in Japanese: despite being infrequent in the input, scrambling is acquired early. To substantiate the poverty of input claim, they cite Iwasaki (2003, p.297) as reporting corpus study evidence that OSV is a very infrequent word order in Japanese. To substantiate the early acquisition claim, they cite Murasagi and Kawamura (2005), who find that children as young as two years of age perform well on interpretation tasks involving both SOV and scrambled OSV sentences. Murasagi and Kawamura propose that canonical order and scrambling may be acquired at the same time. If adult L2A is similar to L1A, then with sufficient and clear input, the developmental path of adults should mimic that of children. In particular, adults should acquire scrambling when they acquire canonical word order. W&K propose that the simultaneous acquisition of canonical and scrambled order is due to a ‘parameter clustering’ effect, in which successfully setting the head-direction parameter to head-final enables the associated phenomenon of movement to the left.

The particular research questions and predictions are as follows. After completing a learning phase,

- Do L1 English learners of Japlish *accept* canonical (head-final) SOV word order?
- Do L1 English learners of Japlish *accept* leftward scrambling, including in constructions that they have never seen before?
- Do L1 English learners of Japlish *reject* SVO (English canonical) word order?
- Do L1 English learners of Japlish *reject* scrambling to the right?

If the answer to all of the above is yes, then learners are displaying target-like behavior. If UG is active in L2A, then it is predicted that learners will accept what is grammatical and reject what is ungrammatical in the target language. Learners will demonstrate a grammaticality contrast between target-like head-final SOV order and the L1-like head-initial SVO order. Learners will accept scrambling of objects both clause-internally and across a finite clause boundary, as long as the scrambling is to the left. On the basis that movement to the right is permissible in head-initial languages like (the L1) English, W&K argue that successful acquisition of L2-like scrambling (as opposed to L1 transfer of something like heavy-NP shift) requires that learners be able to reject rightward movement.<sup>62</sup> If learners are accurate on constructions that are new in their experience, this will support the claim that they are relying on an internalized rule of some kind rather than on memory.

#### **EXPERIMENTAL METHOD**

The human subjects in the first experiment were 41 undergraduates and postgraduates at the University of Cambridge. All were native speakers of English “who had no knowledge of Japanese” (W&K, p.530). They were divided into two groups: an

---

<sup>62</sup> W&K also test two additional constructions: multiple applications of scrambling, and superiority effects in *wh*- movement. The research questions and results for these constructions are beyond the scope of this Section.

‘exposure’ group that participated in a learning phase prior to the testing phase, and a ‘no exposure’ group that did not. The exposure group comprised 25 individuals with a mean age of 25; the no-exposure group comprised 16 individuals with a mean age of 21.

The learning phase (‘exposure phase’) of the experiment consisted of a semantic plausibility task. The task was designed to focus the learners on the meaning rather than on the syntax of the stimuli, so that any learning of the syntax could be considered incidental. The stimuli consisted of a total of 194 isolated sentences, presented both aurally and in written form on a screen. The plausibility judgment was entered by pressing one of two keys. All sentences were grammatical, so, every sentence and every embedded clause was verb-final. Half of the sentences were semantically plausible, half were semantically implausible. In some cases the implausibility could be gleaned from the lexical semantics (*Simon-ga which bowl-o ate?*), and in some cases the learner needed to rely on case marking to detect the implausibility (*Applicant-ga company-ni which job-o offered?*). One third of the sentences contained scrambled word order. The scrambled element was always a Direct Object scrambling to sentence-initial position. Exposure to the Japlish sentences was controlled so that learners experienced input of increasing complexity as the task progressed. Sentences were presented in blocks. The sentences in the first block were all simple canonical; the second block contained all complex canonical examples; the third block contained a mixture of both, and the fourth and final block contained all those plus scrambled sentences.

Learners in the exposure group were instructed in the function of the case markers (-ga for Subject, -o for Object, -ni for Indirect Object) at the beginning of Block 1. At the beginning of Block 2 they were given an example of a complex canonical sentence and

its interpretation (W&K, p.533): they were told that in the sentence *Fred-ga John-ga apple-o ate that said*, Fred did the saying and John did the eating. At the beginning of Block 4, they were told to expect to see different word orders in the example sentences, and the instructions suggested that learners use case markings to work out the meaning.

The test phase of the experiment consisted of a GJ test. Learners in both groups were asked to indicate if a test sentence was ‘likely to be grammatical’ or ‘unlikely to be grammatical’, again by pressing keys. Sentences were presented visually and auditorily as in the exposure phase, but the GJ procedure had the following differences: first, the sentences were preceded by an on-screen diagram of stick figures, words and arrows that conveyed the meaning of the sentence to be presented next. The diagram could be viewed as long as desired. Second, with the intention of pressuring the learner to make a quick decision, the visual presentation of the sentence disappeared from the screen when the auditory presentation was complete. Third, the instructions to both groups called out the word order patterns as the phenomenon of interest. For instance, instructions to the no exposure group said (W&K, p.534): “we are interested in your intuitions about which word order patterns seem to you to be more likely to be grammatical in a language that you do not actually know.”

Some of the GJ test sentences were grammatical, and some ungrammatical. All were semantically plausible. Several of the test constructions in the GJ test had not been presented in the exposure phase. Table 4.7 below details the various constructions included in the exposure phase task and the test phase task, showing which test phase constructions the learners who were in the exposure group had seen and which they hadn’t. Constructions that were new to learners in the GJ test included: movement of the

direct object to the initial position of an embedded clause; movement of an Indirect object; and all ungrammatical sentences. (Of course, for the ‘no exposure’ group, all GJ test items were new.)

**Table 4.7: Construction Types Included in Experimental Tasks**

<b>Construction Type</b>	<b>Exposure Task</b>	<b>GJ Test (trained items)</b>	<b>GJ Test (new items)</b>
<i>Grammatical</i>			
Simple canonical	SV, SOV, SIOV		
Complex canonical	S [S (I) O V] V	S [S (I) O V] V	
Scrambled short IP short VP long IP complex short	O <sub>1</sub> S (I) t <sub>1</sub> V S O <sub>1</sub> I t <sub>1</sub> V O <sub>1</sub> S [S (I) t <sub>1</sub> V] V	O <sub>1</sub> S (I) t <sub>1</sub> V O <sub>1</sub> S [S (I) t <sub>1</sub> V] V	I <sub>1</sub> S t <sub>1</sub> O V I <sub>1</sub> S [S t <sub>1</sub> O V] V S [O <sub>1</sub> S (I) t <sub>1</sub> V] V
<i>Ungrammatical</i>			
Canonical English			*S V (I) O
S to the right			*t <sub>1</sub> I O V S *S [t <sub>1</sub> O V S <sub>1</sub> ] V
O to the right			*S I t <sub>1</sub> V O <sub>1</sub> *S [S t <sub>1</sub> V O <sub>1</sub> ] V

## RESULTS

The no exposure group accepted all examples of short (within-clause) scrambling at above chance, including in complex sentences. However, their acceptance of long scrambling was much lower: 35% for object-initial scrambling and 39% for indirect-object scrambling to sentence-initial position. The no exposure group accepted both canonical Japlish and canonical English word orders at well above chance, 71% and

85%, respectively. These results provide a baseline for comparison to the results from the exposure group.

The exposure group performed well (88% accuracy) on the semantic plausibility task, with high performance on both canonical and scrambled sentences. These results indicate that the subjects had good comprehension of the Japlish sentences. On the GJ test, there was high variability across participants, with some performing fairly poorly, only at chance on almost all non-canonical constructions. W&K therefore restrict their attention to a higher-performing subgroup of the exposure group: participants who accepted short scrambled sentences at or above 75% (note that the group had trained on this construction). Under this criterion there are 14 individuals in the exposure subgroup. For the sake of comparison W&K apply the same criterion to select a no exposure subgroup, resulting in a group of 11 individuals. Results from this point on are reported for these subgroups.

W&K compare the exposure subgroup to the no exposure subgroup in order to ascertain if there is an effect of input. If there is an effect of input, one would expect the exposure group to appear more target-like than the no exposure subgroup. The results showed that the exposure subgroup's acceptance rates were significantly higher than those of the no exposure subgroup on all of the grammatical test sentences with the exception of the new simple short sentences (ISOV). This is perhaps not surprising given that the assignment of individuals to subgroups was made on the basis of having similar performance on the trained simple short sentences. Overall, the exposure subgroup outperformed the no exposure group on grammatical sentences.

With regard to the ungrammatical sentences, the exposure subgroup accepted English canonical order (SV(IO)) at significantly lower rates than did the no exposure subgroup. However, the exposure subgroup did not appear target-like in that they did not clearly reject the English canonical order, rather, they accepted it at chance. On the rightward movement sentences, the exposure subgroup was significantly different (in a target-like direction) from the no exposure subgroup on simple sentences but not on complex sentences.

### **CONCLUSIONS**

W&K conclude that UG is not active in L2A. They argue first that the no exposure group showed evidence of L1 transfer but not evidence of L2 acquisition. Because this group does not appear to have settled on head-final order, W&K suggest that the participants have not yet acquired canonical Japlish order and may be demonstrating some L1 transfer of operator-type movement rather than scrambling to the extent that they appear target-like on some of the scrambled sentences.

As for the exposure (sub)group, although they did accept scrambling at above chance, they did not clearly reject English word order nor did they master constraints on scrambling in the form of rightward movement. Again, the ‘cluster’ of phenomena was not acquired together. Since this developmental path is different from that seen in child L1 acquisition, W&K conclude that L2A is not like L1A. From the differences between the exposure and no exposure subgroups, W&K conclude that there was an effect of input.

#### ***4.3.2.2 Experiment 2: A Role for Statistics?***

##### **RATIONALE**

Having concluded that the evidence from the first experiment does not support UG-guided second language acquisition, W&K consider whether an emergentist account of the learner data is more successful. They adopt the position that connectionist models compute statistics in a psychologically plausible way. On that assumption, they propose to compare their human subject performance to the output of a connectionist model.

W&K's specific research question is:

- Do the acceptance rates produced by human learners correlate with the output of a connectionist model that is trained and tested on the same data?

If there is a positive correlation, this would support the view that human learners rely on guidance from statistical calculations over the accumulated history of their linguistic input experience.

##### **EXPERIMENTAL METHOD**

Section 3.5 above gave a very brief description of connectionist models. Connectionist models posit a set of units with defined activation levels. The units exist in layers: an input layer, an output layer, and a 'hidden' layer that enables the model to transform the input representations as needed. W&K's model uses a particular architecture known as a simple recurrent network (SRN) (Elman 1990; Elman et al. 1996). This architecture provides an additional layer of 'context' units that store the activation levels of the hidden units for an additional iteration of the model. These activation levels are then fed back into the hidden units, in a way 'recycling' the information from previous steps in the model. Effectively, the context layer makes the model sensitive to the whole history of an input sentence. The connections between the units are defined by a set of weights. During

a simulation, the assigned weights can change in response to the model input and according to some set of learning rules.

W&K chose to set the initial connection weights to random values.<sup>63</sup> They then submitted their model to a ‘learning’ phase. For the model, learning consists of adjusting the weights on the connections in the network. What the model learns, effectively, is how to predict the next element in a sequence. As W&K (2008, p.542) describe it: “For example, for the sentence “John loves Mary”, “John” is presented as a pattern of activation over the input layer. Activation spreads through a layer of hidden units to a layer of output units, and the output is compared to the correct prediction, in this case “loves”....”Loves” is then presented, and the network is taught to predict “Mary”.” The input to the model will determine the predictions that the model learns to make. For frequent sequences, the model will eventually learn to produce strong activation patterns. For less frequent sequences, the model will learn to produce weaker activation patterns. For non-occurring sequences, since the connection weights were originally set to random values rather than zero, it is possible that the model will produce a non-zero activation pattern even after extensive training.

For this particular experiment, the input sentences that the human subjects were exposed to in the semantic plausibility task described above were coded in terms of Subject, Object, Indirect Object, and Verb. W&K note that they could have equivalently coded for thematic roles or case markers. The coded representation of every sentence was then given to the model as input, although in random order rather than in blocks as for the

---

<sup>63</sup> This setting is more compatible with the initial state of L1 rather than the initial state of L2. An L2 learner would be expected to have the L1 weights instantiated (Ellis 2006). See further discussion in Chapter 5 below.

human subjects. Note that the frequency of each construction type was preserved. The model trained on a total of 50 cycles of the entire set of sentences.

Having completed this training phase, the ‘learning’ mechanism was turned off and the model was placed in ‘testing phase’. For the testing phase, for each (coded) test sentence presented, the model simply reported the activation pattern of its output units after each segment of the test sentence. The model did not make any additional adjustments to the connection weights based on the test sentence input.

As a follow-up to this simulation, W&K explored the question of later-stage learner development by exposing the model to 5000 cycles of training rather than 50. The results for both of these simulations will be discussed below.

## **RESULTS**

The goal of this study was to ascertain whether a statistical learning procedure gives a good account of human L2A by comparing model performance to human performance. W&K computed a measure of the SRN model’s average output strength for each construction type. Intuitively, the average output strength represents how well the model did at predicting the given sequence of elements. For each construction, they compared this measure to the acceptance rates on that same construction that were given by the beginner human learners in Experiment 1. Intuitively, the learners’ acceptance rates represent how likely it was that the learners thought a given construction would be grammatical.

The specific comparison W&K make is between the model’s output strength and the acceptance rates as given by the exposure subgroup of learners. These are the learners who had been exposed to the same set of test sentences that the model was trained on, and who were successful on grammaticality judgments on short scrambling. Since

performance on short scrambling (OSIV) was the diagnostic for selection to the exposure subgroup, that particular test structure was left out of the comparison. Putting aside long scrambling structures, W&K found a linear relationship between the model performance and the human subject performance on all other test constructions. The model produced lower output strengths on ungrammatical constructions that learners accepted at lower rates. And the model produced higher output strengths on complex canonical constructions and other short scrambling structures such as ISOV. The correlation between learner performance and model output was strong (Pearson  $r = 0.995$ ).

Although the model and the people performed similarly, this is not the same as saying that either did well. Recall for instance, that even the exposure subgroup did not reliably reject ungrammatical sentences. All of the test subjects so far (human and computer!) were early-stage learners. When the model was trained extensively in order to simulate later-stage learning, its performance did not improve. For instance, output strength for ungrammatical sentences increased slightly.

#### **CONCLUSIONS**

W&K conclude that the simulation model is a good descriptor of learner performance. On this basis they propose that learners are sensitive to statistical patterns in the linguistic input. However, the fact that the model performance did not improve with additional training may indicate that statistical learning alone is not enough.

#### **OVERALL CONCLUSIONS**

The overall goal of the W&K experiments was to see which characterization of L2A provided a better fit to learner performance in an experimental setting: a UG characterization or an input-based characterization. Learners were tested in two groups: a no exposure (sub)group who represented the initial state of L2 learning, and an exposure

(sub)group who represented early-stage learners. W&K concluded that UG was not active in L2A for either group, based on their inability to correctly reject word orders that are not permissible in their ‘L2’, the semi-artificial language Japlish. It should be noted that the exposure subgroup did correctly accept scrambled sentences after only their brief exposure (approximately 30 minutes, 194 sentences). However, it should also be noted that many of the original exposure group did not appear to have acquired scrambling at all.

W&K concluded that their computer simulation model provided a good fit for learner performance, and thus argue that L2A is fundamentally different from L1A in its dependence on statistical learning. However, on the basis of their model’s lack of improvement after considerably more training, they propose that statistical learning must be supplemented with a symbolic learning mechanism that is not necessarily language-specific.

#### **4.4 Scope Studies**

The third and final section of this chapter reviews studies of scope phenomena.

Informally, the notion of scope concerns relationships between certain types of elements in a sentence. Consider the following example, which contains quantifiers in both the subject NP and the object NP:

(16) Someone read every book. Marsden (2009, p.137)

In English, there is a reading of (16) which fixes the interpretation of *someone* as a single individual and evaluates whether or not that particular person read every relevant book.

This interpretation is known as the ‘subject-wide’ or ‘surface-scope’ interpretation, and the object *every book* is said to have ‘narrow scope’. Also in English (but not necessarily

other languages), there is a reading of (16) which fixes the interpretation of *every book* and evaluates whether or not, for each one of those books, there is some person who read it. Under this reading, example (16) can be true if a different person read each book, as long as all of the books were read. This interpretation is known as the ‘object-wide’ or ‘inverse scope’ interpretation, and the subject *someone* is said to have narrow scope.

Besides quantifiers, other elements that can participate in scope relations include negation, interrogatives and modals. In general, the element whose interpretation is fixed first is said to take wide scope, and the element whose interpretation depends on the fixed element is said to take narrow scope. Consider the following example, where the existing scope relationship is between negation and a universal quantifier:

(17) Mike didn’t eat all the cookies. O’Grady (2013a, p.261)

Giving negation wide scope results in an interpretation that says it is not the case that all the cookies were eaten. That is, some of the cookies could have been eaten. If on the other hand the universal quantifier is given wide scope, then the interpretation fixes a particular set of relevant cookies and says that for each of those cookies, it is not the case that it was eaten. In other words, all of the cookies remain uneaten.

The presence of scope-taking elements in a sentence provides the potential for ambiguity. The sets of possible interpretations that are actually realized for sentences containing those elements vary cross-linguistically. Thus, scope phenomena provide interesting potential poverty of the stimulus problems for L2A. The remainder of this section reviews a generative study and two emergentist studies of the acquisition of different scope phenomena in a variety of language pairs. All of the studies use truth value judgment tasks (TVJ) to assess learner comprehension.

#### 4.4.1 Marsden (2009)

##### RATIONALE

Making use of the same general methodology as the Kanno and Hopp studies presented in previous sections, Marsden (2009) investigates the possible role of UG in L2A by looking at L2 acquisition of what she argues is a POS phenomenon. The phenomenon in question involves scopal relations between quantifier phrases (QPs) in subject and object positions.<sup>64</sup> The next few paragraphs summarize the relevant facts that Marsden (2009, p.137-139) sets out, on which her experiment is based. Marsden also presents Korean data showing that Korean patterns with Japanese. Only English and Japanese data are presented here. Following the data, Marsden's hypotheses and research questions are stated.

Sentences containing multiple QPs (hereafter QP-QP sentences) are potentially ambiguous because the QPs have the potential to be in different scopal relationships. The availability of different interpretations for such sentences varies both within a language and cross-linguistically. Within a given language, available interpretations can vary with the particular choice of quantifier, or with differences in word order. The former will be illustrated with English data and the latter with Japanese data. Compare the following two English sentences that use different quantifiers in object position (example (16) is repeated from above):

(16) Someone read every book.

Interpretations:

a. There is some person  $x$  such that  $x$  read every book. ( $S > O$ )

---

<sup>64</sup> I follow Marsden in using the term 'quantifier phrase', thereby remaining neutral on the question of Noun Phrases (NPs) versus Determiner Phrases (DPs).

- b. For each book *y*, some person read *y*. (O > S)
- (18) Someone read all the books.
- a. There is some person *x* such that *x* read every book. (S > O)
- b. ??/\*For each book *y*, some person read *y*. \*(O > S)

Sentence (16) has two available interpretations, as shown: either a single person read the entire relevant set of books, or, possibly distinct people read each distinct book. The latter interpretation is less available, if at all, for example (18). It is more difficult in English to construe *all the books* as individual elements that could be read by different people than it is to construe *every book* in that way.

For an example of within-language variation in available interpretations based on word order, consider Japanese. Recall from Section 4.2 that Japanese permits scrambling. Example (19) contains canonical SOV word order, and example (20) contains scrambled OSV word order:

(19) Dareka-ga dono hon-mo yonda.

someone-NOM every<sup>65</sup> book read

‘Someone read every book.’

Interpretation:

a. There is some person *x* such that *x* read every book. (S > O)

b. \*For each book *y*, some person read *y*. \*(O > S)

(20) Dono hon-mo dareka-ga yonda.

every book someone read

‘Someone read every book. (scrambled)’

---

<sup>65</sup> Marsden (2009, p.138-139) presents arguments for glossing *dono N-mo* as ‘every N’ rather than ‘all N’.

Interpretation:

- a. There is some person  $x$  such that  $x$  read every book. ( $S > O$ )
- b. For each book  $y$ , some person read  $y$ . ( $O > S$ )

The canonical order QP-QP sentence does not permit the same range of interpretations as the scrambled order QP-QP sentence does.

Note that in addition to illustrating within-language variation, these examples in English and Japanese illustrate a cross-linguistic variation in available interpretations based on scope. On the analysis of *dono N-mo* as corresponding to English *every N*, example (19) in Japanese is the canonical equivalent to example (16) in English. However, the Japanese sentence lacks the object-wide scope reading that is available for the English sentence, as listed next to the examples repeated below for convenience:

- (16) Someone read every book.  $S > O; O > S$
- (19) Dareka-ga dono hon-mo yonda.  $S > O; *O > S$

Marsden argues that the difference between the QP-QP sentences (16) and (19) constitutes a POS problem for English L1 learners of L2 Japanese, who must acquire a constraint in the L2 that is not present in the L1. First, the L1 cannot provide the knowledge that the L2 lacks the  $O > S$  interpretation, because the L1 *permits* the object-wide scope reading. Second, she argues that the L2 input is insufficient for acquisition on the basis that the simple non-occurrence of (19) in object-wide scope contexts is not sufficient for a learner to induce impossibility. Finally, based on surveys of textbooks and teachers, she argues that acquisition of knowledge of the lack of a wide-scope interpretation for (19) does not come from instruction. On the other hand, since Korean patterns with Japanese in disallowing an object-wide scope reading for the Korean

equivalent to sentence (19), Korean L1 learners of L2 Japanese would not face a POS problem.

Given these facts and POS arguments, Marsden's broad research questions can be formulated in terms of developmental path and ultimate attainment with respect to the acquisition of the interpretation of canonical word order QP-QP sentences of the type exemplified in (19) above:

- Do Korean L1 and English L1 learners of L2 Japanese follow the same developmental path, or does their performance diverge, particularly in earlier stages?
- Do English L1 learners of L2 Japanese ultimately overcome the POS problem to successfully acquire all and only the correct interpretation(s)?

Marsden adopts the Full Transfer/Full Access hypothesis (Schwartz and Sprouse 1996), which sets up the following specific research questions and the predictions set out below:

- Do English L1 learners of L2 Japanese disallow object-wide scope readings for canonical (SOV) order QP-QP sentences with object *dono N-mo* 'every N'?
- Do Korean L1 learners of L2 Japanese disallow the same?

Under Full Transfer/Full Access, English L1 learners at earlier stages are predicted to behave in an L1-like manner due to L1 transfer, *allowing* the object-wide scope reading. English L1 learners of advanced proficiency are predicted to behave in a target-like manner, accessing UG and overcoming the POS problem, *disallowing* that interpretation. In contrast, Korean L1 learners of L2 Japanese are predicted to disallow the object-wide scope reading due to L1 transfer, showing target-like behavior at all proficiency levels. In

summary, the developmental paths of English and Korean learners will diverge, but the ultimate attainment of both groups will be consistent and target-like.

#### EXPERIMENTAL METHOD

Marsden's experiment tests English L1 and Korean L1 learners of L2 Japanese, in addition to control groups of native speakers for all three languages. L2 learners in both L1 language groups were university students. Participants were assigned to 'intermediate' and 'advanced' subgroups in each L1 based on their performance on a cloze test. Native controls were also university students, with the Japanese native speakers residing in Japan, the Korean native speakers in Korea, and the English native speakers in the UK.

The task was a Truth Value Judgment test (TVJ). Participants were presented with a context in the form of a picture that determined either a subject-wide scope or an object-wide scope for an associated sentence. After viewing the picture, the corresponding test sentence was presented aurally and visually. Participants were asked 'Does the picture match the sentence?', and judgments were solicited on a scale of -2 (no, definitely not); -1 (not exactly); +1 (yes, kind of) or +2 (yes, perfectly). Participants had the option to indicate 'x' for 'can't decide'.

All speaker groups were tested on 10 examples of each of three different types of QP-QP constructions, presented in two different contexts (subject-wide scope and object-wide scope), for a total of 60 test items. Fourteen distractor items were included.

Example (21) below is represents the crucial POS construction:

(Given a picture of a girl holding three cats on her lap and petting them, OR

Given a picture of two girls and a boy, each of whom is petting a different cat):

(21) Canonical (SOV) order with *dono-N mo* ('every N') in object position

Dareka-ga dono neko-mo nadeta

Someone-NOM every cat stroked

‘Someone stroked every cat.’

Two additional related sentence types were tested for comparison. Neither of these constructions poses a POS problem according to Marsden; however, both may provide additional information on L1 transfer and provide perspective on the findings with respect to the POS problem. The first additional type is the scrambled version of (21), shown in (22):

(22) Scrambled (OSV) order with *dono-N mo* (‘every N’) as object

Dono neko-mo dareka-ga nadeta

Every cat someone-NOM stroked

‘Someone stroked every cat. (scrambled)’

Marsden predicts that all learners at all proficiencies will allow both subject-wide and object-side readings for this sentence type. Under Full Transfer/Full Access, English L1 intermediate learners whose interlanguage permits both readings for canonical sentences on the basis of L1 transfer will also permit both readings for scrambled sentences. On the other hand, learners whose interlanguage has become target-like for canonical sentences will also have target-like performance on scrambled sentences. In either case, learners will permit both scope readings for scrambled sentences. Korean L1 learners will pattern with Japanese native speakers on the basis of L1 transfer, because Korean patterns with Japanese.

The second type of comparison sentence uses canonical word order with a different quantifier, equivalent to English ‘all the N’. In this case (as in the ‘every N’ case

with canonical word order), a subject-wide scope reading is permitted, but not an object-wide scope reading:

(23) Canonical (SOV) order with *subete no-N* ('all the N') as object

Dareka-ga subete-no suutukeesu-o hakonda

Someone-NOM all-GEN suitcase-ACC carried

'Someone carried all the suitcases.'

Marsden predicts that all learner groups of all proficiencies will reject sentences of the type as in (23), because it is cross-linguistically difficult to get an object-wide reading when the quantifier has a collective interpretation.

Slight modifications were made for the English and Korean native speaker control groups. The test sentences were presented in the native language. The English test was shorter as it did not include sentences with scrambling. The Korean test only included the 'all the N'-type examples (corresponding to (23) above); data on the other two constructions was collected informally.<sup>66</sup>

## RESULTS

For the purpose of reporting results, Marsden aggregated '-2' and '-1' responses as rejections, and '+1' and '+2' results as acceptances. 'x' for 'can't decide' was rarely used.

Tables 4.8 and 4.9 below summarize the results for native speaker (NS) groups for subject-wide contexts and object-wide contexts, respectively. In general, the predicted judgments are attested. A subject-wide scope reading is available for all three test

---

<sup>66</sup> See Marsden (2009, p. 145).

constructions in all three languages. An object-wide scope reading is relatively acceptable for the English canonical ‘every’ construction and the Japanese scrambled construction, and relatively unacceptable elsewhere. English NS accept object-wide scope readings for canonical order sentences with ‘every N’ as object significantly more than Japanese NS accept the object-wide reading in Japanese canonical order sentences with corresponding ‘*dono N-mo*’ as object. Note that Korean NS data was gathered through a variety of means, not simply the same TVJ task. Marsden (2009, p.147) reports that Korean NS judgments pattern with Japanese native speaker judgments. Marsden takes this data to confirm the posited cross-linguistic differences and similarities and thus to confirm the L2 acquisition problem for L1 English learners of Japanese.

**Table 4.8: Acceptance Rates (%) by Native Speaker Group in Subject-Wide Scope Contexts**

	<b>Canonical ‘every’ (see (6) above) S &gt; O</b>	<b>Scrambled ‘every’ (see (7) above) S &gt; O</b>	<b>Canonical ‘all’ (see (8) above) S &gt; O</b>
<b>English NS</b>	98.00	N/A	99.60
<b>Japanese NS</b>	87.50	80.50	90.00
<b>Korean NS</b>	---	---	77.30

Note: Korean data collected informally, hence not reported here. Data from Marsden (2009, p.146).

**Table 4.9: Acceptance Rates (%) by Native Speaker Group in Object-Wide Scope Contexts**

	<b>Canonical ‘every’ O &gt; S</b>	<b>Scrambled ‘every’ O &gt; S</b>	<b>Canonical ‘all’ O &gt; S</b>
<b>English NS</b>	67.50	N/A	*21.30
<b>Japanese NS</b>	*16.00	81.50	*16.50
<b>Korean NS</b>	---	---	*20.00

Note: \* indicates reading was predicted to be disallowed. Data from Marsden (2009, p.146).

Table 4.10 below summarizes the learner results by group for subject-wide scope contexts. All learner groups (all languages, all proficiencies) show high acceptance rates on all subject-wide scope conditions regardless of quantifier type or word order. The L1 English learners have no difficulty allowing the subject-wide interpretation of the scrambled sentences, even at intermediate proficiency.

**Table 4.10: Acceptance Rates (%) by Learner Group in Subject-Wide Scope Contexts**

	<b>Canonical ‘every’ S &gt; O</b>	<b>Scrambled ‘every’ S &gt; O</b>	<b>Canonical ‘all’ S &gt; O</b>
<b>English Int</b>	97.37	90.00	95.79
<b>English Adv</b>	95.83	83.33	100.00
<b>Korean Int</b>	89.50	82.50	93.50
<b>Korean Adv</b>	95.34	94.67	98.67

Note: Data from Marsden (2009, p.148)

The acceptance rates for the learner groups in the object-wide scope condition are more nuanced. Table 4.11 below summarizes the learner results by group for the object-wide scope contexts. Recall that the canonical ‘every’ construction (example (21) above) was argued to present a POS problem for English L1 learners but not for Korean L1 learners. With respect to the question of developmental paths, comparing the earlier of the proficiency levels across the two learner language groups, Marsden found that the English L1 learners were significantly different from the Korean L1 learners. The intermediate English L1 group allows an object-wide interpretation at a significantly higher rate than the intermediate Korean L1 group does, and significantly higher than Japanese NS speakers. In this respect, learners are performing as predicted, with English L1 learners showing L1 transfer effects and Korean L1 learners performing in a more

target-like manner earlier in their development. With respect to the question of ultimate attainment, Marsden finds that English L1 advanced learners are not significantly different from the English L1 intermediate group. In fact, neither group performs very far off of chance. So far, these results do not strongly support a role of UG in the acquisition of this particular scope phenomenon.

**Table 4.11: Acceptance Rates (%) by Learner Group in Object-Wide Scope Contexts**

	<b>Canonical ‘every’ O &gt; S</b>	<b>Scrambled ‘every’ O &gt; S</b>	<b>Canonical ‘all’ O &gt; S</b>
<b>English Int</b>	57.90	77.37	*48.93
<b>English Adv</b>	*43.33	65.00	*28.33
<b>Korean Int</b>	*30.50	70.00	*19.50
<b>Korean Adv</b>	*17.34	71.34	*7.34
<b>Japanese NS</b>	*16.00	81.50	*16.50

Note: \*indicates reading was predicted to be disallowed. Japanese NS results repeated from Table 4.9 for convenience. Data from Marsden (2009, p.146 and 148).

With respect to the two other constructions that did not present POS problems, performance from all learner groups is in the direction predicted. Acceptance of object-wide scope readings in scrambled ‘every N’ constructions is higher than in canonical constructions. In general, acceptance of object-wide scope for ‘all the N’ sentences is lower than for ‘every N’ sentences, however, acceptance by the English L1 intermediate group hovers at chance.

Because the English L1 learners as a group performed near chance on multiple constructions/context pairs, Marsden investigated the consistency of individual results.

She analyzed native speaker groups as well as learner groups. She categorized ‘consistent’ and ‘inconsistent’ as follows:

- Consistent Acceptance: ‘+1’ or ‘+2’ response on 8 out of 10 sentences in a given construction type
- Consistent Rejection: ‘-1’ or ‘-2’ response on 8 out of 10 sentences in a given construction type
- Inconsistency: neither of the above

For present purposes we will highlight the individual consistency of performance on the crucial POS problem, that is, object-wide scope interpretations in canonical order ‘every-N’ type sentences. Table 4.12 below summarizes the consistency data. Close to half (10 out of 24) of the English native speaker group was inconsistent on this construction in this context. The English NS who were consistent at all were mostly consistent in accepting the object-wide interpretation. Although several Japanese native speakers were also inconsistent, all of the Japanese NS who were consistent were consistent in their rejection of the reading. The English L1 intermediate learner group had a high proportion of inconsistent individuals (similar to the English NS group, in fact). The English L1 advanced learner group had only 1 out of 12 who performed inconsistently, however, the consistent performers were almost evenly split between acceptance and rejection of the object-wide reading.

**Table 4.12: Individual Consistency (number of individuals in each category) of Performance on Canonical ‘every N’ Sentence Type, Object-Wide Scope Reading**

	<b>Consistent Acceptance</b>	<b>Consistent Rejection</b>	<b>Inconsistent</b>
<b>English NS</b> (n=24)	12	2	10
<b>Japanese NS</b> (n=20)	0	14	4
<b>English Int</b> (n=19)	8	3	8
<b>English Adv</b> (n=12)	5	6	1

Note: Data from Marsden (2009, p.152)

Examination of the individual performance within the groups reveals that some individuals are inconsistent in their responses. Considering the consistent performers, the trend goes from a higher percentage of L1-like responses at the intermediate stage to a higher percentage of target-like responses at the more advanced stage.

#### **CONCLUSIONS**

Marsden concludes that the evidence from her experiment supports both L1 transfer and UG access. The English L1 intermediate learner group and the Korean L1 intermediate learner group were significantly different, with each demonstrating L1-like performance on the crucial POS construction. This evidence supports Full Transfer. Some English L1 advanced learners performed in a target-like manner, consistently rejecting object-wide scope interpretations for canonical order *dono N-mo* ‘every N’ sentences. Marsden argues that the fact that some learners overcame the poverty of the stimulus challenge is evidence for Full Access to UG for L2 learners.

We now turn to examine studies of scope phenomena from an emergentist perspective.

#### **4.4.2 O’Grady, Lee and Kwak (2009)**

O’Grady, Lee and Kwak (2009) is one of a suite of related studies investigating first and second language acquisition of scope phenomena. With respect to L2A, their study looks at L1 Korean speakers learning L2 English. An experiment by O’Grady (2013a, discussed in Section 4.3.3 below) investigates the reverse direction, that is, L1 English speakers learning L2 Korean. These experiments build on previous work (Hawkins 2004; O’Grady 2005; among others) focusing on the possible role of a general processor rather than a dedicated UG-type language module driving language acquisition. Since these two studies share a common theoretical framework and test the same linguistic phenomenon, their common rationale will be presented in this section.

#### **RATIONALE**

This section will first discuss the theoretical goals of these studies in general terms, and then present the specific phenomenon and hypothesis tested in the experiments. The research questions for the individual experiments will be presented in different subsections below.

The overall theoretical goal of these two studies is to provide evidence on a general hypothesis that language acquisition emerges from the interaction of non-linguistic factors. The factors of interest to O’Grady and his colleagues are a general information human processor, and frequency. The authors start from a basic assumption about the job of processing (processing anything, not just language): the human processor seeks to be maximally efficient. More concretely, the processor seeks to minimize the burden on working memory, that is, to minimize the amount of information stored

temporarily and to minimize the number of operations performed on that information.<sup>67</sup>

For the processing of language, the authors adopt the following “efficiency assumptions” given in (24) and (25) below. The descriptions of the assumptions are based on O’Grady et al. (2009, p.75).

(24) Immediate Interpretation

As the processor “works its way through a sentence”, it assigns each NP an interpretation immediately on encountering it.

(25) Avoid Revisions

An assigned interpretation will be retained unless clearly contradicted. Revisions are costly and are to be avoided.

In the view of these authors, language acquisition does not consist of the acquisition of rules, rather, the acquisition of processing routines. O’Grady (2013a, p.254) goes so far as to say that “...what *appears* to be language acquisition is in fact an accidental side-effect of something very different and far more fundamental — attempts by the processor to facilitate its own operation.” Successful (native-like) acquisition involves the automatizing of these routines, a procedure that requires strengthening through repetition. For this reason, frequency of input is an important factor in acquisition of processing routines.

The authors argue that the processing-routine acquisition proposal sets up general predictions about the developmental path for adult L2 language learners in cases where the L1 and the L2 differ with respect to the processing costs associated with a given

---

<sup>67</sup> For a more detailed presentation, see O’Grady (2005).

linguistic phenomenon.<sup>68</sup> Maintaining the assumption that a processor will always seek to operate in a least-cost manner, the authors propose that an L1 processing routine will transfer to the L2 (and therefore an L1 interpretation will transfer to the L2) “if and only if it does not have a greater processing cost in the L2.” (O’Grady et al. 2009, p.83). We will see below how this proposal works out more specifically in the case of the scope phenomenon under investigation in these studies.

Before turning to a discussion at a more specific level, it is important to note that these authors break with many of their otherwise fellow-emergentists in that they do not deny the existence of POS phenomena. O’Grady and colleagues argue (see for instance 2009, p.73) that POS problems are overcome by the efficiency goals of the processor rather than by grammatical rules.

Within the overarching theoretical goal presented so far, both studies seek to test this processing account of acquisition by examining a particular scopal interaction between negation and quantifiers. Consider again example (17) from the beginning of Section 4.3, repeated here as (26), and its Korean equivalent in (27):

(26) Mike didn’t eat all the cookies.

(27) Mike-ka motun kwcac-lul an mekessta.

Mile-NOM all/every<sup>69</sup> cookie-ACC not ate

As we have seen earlier, the presence of the two scope-bearing elements (in this case, negation and universal quantification) sets up a potential ambiguity. If the universal

---

<sup>68</sup> O’Grady et al. (2009, p. 83) mentions Pienemann (1998) as having previously suggested a role for processing considerations in issues of L1 transfer.

<sup>69</sup> O’Grady (2013, p.10) states that *motun* “falls somewhere between English *every* and *all*”. O’Grady chooses to translate *motun* as *all* on the basis of its being compatible with plurals. I will simply adopt O’Grady’s analysis here. A full discussion of the correspondence between the quantifiers is beyond the scope of this paper.

quantifier is given wide scope over negation, then we have the interpretation where all the cookies are such that they were not eaten. O’Grady and colleagues refer to this as the “full set interpretation”. If negation is given wide scope over the universal quantifier, then we have the interpretation that it is not the case that all the cookies were eaten. That is, some of the cookies may have been eaten. O’Grady and colleagues refer to the ‘some’ interpretation as the “partitioned set interpretation”. From this point forward I will adopt their terminology.

In English, both the full set and the partitioned set interpretations are available for sentence (26), although the partitioned set interpretation is preferred. In Korean, sentence (27) can only be given the full set interpretation, that is, none of the cookies were eaten.

O’Grady and colleagues propose a processing account for these cross-linguistic differences on the assumption that sentences are processed in linear order. For English, note that the negation element is encountered before the universally quantified direct object NP. The authors argue that the fact that the negation operator is encountered first means that the processor is free to assign either a full set or a partitioned set interpretation to the direct object when it is encountered later. Either interpretation can be set up at comparable (low) cost. Over time, the relative frequency of the partitioned interpretation in the language input automatizes the partitioned interpretation. Hence the partitioned interpretation becomes the preferred one.

For Korean, the negation operator is attached to the verb in final position, therefore the universal quantification is encountered first. The authors argue that this encounter immediately sets up the full set interpretation. A partitioned set interpretation

could only be accomplished through revision later, which is costly. Given that processors avoid revision, the full set interpretation will be dominant in Korean.

Let us return to the idea of processing routine transfer to see what predictions these processing arguments make for L2A acquisition. Recall the authors' proposal above that processing routines will transfer from the L1 to the L2 if and only if the particular routine has a comparable or lower cost in the L2 than in the L1. How does this prediction work out for L2 acquisition between English and Korean in either direction? In Korean, the partitioned set interpretation is more costly than the full set interpretation. In English, both the full set and the partitioned set interpretations are relatively low cost. Therefore, an L1 Korean learner of English is predicted to transfer their low-cost full set interpretation routine into the L2, where the full set interpretation is of comparably low cost. On the other hand, an L1 English learner of L2 Korean is predicted *not* to transfer their preferred partitioned set interpretation into the L2. Transfer is blocked because the partitioned set interpretation in the L2 incurs higher costs.

We now turn to the discussion of the individual experiments. O'Grady et al. (2009) investigates the performance of Korean L1 learners of L2 English. Their specific research questions are:

- Do L1 Korean learners of L2 English accept a full set interpretation for universally quantified Direct Object NPs in negated sentences?
- Do L1 Korean learners of L2 English accept a partitioned set interpretation for universally quantified Direct Object NPs in negated sentences?

O'Grady et al. (2009, p.83) predict that learners will prefer the full set interpretation in the L2, based on its comparable low cost in both languages.

## EXPERIMENTAL METHOD

The L2 English learner subjects were 42 students, native speakers of Korean at a university in Seoul, Korea, who were enrolled in a linguistics class. Based on the participants' English-language coursework history<sup>70</sup>, O'Grady et al. categorized the group as Intermediate to High-intermediate proficiency in the L2. There was no separate control group of native speakers of Korean. Instead, native speaker intuitions were investigated by giving the same subject group a Korean version of the experimental task one week after they had been tested on the English version (described below). To investigate English native speaker intuitions, a small control group of 6 English native speakers completed the English version of the test.<sup>71</sup>

The task was a Truth-Value Judgment test (TVJ) comprising 8 test items and 10 filler items with 2 practice items at the beginning. Each test item consisted of a detailed context situation followed by a test sentence. The context determined either a full set or a partitioned set interpretation for the corresponding test sentence. For the English version of the test, the context was presented simultaneously aurally (played from a recording) and in writing. For the Korean version, contexts were presented in written form only. In addition, subjects were presented with a picture illustrating the end result of the situation. Example (28) gives a sample test item<sup>72</sup> (O'Grady et al. 2009, p.82):

(28) Context:

---

<sup>70</sup> No specific information is given regarding length or level of study.

<sup>71</sup> O'Grady et al. (2009, p.84-85) mention a small (n=5) pilot study with subjects who are English L1 learners of Korean. Rather than give details on this pilot study, which the authors term 'no more than mildly suggestive', the discussion of English learners of Korean will be postponed to the presentation of the slightly larger study in O'Grady (2013) below.

<sup>72</sup> The reader may recall this example from Section 3.3.2 above, where it was presented as an example of a TVJ test item.

Tom is at his uncle's repair shop. Tom's uncle is about to go out for lunch. He asks Tom to fix three radios and three computers before he returns. Tom promises to do so.

Tom fixes the three radios early. Then, Tom examines the first computer. But, he can't fix it. He decides to wait until his uncle comes back. Then, Tom looks at the second computer. There is something wrong with the sound, but he can't fix it. Finally, Tom comes to the third computer. There is something wrong with the screen. Screens are very hard to fix. But, Tom manages to fix it.

Test sentence: Tom didn't fix all the computers.

The accompanying picture would show one fixed computer and two broken computers, illustrating the partitioned set interpretation.

An alternative context could support the full set interpretation by substituting the text in (29) for the corresponding final paragraph in the context in (28) above:

(29) Finally, Tom comes to the third computer. There is something wrong with the screen. He thinks that he can fix it quickly. However, after Tom works on it for a while, he gives up.

In this case, the accompanying picture would show all three computers still broken, visually illustrating the full set interpretation. Only one of the two contexts would be presented to a given subject.

Following the presentation of the context, subjects were given 10 seconds to provide a truth judgment on the corresponding test sentence in a test booklet. English

native speakers were given a follow-up questionnaire, however, there is no mention of a corresponding questionnaire for the Korean L1 group.

## RESULTS

Table 4.13 below summarizes the results by subject group. The numbers indicate the percentage of times that the subject group judged the test sentences to be True in the relevant contexts. English native speakers accepted the test sentences under the full set interpretation 67% of the time, but favored the partitioned set interpretation, accepting the test sentences 100% of the time in those contexts. The Korean L1 learners of English did not respond in a similar fashion. Their acceptance rate in contexts determining full set interpretations was higher (93%) than the English NS, and their acceptance rate in contexts determining partitioned set interpretations was quite a bit lower, at 28%. The same Korean L1 subject group gave acceptance rates in their native language on the Korean version of the test one week later that were very comparable to the rates they gave in the L2 English version.

**Table 4.13: Percentage of ‘True’ Responses by Subject Group**

	<b>Full Set Contexts</b>	<b>Partitioned Set Contexts</b>
<b>English NS</b>	67%	100%
<b>Korean learners</b>		
<b>English (L2) version</b>	93%	28%
<b>Korean (L1) version</b>	97%	21%

Note: Data from O’Grady et al. (2009, p.83).

## CONCLUSIONS

O’Grady et al. take the results to confirm their analysis and hypotheses. First, with respect to the native speakers, the authors take both sets of results to confirm the

predictions of the processing account. The Korean native speakers prefer the full set interpretation, that is, they prefer the interpretation that has the lowest processing cost. The English native speakers prefer the partitioned set interpretation, but also allow the full set interpretation at a high rate. The authors take these results to confirm their proposal that both interpretations are relatively low-cost in English, and that the processing costs of the two interpretations are not as different in English as they are in Korean.

Second, with respect to the L2 learners, the authors take the performance of the Korean L1 learners on the English version of the test as again confirming the processing account. On the assumption (supported, they argue, by the English NS results) that the full set interpretation in English is relatively low cost as it is in the L1 Korean, the authors predicted that Korean learners of English would simply transfer their L1 processing routines and replicate their L1 preferences in the L2. O'Grady et al. conclude that the Korean L1 learners of English performed as expected.

These proposals can be further tested by looking at L2A in the other direction: English L1-ers learning L2 Korean. The experiment described in the following section does just that.

#### **4.4.3 O'Grady (2013a)**

This more recent study by O'Grady is conducted under the umbrella of the theoretical approach and rationale given above in Section 4.3.2. The research questions specific to this study are:

- Do L1 English learners of L2 Korean accept a full set interpretation for
- universally quantified Direct Object NPs in negated sentences?

- Do L1 English learners of L2 Korean accept a partitioned set interpretation for universally quantified Direct Object NPs in negated sentences?

O'Grady (2013a) predicts that English learners will not transfer their native L1 routines to the L2, because the partitioned interpretation is more costly in Korean than in English. English L1 learners of Korean are instead predicted to perform in a target-like manner early on. That is, they will accept the full set interpretation and reject the partitioned set interpretation for the direct object NP.

#### **EXPERIMENTAL METHOD**

O'Grady's subjects were 10 students at a university in the United States who were enrolled in a fourth-semester Korean language course. O'Grady (2013a, p.274) categorizes the subjects as low- and mid-intermediate learners. All were native speakers of English in their late teens and early twenties. An additional group of 10 native speakers of English who had never studied Korean were tested on an English version of the task.

The written task comprised 13 test items total: 5 target constructions and 9 distractors. Each test item consisted of a test sentence followed by two contexts. For the target constructions, one context favored a full set interpretation, and one context favored a partitioned set interpretation. Subjects were required to say which context was better described by the sentence. Although O'Grady is not specific, it appears that choosing 'both' was not an option. Contexts were presented in English to avoid difficulties with vocabulary. Example (30) gives a sample test item (O'Grady 2013a, p.275). The gloss and translation of the test sentence are provided here, however, they were not provided in the test.

(30) Tom-i motun chayk-ul an ilkesseyo.

Tom all the books not read

‘Tom didn’t read all the books.’

I gave Tom all the books that he  
was supposed to read, but he didn’t  
read any of them.

(full set interpretation)

I gave Tom all the books that he  
was supposed to read, but he read  
only some of them.

(partitioned set interpretation)

## RESULTS

The English L1 learning Korean subjects chose the full set interpretation 100% of the time, and never chose the partitioned set interpretation. In contrast, the English NS who had not studied Korean chose the partitioned set interpretation 96% of the time.

At this point, it is worth noting the results of the small (n=5) pilot study of English L1 learners of L2 Korean conducted by O’Grady et al. (2009) that was mentioned in footnote 2 above. The five subjects were ‘relatively advanced’ learners (2009, p.85) who completed the Korean version of the task from O’Grady et al. (2009). Recall that in that task, the subject was presented with a (unique) context first, then presented with the test sentence, and asked to judge whether the sentence was true in the given context. On this substantively different task, the results differed from those just reported for O’Grady (2013a). Although the full set interpretation was accepted 100% of the time, it was also the case that the partitioned set interpretation was accepted 50% of the time. Chapter 5 will discuss the relative merits of the two tasks and the possible influence of design on results.

## CONCLUSIONS

Confining attention to the results of the (2013a) study only, O'Grady argues that his results are consistent with a processing theory account. The processing account predicts that English learners of Korean will not transfer their native processing routine favoring the partitioned set interpretation because the partitioned set interpretation is more costly to compute in Korean. "Cost blocks transfer" (2013a, p.275). In this experiment, low- to mid-intermediate learners display target-like performance in the L2. Although early acquisition of a POS phenomenon is often used to argue for UG, O'Grady instead argues that a UG-type explanation is not necessary.

This concludes the literature review. Chapter 4 has outlined three sets of empirical investigations of L2A poverty-of-the-stimulus phenomena. Each set of investigations contained experiments that approached acquisition from a generative and an emergentist point of view. Although this review is not exhaustive, it does offer a good inventory of sets of studies that are comparable on the basis of the linguistic phenomenon under investigation. This review also offers much food for thought, and we now turn to evaluations and commentary in Chapters 5 and 6.

## **CHAPTER 5: CRITICAL EVALUATION**

### **5.1 Overview**

Chapter 4 presented three sets of studies. Each set investigated the acquisition of a particular linguistic phenomenon that has been argued to present a poverty of the stimulus problem for language learners. These phenomena—the Overt Pronoun Constraint, scrambling constraints, and scope interpretation constraints—are challenging for learners to learn and for theories to explain.

Chapter 5 provides a critical assessment of the studies presented in Chapter 4, examining the studies on a detailed level. The discussion is structured around empirical and theoretical issues in the experimental designs and arguments. More specifically, Section 5.2 considers empirical issues around the adequacy of the experimental methodologies, with attention to characteristics of the subject population and to the design of the experimental tasks. Section 5.3 then considers theoretical issues around the rationale, research questions, results and conclusions drawn by the authors, with attention to the appropriateness of the authors' conclusions to the evidence and the extent to which those conclusions are theory-dependent. Studies will be rated 'adequate' or 'inadequate' on the basis of their satisfactory or unsatisfactory empirical design and/or theoretical argumentation. On some criteria, adequacy seems to be a matter of degree rather than strictly binary, and a middle-ground rating may be given. The discussion in Chapter 5 will form the basis for a macro-level discussion in Chapter 6, where broader themes will be taken up and the aggregate evidence will be considered.

## 5.2 Micro-Level Evaluation: Empirical Issues

### 5.2.1 Subject Characteristics

Section 5.2.1 begins with a consideration of the adequacy of the studies on the basis of the number of subjects tested and on subjects' proficiency. Table 5.1 below summarizes the basic facts from each of the studies. The numbers given for Hopp's as well as Marsden's experiments reflect the fact that those researchers investigated L2 learners from two different L1 backgrounds.

**Table 5.1: Empirical Assessment—Subjects**

	<b>Number of L2 Subjects</b>	<b>Proficiency(ies)</b>	<b>Basis for Proficiency</b>	<b>Number of Native Speaker Controls</b>
<b>Kanno 1997</b>	28	Low	course level	20
<b>Kanno 1998</b>	29	Low	course level	12
<b>Pérez-Leroux &amp; Glass 1999</b>	39	Elementary	course level	20
	21	Intermediate	course level	
	18	High	experience	
<b>Sheen 2000</b> Exp 1 Exp 2 Exp 3	33	no information	no information	(relied on Kanno's) 43
	6	not described	graduate level	
	n/a	n/a	n/a	
<b>Hopp 2005</b>	8	High-Intermed	Cloze test for all	26
	5	Advanced		
	7	High-Intermed		
	13	Advanced		
	6	Very advanced		
<b>Williams &amp; Kuribara 2008</b>	41	True beginners	n/a	n/a (semi-artificial language)
<b>Marsden 2009</b>	21	Intermediate	Cloze test for all	24
	12	Advanced		21
	23	Intermediate		22
	15	Advanced		
<b>O'Grady, Lee &amp; Kwak 2008</b>	42	Intermediate to High-Int	history of coursework	42
				6
<b>O'Grady 2013a</b>	10	Low- and Mid-Intermediate	course level	10

The number of learners for any given subject group ranges from 5 to 41 for L2 learners, and from 6 to 43 for native speaker control groups. The numbers at the lower ends of these ranges are uncomfortably low. Low numbers for native speaker groups may be less problematic than for learner groups. As mentioned in Chapter 3 above, native speaker populations are expected to be more consistent in their performance and therefore the performance of a small group might not unreasonably be generalizable. However, L2 learners, especially at lower proficiencies, are expected to show variable performance. The smaller the sample size, the more influence each individual has on statistics such as the mean. Generalizations from a small sample of highly variable subjects may not be reliable.

That being said, for the purpose of these experiments the average performance of any learner subject group is less important than the performances of individual subjects, in the following sense: if any subjects at all show acquisition of the phenomena in question, this fact needs to be explained. From this point of view, small subject sizes can be acceptable but results need to be interpreted with caution. In terms of the number of subjects, most of the studies are adequate. The exceptions are Sheen's second experiment and O'Grady's 2013a experiment which have low numbers of subjects (6 and 10, respectively) and do not report individual results. While most of Hopp's subject groups and one of Marsden's subject groups are also small, both researchers report checking individual level responses, thus providing more understanding of the reliability of their group results.

Turning to the matter of subject proficiency, two items of interest are the particular proficiency levels tested, and the method of assessing proficiency. With regard

to the particular levels studied, Table 5.1 above shows that some researchers tested learners at one proficiency level only, and others tested learners at two or three different levels of proficiency. These different choices will provide different information. When testing high proficiency subjects only, one is more likely to find learners who behave in a native-like manner, fulfilling the acquisition assumption and being able to apply the POS argument. (None of the studies in this set test only high-proficiency learners.) On the other hand, the higher the subjects' proficiency, the more difficult it is to make a POS argument on the basis of rarity in the input. The larger amount of input to which the subjects must have been exposed inherently increases the difficulty of reliably characterizing their input as being insufficient. We will return to the question of sufficiency of input in Chapter 6.

If intermediate-level subjects only are tested (as in O'Grady, Lee and Kwak as well as O'Grady), it is less likely that these subjects will demonstrate mastery of a given phenomenon than high proficiency subjects will. However, should mastery of arguably POS phenomena be found in an intermediate-level population, the argument for UG is correspondingly stronger than if mastery is found in a high proficiency population. Learners would have attained their native-like performance on the basis of less input.

When testing low proficiency subjects only (as with Kanno, or Williams and Kuribara), if low proficiency learners do acquire (POS) phenomena, then the argument for UG is even stronger than it would be if intermediate or high proficiency learners are successful. But if the low proficiency subjects do *not* demonstrate native-like behavior, then no conclusions can be drawn about the existence and role of UG, or the role of input. The subjects' lack of success could be due either to L1 transfer interference, or to an

insufficient amount of input, or both. Therefore, neither the generative nor the emergentist approach could be supported or contradicted. In every case in which only one proficiency level is tested, the possible conclusions based on results may be limited.

The most valuable studies are those which test subjects at multiple proficiency levels. Importantly, studies conducted across proficiency levels may reveal differences in the developmental paths of learners. Recall that the generative and emergentist approaches make different predictions about learners' developmental paths: the former predicts a relatively sudden, relatively early acquisition and stability of performance once acquisition occurs; the later predicts a longer, incremental path with possibly variable performance at all levels. Only by including learners of various proficiencies can a study provide evidence for one approach or the other on the basis of developmental path. In light of the above remarks, the strongest studies are those by Pérez-Leroux and Glass, by Hopp, and by Marsden, all of which test learners at multiple proficiencies, and all of which include high proficiency learners.

The second item of interest with respect to subject proficiency is the way in which it was assessed and subjects categorized. Sheen's L2 experiments cannot be evaluated due to a lack of information. Similarly, O'Grady, Lee and Kwak give only the briefest remark about having taken subjects' coursework history into account when they decided to characterize their learner group as 'intermediate to high-intermediate'. There is no specific information given regarding the number of courses or levels of courses completed. Kanno (both studies), Pérez-Leroux and Glass, and O'Grady all rely on course levels as proxy for proficiency assessment. In particular, Kanno's, O'Grady's, and Pérez-Leroux and Glass's low and low-intermediate groups were all drawn from students

enrolled in 4<sup>th</sup> semester language classes. Kanno remarks (1997, p.269) that her subjects likely represent a good cross-section of language learners. Given that students at the relevant university (University of Hawai'i) were required to take at least 5 semesters of language, the 4<sup>th</sup> semester class would contain students with a complete range of interests and aptitudes. However, without an independent assessment, Kanno could be underestimating her subjects' proficiency on two counts. First, students sometimes enroll in courses that are below their abilities. Second, Kellerman and Yoshioka (1999) observe that the high proportion of students of Japanese heritage at the University of Hawai'i increases the likelihood that student subjects have been exposed to native Japanese (the L2 in Kanno's studies), possibly a significant amount, even if they have not lived with a native speaker of Japanese. Coursework alone cannot guarantee an appropriate proficiency categorization.

The strongest studies with respect to proficiency assessment are those by Hopp and by Marsden, both of which categorize their subjects on the basis of a cloze test in addition to collecting language history information. Even though each study on its own gets high marks for conducting an independent proficiency assessment, caution is in order when comparing results across the two studies, as the scoring methods were not completely consistent. Table 5.2 below shows the mean scores on the cloze tests for each learner group.

**Table 5.2: Mean Scores on Cloze Test Proficiency Assessment by Learner Groups**

	<b>English L1</b>	<b>Japanese L1</b>	<b>Korean L1</b>
<b>Hopp</b>			
<b>High-Intermed</b>	19.3 out of 40	18.8 out of 40	
<b>Advanced</b>	28.7 out of 40	30.4 out of 40	
<b>Very advanced</b>	39.2 out of 40		
<b>Marsden</b>			
<b>Intermediate</b>	7.2 out of 42		6.78 out of 42
<b>Advanced</b>	14.3 out of 42		18.0 out of 42

Marsden’s cloze test used an exact-word scoring method, which means that even syntactically and/or semantically appropriate responses could be marked wrong if they were not identical to the test answer key. Marsden had a group of 30 native speakers take the cloze test, and some of the NS scored as low as 12. For this reason, Marsden assigned any learner who got at least a 12 to her advanced group. Hopp’s native (German) speakers averaged 37.2 out of 40 on his cloze test, which was not an exact-word test. He does not explain why he assigned the scores he did to each different category. Although each study benefits from the independent assessment of proficiency, it is difficult to know how to compare groups across studies. For instance, is Hopp’s advanced group comparable to Marsden’s advanced group? In the conclusion of Section 5.2 below, when summarizing the empirical facts from the entire set of studies, it must be kept in mind that some proficiency groups may be more comparable across studies and some may be less comparable.

Table 5.3 below summarizes the ratings for the studies on the basis of the number of subjects and the subjects’ proficiencies.

**Table 5.3: Adequacy Assessment—Subject Characteristics**

	# of L2 subjects	Proficiency: Levels, Assessment	
<b>Kanno 1997</b>	✓	?	?
<b>Kanno 1998</b>	✓	?	?
<b>Pérez-Leroux &amp; Glass 1999</b>	✓	✓	?
<b>Sheen 2000</b>			
Exp 1	✓	x	x
Exp 2	x	x	x
Exp 3	✓	n/a	n/a
<b>Hopp 2005</b>	✓	✓	✓
<b>Williams &amp; Kuribara 2008</b>	✓	?	✓
<b>Marsden 2009</b>	✓	✓	✓
<b>O’Grady, Lee &amp; Kwak 2008</b>	✓	?	?
<b>O’Grady 2013a</b>	?	?	?

✓ = adequate; ? = caution required; x = inadequate

### 5.2.2 Task Design

Next we will consider empirical issues around task design. All of the studies under consideration use either a grammaticality judgment task (GJ) or a truth-value judgment task (TVJ) with the exception of Pérez-Leroux and Glass (hereafter P-L&G), who use a translation task. As discussed in Chapter 3 above, production tasks in general are not ideal for studying complex and/or infrequent constructions. However, in this case, production in the form of translation is quite controlled and P-L&G are largely successful in eliciting the target construction. The P-L&G study will therefore be included in the discussion in this section. After briefly reviewing the general characteristics that constitute best practice in task design, the studies will be evaluated for their conformity to those standards.

Chapter 3 considered a number of other design characteristics in addition to the production versus comprehension option mentioned above. Two of the characteristics will be put aside for the purpose of this section: the number of test items; and the use of a rating scale versus a binary choice for reporting judgments. The number of test items used in these studies ranges anywhere from 14 to 194. In the absence of a clearly defined minimum or maximum standard, all studies will simply be considered adequate on this basis, although more data should in principle provide greater statistical power. From a theoretical perspective, there could hardly be a maximum. From a practical standpoint, the experimenter must be concerned with subjects becoming fatigued. The preference between a rating scale versus binary choice, as mentioned in Chapter 3, is not clearly indicated. Again, all studies we are examining here will be considered adequate no matter which reporting method they use; recall that these POS studies had already been preselected for the appropriateness of their investigations and on the strength of their methodologies.

The desirable design characteristics presented in Chapter 3 that will be examined here are: distracting the subject from the target construction; providing context for the test sentences; and presenting the subjects with one question at a time. Recall that when subjects do not recognize which particular construction is the target of the task, they are less likely to use conscious strategies to arrive at their judgments and therefore the judgments are more likely to reflect their underlying competence. A task has a better chance of masking the target from the subject if the task contains distractor items and if the instructions to the subjects are of a general nature. The provision of context for test sentences is desirable because it increases the control of the experiment. Context may

facilitate consistency of responses across subjects, and accuracy of responses within subjects. Context can be provided via written stories or pictures, the latter being especially helpful. Finally, clear choices facilitate accurate judgments and elicitation of all possible judgments. The subjects' choices are clearest when a single context or interpretation is matched with a single test sentence. Examples of potentially confusing test designs will be presented below.

With regard to distracting subjects from the target construction, about half the studies include distractor items in their tasks and are therefore considered adequate in this respect: Hopp, Marsden, O'Grady, Lee and Kwak (hereafter OLK), and O'Grady. The proportion of distractor items ranges from approximately 20% (Marsden) to 60% (O'Grady). The other studies (Kanno, P-L&G, Sheen, and Williams and Kuribara) include only target test items. Worse yet, Williams and Kuribara offer instructions that draw the learners' attention to the word order phenomena whose acceptability the researchers are measuring.

With regard to providing context, all of the studies except Kanno's and Sheen's are adequate in that they do provide context to the test sentences. P-L&G and Hopp provide written context; Williams and Kuribara as well as Marsden provide visual context in the form of a stick drawing or a line drawing; OLK provide both written and visual context. O'Grady provides a written context, however, the context follows the test sentence rather than precedes it. For this reason, the O'Grady study has inadequate design with respect to context. The sequencing in which the context is provided does not serve to help control the experiment, in that subjects have the opportunity to imagine their own

context and to form opinions about the test sentence prior to reading the experimentally given contexts.

Finally, with regard to presenting the subjects with one question at a time, all of the studies are adequate with the exception of Kanno, Sheen, and O'Grady. Kanno's experimental task was described in Chapter 4, Section 4.2.1 above. For each test sentence, the learner was given two choices for the interpretation of the embedded pronoun: a sentence-internal antecedent, or a sentence-external antecedent. This means that interpretations that are possibly available but dispreferred are in some sense in 'competition' with the preferred readings. Although the learners were given the option to indicate that both interpretations were possible, it is plausible that any dispreferred readings could be overlooked. In comparison with the simultaneously-presented preferred reading, the dispreferred reading might seem unacceptable to the learner. Since Sheen duplicated Kanno's task design in all three of his experiments, the same criticism applies. The issue with the O'Grady study is slightly different. As mentioned earlier, O'Grady presents the test sentence and then presents two different possible contexts: one which supports a full set interpretation, and one which supports a partitioned set interpretation (for a sample test item, see Chapter 4, Section 4.4.3). The learner was asked to choose which context best suits the test sentence. In this case, the learner was not offered the opportunity to indicate that both contexts are possible, further increasing the risk that a possible but dispreferred interpretation would never be recognized in the test results.

Table 5.4 displays the adequacy ratings for each of the studies on the three criteria discussed.

**Table 5.4: Adequacy Assessment—Task Design Characteristics**

	<b>Distracts from Target</b>	<b>Provides Context</b>	<b>Presents one Choice at a Time</b>
<b>Kanno 1997</b>	x	x	x
<b>Kanno 1998</b>	x	x	x
<b>Pérez-Leroux &amp; Glass 1999</b>	x	✓	✓
<b>Sheen 2000</b>			
<b>Exp 1</b>	x	x	x
<b>Exp 2</b>	x	x	x
<b>Exp 3</b>	x	x	x
<b>Hopp 2005</b>	✓	✓	✓
<b>Williams &amp; Kuribara 2008</b>	x	✓	✓
<b>Marsden 2009</b>	✓	✓	✓
<b>O’Grady, Lee &amp; Kwak 2008</b>	✓	✓	✓
<b>O’Grady 2013a</b>	✓	x	x

✓ = adequate; x = inadequate

Table 5.4 shows that only three studies are adequate on all three criteria: Hopp, Marsden, and OLK. For the purpose of creating an aggregate rating for task design, we will designate these three studies as adequate. Studies which earned two check marks will be rated ‘caution required’, and studies with one or zero check marks will be rated inadequate. Table 5.5 summarizes the adequacy assessment on empirical factors, subject characteristics as well as task design.

**Table 5.5: Adequacy Assessment—Empirical Factors**

	<b>Number of L2 Subjects</b>	<b>Proficiency: Levels, Assessment</b>		<b>Task Design</b>
<b>Kanno 1997</b>	✓	?	?	x
<b>Kanno 1998</b>	✓	?	?	x
<b>Pérez-Leroux &amp; Glass 1999</b>	✓	✓	?	?
<b>Sheen 2000</b>				
<b>Exp 1</b>	✓	x	x	x
<b>Exp 2</b>	x	x	x	x
<b>Exp 3</b>	✓	n/a	n/a	x
<b>Hopp 2005</b>	✓	✓	✓	✓
<b>Williams &amp; Kuribara 2008</b>	✓	?	✓	?
<b>Marsden 2009</b>	✓	✓	✓	✓
<b>O’Grady, Lee &amp; Kwak 2008</b>	✓	?	?	✓
<b>O’Grady 2013a</b>	?	?	?	x

✓ = adequate; ? = caution required; x = inadequate

To close the critique of empirical factors, it is worth noting that two of the experiments in the studies confirmed the caveats presented in Chapter 3 above regarding the reliability and validity of judgment tasks. Kanno’s 1998 experiment tested her learner groups twice, the second time being twelve weeks later than the first. Many of her subjects showed inconsistent responses across the two tests, indicating a possible problem with reliability of the task. One of Sheen’s experiments investigated the reasons behind learner judgments. Sheen found that a majority of learners were consciously implementing a prescribed rule rather than relying on intuition. This indicates a possible problem with the validity of judgment test responses. These results will be kept in mind when we summarize the empirical facts gleaned from these studies.

### **5.3 Theoretical Issues**

This section considers and evaluates the theoretical argumentation offered in the set of studies examined in Chapter 4. The discussion centers on the research questions, the rationales, and the reasoning and conclusions that the authors provide around their empirical results. The review will cover the generative studies first, followed by the emergentist studies.

#### **5.3.1 Generative Studies**

##### ***5.3.1.1 Generative Studies: Research Questions***

The generative studies share a common general research question: does UG play an active role in 2LA? The specific research questions vary with the specific poverty-of-the-stimulus phenomena investigated. The one exception is Kanno's 1998 study, which investigates adult L2 learner variability across learners at a certain time as well as within learners across time. As mentioned at the close of the previous section, since her study examines L2 learner performance on a POS phenomenon, her experiment is relevant to the assessment of the findings of the other POS studies. In addition to investigating the role of UG, two of the generative POS studies (Hopp 2005 and Marsden 2009) also investigate the role of L1 input and L1 transfer by testing two different L2 learner populations: one whose L1 instantiates the target construction, and one whose L1 does not. All of these studies are adequate on the basis of their research questions.

##### ***5.3.1.2 Generative Studies: Rationales***

For their rationales, the generative POS studies employ the standard POS argument: If learners can be shown to have acquired a target construction for which they experienced insufficient input, then UG must be active in their acquisition process. The adequacy of

this rationale rests on the strength of the argument that the input is impoverished. Chapter 2 above presented an inventory of ways in which linguistic input has been argued to be insufficient for learning. Briefly: Some phenomena appear too rarely or are too complex to be gleaned from exposure to typical input. Some strings of words are ambiguous as to their structure, therefore the input is under-informative. Some patterns of sentences may mislead a learner into incorrect predictions about the grammaticality of related sentences. Typical input does not necessarily contain negative evidence that would allow a learner to retreat from overgeneralizations. For adult L2 learners, two additional possible sources of input must be considered: the L1, and instruction. If the L1 does not instantiate the construction and if learners do not receive explicit instruction on the target construction, then their input from those potential sources is insufficient.

As a tool for evaluating the POS arguments and therefore the study rationales, Table 5.6 reports which of the above-mentioned factors are cited as part of the stimulus poverty arguments in each study. For instance, Kanno is listed as having argued that the Overt Pronoun Constraint is a poverty-of-the-stimulus phenomenon on the basis that the input is misleading, that learners are neither taught nor directly corrected, and that the OPC is not instantiated in her learners' L1 (English). (For more complete descriptions of the arguments, see the Rationale sections for the relevant studies in Chapter 4.) In Table 5.6, a 'y' indicates that an argument of the category in the column heading was presented in the study listed in the row heading; no mark means that that particular type of argument was not offered for a given study. Several of these POS arguments merit scrutiny, the reasons for which will be explained below.

**Table 5.6: POS Arguments Offered in Each Study**

	<b>Complex/ Rare</b>	<b>Uninformative</b>	<b>Misleading</b>	<b>No Neg Evidence</b>	<b>Not Taught</b>	<b>Not in L1</b>
<b>Kanno 1997, 1998</b>			y	y	y	y
<b>Pérez-Ler. &amp; Glass 1999</b>	y				y	
<b>Hopp 2005</b>	y	y	y	y	y	y
<b>Marsden 2009</b>		y			y	y

y = argument was presented in the study.

Kanno argues that L1 English learners of L2 Japanese could be misled by the input, given characteristics of the L1. Embedded null subject pronouns can take both referential and quantified antecedents, so learners could reasonably expect that embedded overt subject pronouns could do the same, especially considering “the unreliability of negative evidence” (1998, p.1127). Kanno’s remark about negative evidence refers to direct negative evidence, that is, direct correction. While research supports her contention that direct correction is not reliably available to all L2 learners (see Chapter 2 above), recall (also from Chapter 2) that other input may function as negative evidence. In particular, it has been suggested that unmet expectations can function as negative evidence (see for instance Ramscar, Dye, and McCauley 2013 and references therein). If Kanno is correct that learners will expect grammatical quantified antecedents for embedded overt pronouns, and if the expectations proposal is on the right track, then the fact that quantified antecedents are not found in the L2 input will serve to inform learners of the ungrammaticality (but see Pinker 2004 for a different point of view on indirect negative evidence).

Kanno also argues that the OPC is not instantiated in her subjects' L1 because English lacks null pronouns of the Japanese type.

Pérez-Leroux and Glass (1999), on the other hand, base their POS arguments on the lack of instruction on the OPC and on the rarity of occurrence of input that illustrates the OPC. However, their argument from rarity is difficult to assess in the absence of any quantitative evidence. Their evidence consists of estimations such as (1999, p.230): “Potential OPC environments are a very small subset of all pronouns used.” While such a statement may be true, it is not conclusive with respect to sufficiency of input, on two counts: first, relative infrequency does not entail absolute rarity; and second, there is no established minimum number of exposures required for acquisition. Unless one can claim that the number of exposures is less than some established minimum, the rarity argument is ‘soft’ (see Schwartz and Sprouse 2013 for more comments on ‘soft’ POS arguments).

Hopp (2005) provides a number of POS arguments concerning his target construction, remnant scrambling; we will question three of them. First, he does cite corpus studies in an effort to bolster his POS argument on the basis of infrequent input. However, two of the three corpus studies he cites do not provide direct evidence on remnant scrambling, thus weakening his quantitative argument.<sup>73</sup> Schlesewsky et al. (2000, p.67-68) discusses the relative frequency of subject-initial declaratives versus object-initial declaratives and finds that for NP-initial sentences, 90% are subject-initial. However, this information is relevant to topicalization rather than scrambling (recall that scrambling in German takes place in the complement field). Bornkessel et al. (2002), presents information on the appearance of a nominative-marked constituent versus an

---

<sup>73</sup> I have not been able to verify the third (Hoberg 1981).

accusative- or dative-marked constituent after a complementizer. The former is always indicative of canonical word order whereas that latter two are indicative of scrambled orders. The corpus indicates that nominative-marked constituents occur eight times more often than accusative- or dative-marked constituents. Again, this is a measure of relative frequency, not absolute frequency. But the point of interest here is that there are a number of sentence patterns in the remnant movement paradigms tested by Hopp that would not show up in a count as performed by Bornkessel et al. Hence their frequency analysis may not be applicable to Hopp's target constructions.

Recall from Chapter 4 that Hopp (2005) also presents an argument that L2 input could be misleading, in the sense that if learners only pay attention to the main elements of sentences, surface order is not sufficient to distinguish grammaticality from ungrammaticality. Examples (1) and (2) below (repeated from (5b) and (6) in Chapter 4) illustrate the similarity:

(1) [t<sub>1</sub> zu reparieren]<sub>2</sub> hat Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht.

to repair                      has Peter the car                      already                      tried

'Peter has already tried to repair the car.'

(2) \*Ich glaube, dass [t<sub>1</sub> zu reparieren]<sub>2</sub> Peter [den Wagen]<sub>1</sub> schon t<sub>2</sub> versucht hat.

I think that to repair                      Peter the car                      already tried                      has

'I think that Peter has already tried to repair the car.'

Such an argument ignores the many other surface (not to mention structural) differences between the two sentences. Unless it has been established independently that learners do in fact attend to main elements, this argument is not conclusive.

Finally, there is an issue with Hopp's argument with respect to whether or not English as an L1 can contribute to the acquisition of the constraints on remnant scrambling in German. Hopp argues (2005, p.37, citing Oka 1996 and Sauerland 1999) that "...whether a language instantiates scrambling reduces to the availability of a strong uninterpretable scrambling feature [scr] in the language-particular functional lexicon, Lex<sub>FF</sub>." Hopp then argues that a universally available Principle, the Principle of Unambiguous Domination, allows learners to acquire the *constraints* on remnant movement of all kinds 'for free'. In other words, an English L1 learner would receive input through the L1 that is relevant to constraints on remnant scrambling. This analysis is problematic for Hopp's POS argument on the following basis: Hopp does not claim that scrambling is rare, only that remnant scrambling is rare. Therefore, it is plausible that learners could acquire the scrambling feature on the basis of positive data, then acquire the movement constraint on the basis of the L1. Hopp's analysis of what the learner must acquire seems to run counter to a POS claim.

Marsden (2009) seems to offer the strongest POS argumentation in that the three arguments that she offers are solid.

Table 5.7 summarizes the adequacy ratings for the study research questions and rationales. The ratings for the rationales are based on the strength or weakness of the POS arguments, as discussed above. Chapter 6, the macro evaluation, will return to the question of the efficacy of POS argumentation from a broader perspective.

**Table 5.7: Adequacy Assessment, Generative Studies—Theoretical Factors, Part 1**

	<b>Research Questions</b>	<b>Rationale</b>
<b>Kanno 1997</b>	✓	?
<b>Kanno 1998</b>	✓	?
<b>Pérez-Leroux &amp; Glass 1999</b>	✓	?
<b>Hopp 2005</b>	✓	?
<b>Marsden 2009</b>	✓	✓

✓ = adequate; ? = caution required; x = inadequate

### ***5.3.1.3 Generative Studies: Results and Conclusions***

The assessment of the theoretical issues in the generative studies will conclude by turning to consider the various study results and conclusions. In particular, we will benchmark the authors' arguments against their empirical findings.

Kanno's 1997 Overt Pronoun Constraint study results seem remarkably definitive on first look. Recall that Kanno reports that as a group, her low-proficiency subjects are not significantly different from native speakers. She verifies the group results by checking individual performance, and finds that 86% of the L2 learners perform at a native-like level, consistently (4 out of 5 times) disallowing a quantified matrix antecedent. However, these results may be artificially strong for two reasons. First, as mentioned in Section 4.2.1 above, Kanno's task design placed both possible interpretations for the target pronoun in competition with one another on each test item. Under this task design an existing but dispreferred reading may have been underrepresented. That is, if learner competence did actually include some level of acceptance of matrix quantified antecedents for overt embedded subject pronouns,

Kanno's test would not necessarily reveal it. Second, Kanno may have underestimated the proficiency of her subject population. Higher overall proficiency would be expected to manifest more native-like performance. Both of these factors could have contributed to exaggerating the learner results in the direction of native speaker performance.

Kanno's conclusion that her empirical results support a role for UG in L2 learning is appropriate to her data as found. Early learning of a POS phenomenon is considered good evidence for UG-controlled acquisition. However, we have seen that both the early learning claim and the POS argument can be criticized. An additional caveat comes from the results of Kanno's 1998 study, which will be discussed next.

The 1998 study replicates the task of the 1997 study and therefore comes with the same performance measurement caveats. In addition, the 1998 study measured learner performance across time in two different sessions twelve weeks apart. Individual learner performance was mostly inconsistent across the two sessions. Only 31% of the L2 learners obeyed the OPC in both sessions. On the one hand, the evidence from these nine learners supports the argument in favor of UG-directed learning.<sup>74</sup> On the other hand, the fact that almost 70% of the learners were inconsistent from one session to the next brings up the possibilities that the task results are not completely reliable, or the path of learner acquisition is not monotonically increasing.

Kanno suggests (1998, p.1139) that "One possibility worth exploring is that consistency in the use of UG principles "improves" with increased exposure to sentences in which they are unambiguously attested." In other words, a UG principle may be unstable until it is well-established by exposure to a sufficient amount of relevant input.

---

<sup>74</sup> Again, the argument rests on having a solid POS argument for the Constraint being studied.

This argument is more in the spirit of an input-driven argument than a UG-driven argument. Recall the characteristics of the two arguments. As seen in Chapter 1 above, input-driven learning is compatible with a period of inconsistent performance during which learners are sorting through rich and complex input. Input-driven learning is also compatible with the idea of being sensitive to the quantity of input received. On the other hand, UG-driven learning is supposed to be less sensitive to the quantity of input than to its quality. And, input should act as a trigger; once the trigger is discovered, the constraint is acquired. Kanno's proposal for an unstable constraint that stabilizes based on increased input is not consistent with this picture. An alternate proposal that is consistent with both the generative predictions and Kanno's data is that learners are unable to perform perfectly despite having acquired the underlying competence.

Pérez-Leroux and Glass's results also appear strong. Testing three different proficiency levels, P-L&G find no significant difference between learners and native speakers at any proficiency level. All learners distinguish between null and overt pronouns with respect to the availability of matrix quantified antecedents. The percentage use of embedded null pronouns increases with increasing proficiency in both bound variable contexts and referential contexts. However, all proficiency groups used null pronouns more often than overt in bound variable contexts, and less often than overt in referential contexts. There is one exception: The Advanced group used more null pronouns than overt pronouns in referential contexts. This anomaly is unexplained.

P-L&G's task does not facilitate the same possibilities for misrepresentation of learner competence that we saw in Kanno's task, in that learners were offered one interpretive choice at a time. And although proficiency levels were not independently

assessed, the criteria that P-L&G used to group their subjects can be taken to reliably establish a relative proficiency level across the groups, with the Advanced group having higher proficiency than the Intermediate group, etc. In this sense, developmental trends can be discussed even though it cannot be established that P-L&G's Elementary learners were of comparable proficiency to Kanno's Elementary learners. A different caveat applies to the P-L&G results, however. Because the task was focused on one target construction with no distractor items, and because subjects were given lots of time to complete their translation task, it is more likely that the subjects engaged general cognitive and test-taking strategies instead of or in addition to their linguistic competence. Another caveat to keep in mind is that P-L&G do not report individual results, only group results. From a theoretical perspective, and given Kanno's 1998 findings, individual results should be investigated.

On the basis of a POS argument, P-L&G conclude that their empirical results support UG-driven learning. However, their POS argument is not strong. Their results still provide empirical evidence for a UG account of L2A, in three ways. First, learners obeyed the OPC at early proficiency and continued with consistent performance as proficiency increased. These results are compatible with an input-as-trigger account, where quality of input matters more than quantity. Second, P-L&G tested a different language pair than Kanno did. To the extent that results from studies of two different language pairs both support a UG account, the account is strengthened further. Third, P-L&G tested multiple populations with different proficiencies. Their evidence thus supports a UG account on the basis of the demonstrated developmental path. Although P-L&G seem to present good evidence and logical arguments for UG-driven learning, we

must remember that Kanno showed the importance of verifying individual results and of testing subject populations at multiple intervals. Since neither of these two things were done by P-L&G, we must reserve full endorsement of their study.

Hopp (2005) reports results with respect to two different sets of criteria. First, he reports relative contrasts, that is, comparisons between learner judgments on licit versus illicit sentence types. Recall that the predicted licit sentence types are intact scrambling, and remnant movement when the two movements are of two different types: topicalization over scrambling. The predicted illicit sentence types are those where remnant movement involves the same type of movement twice: scrambling over scrambling. Second, Hopp reports between-group results, that is, comparison of judgments between learner groups as well as between learner groups and the native speaker group. We will consider each in turn.

Hopp presents the relative contrast results as being uniform and definitive. Acquisition of the predicted relative contrasts is robust across L1s and across proficiencies, in the following sense: Hopp verifies that almost every individual learner distinguishes grammatical from ungrammatical sentences by a margin of at least 30%.<sup>75</sup> No learner accepts ungrammatical sentences at a higher rate than grammatical sentences. Such uniformity of results as presented does argue rather definitively for UG-driven learning rather than input-driven learning.

However, a few more pieces of information could affect the perception of the results as uniform. For example, the rates at which learners distinguish the licit from the illicit sentence types range from 10% to 100%. Hopp supports his argument for UG on

---

<sup>75</sup> Hopp adopts the 30% criterion from Dekydsprotter et al. (1997) as an arbitrary threshold for a reliable distinction rate.

the basis of unidirectional learner performance. The argument for UG is strongest when learners demonstrate consistent native-like performance after some triggering proficiency level. But there is a wide range of learner performance in this case. Therefore, there is room for an input-driven argument if one could correlate increased distinction rates with increased learner proficiency (on the assumption that higher proficiency learners have received more L2 input). Unfortunately we are not in a position to assess this correlation, as Hopp does not identify the L1s or proficiencies of the learners at each distinction rate. For instance, we do not know if the approximately 10% of learners who distinguished licit from illicit sentences at 20% or less were all Intermediate learners, or all English L1s or all Japanese L1s, etc. The aggregate data that Hopp presents could mask underlying patterns, and some potential underlying patterns could indicate a role for input-driven learning or for L1 effects. Considering the range of learner distinction rates, the results cannot necessarily be considered completely uniform nor can the UG argument, although strong, be considered definitive.

The second set of criteria Hopp used to report learner performance involves comparisons between subject groups. In this case, the applicability of results to UG arguments can be less clear. In the first place, native speaker performance was somewhat off of predictions. In the infinitival paradigm, remnant topicalization was accepted at 59.6% and long remnant topicalization was accepted at 42.3% although both sentence types were predicted to be grammatical. Remnant scrambling over scrambling, predicted to be ungrammatical, was accepted at 24.4%. In the DP paradigm, intact scrambling is only accepted at 43.6%, and remnant topicalization at 62.3%, although long remnant topicalization was accepted at a higher absolute level of 75.6%. These unexpected results

present challenges in assessing the comparisons between learner performance and native performance in the study.

Hopp argues that native judgments on optional constructions are unstable due to a conscious or unconscious preference ranking in comparison with other alternatives. For this particular target construction (Hopp 2005, p.48), "...rejection of a given linearization need not reflect its ungrammaticality *per se*, but rather its decreased acceptability compared to a less marked variant." Hopp's statistical tests report that learner group performance is comparable to native speaker performance on the infinitival paradigm at least (all learner groups, all proficiencies, with the exception of one sentence type for English learners; see Hopp 2005, p.52-54). However, the UG argument from these comparisons is not straightforward when based on 'unstable' native speaker judgments. It is not clear which of the following possibilities could have occurred: first, learners could have acquired the whole inventory of word order options, and assigned the same rather low level of acceptability to the test items based on the same relative ranking that native speakers assign; or second, learners could simply be confused or uncertain; or third, UG may not be active in L2A. Sheen's (2000) point that learners and natives may arrive at similar judgment rates for different reasons may be relevant here, as it was for Kanno's experiments, exacerbated by the idea of gradient or unstable native judgments. The general issue of native performance as a benchmark for L2 performance was raised in Chapter 2 above and will be considered further in Chapter 6 below.

Returning to the results for the DP paradigm, we have already seen that NS acceptance of intact scrambling was unexpectedly low. Hopp also reports (2005, p.54, 56, 60) that L1 English learners (taken together) and L1 Japanese learners (also taken

together) differ significantly from native speakers and from each other on intact scrambling in the DP paradigm. L1 Japanese learners accept intact scrambling significantly more than L1 English learners, who accept the same significantly more than the NS group does. From a certain point of view, the learner groups perform in a more native-like manner than the native speakers, based on the original predictions.

Hopp (2005, p.66-67) suggests an additional rationale for these results (beyond the instability of NS judgments mentioned above). He cites Abraham (1986), suggesting that scrambling of DPs in German is subject to a definiteness constraint whereby scrambling of indefinites is marginal. (Definiteness was not discussed as a factor in the original study rationale.) Since the DP paradigm utilized an indefinite DP, a definiteness constraint would indeed help account for the ambivalence of native speakers to the licit intact scrambling test sentence. Such a constraint is also useful in accounting for the higher acceptance rate among L1 Japanese learners, whose L1 does not grammatically encode definiteness. On the other hand, L1 English learners already have a definiteness distinction in the L1. It is not clear, and Hopp does not address, why L1 English learners would accept scrambling at higher rates than native speakers do, when the learners should already be sensitive to definiteness based on their L1.

So we see that the interpretation of learners' results has become a little confounded by the initially incomplete account of what must be learned. Without consideration of the definiteness constraint, Hopp argues that the higher acceptance of scrambling by L1 Japanese learners compared to L1 English learners is due to L1 effects in that English learners need to activate a scrambling feature that Japanese learners

already have active in their L1. This is a reasonable argument in support of a role for L1 transfer effects rather than input-driven learning.

Hopp's conclusions based on his empirical results are generally strong, especially with respect to the arguments based on learners' acquisition of relative contrasts. However, his study raises interesting issues in the assessment of acquisition, particularly the issue of use of native speaker performance as a benchmark.

The final generative study under consideration is that by Marsden, concerning scope phenomena. Like Hopp, Marsden's research questions concern both the role of UG and the role of L1 transfer, testing learners from typologically different L1s (English and Korean) learning Japanese, which is similar to Korean and different from English on the target construction. We will concentrate here on her results as with respect to the possible role of UG.

Unlike in Hopp's study, Marsden's native speaker groups performed as predicted. By a loose standard, both the intermediate and advanced proficiency L1 English learner groups also performed according to the research hypotheses. Readings that were predicted to be available were accepted at least 50% of the time; readings that were predicted to be unavailable were rejected at least 50% of the time. Although both groups were on the 'correct' side of predictions, Marsden instead characterized the group performance as too near chance to provide strong support for UG-driven learning. For instance, L1 English advanced learners accepted an object-wide scope reading for canonical order Japanese sentences 43% of the time. For Marsden, this is not a strong

rejection, and does not demonstrate that learners have overcome L1 transfer to arrive at acquisition of a POS construction.<sup>76</sup>

Marsden examined individual learner results and discovered that about half of the learners in the L1 English intermediate group were inconsistent in their judgments on the crucial POS target sentence. The approximately half who were consistent in their judgments were consistent in incorrectly accepting an O > S reading for canonical word order Japanese sentences. On the other hand, Marsden found that of the L1 English advanced group, only one learner out of twelve was inconsistent. Of the eleven learners whose judgments about this sentence type were consistent, five incorrectly accepted the O > S reading, and six correctly rejected it. The proportion of subjects whose performance is native-like increases with increasing proficiency. Based on this trend, Marsden concluded that her data do support UG-driven learning after all. But it is interesting that such a high proportion of the advanced learners are consistently non-native-like.

Marsden's findings are not as robust as those in other experiments. Her best evidence for a role for UG in adult 2LA is the existence of a few advanced proficiency learners whose performance appears target-like. However, her data do not show the predicted developmental path of UG-driven learning.

In the rightmost column, Table 5.8 gives the adequacy ratings for the generative studies' results and conclusions as based on the above discussion.

---

<sup>76</sup> In this matter, she is consistent with Hopp (2005) in interpreting absolute results. Recall that Hopp construed an acceptance rate as low as 43% as evidence of (native speaker) acceptance of a certain sentence type.

**Table 5.8: Adequacy Assessment, Generative Studies—Theoretical Factors**

	<b>Research Questions</b>	<b>Rationale</b>	<b>Results and Conclusions</b>
<b>Kanno 1997</b>	✓	?	✓
<b>Kanno 1998</b>	✓	?	?
<b>Pérez-Leroux &amp; Glass 1999</b>	✓	?	✓
<b>Hopp 2005</b>	✓	?	✓
<b>Marsden 2009</b>	✓	✓	?

✓ = adequate; ? = caution required; x = inadequate

Evaluation of the emergentist studies on the same three criteria will round out this micro-evaluation chapter.

### **5.3.2 Emergentist Studies**

#### ***5.3.2.1 Emergentist Studies: Research Questions***

The research questions across the emergentist studies lack the uniformity that was found in the generative studies. In part, the variety reflects the lack of consensus within the emergentist literature, which in turn is due in part to the fact that the body of empirical investigation of second language acquisition within an emergentist framework is relatively small and relatively new.

Sheen (2000) conducts experiments specifically aimed at rebutting Kanno's (1998) Overt Pronoun Constraint study, and the research questions reflect that specificity. Of the three experiments discussed in Chapter 4 above, the first investigates the reasons behind L2 learner judgments. The second experiment asks if L2 learners whose L1 is typologically similar to the L2 demonstrate knowledge of the L2 target construction earlier than learners whose L1 is typologically dissimilar. The third seeks to verify native

speaker intuitions. These specific questions all contribute to the broader question of the role of UG in 2LA, as will be discussed in the remarks on study rationales below. The rest of the emergentist studies all compare UG accounts for 2LA to alternate, domain-general accounts. Williams and Kuribara (2008) (henceforth W&K) use two separate experiments. One asks if UG plays an active role in 2LA. The other asks if linguistic input, and statistical calculations on that input, play an active role in 2LA. The O'Grady studies test the Amelioration Hypothesis, that is, the hypothesis that a processor that is attempting to be maximally efficient is responsible for linguistic behavior that only appears to be rule-governed. All of these studies are adequate with respect to their broad research questions.

#### ***5.3.2.2 Emergentist Studies: Rationales***

The adequacy of the rationales behind the research questions varies. The rationales for Sheen's three experiments are well-designed for his purpose, which is to challenge the evidence that Kanno presents in arguing for a role for UG in 2LA. Recall that Kanno used a grammaticality judgment (GJ) task to investigate L1 English acquisition of L2 Japanese, where English and Japanese differ typologically, the former being a non-null subject language, the latter being a null subject language. The rationale for Sheen's first experiment is to challenge the validity of the GJ task results. If L2 learners arrive at their judgments in a different way than native speakers do, then learner performance similar to native speaker performance on the GJ task does not necessarily reflect native-like acquisition, and such performance might not necessarily qualify as evidence for UG. In particular, if L2 learners make conscious use of instructed rules, then it is arguable whether or not any acquisition has even taken place. The rationale for Sheen's second

experiment is to challenge the UG account of Kanno's L2 learner performance. If L2 learners whose L1 instantiates the OPC (i.e. Korean) are more successful than Kanno's English L1 learners at similar proficiency levels, then the UG account is supported. If the Korean L1 learners are the same or worse, then the UG account is weakened, because it does not appear that the Korean learners have differential, advantaged access to the supposed UG principle. The rationale for Sheen's third experiment is to challenge Kanno's argument that L1 transfer did not play a role in her subjects' performance. If native speakers of English can be shown to have similar intuitions to Japanese native speakers, then L1 transfer may account for native-like L2 learner performance. The rationales for Sheen's experiments are logical.

The rationales for the two experiments by W&K are more problematic. We will take each experiment in turn. In the first experiment, adult L2 learners whose L1 does not license scrambling are tested on scrambling in a semi-artificial language. The rationale for this experimental design as a means to investigate the role of UG in 2LA has several steps. First, W&K assume that child L1 acquisition is UG-driven. More specifically, W&K posit that the simultaneous acquisition of both canonical and scrambled word order by child L1 learners of Japanese is explained as the acquisition of a head-direction parameter. They then argue that if adult L2 acquisition is UG-driven, then the developmental path of adult acquisition will demonstrate the same 'clustering effect' seen in the child's developmental path. In other words, adults will acquire canonical word order and scrambling at the same time. If on the other hand adults do not evidence this clustering effect, then their acquisition is not an L1-like process of parameter (re)setting, but rather 'fundamentally different' from child L1A, and not UG-driven. The clustering

approach provides a nice empirically testable hypothesis, and has been used in a number of studies (for overviews, see Herschensohn 2000 or White 2003a).

There are a few issues with these assumptions and this chain of reasoning. First, there is little evidence in the L2 acquisition literature in general for clustering effects, whereby phenomena which are underlyingly related to each other are acquired simultaneously.<sup>77</sup> It is somewhat problematic to depend on finding evidence of clustering in this particular instance when it has been difficult to find evidence of clustering in general. Finally, the reasoning does not acknowledge the known difference between child L1 learners and adult L2 learners, that is, that adults have an L1 instantiated that can alter the course of adult acquisition compared to child acquisition. L1 transfer effects, unlike clustering effects, have been well-documented and should be expected. This is especially true for beginning learners, and W&K are studying absolute beginners whose exposure to the semi-artificial language is fairly limited prior to their taking the GJ test. The rationale for this first experiment is not clearly adequate. On the other hand, if one ignores the problematic theory-dependent aspects of the rationale, the basic methodology of testing learners on constructions to which they have had little exposure is standard.

In the second experiment, a computer simulation model is given the same input as the human subjects in the first experiment were given (albeit a quite larger amount) and then tested on the same sentences. The rationale for this experiment is again based on comparison between two learner ‘populations’: in this case, human and machine. Given that the computer simulation (a connectionist model) is clearly input-driven, then if the human performance and the computer simulation performance on the same GJ test are

---

<sup>77</sup> For a classic study of the Null Subject Parameter cluster, see White (1985). For a study of the Verb Movement Parameter, see Ayoun (1999).

correlated, the human acquisition process may also be input-driven. Keep in mind that the human population under study is one of adult L2 learners.

Although the given reasoning is not necessarily problematic on its own, the particular implementation in this particular experiment has issues. First, W&K set up their computer simulation with the initial weights on the connections between nodes set to random values. A random-value initial state for learning is more representative of child L1 learning than of adult L2 learning. As mentioned in the discussion of Sheen above, adult learners will have an L1 instantiated. From a computer simulation standpoint, more plausibly, the initial weights on the connections in the model should reflect what the L1 weights would be. Second, the training phase for the computer was conducted differently than for the people in two important ways: the amount of input, and the structure of the input. The computer model was given 50 cycles of the 194 sentences that the people were trained on. An additional experiment to simulate high proficiency learners gave the computer 5,000 cycles of input. These numbers are unrealistically high. The sentences were presented to the computer in random order, whereas the sentences presented to the people were structured in blocks of increasing complexity. A third design detail that presents problems is that the GJ test given to the people and the one given to the computer were different. The people saw a stick-figure depiction of the meaning of each test sentence before being asked to judge its grammaticality. The computer was simply given a string of items, coded as Subject or Object or Indirect Object or Verb. The computer did not have access to the meaning of the sentence it was evaluating and thus lacked clues as to the likelihood of the sentence being simple or complex. For all of these

reasons, the computer simulation and the human experiment are not necessarily sufficiently comparable to form a basis for the study conclusions.

The final emergentist study rationale to consider is that presented in the studies by O’Grady, Lee and Kwak (2009) (hereafter OLK) and O’Grady (2013a). Since both studies share the same assumptions and predictions, the remarks in this section apply to both, and the adequacy ratings will be the same for both. Recall that these studies propose and test a processing account for native speaker and L2 learner performance. (Section 4.3.2 in Chapter 4 above gives the details of the study rationale.) The rationale depends crucially on the processing account as proposed, and on the predictions for the developmental paths of adult L2 learners that the authors ascribe to their proposed processor. The theoretical rationale suffers from several difficulties. First, the processing account in general can be criticized. Second, the particular processing account as applied to English by OLK and O’Grady is problematic.

O’Grady’s Amelioration Hypothesis does not offer a detailed description of the human processor as applied to language; however, the principle of Unidirectionality (O’Grady 2103, p.13) provides some fundamental guidance for efficient processing, as follows: interpret each item<sup>78</sup> as soon as possible, and avoid making revisions. One difficulty with this general description is that under Unidirectionality, word order becomes determinative of interpretation. As O’Grady describes it, when for instance a universally quantified NP is encountered, the NP should receive a full-set interpretation, with no revisions as a result of subsequently-encountered elements. However, Crain (2013) points out that in some cases the same surface word order can yield two different

---

<sup>78</sup> by ‘item’, O’Grady seems to mean ‘constituent’ sometimes, and ‘word’ sometimes. For instance, the discussion of scope phenomena concentrates on the interpretation of NPs and negation.

preferred interpretations cross-linguistically. Consider the following examples from English and Mandarin Chinese, taken from Crain (2013, p.302), both of which contain negation prior to disjunction:

(3) Ted didn't order pasta or sushi.

(4) (Wǒ cāi) Tàidé méiyǒu diǎn yìdàlimiànshì huòzhě shòusī.

(I guess) Ted not order pasta or sushi

'It's either pasta or sushi that Ted did not order.'

The English example (3) has the reading that Ted did not order pasta and Ted did not order sushi. In other words, (3) has the surface order scope interpretation NEG > OR. However, the Mandarin example (4) carries the inverse scope interpretation OR > NEG. That is, (4) has the reading that either Ted didn't order pasta, or Ted didn't order sushi. Such differences in possible interpretations are problematic for the Unidirectionality principle.<sup>79</sup>

Another difficulty with the general idea of a universal, uniform processor (regardless of the specific principles that are involved) is that recent research reveals individual differences in processing, both quantitative and qualitative. Tanner (2013, p.351) cites a number of studies that indicate that individual differences in working memory capacity correlate with quantitative differences in processing efficiency. Individuals with higher working memory capacity have been shown to process lexical and syntactic information faster than individuals with lower working memory capacity (see for example King and Just 1991). In addition to quantitative differences, qualitative

---

<sup>79</sup> O'Grady (2013b, p.378) argues that disjunction works differently from quantification with respect to negation, and that his account covers these examples. However, since it is still the case that there are multiple possible interpretations and two different surface orders, it is not clear to me why revisions would not be required when the surface order and the interpretation do not match.

differences have been observed both between individuals (see for instance Nakano, Saron and Schwab 2010) and within individuals across time (see for instance Osterhout et al. 2006). Testing subject-verb agreement, Osterhout et al. found that beginning learners responded to anomalies with N400 effects, however, by the end of a year, learners' brains responded to subject-verb violations by displaying P600 effects.<sup>80</sup> If the human language processor is not uniform across individuals or across time, O'Grady's processing proposal is an oversimplification.

Turning to O'Grady's more specific proposals for the particular scope phenomena in the studies under consideration here, we find that his processing proposal may not account for English native speaker behavior very well. O'Grady proposes that for English sentences such as *Mike didn't eat all the cookies*, the occurrence of negation early in the sentence allows the processor to create a partitioned-set interpretation ('Mike ate some of the cookies') for the universally quantified direct object NP on first encounter, as opposed to a full-set interpretation ('Mike didn't eat any of the cookies'). Thus the preferred partitioned-set interpretation comes about without any need to revise an initial full-set interpretation; the partitioned-set interpretation is no more costly than a full-set interpretation would be. But as pointed out even in OLK (p.81, fn 6), allowing the partitioned-set interpretation does not explain why that interpretation is *preferred*. OLK and O'Grady both mention that pragmatic factors might be involved. If this is so, then the processing account should not be solely responsible for predictions for learner behavior.

---

<sup>80</sup> N400 and P600 are measurements of electrical activity of the brain, taken via a set of electrodes on the scalp. An N400 is an event in the negative direction that occurs 400 milliseconds following a stimulus. N400 is associated with semantic processing (Kutas and Hillyard 1980; Kutas and Federmeir 2000). P600 is an event in the positive direction that occurs 600 milliseconds following a stimulus. P600 is associated with syntactic violations (Osterhout and Holcomb 1992; Kaan, Harris, Gibson and Holcomb 2000).

Moreover, there are exceptions to the preference for the partitioned-set interpretation, as pointed out to O’Grady and colleagues by the following example offered by K. Gregg, and given in OLK (p.81):

(5) Max didn’t consider all the people who would be inconvenienced by his decision. A full-set interpretation is preferred for example (5). Here we have a within-language example of Crain’s critique of O’Grady’s proposal, where the same surface word order yields two different interpretations on different occasions. Such differences should not be possible with a uniform, consistently optimal processor.

Table 5.9 below summarizes the adequacy assessment for the emergentist studies in the categories of Research questions and Rationale.

**Table 5.9: Adequacy Assessment, Emergentist Studies—Theoretical Factors, Part 1**

	<b>Research Questions</b>	<b>Rationale</b>
<b>Sheen 2000</b>	✓	✓
<b>Williams &amp; Kuribara 2008</b>		
<b>Human L2</b>	✓	?
<b>Computer</b>	✓	x
<b>O’Grady, Lee &amp; Kwak 2009</b>	✓	?
<b>O’Grady 2013a</b>	✓	?

✓ = adequate; ? = caution required; x = inadequate

### **5.3.2.3 Emergentist studies: Results and conclusions**

We now turn to consideration of the results and conclusions of the emergentist studies.

Sheen’s three experiments were designed to cast doubt on Kanno’s results, and were for the most part successful in making their points. In the experiment that explored the

reasons behind L2 learner judgments, L2 learners did indicate that they often were consciously obeying a rule. In other words, although their performance may have appeared target-like, it may not have been native-like, as native speakers do not consciously access rules.

In the experiment that explored learner performance when the L1 was supposed to be typologically similar to the L2, Korean L1 learners of L2 Japanese performed worse than Kanno's English L1 learners did. Sheen concluded that Korean may not instantiate the Overt Pronoun Constraint, thus challenging the OPC as a linguistic universal and a candidate for UG.

In the experiment that explored native English speaker intuitions about pronoun interpretation, Sheen reported that NS English preferred an extra-sentential antecedent for embedded overt pronouns when the matrix subject is quantified, and preferred an intra-sentential antecedent when the matrix subject is referential. He concluded that native English speaker intuitions align with Japanese intuitions, in which case Kanno's learners could have simply been demonstrating L1 transfer rather than L2 acquisition. Marsden (2002, p.72) points out that in reporting his results, Sheen counted a learner's response of 'either intra-sentential or extra-sentential' as expressing a preference for one or the other, when in fact no such preference can be inferred. A learner who answers that both antecedents are available has no means in this task design to indicate whether one or the other reading is preferred or marginal. It is incorrect of Sheen to generalize to a preference for one particular antecedent on the basis of such an answer.

Sheen (2000) proposed an alternate, non-UG account of learner performance. Sheen proposed (p.803) a '*kare* rule' as follows: the antecedent of the overt pronoun *kare*

is always extra-sentential. Later in his paper he expands the rule to include the converse (p.812): "...that rule being that the null pronoun indicates coreference [with the intra-sentential antecedent – LP] and the overt pronoun indicates contra-reference [someone other than the matrix subject – LP]". Sheen notes that pedagogical materials have many examples of monoclausal sentences with *kare*, and in all cases *kare* takes an extra-sentential antecedent. He proposes that L2 learners generalize from monoclausal examples to bi-clausal occurrences of *kare*. Marsden (2002) notes that Sheen's rule does not account for the ability of embedded overt subject pronouns to take referential matrix subject antecedents. And she further notes that in monoclausal sentences, pronouns must have extra-sentential antecedents, making the *kare* rule redundant with broader principles of binding. On these bases and others (for more details, see Marsden 2002), Sheen's non-UG account of the OPC does not hold up.

Turning to the next emergentist study, Williams and Kuribara (W&K) (2008) maintain that the evidence from their two experiments supports that adult L2A is input-driven. I will argue instead that their evidence is inconsistent with input-driven learning, but could support UG-driven learning. Consider the experiment with human subjects, some of whom received exposure to Japlish input prior to the grammaticality judgment test and some of whom did not. Recall from Chapter 4 that because of the amount of variability in individual performance, W&K restricted their attention to those individuals who were successful (75% acceptance rate or greater) on grammatical Japlish short scrambled sentences. W&K characterized the performance of the exposure group 'scramblers' as target-like relative to that of the no-exposure group scramblers, because the exposure subgroup accepted grammatical Japlish sentences at significantly higher

rates than the no-exposure subgroup and accepted two types of ungrammatical sentences at significantly lower rates than the no-exposure subgroup did. On this basis, W&K argued that input made a difference, and input-driven learning is supported. However, W&K did not characterize the exposure subgroup performance as target-like in an absolute sense, because the exposure subgroup accepted English canonical word order (ungrammatical in Japlish) at chance, and accepted ungrammatical complex scrambled sentences at rates not significantly different from the no-exposure subgroup. In other words, learners did not show native-like performance on the entire cluster of behaviors associated with resetting a head parameter. On this basis, W&K argued that UG-driven learning is not evidenced and therefore not supported.

W&K's arguments rest on the assumption that clustering effects as well as native-like performance are necessary for demonstrating acquisition. If one relaxes these standards then one comes to different conclusions on the basis of the same empirical data. It is quite reasonable to relax the clustering requirement. The discussion of the rationale for this experiment (see p.220 above) has already highlighted that there is evidence against simultaneous acquisition even in L1A, much less L2A. It is also reasonable to relax the native-like performance requirement. Recall, for instance, Hopp's (2005) reliance on relative contrasts between judgments on grammatical and ungrammatical test items as the basis for his conclusions, rather than comparisons to native-speaker results.

Reinterpreting the data from W&K's human subjects experiment, UG-driven learning receives support from two observed outcomes. First, very early learners went beyond the controlled input they received, accepting scrambled sentence types to which they had not been exposed. Notice that this fact may also provide evidence against input-

driven learning. Second, these same very early learners are trending away from L1-like performance by accepting English canonical word order and one type of rightward movement only at chance, rather than robustly. From this point of view, these data provide support for UG-driven learning through classic POS argumentation. W&K's argument that input affected performance does not contradict a UG account. Even under a UG account, input is required so that learning can be triggered. The question of why this certain amount of input triggered acquisition in some subjects but not others is beyond the scope of this dissertation. However, in the absence of the investigation of multiple proficiency levels of learners, one cannot distinguish between input as trigger and input as determinative. We return to this point in the discussion of W&K's computer simulation experiment, next.

W&K found a positive correlation between the performance of their computer simulation model and the performance of their human learners on the same grammaticality judgment (GJ) test.<sup>81</sup> W&K argue that similar performance indicates similar acquisition mechanisms. The computer model was designed so that all of its learning resulted from exposure to input, with no initial constraints. Therefore, human learners must have relied on input without recourse to innate constraints, just as the model did.

In order for this logic to hold, the human learners and the computer simulation model must be fundamentally similar in relevant ways, not just in performance. However,

---

<sup>81</sup> The question of whether or not the model's average output strength on a given sentence type is a good comparison to human learner acceptance rates on that same sentence type is beyond the scope of this work. I also put aside for now questions about why W&K chose to present their correspondence data in two pieces: comparing model output to the exposure group 'scrambler' acceptance rates, and separately comparing model output to exposure group 'non-scrambler' acceptance rates, that is, those subjects who did not accept even simple scrambling at very high rates. Both subject groups show similarities to the computer simulation output.

the human subjects and the model were fundamentally different. The model was not comparable to adult L2 learners in general in that its initial state was random, whereas adult learners have an L1 instantiated. The model was not comparable to these particular adult L2 human learners in that the structure and amount of its exposure to Japlish differed significantly from the human experience. Also, the GJ test differed in that human learners had access to the meaning of the sentence whose grammaticality they were judging, whereas the model did not.<sup>82</sup> Given so many important differences in initial state, exposure, and testing, it is difficult to grant to W&K that similarities in acquisition processes can be inferred from similarities in performance. As a result, the model simulation cannot support either input-driven learning or UG-driven learning in adult L2 learners. And by their own admission, W&K's argument for input-driven learning based on their computer simulation results breaks down when they consider the fact that the computer performance did not improve with increased input and training. W&K suggest that there may be a role for rule-based learning, but that the 'rules' do not have to be language specific.<sup>83</sup>

There is a way by which W&K could have used their model results to support an input-driven account of learning. If they were to assume that L1A is input-driven, with no language-specific constraints at work, then the model in its initial state would be more comparable to child L1 learners than to adult L2 learners. W&K could have argued that the L2 learners perform similarly to the computer model, and the model is similar to L1 learners, therefore L2 learners are similar to L1 learners (both populations relying on input rather than UG). But this is not the argument they make. In fact, whereas other

---

<sup>82</sup> This procedural difference is described on page 222, above.

<sup>83</sup> O'Grady would surely agree.

emergentist approaches assume that neither L1 nor L2 are UG-driven, W&K begin their article with the assumption that child L1A is UG-driven.

In summary, the W&K (2008) experiments do not provide strong support for input-driven L2A, but in fact could be construed to support a UG account. The discussion highlights the issue of how to determine if acquisition has occurred. We will return to this discussion in Chapter 6.

We conclude Section 5.2 with a discussion of the results and conclusions from the studies by O'Grady (2013a) and O'Grady, Lee and Kwak (2009) (henceforth OLK). I will use 'O'Grady and colleagues' to refer to both studies together. The crucial assumption for these particular studies concerns the relative processing cost of two different scope interpretations, where processing costs are assumed to be higher when initial, immediate interpretations need to be revised. The critical predictions of these studies testing O'Grady's processing account of acquisition concern the developmental path of learners whose L1 and L2 are typologically different.

The results of both studies aligned nicely with predictions: Korean L1 learners of L2 English displayed L1-like behavior, preferring full set interpretations (wide-scope for the quantifier over negation, *all > not*); English L1 learners of L2 Korean accepted full set interpretations of Korean test sentences fully 100% of the time, despite the fact that English native speakers prefer partitioned set (*not > all*) interpretations for comparable English sentences. O'Grady and colleagues' conclusions aligned nicely with their results: they argued that L1 Korean learners initially transfer their dominant processing routine to the English L2, because the full set interpretation is relatively low cost in English as well as Korean. O'Grady and colleagues further argue that L1 English learners do not transfer

their dominant partitioned set processing routine to the Korean L2, because the *not > all* interpretation is relatively high cost in Korean. “Cost blocks transfer”, is the slogan. The developmental path of acquisition is accounted for by reference to constraints imposed by a processor seeking to maximize its efficiency as it encounters sentence elements in linear order, and notably not by reference to abstract underlying structure, in particular, anything like c-command. All that being said, note that the L1 English-L2 Korean learner performance cannot be characterized as L2-like, because the OLK experiment showed native speakers of Korean to accept a partitioned set interpretation on 21% of the test sentences.

There are a few issues to consider with respect to the processing account. One concern was touched on in the Rationale discussion above. O’Grady’s processing proposal relies completely on surface linear order of elements in the sentences. In English, an SVO language, negation precedes the quantified object. Recall that O’Grady holds that this precedence relation allows both scope interpretation possibilities to arise at comparable cost in English. In contrast, Korean is an SOV language, with the quantified object preceding negation. O’Grady contends that in Korean only the full set interpretation arises when *all* is encountered before negation; the partitioned set interpretation would involve revision and is therefore significantly more costly. This account is certainly an oversimplification. For one thing, how does a language user know that a preceding negation in English allows multiple interpretations to arise for an object? Or from another point of view, especially given that *all* entails *some*, what blocks the partitioned set reading from arising in Korean when negation is not present?

For another thing, consider the factor of morphological case. Korean marks accusative case on objects. An alternate plausible processing account might say that when a hearer encounters an object, interpretation of the NP is delayed until it is known whether or not negation is present in the sentence. This procedure would not be maximally efficient in that it violates the goal of immediate interpretation. The point is that overt morphological case, its existence and its linear precedence relations to other elements in the sentence are not a part of O'Grady's proposal despite his emphasis on the importance of surface elements and surface order.

Another factor, by their own admission wrongly put aside, is pragmatics. Chung (2012, p.303) characterizes scope of quantification and negation as "...a linguistic phenomenon that requires integration of multiple types of information." In particular, she maintains (p.287ff) that computation of scope interpretations involves structural properties, knowledge of entailment relations, and scalar implicatures. O'Grady's proposal for this particular linguistic phenomenon does not take all these factors into account.<sup>84</sup> Yet he implicitly acknowledges the role of context in interpretation in that his account relies on extra-sentential contextual information to determine which of two available readings is chosen in English sentences. The huge and as yet unresolved question of what information is used at what points during sentence processing is crucial

---

<sup>84</sup> Interface phenomena have been argued elsewhere to present additional difficulty for L2 learners (see Sorace and Filiaci 2006; Sorace and Serratrice 2009; among others). In particular, discourse and pragmatic factors have been argued to be acquired later than non-interface phenomena. That discussion is rooted in the generative approach, relying on modularity of grammar and abstract grammatical principles, so is somewhat orthogonal to O'Grady's approach here. But despite O'Grady's different characterization of learner knowledge, he must still account for pragmatics and context.

to characterizing the relative costs of processing comparable sentences cross-linguistically.

In addition to oversimplification, a second concern centers on the proper assessment of learners' developmental path. Strictly speaking, a path should be evaluated at multiple points. O'Grady and colleagues tested one subject group in each study, and the two learner groups were not characterized as equal in proficiency to each other. The English L1 learners of L2 Korean were of low-to-intermediate proficiency, and the Korean L1 learners of L2 English were of intermediate- to high-intermediate proficiency. These are not huge concerns in that O'Grady and colleagues did not over-represent their data. They drew conclusions about early learning without speculating on ultimate attainment or the details of whether the learning path seems steady or variable. Of greater concern is the lack of precision in assessing proficiency. As mentioned in Chapter 4, very little information is given about the subjects' language history and there was no independent proficiency assessment. Lee (2009) and Chung (2012), both 2LA studies of scope phenomena as well, found that performance varied with proficiency. There are at least two factors that might have a role in producing such variation. First, as just mentioned, sources for judgments may change over time as learners learn to attend to more of the components of the interpretation. Second, and perhaps partially dependent on the first, there can be qualitative as well as quantitative differences in processing between individuals and within individuals across time. This variability along several dimensions entails that learner proficiency should be well-understood when interpreting performance results. Also, learners of varying proficiencies should not be grouped, as the group results may mask developmental patterns.

One might argue that the categorical performance of the English L1 learners of Korean obviates the need for such individual analysis. However, one could also argue that these results of the O’Grady (2013a) study seem to confirm the concerns about the task design discussed in Section 5.2.2 above. The English learners’ acceptance of the full set interpretation in Korean at 100% and more importantly, their complete rejection of the partitioned set interpretation (0% acceptance) makes it plausible that the task itself allowed the full set interpretation to overwhelm the partitioned set interpretation. (That being said, the rejection of the partitioned set interpretation does not alter the fact that the full set interpretation was robustly accepted, in support of the processing account.)

Finally, note that according to O’Grady (2013b, p.380), for L1 Korean learners of L2 English, this scope interpretation issue is not a POS phenomenon. According to O’Grady, “...anecdotal evidence and curricular materials point to an intense pedagogical effort to impose the *not > all* interpretation.” The fact that learners persisted in their non-target-like behavior despite such exposure strengthens the L1 transfer account, and does not provide support for a UG account in which some relatively smaller amount of input could be enough to trigger acquisition.

This concludes the discussion of the adequacy of the studies under consideration. Table 5.10 summarizes the adequacy assessments for theoretical issues in the emergentist studies, based on the discussions above. Note: the ‘x’ in the top row of the column for Sheen (2000) indicates that independent of his experimental results, Sheen’s non-UG account of the OPC was insufficient. (His proposed *kare* rule did not account for matrix referential antecedents for overt embedded pronouns.)

**Table 5.10: Adequacy Assessment, Emergentist Studies—Theoretical Factors**

	<b>Research Questions</b>	<b>Rationale</b>	<b>Results and Conclusions</b>
<b>Sheen 2000</b>			x
<b>Experiment 1</b>	✓	✓	✓
<b>Experiment 2</b>	✓	✓	✓
<b>Experiment 3</b>	✓	✓	?
<b>Williams &amp; Kuribara 2008</b>			
<b>Human L2</b>	✓	?	x
<b>Computer</b>	✓	x	x
<b>O’Grady, Lee &amp; Kwak 2009</b>	✓	?	?
<b>O’Grady 2013a</b>	✓	?	?

✓ = adequate; ? = caution required; x = inadequate

Table 5.11 below summarizes the adequacy assessments for all studies on all (micro) criteria.

**Table 5.11: Adequacy Assessment—Empirical and Theoretical Factors**

	Number of L2 Subjects	Proficiency: Levels, Assessment		Task Design	Research Q's	Rationale	Results & Concl.
<b>Kanno 1997</b>	✓	?	?	x	✓	?	✓
<b>Kanno 1998</b>	✓	?	?	x	✓	?	?
<b>Pérez-Leroux &amp; Glass 1999</b>	✓	✓	?	?	✓	?	✓
<b>Sheen 2000</b>							x
<b>Exp 1</b>	✓	x	x	x	✓	✓	✓
<b>Exp 2</b>	x	x	x	x	✓	✓	✓
<b>Exp 3</b>	✓	n/a	n/a	x	✓	✓	?
<b>Hopp 2005</b>	✓	✓	✓	✓	✓	?	✓
<b>Wms &amp; Kuribara 2008</b>							
<b>Human</b>	✓	?	✓	?	✓	?	x
<b>Comp.</b>	n/a	✓	n/a	?	✓	x	x
<b>Marsden 2009</b>	✓	✓	✓	✓	✓	?	?
<b>O'Grady, Lee &amp; Kwak 2008</b>	✓	?	?	✓	✓	?	?
<b>O'Grady 2013a</b>	?	?	?	x	✓	?	?

✓ = adequate; ? = caution required; x = inadequate

Chapter 6, next, will rely on these micro-level assessments to draw some macro-level conclusions about the current state of the evidence in favor of or against the generative account and the emergentist account of L2A.

## **CHAPTER 6: REVISITING POVERTY OF THE STIMULUS: DISCUSSION AND CONCLUSIONS**

### **6.1 Overview**

Following this brief overview, Section 6.2 revisits the POS argument, with particular attention to the two assumptions first examined in Chapter 2 above: successful learning and stimulus poverty. Both assumptions are difficult to verify in practice. Verification of successful learning (Section 6.2.1) is clouded by issues around native speaker performance and the comparison of learner performance to native speaker performance. Verification of stimulus poverty (Section 6.2.2) is made difficult by practical issues in input documentation as well as theoretical unknowns with respect to understanding what quality and quantity of input suffices for learning. Section 6.3 weighs generativism versus emergentism in light of the evidence and arguments presented in the studies in Chapter 4. Based on these studies, the generativist approach receives the best support overall. However, in light of the practical and theoretical issues around the POS argument, it is difficult to substantiate extreme versions of either a generative or an emergentist account. Section 6.4 reflects on the extent of the common ground between the two approaches, on the utility of the POS argument, the utility of studying complex syntactic or semantic phenomena, and the utility of studying adult L2 learners for providing evidence on the question of innateness. Section 6.5 concludes the dissertation.

### **6.2 Difficulties with Implementing the POS Argument**

Section 2.3.1 of Chapter 2 above discussed the fact that in order to apply the POS argument (for L1 or L2 acquisition), its two premises need to be satisfied. First, acquisition must have taken place. And second, the linguistic input must underdetermine the acquired phenomena. Both premises are difficult to verify. Although strictly speaking

only the generative studies seek to apply POS argumentation, data from emergentist studies will also be mentioned where relevant for the purpose of illustration.

### **6.2.1 Determining Acquisition**

Note that the title of this section does not specify what is being acquired. As a general methodological point, we need to keep in mind that judgment tests such as those used in the set of studies under consideration do not on their own demonstrate what particular grammar an L2 learner has acquired.<sup>85</sup> Judgment tests simply quantify native speaker performance and L2 learner performance. Inferring acquisition based on these data depends on the proper characterization of native speaker (NS) performance, of L2 learner performance, and the relation between the two.<sup>86</sup> One must at least show that L2 learner performance is on a path away from the learners' L1-like behavior and toward the target (NS) behavior. The studies under consideration have shown that difficulties arise in the characterization of both NS and L2 performance.

#### ***6.2.1.1 Characterizing Native Speaker Performance***

Two potential problems can occur with NS results: NS performance may not be clear-cut, and NS performance may not match theoretical predictions. Examples of each of these can be found in Table 6.1 below, which summarizes the NS performance in all studies except Williams and Kuribara (2008) (since by definition there are no native speakers of the artificial language that they constructed for their study).<sup>87</sup> The chart purposely does

---

<sup>85</sup> For examples of L2 studies whose learner results are compatible with multiple theoretical (syntactic) grammar proposals, see Schwartz and Sprouse (2000). And as we have seen, O'Grady (2005, 2013a,b) would argue that no grammar is being acquired at all.

<sup>86</sup> In Section 2.3.1.1 above I adopted a 'trend toward native-like performance' rather than attainment of native-like performance as the appropriate standard for successful acquisition.

<sup>87</sup> Strictly speaking, only the generative studies are relevant to the application of POS arguments. However, the results of all studies provide evidence on the difficulty of characterizing NS performance. Hence (almost) all studies are included in the inventory in Table 6.1. and will likewise be discussed in Section 6.2.2 concerning input.

not provide examples of the specific constructions. Abstracting away from the specific test sentences, the information of interest comprises the absolute levels of NS acceptances, and the grammaticality predictions (an asterisk indicates that the construction was predicted to be ungrammatical in the relevant language). Note that the Spanish NS data from the Overt Pronoun Constraint (OPC) study are production data, not comprehension data, and should not be compared to the Japanese NS OPC data even though both languages are predicted to pattern together. The three sets of data for English NS in the OPC studies reflects the fact that Sheen (2000) tested two NS groups (one American, one British) and Kanno (1998) tested one group (American). Discussion of entries illustrating the difficulties of characterizing NS performance follows Table 6.1.

**Table 6.1: Native Speakers—Percentage Acceptance of Test Constructions**

Construction		Native Languages		
<b>Overt Pronoun Constraint</b>		<b>English</b>	<b>Spanish+</b>	<b>Japanese</b>
	QP antecedent/null pronoun	na/na	85	83
	QP antecedent /overt pr.	70/20/85	*13.7	*2
	Referential ant./null pr.	na/na	31.3	100
	Referential ant./overt pr.	99/100/85	67.5	47
<b>Scrambling</b>		<b>German</b>		
	simple		80.5	
	remnant top over scr		59.6	
	rem. top over scr (finite clause boundary)		42.3	
	*scr over scr (over Adverb)		*12.8	
	*scr over scr (over Subject)		*24.4	
<b>Scope</b>		<b>English</b>	<b>Korean</b>	<b>Japanese</b>
	Subject-wide			
	canonical <i>every</i>	98	n/a	87.5
	scrambled <i>every</i>	n/a	n/a	80.5
	canonical <i>all</i>	99.6	77.3	90
	Object-wide			
	canonical <i>every</i>	67.5	n/a	*16
	scrambled <i>every</i>	n/a	n/a	81.5
	canonical <i>all</i>	*21.3	*20	*16.5
	Partitioned-set reading	100	*21	
	Full-set reading	67	97	

Note: Exception: + indicates production data rather than comprehension (percentage of pronouns produced in contexts favoring the matrix subject antecedent of the given type). Asterisk indicates target construction predicted to be ungrammatical.

We first note and put aside the surprising difference between the 70% and 20% OPC acceptance rates for QP antecedents for overt embedded pronouns in Sheen (2000)'s two English NS groups. We do not have enough information about the subjects or the test items to speculate on possible causes for this discrepancy. Beyond that, there are several examples where the NS performance hovers in an uncertain range, say, from 40% to 80% acceptance for a test construction that was predicted to be grammatical. For instance, for the OPC, Kanno's (1998) Japanese NS accepted referential matrix subject antecedents for

embedded overt pronouns 47% of the time. Hopp's (2005) German NS accepted simple scrambling at only 80.5%, and remnant topicalization over scrambling at only about 60%.

Some acceptance rates are low enough that the performance is difficult to construe as matching the theoretical predictions. For instance, Hopp's German NS accepted long remnant topicalization at only 42.3%. This is a theoretically grammatical construction, but not strongly accepted by NS. Other theoretically ungrammatical constructions were not strongly rejected. For instance, Hopp's German NS accepted scrambling over scrambling, a purported violation of the Unambiguous Domination Principle, almost a quarter of the time. All of the scope interpretations that were predicted to be unavailable were accepted at least 16% of the time, with some of them accepted at 20% or even slightly more.

In one case at least, less-than-definitive acceptance rates can be attributed to factors that were not accounted for in the original predictions, and therefore not controlled for in the test sentences. Hopp (2005) points out that scrambling in German is subject to definiteness effects. When unintended factors can be identified as having influenced results, test instruments can be refined and native speaker performance perhaps characterized more precisely. Otherwise, unexpected acceptance rates, particularly lower-than-expected acceptance rates, may be more difficult to clarify. Some have argued that complex phenomena are harder to process, with increased processing load accounting for more errors and uncertainty. From a generative point of view, this is a matter of performance rather than grammaticality per se that would not disqualify a weak acceptance rate as indicative of grammaticality.

Other accounts of weak acceptance or rejection rates cite things like gradience in grammars (see for instance Aarts 2007), or prototype effects (for discussion, see Schütze 1996, Sections 3.3 and 6.2). As an example, Dąbrowska (2004, p.198) describes an experiment eliciting English native speaker judgments on interrogatives with long distance dependencies (LDDs). Prototypical LDDs (*Where do you think they sent the documents?*) were rated significantly higher than unprototypical LDDs (*Where will the customers remember they sent the document?*). It is worth noting that unprototypical LDDs were rated significantly higher than ungrammatical control sentences (*\*Who do you think that left?*). These two types of accounts, that is, gradience effects and prototypicality effects, impinge on our ideas of what constitutes grammar, muddying the picture even further.

In summary, native speaker performance is not always definitive, at times being uncertain and/or unexpected relative to predictions. These facts create challenges for verifying L2 learner success, thereby creating challenges for applying a POS argument to an L2 acquisition investigation.

#### ***6.2.1.2 Characterizing L2 Learner Performance***

Even more than NS performance, L2 learner performance is prone to uncertainty and variability (see Chapter 3, Section 3.2.4 for brief discussion). In this section, we acknowledge all of the previously discussed issues with characterizing L2 judgments on their own, and focus on how to characterize L2 learner performance in relation to native speaker performance. Some such comparison between these two (possibly at issue) data sets must be made in order to evaluate learner progress or acquisition.

The studies at hand exemplify two types of argumentation used to justify a conclusion of learner acquisition. The first is comparison between NS and L2 learner performance on an absolute basis. In other words, acquisition is inferred if L2 learners reject (or accept) target sentences at rates that are not significantly different from NS acceptance rates. Kanno (1997), for instance, reports results this way, concluding that L2 learners were similar to native speaker controls. Marsden (2009) compares L2 performance across her different L1 learner groups, concluding that L1 Korean learners of L2 Japanese were different from L1 English learners of L2 Japanese. This is a stricter standard than I adopt in Chapter 2 above.

A second type of argument is from shared relative contrasts rather than absolute acceptance rates. On this approach, if native speakers demonstrate a grammaticality contrast, and L2 learners also distinguish that contrast, this is considered sufficient evidence for acquisition. This style of argument was used by Hopp (2005) (whose NS results were somewhat indeterminate). L2 learners were argued to distinguish the predicted grammatical sentences from the predicted ungrammatical ones and therefore to have acquired the relevant contrast. This is his conclusion despite the fact that the absolute acceptance rates between NS and L2 learners on certain test constructions differed by as much as 30 percentage points. L1 English High-Intermediate learners of German accepted long remnant topicalization at 71.4%, whereas native speakers accepted the same sentence type at 42.3%. The relative contrast argument is from a certain point of view a weaker standard of acquisition, and as such, is vulnerable to the criticism of overestimating learner acquisition. However, in my opinion the weaker standard is sufficient for satisfaction of the premise of the POS argument, because underlying

knowledge can be reasonably inferred from a statistically significant difference in performance (acceptance rates) between grammatical and ungrammatical sentences .

The discussion so far applies to reported results on a group level. But we have seen that group level averages can mask underlying patterns. For instance, Marsden (2009) investigates the performance of each individual in her learner groups in order to account for group average performance that hovered around chance. As discussed in Sections 4.4.1 and 5.3.1.3 above, Marsden finds subgroups of learners with performance varying by group. These results highlight the value of individual-level reporting (the same could be said for reporting results of individual test items, however, none of the studies at hand went into that much detail). The strongest evidence arises from results that are consistent across individuals and across time. On the other hand, if group-level performance averages at chance, and individual-level performance is revealed to be consistent in that regard, the data are not necessarily uninformative. For instance, if you have a baseline lower-proficiency learner group performance at higher than chance (in other words, a stronger rejection or stronger acceptance), then movement away from categorical behavior can be argued to be progress toward an opposite target. Without a baseline comparison, however, such an inference cannot be made. Performance at chance could easily be an artifact of confusion.

We have seen that the POS requirement to demonstrate acquisition is difficult to satisfy in practice. Next we consider the second task: demonstrating that input is insufficient to the evidenced acquisition.

## 6.2.2 Determining (In)Sufficiency of Input

As was the case with the acquisition assumption, there is a gap between the theoretical goal of verifying the stimulus poverty assumption and the practical ability to do so. It must be shown that the learner's environment lacks sufficient quantity of relevant input. Recall that three sources of input are available to the L2 learner: the L1; specific relevant instruction; and general L2 input, the richness of which will vary with the learners' circumstances. For instance, some learners receive input primarily from classroom instruction, and some are immersed in a foreign language environment. The L1 and instruction sources are easier to verify than the general L2 input is.

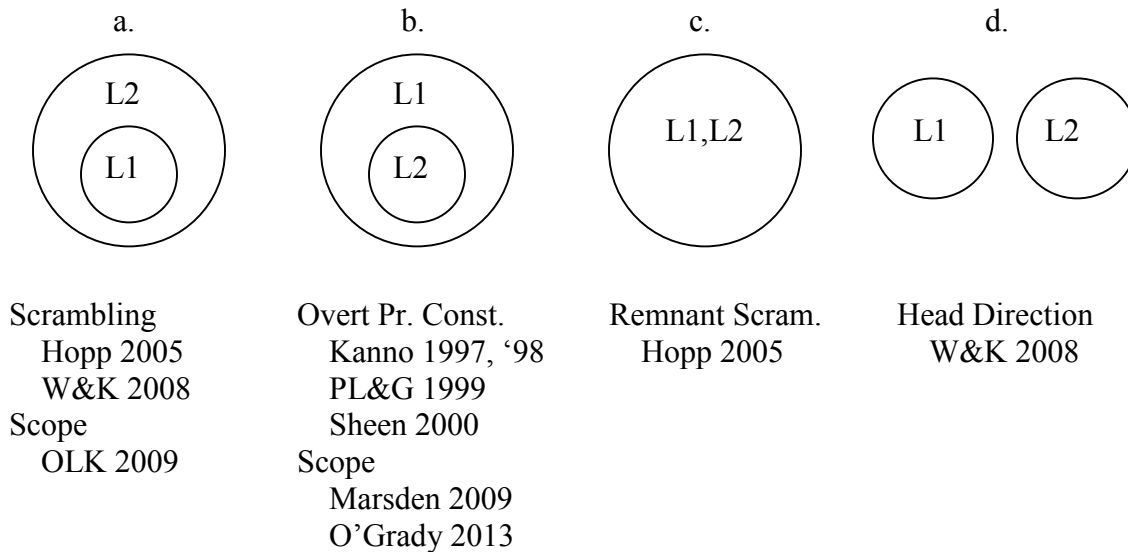
We will consider the sources of input that were available to the learners in the studies under consideration in this work, and the input poverty arguments given. First, concerning the L1 as a source of relevant input, Figure 6.1 below summarizes the relationship between the learners' L1 and the target L2 with respect to the target phenomena for all the studies. In some cases, a single study essentially investigated the acquisition of more than one phenomenon and therefore appears in multiple places in the chart. Each placement will be explained below.

For acquisition of simple scrambling (English L1 learning L2 German or L2 Japlish), and for acquisition of the partitioned set interpretation<sup>88</sup> for universally quantified direct object NPs (Korean L1 learning L2 English), the target phenomena were not instantiated in the L1. This is represented in Figure 6.1(a) as a subset-superset relationship for the L1 and L2, respectively. In this case, the L1 does not provide evidence for the target phenomenon in the form of example sentences or interpretations.

---

<sup>88</sup> That is, the reading of *Mike didn't eat all the cookies* in which Mike eats some of the cookies.

For the acquisition of the Overt Pronoun Constraint (English L1 learning L2 Japanese or Spanish) and the other scope phenomena (English L1 learning L2 Japanese and English L1 learning L2 Korean), the situation is reversed. The L1 contains grammatical sentences and interpretations that are not available in the L2s. In other words, the L1 is a source of misleading input; the learner must acquire a constraint in the L2 that does not exist in the L1. This is represented in Figure 6.1(b) as a superset-subset relationship between the L1 and the L2, respectively. For remnant scrambling, Hopp (2005) argued that the constraint on multiple movements of the same type was universally instantiated, therefore Figure 6.1(c) shows the L1 and L2 (English and German) coinciding.<sup>89</sup> For head direction, Japlish is designed to be head-final while the L1 English is head-initial, so the languages are represented in Figure 6.1(d) as not intersecting.



**Figure 6.1: Subset-Superset Relationship Between L1 and L2: All Studies**

Figure 6.1 visually summarizes the potential contribution of an L1 as a source for the L2 learner. The L1 contribution can range from completely helpful to completely

<sup>89</sup> See Section 5.3.1.2 above for a discussion of the logic of Hopp's POS argument.

deficient. In situation 6.1(c), the L1 could essentially eliminate the acquisition process, transferring the phenomenon ‘for free’. In situations 6.1(a) and 6.1(d), the L1 input is insufficient in that no positive evidence is available for the target phenomenon. In situation 6.1(b), the L1 will contain examples of sentences and interpretations that are not available in the L2, and crucially will *not* provide (positive) evidence indicating the existence of a constraint. To summarize, all situations except 6.1(c) provide support in fulfilling the assumptions of a POS argument.

The second input source for L2 learners is instruction. As a general point, it is in principle possible to assess the type and amount of instruction delivered to L2 learners, making this input source relatively verifiable. It is of course very difficult to document complete input, and impossible to guarantee that a topic that is not in the lesson plans was never covered informally. But for input to ‘count’, it must be available to all learners; occasional and individual deviations from the standard course do not constitute evidence of instructional input that is available to all learners. From that point of view, assessment of textbooks and interviews with teachers constitute sufficiently reliable information about the content of L2 instruction.

For these studies, with the exception of the artificial language study, each researcher attempted to verify that formal instruction on the target phenomenon did not take place. The investigation usually consisted of a sampling of textbooks and sometimes included personal contact with a sample of language teachers. In most cases, on the basis of this type of survey, instruction was ruled out. The exception was noted in Section 5.3.2.3, where O’Grady (2013b) found that the full set interpretation of the scope

ambiguity (For all the cookies, Mike didn't eat them) was in fact taught to English L1 learners of L2 Korean.

The third and most challenging source of input for L2 learners that needs to be assessed is the general L2 input that they receive. Williams and Kuribara (2008) was a special case in which learners were given input in the artificial L2 language that was completely controlled for quality and quantity. By design, learners were offered a certain set of sentences and then tested on an expanded set of sentences. In other words, the input was designed to be sufficient for some types of scrambling and insufficient for other more complex examples. The clarity of the sufficiency of the input argument was clouded by a set of instructions that guided the learners' attention to issues of word order and morphology, and specifically informed the learners that different languages accept different word orders.

For the rest of the studies, the participants were university students or instructors. In many cases the subjects' main L2 language experience came from the classroom. This means they primarily experienced a structured environment, with information presented in leveled increments, often emphasizing written input to a larger extent than provided in a naturalistic environment. For two studies, learner populations of advanced proficiencies had had some residency in the primary country in which the L2 language is spoken. These learners had access to an immersion environment in addition to instructed environments.

In order to verify the assumption that L2 input is insufficient to acquisition, the quantity and type of input must be assessed. Regarding quantity, only two of the studies argued that examples identical to the target phenomenon were rare in the L2 input, and

only one of these (Hopp 2005) cited corpus evidence. There were some issues with the corpus evidence. First, on investigation, the corpus statistics were not tallied on examples that were necessarily exact matches for Hopp's test sentences. Second, the evidence reported involved the relative frequencies of different constructions, not the absolute frequencies. Additionally, the relevance of naturalistic corpora as an analogy to instructed L2 input is not clear. These examples illustrate the difficulty of assessing the quantity of direct (in this case, exactly matching) L2 input available to the L2 learner.

But beyond the question of quantity of direct input that matches the target construction exactly, there is the question of what other types of input might be relevant and helpful to the acquisition of that construction. Assessing the sufficiency of L2 input cannot be done without a better understanding of what counts as relevant evidence.

Marsden (2009) (among others) argues that in situations such as in Figure 6.1(b) above, where the L2 is a subset of the L1 with respect to the target construction, the *absence* of certain sentences and interpretations in the L2 does not constitute sufficient evidence of ungrammaticality of those sentences and interpretations. This argument is disputed by those who claim that a learner's unmet expectations can serve as negative evidence.

Getting back to the question of overt evidence, some researchers proposed that inexact but related input could contribute to acquisition of certain target constructions. Recall from Chapter 2, for instance, that Pullum and Scholz (2002) argued that *when* questions can provide evidence on structural dependence in polar interrogatives. Along the same lines, Sheen (2000) argued that mono-clausal sentences provide evidence for the OPC in bi-clausal sentences. Hopp (2005) argued that the surface order of major elements in grammatical sentences could provide misleading evidence for the grammaticality of

other (ungrammatical) sentences in which the same order had been derived through illegal movement operations.

These considerations of related but not exact constructions in the assessment of the utility of the L2 input highlight an important missing piece in the discussion. There is no agreed-on independent characterization of what input is essential to a particular acquisition task, especially with respect to the complex syntactic and semantic phenomena under study in these investigations. It is an open question whether absence counts as ‘input’. It is an open question as to what extent the positive input needs to match the target input in order to be relevant. It is an open question whether linguistic input is the only relevant input (a traditionally generative point of view), or whether other sources such as lexical semantics, social cues, or stochastic analyses are also relevant (as many emergentists would argue). There are additional questions around the difference between what input is available to the learner and what input is actually taken up. These questions bring up issue of the role of proper (that is, efficacious) structure and timing of the presentation of input to the learner. There are many theoretical and practical unknowns in the specification of the essential set of input for a particular target. As Clark and Lappin (2011, p.40 ) put it: “It is simplistic, then, to condition the learning of a given linguistic generalization on a particular set of sentences, which are construed as indispensable to acquisition of the principle.” Without a characterization of what input is essential, it is especially difficult to document deficiencies.

The goal of identifying what input is essential to a given learning task is challenging, in that such a characterization can only be made rigorous when made in conjunction with a characterization of the learning mechanisms that an L2 learner

employs. As a simplistic example, if learners can only learn by imitation, then they can only learn what is presented to them exactly in the input. Learners with a larger inventory of learning mechanisms do not need to experience a particular sentence in order to produce or comprehend it.

Studies within the emergentist approach can make inroads on these questions. By necessity, computer simulations such as that done by Williams and Kuribara (2008) do make assumptions about learning mechanisms and specific characteristics of the L2 input. Computer simulations are constructed with a set of hardware and a set of software, each of which instantiates an architecture that delimits the input that can be accepted and the manipulations that can be performed on that input. In this way, connectionist-type emergentist investigations can be helpful in creating a characterization of what input and in what amount is essential to the acquisition of a construction under certain assumptions on learning. Of course, the neural and psychological plausibility of any given computer simulation model is an open but equally essential question in the effort to characterize human learning. Unless a model is known to accurately model human language processing, the quantity and quality of input required for the model to perform well may not correspond to what a human learner would require.

In summary, Section 6. 2 has shown that the studies under consideration found that L2 learners made progress toward native-like performance on a set of complex phenomena. However, the studies struggled to supply definitive evidence regarding insufficiency of input.

### **6.3 Summarizing the Empirical Results of these L2 Studies: More Support for Generativism than Emergentism**

Chapters 5 and 6 so far have identified a number of flaws with many of the studies, casting some doubt on their utility. But not all of the flaws are fatal, so this section will proceed to assess the overall conclusion supported by this body of studies with respect to which approach is best supported: generativism or emergentism. We will find that on balance, the predictions of the generative approach receive more support than the predictions of the emergentist approach.

First, I argue that the POS argument will not assist us in concluding which approach is best supported. None of the studies presented are able to definitively establish the stimulus poverty premise of the POS argument. Of the three sources of input to L2 learners, the lack of instruction was best supported. With respect to the possible contribution from the L1, the strongest studies are those in which the L2 is a subset of the L1 (Figure 6.1(b) above), where the L1 would lead a learner to hypothesize that certain sentences are grammatical in the L2 on the basis that they are grammatical in the L1. In other words, the learner must acquire a constraint. These are the OPC studies and the quantifier scope study. But given the questions around whether or not absence of occurrence provides evidence for ungrammaticality, it is still not definitively established that positive evidence from the L2 is insufficient for acquisition. Further, with respect to positive L2 evidence, it is not the case for any of the studied phenomena that we know what quantity of relevant input is required for acquisition, nor do we know which constructions constitute relevant input, that is, what quality of input is necessary. On the basis of this discussion, I argue that these studies cannot argue for a role for UG in L2A on the basis of the POS argument.

Let us turn to a different possible source of evidence on the question of innateness. Recall from Chapter 1 that the generative and emergentist approaches make predictions regarding learners' *ultimate attainment* and *developmental paths*. Regarding *ultimate attainment*, generative and emergentist predictions are not that far apart. Both would say that in principle, attainment of native-like performance is possible, although each would attribute that attainment to different acquisition mechanisms. As mentioned in Chapter 1, it is well known that L2 acquisition is typically variable and non-convergent, and as mentioned in Chapter 3, even expert learners are not fully native-like. Proposed explanations for these empirical observations are beyond the scope of this chapter. The point here is that neither the generative nor the emergentist approach rules out ultimate expert learning in principle<sup>90</sup>, and therefore the approaches are better distinguished on other grounds.

The two approaches make different predictions for learners' *developmental paths*. Under the generative approach, the theoretical prediction is that learners who have encountered the appropriate triggering input will show (Chomsky 1980; Meisel 2011):

- sudden progress toward native-like behavior, possibly early on (G1);
- no backsliding: consistent use of the target phenomenon once attained (G2);
- uniform attainment across individual learners (G3).

A few points of clarification are in order. First, the (G1) prediction does not state that acquisition is necessarily achieved early on. Generativists acknowledge that learning takes time, requires input, and that the native language (and for that matter other

---

<sup>90</sup> Researchers of all stripes acknowledge that in practice, L2 learner attainment falls short of being expert, and is never fully native-like. A discussion of the proposed reasons for the differences between theoretical potential and actual achievement (such as L1 interference, affect, blocking of input, among others) is beyond the scope of this work.

previously learned languages, if any) affects the learning trajectory. However, the generative approach posits that the *right* kind of input is necessary, not that a *substantial* amount of input is necessary. Therefore it is possible that some target phenomena could be acquired after relatively little input or experience. Second, the (G2) prediction concerns L2 competence, not performance. On the view that L2 learners who demonstrate a grammaticality contrast have acquired knowledge of the relevant distinction, the (G2) prediction is fulfilled as long as the contrast is demonstrated. L2 learner performance does not have to be perfect and performance can vary across individuals as long as the grammaticality contrast at issue is evident in each individual's judgments.

Under the emergentist approach, the theoretical predictions are that learners who have experienced relevant input will demonstrate (Ellis 2003, 2012):

- gradual progress<sup>91</sup> toward native-like behavior, possibly requiring a substantial amount of input (E1);
- possible backsliding: prolonged inconsistent use of the target phenomenon (E2);
- variation in attainment across individual learners (E3).

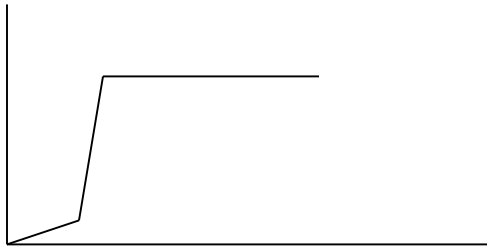
Again as a point of clarification, not all emergentist approaches predict that significant input will be required for acquisition. Prediction (E1) best characterizes an input-frequency-type emergentist account. However, O'Grady's processing-type emergentist account allows for early acquisition of certain target phenomena for which the L1 and L2 processing routines are appropriately aligned. Thus a processing-type emergentist

---

<sup>91</sup> The Power Law of Learning (Newell & Rosenbloom 1981) states in part that error rates decrease as a function of practice, with practice being most effective early on. The effect of practice diminishes with additional experience, gradually becoming asymptotic.

account may be consistent with (G1) rather than (E1) with respect to the acquisition of certain phenomena.

Figures 6.2 and 6.3 below illustrate the theoretical predictions for the idealized learning paths under the generative and emergentist approaches, respectively.



**Figure 6.2: Learning Path, Generative Prediction**



**Figure 6.3: Learning Path, Emergentist Prediction**

Using these sets of predictions to distinguish the two approaches, the following discussion categorizes the empirical evidence from the 2LA studies under consideration. After providing an inventory of which studies support which of the above predictions, the section concludes by arguing that the generative approach receives more support.

First, with respect to the shape and timing of progress toward attainment (predictions G1 and E1), several studies supported the generative approach in that they

found evidence of robust acquisition by relatively low proficiency learners. Kanno (1997) found that her low proficiency learners' performance as a group (and individually, see discussion below) was not significantly different from that of native speakers on the Overt Pronoun Constraint (OPC). Pérez-Leroux and Glass (1999) (henceforth PL&G), testing a different pair of languages, also found that elementary learners demonstrated mastery of the OPC. Williams and Kuribara (2008) (henceforth W&K), in their human subject experiment, found that learners who had brief exposure to scrambling in an artificial language performed well, even on test items containing structures to which they had not been exposed. O'Grady (2013), whose learners were characterized as low-to-mid-intermediate, performed in a native-like manner on constraints on scope interpretations. The results from these four studies are consistent with sudden and early acquisition, supporting generative prediction (G1).

On the other hand, a few studies found evidence for the emergentist approach. The O'Grady (2013) results, just mentioned above, are also consistent with the processing-type emergentist account that he advocates. Other studies found evidence for an input-frequency-type account in that higher proficiency learners did not achieve mastery. Considering her English L1 learners (for whom the target construction was argued to be a POS phenomenon), Marsden (2009) found that almost half of her intermediate learners were inconsistent on interpretation of scope relations between quantifiers, sometimes accepting and sometimes rejecting sentences of the same construction type. O'Grady, Lee and Kwak (2009) (henceforth OLK), testing quantifier/negation scope relations, found that their intermediate-to-high-intermediate learner group maintained their L1 interpretations, with little progress towards the target

interpretation. The results of these studies are consistent with later acquisition, supporting the first emergentist prediction (E1) which says that substantial input may be required for acquisition. However, the W&K computer simulation did not support the emergentist prediction, in that the model's performance did not improve with a substantial amount of additional training.

The second set of predictions (G2 and E2) concern learner performance across time. The generative approach predicts that once learners demonstrate mastery, they will maintain consistent performance. The emergentist approach predicts that learners may perform inconsistently even after having demonstrated some degree of mastery. Three studies tested learners at multiple proficiency levels: PL&G (1999), Hopp (2005), and Marsden (2009).<sup>92</sup> PL&G (1999) tested elementary, intermediate, and high proficiency groups, finding that learners at every level were not significantly different from native speakers. Hopp (2005) tested high-intermediate, advanced, and very advanced learners, all of whom reliably distinguished between grammatical and ungrammatical test sentences. These studies support the generative prediction (G2) that once learners demonstrate acquisition, they will maintain their good performance.

Marsden (2009) tested intermediate and advanced learner groups. As mentioned above, a large proportion (42%) of her English L1 intermediate learner group judged test items inconsistently. An additional 42% incorrectly accepted the object-wide scope reading for the target sentence type. As for her English L1 advanced group, most were consistent in their judgments of a given test item, however, almost half consistently accepted the object-wide scope reading, again, performing in a non-native-like manner.

---

<sup>92</sup> Longitudinal studies would be more reliable. Cross-sectional studies are only an approximation.

To summarize, Marsden's learner groups achieved some mastery by the intermediate proficiency level; however, they were not shown to maintain their mastery across time. Marsden's results support the emergentist prediction (E2) more than the generative prediction (G2).

The third and final set of predictions concerns variation across individuals. Three studies analyzed individual results: Kanno (1997), Hopp (2005), and Marsden (2009). Kanno (1997) found that 86% of her individual subjects performed in a native-like manner. Hopp (2005) found that all subjects reliably distinguished between licit and illicit sentences. These studies support the generative prediction (G3). Marsden (2009), in contrast, found inconsistent performance across individuals. These results are more consistent with an input-dependent, emergentist-type approach than a generative approach.

Table 6.2 summarizes the above discussion, showing which studies support which predictions for the two approaches. Note that Sheen (2000) has not been discussed and does not appear in the table. Recall that Sheen's experiments had methodological goals. They tested the reasons behind learner judgments, investigated the intuitions of native speakers, and investigated the universality of the OPC. Those experimental results do not inform the predictions we are considering.

**Table 6.2: Which Studies Support Which Predictions**

<b>Approach/Predictions</b>	<b>Studies that support the prediction</b>
<b>Generative</b>	
Sudden, possibly early attainment	Kanno 1997, Pérez-Leroux and Glass 1999, Williams and Kuribara 2008, O’Grady 2013
Consistent performance	Pérez-Leroux and Glass 1999, Hopp 2005
Uniform across individuals	Kanno 1997, Hopp 2005
<b>Emergentist</b>	
Incremental, possibly later attainment	Marsden 2009, O’Grady, Lee and Kwak 2009
Inconsistent use of target form	Kanno 1998, Marsden 2009
Variation across individuals	Marsden 2009

Based on these studies, there is a larger quantity of evidence in support of the generative approach than the emergentist approach, as measured by fulfillment of the predictions concerning learners’ developmental path and competence at time of testing. Furthermore, the generative evidence is stronger in quality in the sense that some of the studies that support the generative approach (for instance Hopp 2005, PL&G 1999, and Kanno 1997) were rated more highly in Chapter 5 than the study that I have argued (contra the author’s own conclusions) provides support for the emergentist predictions (Marsden 2009). W&K (2008) did not receive a high rating in Chapter 5 because it was argued that their conclusions did not match their evidence. However, their evidence can be construed to favor a UG approach. For these reasons, I argue that on balance, the evidence in these studies supports a generative approach over an emergentist approach. Note that this conclusion does not rest on any POS argument per se. The conclusion rests

on the correspondence between the empirical results of the studies and the predictions of each approach with respect to learners' developmental paths.

## **6.4 Reflections**

This section rounds out the discussion and conclusions given in the preceding sections with reflections on a few relevant points. First, we consider the extent to which the generative and emergentist approaches share common ground, thus making the tasks of establishing the existence of domain-specific faculties and teasing apart the different contributions of domain-specific and domain-general faculties (if necessary) even more difficult. Next we reconsider a few of the main tools in the empirical investigations of these questions: the POS argument, the use of complex phenomena as targets in acquisition experiments, and the utility of investigating adult L2 learners.

### **6.4.1 Common Ground**

Up to this point, the emphasis in this work has been on the differences between the generative and the emergentist approaches. It is natural, even necessary, to focus on differences when attempting to distinguish two theories. But it is also necessary to recognize similarities, in order to avoid attributing evidence as supporting a certain position when in fact it might be compatible with both. As presented in Chapter 1, the more extreme versions of the two approaches go something like this: For the generativists, the mind/brain is richly structured, innately endowed with a Universal Grammar (variously characterized) which constrains the possible set of natural languages and which guides the learner in his or her effort to acquire the particular language(s) predominant in the environment. Linguistic input acts as a trigger for acquisition, where acquisition is said to be analogous to natural growth rather than a process of effort.

Acquisition is ultimately uniform across individuals despite wide variation in the robustness of the input each one receives. Input does not need to resemble the acquired grammar in all respects. In fact, the POS argument claims that the linguistic input is impoverished, and missing key information. For the emergentists, the mind/brain is supplied with general cognitive skills, which the learner applies to a rich set of input. Learning takes place through processes of pattern search, distributional analysis and the like, in which the learner actively engages. Input is determinative of the grammar acquired and therefore input resembles what is acquired in important ways. At a very basic level, both approaches acknowledge some role for innate endowment and some role for linguistic input, however, each approach characterizes those roles very differently.

In recent decades, the generativists have softened their position. In particular, the hypothesized content of UG has been sharply curtailed. Berwick et al. (2011, p.1210) declares “...we share the desire to reduce any language-specific innate endowment, ideally to a logical minimum.” The philosophy of the Minimalist Program (Chomsky 1995, 2005 among many others) encourages positing of only the leanest, most general principles, such as structural dependence, hierarchical structure or simply recursion along with a limited set of features (an evolution of the older notion of parameters) for the contents of UG.<sup>93</sup> Reduction of the innate component creates a space for a larger role for other factors. With respect to linguistic input, some generativists now acknowledge a role for input frequency in language acquisition<sup>94</sup>, and many L2A studies document an effect of frequency.<sup>95</sup> Other extra-grammatical factors such as processing, social values and

---

<sup>93</sup> See for instance the proposal by Hauser, Chomsky and Fitch (2002).

<sup>94</sup> For child L1A, see for instance Yang (2002).

<sup>95</sup> See Ellis (2002) for a review.

individual affect or aptitude have been argued to impact L2A.<sup>96</sup> However, generativists have not conceded any ground in the sense that they still maintain that input overall and linguistic input specifically is insufficient to determine acquired grammars.

On the other hand, emergentist assumptions remain fairly stable. It is not always emphasized, but as mentioned in Chapter 1, emergentists do not deny the existence of innate abilities completely; they argue against the existence of *language-specific* innate faculties. Any computer model must be initialized with a set of inputs and a set of ways to manipulate those inputs. For computer simulations of language acquisition, this effectively means that a simulation model assumes a set of primitives (words, categories, thematic roles) and a set of learning mechanisms (connection weight adjustments, frequency analysis), all of which constitute a set of assumptions about a human learner's innate knowledge and capabilities. Although no comprehensive model currently exists that takes an unsegmented stream of input and acquires a grammar (as humans do), progress has been made on separate models of acquisition of word segmentation, word classes, morphology, and syntactic parsing<sup>97</sup>. As computing power and modeling techniques evolve, more comprehensive models will be attempted. Elman et al. (1996, p.361-2) suggest that models are well-suited to the task of exploring the limits of the explanatory power of domain-general learning when they say, "We believe that connectionist models are particularly useful as we try to figure out just how much specificity of content we have to build into our machines, and at what level, in order to simulate something that looks like human learning."

---

<sup>96</sup> Moyer (2004) is an example.

<sup>97</sup> See Clark and Lappin (2011, Chapter 9) for a discussion of achievements in computer modeling of language acquisition to date.

So as the generative view of the contents of a language-specific faculty shrinks, the generative picture appears more similar to the emergentist picture in ascribing a significant role to input and extragrammatical factors in acquisition. Even though the two approaches now share more common ground, they are ultimately incompatible in that there is or is not a language-specific faculty, determined in part on a genetic basis. At the moment, we are far from being able to fully characterize that endowment, if it exists, or from understanding how that endowment is physically instantiated in the brain. Elman et al. (1996, Chapter 1) suggest a number of mechanisms for genetic encoding of innateness for language, including things such as constraints on patterns of connectivity of neurons, constraints on global architecture of brain regions, and constraints on the timing of development. One presumes that the same questions must be addressed for domain-general learning faculties. The point is that ‘both’ does not seem to be a possible answer to the question of generativism versus emergentism.

#### **6.4.2 Other Experimental Tools Reconsidered**

This section offers a few additional reflections on the poverty of the stimulus argument, on studying complex linguistic phenomena, and on studying adult L2 learners.

Section 6.2 discussed the particular types of arguments that were used in the (generative) studies under consideration to verify the assumptions of the POS argument. Section 6.2 also mentioned a few factors specific to assessing the L2 input that are important, but nevertheless were not considered in any of the studies. This section will briefly recap the issues at a general level.

The basic logic of the POS argument is sound. If there is a gap between the input that learners receive and the grammars they acquire, then that gap must be filled by the

learners themselves, that is, by an innate endowment that supplies the necessary information. However, the argument is difficult to implement in practice. The premises of the argument are difficult to verify. The stimulus poverty assumption is particularly challenging. We have seen that researchers try to assess the input on its own merits, but this approach is inadequate for several reasons. First, input cannot be considered simply on its own, but must be considered in conjunction with specification of the assumed learning mechanisms. Unless the learning mechanisms are specified, the quality and quantity of the input required is not known. In particular, it is not known whether the learner needs input in the exact form of the target, or to what extent input that is different from the target can provide relevant information to the learner.

Many of the researchers argued for insufficiency of the input on the basis of the frequency and/or overall number of exposures to the *exact* target phenomenon type that the learner could potentially have experienced. From a generative point of view, this argument should have little force. Although linguistic nativists hold that some input is required in order to trigger innate knowledge, triggering input should not be required to be similar to the thing attained. Hornstein (2005, p.149-150) draws an analogy to a person walking into a dark room and turning on a light switch. Once the lights are on, the person can see the furniture and artwork that was already in the room. But turning on the lights had nothing to do with what the furniture and artwork are. The triggering switch does not have to resemble the acquired knowledge. Arguments from frequency are much more relevant to emergentist studies, where such arguments are used to bolster the sufficiency of the input rather than its insufficiency.

The most compelling arguments for insufficiency of input seem to center on cases where learners must acquire a constraint; they must come to know what is ungrammatical. But even here, recent research shows cases where non-occurrence does seem to serve as evidence of ungrammaticality.<sup>98</sup> This means that positive evidence alone can provide the necessary information; direct negative evidence is not required. It is an open question whether such absence can act as negative evidence for more complex phenomena such as scope interpretations. Marsden (2009, p.139-140) argues that non-occurrence of a construction (in L2 input) can only act as indirect negative evidence when there is an obligatory context for the target construction in the L1. If there were a context in the L1 in which the target construction is obligatory, and the equivalent context in the L2 never makes use of the target, then learners might infer the impossibility of using the target construction in that context. However, if the target construction does not have any obligatory contexts, then the learner would more plausibly simply infer that the target construction is dispreferred in the L2, but is not necessarily ungrammatical. Hence, acquisition of constraints on interpretations may be a good test case for poverty of the stimulus arguments in adult L2 acquisition.

Our next topic is the use of complex constructions as target acquisition phenomena. The discussion so far implies that the consideration of certain types of complex constructions is useful to the inquiry into the existence of a language-specific endowment in humans. To reiterate the argument of Schwartz and Sprouse (2013), first discussed in Chapter 2 above, some POS arguments are softer than others. Some phenomena that are considered complex, for instance on the basis of something like an

---

<sup>98</sup> See for instance Ramscar, Dye and McCauley (2014) and references therein.

intricate inventory of forms, may not be good examples of poverty of the stimulus. That is, complex paradigms may be very ‘surfacey’ in that they are observable in the input. A more appropriate description for the better POS phenomena may be ‘hidden’ rather than ‘complex’; constructions for which structure that is crucial to an interpretation cannot be read off the surface linear order are better targets of study (in the sense that they are better examples of POS). I conclude by mentioning a caveat with regard to studying complex or hidden phenomena. They are more difficult to get accurate judgments on, even when testing native speakers. Hopp (2005) is an example of this.

Our final topic is the utility of studying adult second language learners when looking for evidence in the linguistic nativism debate. In general, testing and comparing different learner populations (adult L2, child L1, child L2, child bilingual, heritage speakers, etc.) provides an opportunity to tease apart different factors that may impact acquisition. Children learning their L1 are immersed in the language, receive leveled and structured input that is appropriate for their intake abilities, and participate in a lot of social interaction, all of which promote learning. But children are cognitively less mature. They must develop abstract concepts, their processing skills are not as well-developed as in adults, and their meta-linguistic skills are under development. Adult second language learners, especially instructed learners (as were most of the subjects in these experiments), receive relatively poorer input in the L2. Adult classroom learners are exposed to L2 input for a limited amount of time per day, some of the input from other L2 students may be ungrammatical, and they have far less social interaction in which they use the L2. However, adults have fully mature cognitive and processing systems and skills, including good meta-linguistic skills, and pragmatic skills.

Studying adult second language learners can therefore be particularly useful in that cognitive and processing factors can be taken out of the equation, and the focus can be on the effect of input on acquisition. Unlike in child L1, where the withholding of input is ‘the forbidden experiment’, it should be ethically possible (with the consent of the subjects) to conduct ‘forbidden’-type experiments, carefully controlling the input given to a certain subject population, even to the point of deliberately offering less input than would be expected to be efficacious. From this point of view, the adult L2 learner population has unique potential to provide evidence to the innateness inquiry.

The critical discussions in Chapters 5 and 6 are certainly not exhaustive. However, I hope they have been extensive enough to motivate further investigation into the question of the existence of an innate language faculty, and to point directions to improvements in study methodologies and argumentation that will increase the reliability and validity of future experiments.

## **6.5 Conclusion**

In this work, I examined three sets of existing second language acquisition studies, each set concerning target phenomena which have been argued to be POS phenomena: the Overt Pronoun Constraint, scrambling phenomena, and scope phenomena. Each set of studies included at least one study conducted in the generative framework, and one conducted in an emergentist framework. After presenting the studies, I gave a critical evaluation of their methodologies, rationales, and the conclusions that were based on the empirical results found. I argued that taken as a whole, the evidence from this body of studies offers more support for a nativist position than for an emergentist position. This evidence comes with a caveat, however, in that grammaticality judgment tasks such as

those used in the experiments under consideration can have reliability issues. Also, cross-sectional studies such as these are only an approximation for longitudinal studies, which would provide more reliable evidence concerning learners' developmental paths.

Note that although the poverty-of-the-stimulus argument for the nativist position was a major focus of this thesis and provided the original impetus for the selection of this particular set of studies for the meta-study, the above conclusion does not rest on the POS argument per se. The conclusion is *not* that the generativist researchers established that the learners went beyond the input they received. The conclusion is rather that the nativist prediction for early, rapid learning with no backsliding, and uniform attainment across individuals at a given proficiency level fits the empirical data better than does the emergentist prediction for gradual, later acquisition with backsliding, and with more variation in competence with respect to a target construction across individuals who are otherwise at a comparable proficiency level. The POS argument itself does not stand up to its reputation. In theory, it is a strong argument. However, it is exceedingly difficult to implement because its assumptions are so challenging to verify.

Adult second language acquisition is a fruitful area of research and can contribute to our progress on the fundamental question of the existence of an innate, language-specific endowment. We should continue to study adult 2LA using phenomena that require cognitive maturity and robust processing resources, and take advantage of the fact that we should be freer to experiment with manipulating linguistic input for adult populations. Also, since a computer simulation model requires the explicit specification of input and learning mechanisms, there is value to pursuing computational modeling as

evidence for the relative contributions of input and language-specific or domain-general learning mechanisms.

The extreme versions of the generative and emergentist positions have moved toward each other. There is some common ground. However, ultimately nativism is an empirical question and one with a definitive answer: there either is or is not something in the human genome that is specific to language.

## BIBLIOGRAPHY

- Aarts, Bas. (2007). *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.
- Abraham, W. (1986). Word order in the middle field of the German sentence. In Abraham, W. and de Meij, S. (eds.), *Topic, Focus, and Configurationality*: 15-38. Amsterdam: John Benjamins.
- Abrahamsson, N. and Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249-306.
- Anderson, Stephen R. and Lightfoot, David W. (2002). *The Language Organ: Linguistics as Cognitive Psychology*. Cambridge: Cambridge University Press.
- Ayoun, Dalila. (1999). Verb movement in French L2 acquisition. *Bilingualism: Language and Cognition* 2: 103-125.
- Bader, Markus and Häussler, Jana. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273-330.
- Baker, Carl Lee and McCarthy, John J. (1981). *The Logical Problem of Language Acquisition*. Cambridge: MIT Press.
- Baker, Carl Lee. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry* 10: 233-280.
- Bard, Ellen Gurman, Robertson, Dan, and Sorace, Antonella. (1996). Magnitude estimation of linguistic acceptability. *Language* 72: 32-68.
- Berwick, Robert C. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.
- Berwick, Robert C., Pietroski, Paul, Yankama, Berach, and Chomsky Noam. (2011). Poverty of the stimulus revisited. *Cognitive Science* 35: 1207-1242.
- Birdsong, David. (1989). *Metalinguistic Performance and Interlinguistic Competence*. Berlin/New York: Springer-Verlag.
- Birdsong, David. (1992). Ultimate attainment in second language acquisition. *Language* 68: 706-755.
- Birdsong, D., Johnson, C., McMinn, J. and Ingmundson, P. (1985). Universals versus transfer revisited. ERIC ED 255 061.
- Bley-Vroman, Robert. (1990). The logical problem of foreign language learning. *Linguistic Analysis* 20: 3-49.

- Bley-Vroman, Robert. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition* 31: 175-198.
- Bloomfield, Leonard. (1933). *Language*. NY: Holt, Rinehart and Winston.
- Bornkessel, Ian, Schlesewsky, Matthais, and Friederici, Angela. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition* 85: B21-B30.
- Bott, Oliver and Radó, Janina. (2007). Quantifying quantifier scope: a cross-methodological comparison. In Featherston, Sam and Sternefeld, Wolfgang (eds.), *Roots: Linguistics in Search of its Evidential Basis*: 53-74. Berlin/New York: de Gruyter.
- Bott, Oliver and Schlotterbeck, Fabian. (2012). Incremental truth value judgments. In Stolterfoht, Britta and Featherston, Sam (eds.), *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure*: 17-42. DeGruyter Mouton.
- Bowerman, Melissa. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In Hawkins, John A. (ed.), *Explaining Language Universals*. Oxford: Blackwell.
- Boysson-Bardies, Bénédicte de. (1999). *How Language Comes to Children*. Cambridge, MA: Cambridge University Press.
- Brenowitz, Eliot. (2004). Plasticity in the adult avian song control system. *Annals of the New York Academy of Sciences* 1016: 560-585.
- Brooks, P. and Tomasello, Michael. (1999). Young children learn to produce passives with nonce verbs. *Developmental Psychology* 35: 29-44.
- Bybee, J. and Thompson, S. (2000). Three frequency effects in syntax. *Berkeley Linguistic Society* 23: 65-85.
- Carroll, J.M., Bever, T.G. and Pollock, C.R. (1981). The non-uniqueness of linguistic intuitions. *Language* 68: 706-735.
- Carroll, Susanne E. (2001). *Input and Evidence: the Raw Material of Second Language Acquisition*. Philadelphia: John Benjamins Publishing Company.
- Chaudron, Craig. (1983). Research on metalinguistic judgments: A review of theory, methods and results. *Language Learning* 33: 343-377.
- Chaudron, Craig. (2003). Data collection in SLA research. In Doughty, Catherine J. and Long, Michael H. (eds.) *The Handbook of Second Language Acquisition*: 762-829. Blackwell.

- Chesi, Cristiano and Moro, Andrea. (2014). Computational complexity in the brain. In Newmeyer, Frederick J. and Preston, Laurel B. (eds.), *Measuring Grammatical Complexity*: 264-280. Oxford: Oxford University Press.
- Chomsky, Noam. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. (1959). A review of B.F. Skinner, *Verbal Behavior*. *Language* 35: 26-58.
- Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (1966). *Cartesian Linguistics*. New York: Harper & Row.
- Chomsky, Noam. (1975). *Reflections on Language*. New York: Random House (Pantheon Books).
- Chomsky, Noam. (1988). *Language and Problems of Knowledge*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (2005). Three Factors in Language Design. *Linguistic Inquiry* 36(1): 1-22.
- Chomsky, Noam. (2005/1980). *Rules and Representations* (forward by Norbert Hornstein). New York: Columbia University Press.
- Chung, Eun Seon (Eunice). (2012). Sources of difficulty in L2 scope judgments. *Second Language Research* 29(3): 285-310.
- Clahsen, H. and Muysken, P. (1986). The availability of Universal Grammar to adult and child learners – a study of the acquisition of German word order. *Second Language Research* 2: 93-119.
- Clahsen, H. and Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research* 5: 1-29.
- Clark, Alexander and Lappin, Shalom. (2011). *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell.
- Cook, Vivian. (2003). The poverty-of-the-stimulus argument and structure-dependency in L2 users of English. *IRAL* 41: 201-221.
- Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language* 63: 544-573.
- Cowart, Wayne. (1994). Anchoring and grammar effects in judgments of sentence acceptability. *Perceptual and Motor Skills* 79: 1171-1182.

- Cowart, Wayne. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage.
- Cowie, Fiona. (2008). *What's Within?* New York/Oxford: Oxford University Press.
- Cowie, Fiona, "Innateness and Language", *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2010/entries/innateness-language/>.
- Crain, Stephen. (2012). *The Emergence of Meaning*. Cambridge: Cambridge University Press.
- Crain, Stephen. (2013). What's parsing got to do with it? *Linguistic Approaches to Bilingualism* 3(3): 301-307.
- Crain, Stephen and McKee, C. (1985). The acquisition of structural restrictions on anaphora. In Berman, S., Choe, J., and McDonough, J. (eds.), *Proceedings of NELS 15*. Amherst: University of Massachusetts, GSLA.
- Crain, Stephen and Nakayama, Minehara. (1987). Structure dependence in grammar formation. *Language* 63: 522-543.
- Crain, Stephen and Pietroski, Paul. (2001). Nature, nurture and Universal Grammar. *Linguistics and Philosophy* 24: 139-186.
- Crain, Stephen and Steedman, Mark. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In Dowty, David D., Karttunen, Lauri, and Zwicky, Arnold M. (eds.), *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives*. Cambridge: Cambridge University Press.
- Crain, Stephen and Thornton, Rosalind. (1998). *Investigations in Universal Grammar*. Cambridge/London: The MIT Press.
- Dąbrowska, Ewa. (2004). *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar*. Washington, D.C.: Georgetown University Press.
- Dekydsporter, Laurent. (2001). The universal parser and interlanguage: Domain-specific mental organization in the comprehension of *combien* interrogatives in English-French interlanguage. *Second Language Research* 17: 91-43.
- Dekydsporter, Laurent. (2009). Second language epistemology: Take two. *Studies in Second Language Acquisition* 31: 291-321.
- Dekydsporter, Laurent and Sprouse, R.A. (2001). Mental design and (second) language epistemology: Adjectival restrictions of *wh*-quantifiers and tense in English-French interlanguage. *Second Language Research* 17: 1-35.

- Dekydsporter, Laurent, Sprouse, Rex, and Swanson, K. (2001). Reflexes of mental architecture in second language acquisition: the interpretation of discontinuous *Combien* extractions in English-French interlanguage. *Language Acquisition* 9: 175-227.
- Dekydsporter, L., Sprouse, R. A. and Thyre, R. (1999/2000). The interpretation of quantification at a distance in English–French interlanguage: domain specificity and second-language acquisition. *Language Acquisition* 8:1–36.
- Demirdache, Hamida and Lungu, Oana. (2008). Sequence of tense in French child language. *Linguistic Variation Yearbook* 8: 101-130. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Descartes, René. (1637/1978). *Discourse on Method*. Translated by John Veitch, LL.D. New York: Dutton.
- Descartes, René. (1985). *The Philosophical Writings of Descartes, vol. 1*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press.
- Descartes, René. (1988). *Descartes: Selected Philosophical Writings*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press.
- Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies* 11(9): 25-45.
- Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research* 168: 49-64.
- Doughty, Catherine J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In Doughty, Catherine J. and Long, Michael H. (eds.), *The Handbook of Second Language Acquisition*: 256-310. Blackwell Publishing.
- Drozd, Kenneth F. (2004). Learnability and linguistic performance. *Journal of Child Language* 31: 431-457.
- Duffield, Nigel and White, L. (1999). Assessing L1 knowledge of Spanish clitic placement: converging methodologies. *Second Language Research* 15(2): 133-160.
- Ellis, Nick C. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition* 13(2): 161-186.
- Ellis, Nick C. (1998). Emergentism, connectionism and language learning. *Language Learning* 48: 631–664.

- Ellis, Nick C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143-188.
- Ellis, Nick C. (2003). Constructions, chunking, and connectionism: the emergence of second language structure. In C. Doughty and M. Long (eds.), *The Handbook of Second Language Acquisition*: 63-103. Oxford: Blackwell.
- Ellis, Nick C. (2012). Frequency based accounts of second language acquisition. In Gass, Susan and Mackey, Alison (eds.), *The Routledge Handbook of Second Language Acquisition*: 193-210. London and New York: Routledge.
- Ellis, Nick C. and Larsen-Freeman, Diane. (2006). Language emergence: implications for applied linguistics – Introduction to the special issue. *Applied Linguistics* 27(4): 558-589.
- Ellis, N. and Schmidt, R. (1997). Morphology and longer distance dependencies: laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition* 19: 145-171.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language; A psychometric study. *Studies in Second Language Acquisition* 27: 141-172.
- Elman, Jeffrey L. (1990) Finding structure in time. *Cognitive Science* 14: 179-211.
- Elman, Jeffrey L., Bates, Elizabeth A., Johnson, Mark H., Karmiloff-Smith, Annette, Parisi, Domenico, Plunkett, Kim (eds.). (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Epstein, S.A. and Flynn, S. and Martohardjono, G. (1986). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences* 19: 677-714.
- Featherston, Sam. (2007a). Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3): 269-318.
- Featherston, Sam. (2007b). Reply. *Theoretical Linguistics* 33(3): 401-413.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge: MIT Press.
- Fodor, J. A. (2000). *The Mind Doesn't Work That Way: Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.
- Fodor, Janet Dean and Crowther, Carrie. (2002). Understanding stimulus poverty arguments. *The Linguistic Review* 19: 105-145.
- Gass, Susan M. and Selinker, Larry. (2001). *Second Language Acquisition: An Introductory Course*. Mahwah, New Jersey: Lawrence Erlbaum.

- Goodall, Jane. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, Massachusetts: Harvard University Press.
- Gregg, Kevin R. (2003). The state of emergentism in second language acquisition. *Second Language Research* 19: 95-128.
- Gregg, Kevin R. (2005). SLA Theory: Construction and assessment. In Doughty, Catherine J. and Long, Michael H. (eds.), *The Handbook of Second Language Acquisition*: 831-865. Blackwell.
- Grüter, Theres. (2103). Individual differences and the nature of the processor. *Linguistic Approaches to Bilingualism* 3(3): 312-315.
- Hale, K. (1996). Can UG and L1 be distinguished in L2 acquisition? *Behavioral and Brain Sciences* 19: 726-728.
- Harley, Trevor. (2001). *The Psychology of Language: From Data To Theory*. 2<sup>nd</sup> Edition. Hove and New York: Psychology Press Ltd.
- Hauser, Marc D., Chomsky, Noam, and Fitch, W. Tecumseh. (2002). The faculty of language. *Science* 198: 1569-1579.
- Hawkins, John A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, Roger and Chan, Y-h. C. (1997). The partial availability of Universal Grammar in second language acquisition: the 'Failed Functional Features Hypothesis'. *Second Language Research* 13: 187-226.
- Hawkins, R. and Hattori, H. (2006). Interpretation of multiple *wh*-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research* 22: 269-301.
- Herschensohn, Julia. (2000). *The Second Time Around: Minimalism and L2 Acquisition*. Amsterdam: Benjamins.
- Herschensohn, Julia. (2007). *Language Development and Age*. Cambridge: Cambridge University Press.
- Herschensohn, Julia. (2009). Fundamental and gradient differences in language development. *Studies in Second Language Acquisition* 31: 175-198.
- Herschensohn, Julia. (2013). Can pattern recognition explain grammatical learning? *Linguistic Approaches to Bilingualism* 3 (3): 316-320.
- Hill, Archibald A. (1961). Grammaticality. *Word* 17(1): 1-10.

- Hoberg, U. (1981). *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache* [Word order in contemporary written German]. München: Hueber.
- Holme, Randal. (2013). Emergentism, connectionism and complexity. In Herschensohn, Julia Rogers and Young-Scholten, Martha (eds.), *The Cambridge Handbook of Second Language Acquisition*: 605-626. Cambridge; New York: Cambridge University Press.
- Hopp, Holger. (2005). Constraining second language word order optionality: scrambling in advanced English–German and Japanese–German interlanguage. *Second Language Research* 21: 34–71.
- Hopp, Holger. (2002). Constraints on word order variation: learnability and UG in advanced English–German and Japanese–German interlanguage. Unpublished MA thesis, University of Durham.
- Hopp, Holger. (2007). Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing. Groningen Dissertations in Linguistics 65.
- Hsu, Anne. S., Chater, Nick and Vitányi, Paul M.B. (2011). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition* 120: 380-390.
- Hume, David. (1739/1978). *A Treatise on Human Nature*. Oxford: Clarendon.
- Inagaki, Shunji. (2002). Japanese learners' acquisition of English manner-of-motion verbs in locational/directional PPs. *Second Language Research* 18: 3-27.
- Inagaki, Shunji. (2006). Manner-of-motion verbs with locational/directional PPs in L2 English and Japanese. In Slabakova, R., Montrul, S., and Prevost, P. (eds.), *Inquiries in Linguistic Development: In Honor of Lydia White*: 41-68. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Iwasaki, Noriko. (2003). L2 acquisition of Japanese: knowledge and use of case particles in SOV and OSV sentences. In Karimi, Simin (ed.), *Word Order and Scrambling*: 273-300. Blackwell.
- Kaan, E., Harris, A., Gibson, E. and Holcomb, P.J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15: 159-201.
- Kanno, Kazue. (1997). The acquisition of null and overt pronominals in Japanese by English speakers. *Second Language Research* 13 (3): 265-287.
- Kanno, Kazue. (1998). The stability of UG principles in second language acquisition: evidence from Japanese. *Linguistics* 36(6): 1125-1146.
- Kellerman, Eric and Yoshioka, Kaoru. (1999). Inter- and intra-population consistency: A comment on Kanno (1998). *Second Language Research* 15(1): 101-109.

- King, J.W. and Just, M.A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30: 580-602.
- Kutas, M. and Fedemeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* 4: 463-470.
- Kutas, M. and Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science* 207: 203-205.
- Lardiere, Donna. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25: 173-227.
- Lardiere, Donna. (2013). Nativelike and non-nativelike attainment. In Herschensohn, Julia R. and Young-Scholten, Martha (eds.), *The Cambridge Handbook of Second Language Acquisition*: 670-691. Cambridge: Cambridge University Press.
- Larsen-Freeman, Diane. (2012). Complexity theory. In Gass, Susan and Mackey, Alison (eds.), *The Routledge Handbook of Second Language Acquisition*: 73-87. London and New York: Routledge.
- Lasnik, Howard and Uriagereka, Juan. (2002). On the poverty of the challenge. *The Linguistic Review* 19: 147-150.
- Lee, S. (2009). Interpreting scope ambiguity in first and second language processing: Universal quantifiers and negation. Ph.D. dissertation, University of Hawaii at Manoa.
- Legate, Julie Ann and Yang, Charles D. (2002). Empirical reassessment of stimulus poverty arguments. *The Linguistic Review* 19: 151-162.
- Leowen, Shawn. (2003). Variation in the frequency and characteristics of incidental focus on form. *Language Teaching Research* 7: 315-345.
- Leowen, Shawn. (2012). The role of feedback. In Gass, Susan M. and Mackey, Alison (eds.), *The Routledge Handbook of Second Language Acquisition*: 24-40. London and New York: Routledge.
- Li, Ping. (2009). Lexical Organization and Competition in First and Second Languages: Computational and Neural Mechanisms. *Cognitive Science* 33: 629-664.
- Li, Ping and Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In Heredia, R. and Altarriba, J. (eds.), *Bilingual Sentence Processing*. Oxford: Elsevier.
- Lightfoot, David. (2005). Plato's problem, UG, and the language organ. In McGilvray, James (ed.), *The Cambridge Companion to Chomsky*. Cambridge: Cambridge University Press.

- Locke, John. (1690/1964). *An Essay Concerning Human Understanding*. Wozzley, A.D. (ed.). Cleveland: Meridian Books.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition* 21: 557-587.
- MacWhinney, Brian (ed.). (1999). *The Emergence of Language*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- MacWhinney, Brian. (2008). A unified model. In Robinson, P. and Ellis, Nick (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- MacWhinney, Brian. (2012). The logic of the unified model. In Gass, Susan and Mackey, Alison (eds.), *The Routledge Handbook of Second Language Acquisition*. London and New York: Routledge.
- Marsden, Heather. (2002). Poverty of the stimulus and UG in Japanese L2 acquisition: A response to Sheen (2000). *Durham Working Papers in Linguistics* 8: 67-80.
- Marsden, Heather. (2004). Quantifier scope in non-native Japanese: a comparative interlanguage study of Chinese, English, and Korean-speaking learners. Unpublished PhD dissertation, University of Durham.
- Marsden, Heather. (2009). Distributive quantifier scope in English–Japanese and Korean–Japanese interlanguage. *Language Acquisition* 16: 135–77.
- Masny, Diana and d'Anglejan, Alison. (1985). Language, cognition, and second language grammaticality judgments. *Journal of Psycholinguistic Research* 14(2): 175-197.
- McCarthy, John J. and Prince, Alan. (1993). Prosodic morphology I: Constraint interaction and satisfaction. *Linguistics Department Faculty Publication Series*, Paper 14.
- Medler, David A. (1998). A brief history of connectionism. *Neural Computing Surveys* 1(2): 18-72.
- Meisel, Jürgen M. (2011). *First and Second Language Acquisition*. Cambridge: Cambridge University Press.
- Montalbetti, M. (1983). After binding: on the interpretation of pronouns. Unpublished doctoral dissertation, MIT.
- Moyer, Alene. (2004). *Age, Accent and Experience in Second Language Acquisition*. Clevedon: Multilingual Matters Ltd.

- Müller, Gereon. (1996). A constraint on remnant movement. *Natural Language and Linguistic Theory* 14: 355-407.
- Müller, Gereon. (1998). *Incomplete Category Fronting*. Dordrecht: Kluwer.
- Murasugi, K. and Kawamura, T. (2005). On the acquisition of scrambling in Japanese. In Sabel, J. and Saito, M. (eds.), *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity, Vol 69*. Berlin/New York: Mouton de Gruyter, p.221-242.
- Myles, Florence. (2013). Theoretical approaches. In Herschensohn, Julia R. and Young-Scholten, Martha (eds.), *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press.
- Nakano, H., Saron, C. and Schwab, T.Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience* 22: 2886-2898.
- Narens, Louis. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology* 40:109–29.
- Nemoto, Naoko. (1995). Scrambling in Japanese, AGRoP, and economy of derivation. *Lingua* 97: 257-273.
- Newell, A. and Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In Anderson, J.R. (ed.), *Cognitive Skills and Their Acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. p. 1-55.
- Newmeyer, Frederick J. (1986). *Linguistic Theory in America*: Second edition. Academic Press.
- Newmeyer, Frederick J. (2003). Grammar is grammar and usage is usage. *Language* 79(4): 682-707.
- Newmeyer, Frederick J. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.
- Newmeyer, Frederick J. (2007). Commentary on Sam Featherston, ‘Data in generative grammar: The stick and the carrot’. *Theoretical Linguistics* 33(3): 395-399.
- Noguchi, T. (1997). Two types of pronouns and variable binding. *Language* 73: 770-797.
- Norris, J. M., and Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50: 417–528.
- Norris, John M. and Ortega, Lourdes. (2012). Assessing learner knowledge. In Gass, Susan and Mackey, Alison (eds.), *The Routledge Handbook of Second Language Acquisition*. London and New York: Routledge.

- O'Grady, William. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, New Jersey/London: Lawrence Erlbaum Associates, Publishers.
- O'Grady, William. (2008). Innateness, Universal Grammar, and Emergentism. *Lingua* 118: 620-631.
- O'Grady, William. (2013a). The illusion of language acquisition. *Linguistic Approaches to Bilingualism* 3: 253-285.
- O'Grady, William. (2013b). Facts and issues: Response to the commentaries. *Linguistic Approaches to Bilingualism* 3(3): 373-383.
- O'Grady, W., Lee, M. and Choo, M. (2001). The acquisition of relative clauses by heritage and non-heritage learners of Korean as a second language: a comparative study. *Journal of Korean Language Education* 12: 283-94.
- O'Grady, William, Lee, Miseon, and Kwak, Hye-Young. (2009). Emergentism and second language acquisition. In Ritchie, W. and Bhatia, T. (eds.), *The New Handbook of Second Language Acquisition*, 69-88. Bingley, UK: Emerald Press.
- Oka, T. (1996). Scrambling in Japanese and English. *Formal Approaches to Japanese Linguistics I*. MIT Working Papers in Linguistics 29: 361-388.
- Ortega, Lourdes. (2000). Understanding syntactic complexity: the measurement of change in the syntax of instructed L2 Spanish learners. PhD dissertation. University of Hawai'i.
- Osterhout, L. and Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31: 785-806.
- Osterhout, L., McLaughlin, J., Pitkänen, I. Frenck-Mestre, C., and Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning* 56(Suppl 1): 199-230.
- Pérez-Leroux, Ana T. (1995). Resumptives in the acquisition of relative clauses. *Language Acquisition* 4: 105-138.
- Pérez-Leroux, Ana T. and Glass, William R. (1999). Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches. *Second Language Research* 15(2): 220-249.
- Peters, Stanley (1972). The projection problem: How is a grammar to be selected? In Stanley Peters (ed.), *Goals of Linguistic Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Phillips, Colin. (2008). Should we impeach armchair linguists? In Iwasaki, S. (ed.) *Japanese/Korean Linguistics 17*. CSLI Publications.

- Piaget, Jean. (1955[1923]). *The Language and Thought of the Child*. Trans. by M. Gabain. Cleveland: Meridian.
- Piatelli-Palmarini, Massimo (ed.) (1980). *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
- Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.
- Pinker, Steven. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Pinker, Steven. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.
- Pinker, Steven and Jackendoff, Ray. (2005). The faculty of language: what's special about it? *Cognition* 95(2): 201-236.
- Pinker, Steven and Prince, Alan. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28: 73-193.
- Preston, Laurel B. (2005). Are grammaticality judgment tests good measures of proficiency? Seminar paper, University of Washington.
- Preston, Laurel B. (2013). Adult L2 Acquisition of Sequence of Tense: Providing Evidence on Two Hypotheses of Adult Second Language Acquisition. General paper, University of Washington.
- Prince, Alan and Smolensky, Paul. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. RUCCS Technical Report, Vol. 2. Piscataway, New Jersey: Rutgers University Center for Cognitive Science.
- Progovac, Ljiljana. (2014). Degrees of complexity in syntax: A view from evolution. In Newmeyer, Frederick J. and Preston, Laurel B. (eds.) *Measuring Grammatical Complexity*: 83-102. Oxford: Oxford University Press.
- Pullum, Geoffrey K. and Scholz, Barbara. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19: 8-50.
- Pullum, Geoffrey K. and Scholz, Barbara. (2002). Searching for arguments to support linguistic nativism. *The Linguistic Review* 19: 185-223.
- Ramscar, Michael, Dye, Melody, and McCauley, Stuart M. (2013). Error and expectation in language learning: The curious absence of *mouses* in adult speech. *Language* 89(4): 760-793.

- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 317–327
- Rebuschat, Patrick. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning* 63(3): 595-626.
- Redington, Martin and Chater, Nick. (1998). Connectionist and Statistical Approaches to Language Acquisition: A Distributional Perspective. *Language and Cognitive Processes* 13: 2-3, 129-191, DOI: [10.1080/016909698386500](https://doi.org/10.1080/016909698386500).
- Roeper, T. (2000). Universal bilingualism. *Bilingual Language and Cognition* 2: 169-186.
- Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland, J. L. , D.E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2. Psychological and Biological Models*: 216-271. Cambridge: Bradford Book/MIT Press.
- Saito, M. (1985). Some asymmetries in Japanese and their theoretical implications. Unpublished PhD dissertation. MIT, Cambridge, MA.
- Sampson, Geoffrey. (2005). *The 'Language Instinct' Debate: Revised Edition*. London: Continuum.
- Sauerland, Uli. (1999). Erasability and interpretation. *Syntax* 2: 161-188.
- Schlesewsky, Matthais, Fanselow, Gisbert, Kliegl, Reinhold and Krems, Josef. (2000). The subject preference in the processing of locally ambiguous *wh*-questions in German. In Hemforth, B. and Konieczny, L. (eds.), *German Sentence Processing*: 65-84. Dordrecht: Kluwer.
- Schmitt, Cristina, and Miller, Karen. (2010). Using comprehension methods in language acquisition research. In Blom, Elma and Unsworth, Sharon, (eds.), *Experimental Methods in language Acquisition Research*: 35-56. John Benjamins Publishing Company.
- Schreiber, T. and Sprouse, Rex. A. (1998). Knowledge of topicalization and scrambling in English-German interlanguage. *McGill Working Papers in Linguistics* 13: 162-172.
- Schütze, Carson T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Schwartz, Bonnie Dale. (1987). *The Modular Basis of Second Language Acquisition*. Unpublished PhD Dissertation, University of Southern California.

- Schwartz, Bonnie D. and Sprouse, Rex A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12(1): 40-72.
- Schwartz, Bonnie D. and Sprouse, Rex A. (2000). When syntactic theories evolve: consequences for L2 acquisition research. In Archibald, J. (ed.), *Second Language Acquisition and Linguistic Theory*: 156-187. Oxford: Blackwell.
- Schwartz, Bonnie D. and Sprouse, Rex A. (2013). Generative approaches and the poverty of the stimulus. In Herschensohn, Julia R. and Young-Scholten, Martha (eds.), *The Cambridge Handbook of Second Language Acquisition*: 137-158. Cambridge: Cambridge University Press.
- Scribner, Sylvia and Cole, Michael. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10: 209-231.
- Sells, P. (1984). *Syntax and semantics of resumptive pronouns*. PhD dissertation, University of Massachusetts.
- Sheen, Ron. (2000). A response to Kanno's "The stability of UG principles in second-language acquisition: evidence from Japanese". *Linguistics* 38(4): 799-816.
- Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Slabakova, Roumyana. (2003). Semantic evidence for functional categories in interlanguage grammars. *Second Language Research* 19(1): 42-75.
- Smith, Linda B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In MacWhinney, Brian (ed.), *The Emergence of Language*: 277-304. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Snow, C.E. and Ferguson, C.A. (eds.). (1977). *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press.
- Snyder, William. (2000). An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry* 31(3): 575-582.
- Sorace, Antonella. (1996). The use of acceptability judgments in second language acquisition research. In Ritchie, William C. and Bhatia, Tej K. (eds.), *Handbook of Second Language Acquisition*: 375-409. San Diego: Academic Press.
- Sorace, A. and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22(3), 339-368.

- Sorace, A. and Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism* 13: 195-210.
- Sprouse, Jon. (2011). A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language* 87(2): 274-288.
- Stevens, S. S. and Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology* 54(6): 377-411.
- Stevens, Stanley S. (1956). The direct estimation of sensory magnitudes: Loudness. *American Journal of Psychology* 69: 1-25.
- Tanner, Darren. (2013). Individual differences and streams of processing. *Linguistic Approaches to Bilingualism* 3(3): 350-356.
- Thomas, Margaret. (2002). Development of the concept of “the poverty of the stimulus”. *Linguistic Review* 19: 51-71.
- Thomas, M.S.C. (1997). Connectionist Networks and Knowledge Representation: The Case of Bilingual Lexical Processing. PhD Dissertation, Oxford University.
- Tomasello, Michael. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Tsujioka, Takae. (2001). Improper remnant A-movement. *Proceedings of NELS* 31: 483-500.
- Von Frisch, K. (1950). *Bees, their vision, chemical senses, and language*. Ithaca, New York: Cornell University Press.
- Warner, John and Glass, Arnold. (1987). Context and distance-to-distance disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language* 26: 714-738.
- Weskott, Thomas and Fanselow, Gilbert. (2011). On the informativity of different measures of linguistic acceptability. *Language* 87: 249-73.
- White, Lydia. (1985). The “pro-drop” parameter in adult second language acquisition. *Language Learning* 35: 47-62.
- White, Lydia. (1989). *Universal Grammar and Second Language Acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- White, Lydia. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research* 7, 133-161.
- White, Lydia. (1992). Subjacency violations and empty categories in L2 acquisition. In Goodluck, H. and Rochemont, M. (eds.), *Island Constraints*: 445-464. Dordrecht: Kluwer.
- White, Lydia. (2003a). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- White, Lydia. (2003b). On the nature of interlanguage representation: Universal grammar in the second language. In Doughty, Catherine J. and Long, Michael H. (eds.), *The Handbook of Second Language Acquisition*: 19-42. Blackwell.
- White, Lydia, Bruhn-Garavito, J., Kawasaki, T., Pater, J., and Prévost, P. (1997). The researcher gave the subject a test about himself: Problems of ambiguity and preference in the investigation of reflexive binding. *Language Learning* 47: 145-172.
- White, Lydia and Gennese, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research* 11: 233-265.
- Williams, John N. and Kuribara, Chieko. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua* 118: 522-553.
- Williams, John N. and Rebuschat, Patrick. (2011). Statistical learning and syntax. In Williams, John N. and Rebuschat, Patrick (eds.), *Statistical Learning and Language Acquisition*: 237-264. Boston: De Gruyter.
- Winter, Bill and Reber, Arthur. (1994). Implicit learning and the acquisition of natural languages. In Ellis, Nick C. (ed.), *Implicit and Explicit Learning of Languages*: 115-146. London: Academic Press.
- Yang, Charles D. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.
- Yang, Charles D. (2004). Universal grammar, statistics, or both? *TRENDS in Cognitive Sciences* 8(10): 451-456.
- Younger, B. (1990). Infants' detection of correlations among feature categories. *Child Development* 61: 614-621.